Troposphere Modeling and Filtering for Precise GPS Leveling

Frank Kleijer

Contents

General summary

Troposphere Modeling and Filtering for Precise GPS Leveling

precision

In the Netherlands, a precision of 5 mm for estimated GPS height differences is required to achieve comparable accuracy as geoid height differences. This precision can be achieved for 24-hour data sets when applying a proper modeling. Precise leveling with GPS requires a judicious modeling of tropospheric delays, which has physical, functional, and stochastic aspects.

modeling
aspects

The physical modeling comprises zenith delays for the hydrostatic and wet component and zenith angle dependent mapping functions. Because the amount of water vapor in the atmosphere fluctuates widely and because the water-vapor induced signal delays affect the height component strongly, a-priori modeling of these delays results in an insufficient precision of height differences. Parameterizing the tropospheric delay in the functional model is therefore necessary, at least for medium and long baselines. The observation model can further be strengthened by pseudo-observations. These pseudo-observations may be spatiotemporal constraints on tropospheric delay differences, or constraints on residual slant delays. With the latter type of constraint the isotropy assumption is loosened. An existing theoretical model is revised to obtain the corresponding covariance matrix. The stochastic modeling of both types of constraints is based on the assumption of Kolmogorov turbulence.

filtering

The observation models can be implemented in a recursive filter like the Kalman Filter or the SRIF. Several variations of these filters are described. For fast computations the most suitable recursive filtering technique is the Kalman–Cholesky Filter with pre-elimination of temporal GPS parameters, such as clock errors and ionospheric delays. Some tests and reliability descriptions are worked out for this filter. For practical implementation the temporal behavior of the zenith wet delay is to be assumed a random-walk process, which gives a fair description. The zenith wet delay can be estimated every epoch or every pre-defined batch of epochs.

impact of
model
components

The effect of several model components on mainly the height is analyzed by simulation software. Special attention is given to the residual-slant-delay model because it is potentially precision and reliability improving. The impact of this model does however depend on the precision level of the observations and it still needs to be validated. Although observations to low-elevation satellites have a large contribution to the precision of the height, the residual-slant-delay model implies a strong down weighting of observations to satellites below ten degrees elevation. The highest accuracy can be obtained when the phase ambiguities are fixed. Even for long observation time spans this makes a difference of up to 15–20%. Batch size and spatiotemporal constraints turned out to have little

effect on the formal precision of the height, but they have a large influence on the precision of the filtered zenith delays. Because overconstraining (with constraints that are too tight) can have a large precision-deteriorating effect, spatiotemporal constraints are not recommended for GPS leveling. Further, short batches are preferred to avoid biases. To prevent the presence of near rank deficiencies, the zenith delays of one station are often fixed. From a precision point of view this is not necessary and is not recommended because this can also introduce biases. Even larger biases can be introduced when the zenith delays of all stations are fixed to their a-priori values. For very short baselines ($< \pm 1$ km), this model is however justifiable because the formal precision improves considerately, keeping the effect of the biases in balance.

Algemene samenvatting

Troposfeermodellering en -filtering voor precieze GPS-waterpassing

precisie

Om in Nederland een vergelijkbare precisie te krijgen als van de Nederlandse geoïde, is voor geschatte GPS-hoogteverschillen een precisie nodig van 5 mm. Deze precisie kan bij adequate modellering gehaald worden voor datasets van 24 uur. Precieze GPS-waterpassing vereist een zorgvuldige modellering van troposferische vertragingen, welke fysische, functionele en stochastische aspecten heeft.

modellerings-aspecten

De fysische modellering omvat zenitvertragingen voor de hydrostatische en natte component en zenithoekafhankelijke 'mapping'-functies. Omdat de hoeveelheid waterdamp in de atmosfeer sterk fluctueert en omdat de hierdoor veroorzaakte vertragingen van de GPS-signalen de hoogtecomponent sterk beïnvloeden, resulteert a priori modellering van deze vertragingen in een onvoldoende precisie van de hoogteverschillen. Parameterizeren van de troposferische vertraging in het functionele model is daarom nodig, althans voor middellange en lange basislijnen. Het waarnemingsmodel kan verder worden versterkt met pseudowaarnemingen. Deze pseudowaarnemingen kunnen spatieel-temporele troposfeer-constraints zijn of constraints op schuine restvertragingen. Met het tweede soort constraints wordt de troposferische isotropieaanname losgelaten. Voor de corresponderende covariantiematrix is een bestaand theoretisch model herzien. De stochastische modellering van beide typen constraints is gebaseerd op de aanname van Kolmogorov-turbulentie.

filteren

De waarnemingsmodellen kunnen worden geïmplementeerd in een recursief filter zoals het Kalman-filter of het SRIF. Diverse versies van deze filters worden beschreven. Voor snelle berekeningen is het Kalman-Cholesky-filter met pre-eliminatie van temporele parameters, zoals klokfouten en ionosferische vertragingen, het geschiktst. Enige toetsen en betrouwbaarheidsbeschrijvingen zijn voor dit filter uitgewerkt. Voor praktische implementatie moet voor het temporele gedrag van de natte zenitvertraging een random-walkproces worden aangenomen, hetgeen een redelijk goede beschrijving geeft. De natte zenitvertraging kan elke epoche of elke vooraf gedefinieerde batch (reeks) van epochen geschat worden.

effect van model-componenten

Het effect van diverse modelcomponenten op met name de hoogte is met simulatieprogrammatuur geanalyseerd. Speciale aandacht is gegeven aan het model voor schuine restvertragingen omdat dit potentieel precisie- en betrouwbaarheidsverbeterend is. De invloed van dit model is echter afhankelijk van het precisieniveau van de waarnemingen en het model moet nog gevalideerd worden. Hoewel waarnemingen naar lage-elevatiesatellieten een grote bijdrage aan de precisie van de hoogte hebben, impliceert het model voor schuine restvertragingen dat waarnemingen naar satellieten met een elevatie lager dan

tien graden een laag gewicht krijgen. Echter, de hoogste nauwkeurigheid kan worden verkregen als de fasemeerduidigheden worden vastgehouden. Zelfs voor lange waarnemingstijden scheelt dit tot 15–20%. Batch-grootte en spatieel-temporele constraints bleken weinig effect te hebben op de formele precisie van de hoogte maar ze hebben een grote invloed op de precisie van de gefilterde zenitvertragingen. Omdat overconstraining (met te strakke constraints) een groot precisie-verslechterend effect kan hebben, worden spatieel-temporele constraints niet aangeraden voor GPS-waterpassen. Korte batches hebben verder de voorkeur, om onzuiverheden te vermijden. Om bijna-rangdefecten te voorkomen worden vaak de zenitvertragingen van één station vastgehouden. Voor de precisie is dit niet nodig, en het wordt ook niet aangeraden omdat ook dit onzuiverheden kan introduceren. Nog grotere onzuiverheden kunnen geïntroduceerd worden als de zenitvertragingen van alle stations worden vastgehouden op hun a priori waarden. Voor hele korte basislijnen ($< \pm 1$ km) is dit model echter te rechtvaardigen omdat de formele precisie aanzienlijk verbetert, hetgeen hiertegenop weegt.

Abbreviations

abbreviation	meaning
AGRS.NL	Active GPS Reference System of the NetherLands
BLUE	Best Linear Unbiased Estimate
DD	Double Differenced
DOY	Day Of Year
GOM	Global Overall Model
GPS	Global Positioning System
GS	Global Slippage
HIRLAM	HIgh Resolution Limited Area Model
IGS	International GPS Service
InSAR	Interferometric Synthetic Aperture Radar
IPWV	Integrated Precipitable Water Vapor
IWV	Integrated Water Vapor
KF	Kalman Filter
LOM	Local Overall Model
LS	Local Slippage
LSQ	Least SQuares
MDB	Minimal Detectable Bias
MSE	Mean Square Error
NAP	Normaal Amsterdams Peil (Amsterdam Ordnance Datum)
msl	mean sea level
NWP	Numerical Weather Prediction
OMT	Overall Model Test (variance of unit weight)
PC	Personal Computer
PPP	Precise Point Positioning
PSD	Power Spectral Density
QR	not an abbreviation: simply Q times R
RINEX	Receiver INdependent EXchange (format)
RMS	Root Mean Square

abbreviation	meaning
SA	Selective Availability
SD	Single Differenced
SRIF	Square Root Information Filter
STD	Slant Total (Tropospheric) Delay
TL2	Treuhaft and Lanyi model, scale height 2 km
UMPI	Uniformly Most Powerful Invariant
UT	Universal Time
VLBI	Very Long Baseline Interferometry
WVR	Water Vapor Radiometer
ZD	Zero Differenced
ZHD	Zenith Hydrostatic Delay
ZTD	Zenith Total (Tropospheric) Delay
ZWD	Zenith Wet Delay

Equal signs and operators

symbol	meaning
$=$	is equal to
$\neq$	is not equal to
$\approx$	is approximately equal to
$\doteq$	is by definition equal to
$\leftarrow$	assignment
$\sim$	is distributed as
$<$	is smaller than
$\leq$	is smaller than or equal to
$\ll$	is much smaller than
$>$	is larger than
$\geq$	is larger than or equal to
$\gg$	is much larger than
$\cdot$	product
$\otimes$	Kronecker product
$\oplus$	vector-space addition
A'	transpose of A
A^{-1}	inverse of A
$\text{diag}[Q_1, ..., Q_n]$	(block-)diagonal matrix with elements $Q_1, ..., Q_n$
$\lvert \cdot \rvert$	absolute value
$<.>$	ensemble/time average
$\mathbf{D}\{.\}$	mathematical dispersion
$\mathbf{E}\{.\}$	mathematical expectation
$\mathcal{F}\{.\}$	Fourier transform
$\dot{x}$	time derivative of x (dx/dt)
X^*	complex conjugate of X
$\lVert x \rVert_{Q_x}$	L2 norm of x $(\sqrt{x'Q_x^{-1}x})$
$\lVert x \rVert$	L2 norm of x $(\sqrt{x'x})$
$\backslash$	substitution
$R = QR(A)$	QR decomposition $R = Q'A$
$R = Chol(A)$	Cholesky decomposition $R'R = A$

GENERAL INTRODUCTION

<div align="right">

Chapter 1

Problem area

</div>

GPS   The Global Positioning System (GPS) is a satellite-based radio positioning system that has been fully operational since 1994. Although GPS was originally designed for (military) navigation purposes, the system has shown to be of great interest for surveying. Where for navigation applications meter-level accuracy is desired using GPS' pseudo-range observations, surveyors aim at centimeter-level accuracy. To obtain this accuracy, they use the GPS phase observations and relative positioning techniques. The highest accuracy can be obtained with long observation times, by resolving the phase ambiguities, and by using accurate orbits in postprocessing.

For most survey purposes, cm-level accuracy is sufficient. For the height component however, the demands are often somewhat stricter than for the horizontal coordinates. Unfortunately, the intrinsic precision of GPS-derived heights is worse than that of the horizontal coordinates. This is caused by a combination of satellite geometry with all satellites on one side (above) the receiver, the presence of receiver clock errors, and disturbances in the neutral

troposphere   atmosphere, also referred to as troposphere, that cause delays of the GPS signals. The first aspect can only partly be overcome by using low-elevation satellites, but the observations to these satellites are of lower accuracy and are more sensitive to error sources like multipath. The second aspect can only be overcome by using accurate atomic clocks, which would be too expensive to make GPS receivers suitable for every-day use. But by a sophisticated modeling of the delays caused by the troposphere, the accuracy of GPS-derived heights can become 'as good as it gets'.

The GPS is expected to be(come) a cost-effective addition to, or replacement of, (part of) the relatively expensive spirit and hydrostatic levelings of Rijkswaterstaat. Levelings have always been used by Rijkswaterstaat to maintain

NAP   the Dutch height reference datum NAP. From the Second Primary Leveling (1926–1940) to the Fifth Primary Leveling (1996–1999), the precision of primary levelings (± 1 mm/$\sqrt{\text{km}}$) has not changed much [*Murre*, 1985]. Although the amount of required manpower is reduced considerably, leveling is still considered a rather expensive technique. In the primary levelings, the heights of underground benchmarks are measured.[1] Because all other NAP heights are derived from these reference benchmarks and because accurate heights are of utmost importance in a low and watery country like the Netherlands, the highest possible accuracy for the underground benchmarks is desired against reasonable costs.

ellipsoidal   GPS heights are geometric (ellipsoidal) heights which only have a mathe-
heights

---

[1]The connection of underground benchmarks to nearby positioned GPS receivers can be done with mm precision.

orthometric heights
matical, but no physical meaning. Leveling-derived heights on the other hand do have a more significant physical meaning. The exact definition of our leveling-derived heights is of minor importance in a flat country like the Netherlands. They may, for example, be orthometric or normal heights. In both cases we need to apply corrections to the leveling observations up to only a few millimeters. In the Second Primary Leveling a maximum orthometric correction of 1.5 mm was applied for the leveling line Roermond–Maastricht [*Waalewijn*, 1979]. Once the corrections are applied, the derived heights basically (very nearly) tell us the direction of flowing water.

deformation analysis
Because we are usually more interested in heights with a physical meaning, the use of GPS for deformation analysis seems to be most appealing, since the velocities we find from repeated GPS and leveling campaigns are (except for possible biases and noise) effectively the same. In the period 1994–1997 several GPS campaigns were conducted in the Groningen area [*Beckers et al.*, 1996], [*Beckers et al.*, 1998]. This area has been subsiding since the start of the gas field exploitation in 1964. The Dutch Oil Company NAM has monitored the subsidence behavior by repeated spirit-leveling campaigns because of possible damage claims. From the additional GPS campaigns a somewhat disappointing baseline precision was found that showed a distance-dependent behavior (±0.5 mm/km) and a relatively large minimum standard deviation (±5 mm). The baselines were up to about 10 km and observation times were used of about one hour; no troposphere parameters were estimated.

Fifth Primary Leveling
The Fifth Primary Leveling of the Netherlands not only included spirit and hydrostatic leveling campaigns, but also a GPS campaign. In this campaign the baselines had a mean length of about 20 km and the observation times were about five days. By using these longer baselines and observation times, one strove toward a precision of the GPS-derived heights that is comparable with that of the leveling-derived heights. One tropospheric zenith-delay parameter per station per two hours was estimated. The tropospheric delay showed to be one of the precision-limiting factors.

AGRS.NL
Since 1996 also the Active GPS Reference System network of the Netherlands (AGRS.NL) started to continuously collect GPS data [*Brouwer et al.*, 1996]. In the period 1997–1999 this network had five operational stations with baseline lengths of about 100–300 km. Daily solutions of the station coordinates were computed using one troposphere zenith-delay parameter per station per two hours. In [*Kleijer*, 2002] a standard deviation of the daily coordinate solutions was demonstrated of ±5 mm in the height component and 1.5–2 mm in the horizontal coordinates. The time series of the coordinates did not only show a stochastic behavior and a linear trend, but also a clear and significant cyclic signal with a wavelength of about a year; see Fig. 1.1. The amplitude is in the order of 1–2 mm for the height component, whereas the horizontal coordinates have amplitudes below 1 mm. Because of the seasonal weather changes, this amplitude may be troposphere related, but other causes may be present.

modernization
The obtainable precision of GPS coordinates has been improving gradually in time [*Bock*, 1998]. Hardware improvement and new modeling techniques have contributed to this. In the near future, GPS will be modernized [*Eisfeller et al.*, 2002] and a European satellite system (Galileo) is likely to see the light (the future of the Russian positioning system Glonass is still uncertain). These satellite
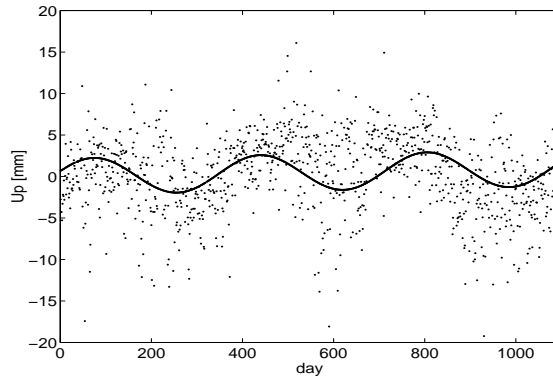
Figure 1.1   Time series (1997-1999) of the height component of the baseline Kootwijk-Delft. The heights are relative to the estimated height at the start of the time series. The plotted cyclic function is the best fit in a network adjustment after outlier elimination.

positioning systems use radio signals penetrating the earth's atmosphere. Ongoing research on the troposphere's behavior is therefore necessary while the precision of an integrated GPS/Galileo will further improve the precision of, for example, the height component [*Van der Marel and De Haan*, 2001].

De Min geoid

While the precision of GPS-derived coordinates is improving, also the precision of the Dutch geoid that forms the link between ellipsoidal and orthometric heights had been improved with the 'De Min' geoid. This geoid has a precision of 1–2 cm and was expected to be improvable to about 5–10 mm [*De Min*, 1996]. The relative precision of two close points was expected to be even better. A new geoid [*De Bruijne and Crombaghs*, 2002], now likely to be named NLGEO2004, will replace the De Min geoid and gives a relative precision of about 5 mm.

hydrostatic leveling

At Rijkswaterstaat one strives toward an increased use of satellite positioning systems for height determination, mainly as replacement of the existing levelings. Especially the technique of hydrostatic leveling is considered too expensive. This technique has been the only accurate technique for leveling large distances over water in the Netherlands since its introduction in the 1950s [*Waalewijn*, 1959]. It has, for example, been used for levelings from the Dutch coast to the Frisian islands and to 'Nulpalen', benchmarks at a distance of about a kilometer from the coast for monitoring sea-level changes. For several decades the Dutch ship 'De Niveau' has been the only ship with which a specialist team conducted these levelings on a regular basis. Now that the ship is out of use [*Scheele*, 2002], it seems unlikely that this technique will ever be used again.

With the aim to replace hydrostatic levelings by GPS, Rijkswaterstaat carried out GPS test campaigns (IJssel, 2000; Huibertsgat en Wierumergronden, 2000/2001; Westerschelde, 2001) to investigate the achievable precision of GPS with existing commercial software. Observation time spans were used of 3–24 hours (and 3×24 hours for the reference stations) for baselines up to 25 km; no troposphere parameters were estimated. The errors of the baseline heights after transformation to NAP were about 2 cm peaking up to 3–4 cm, which is larger than the desired precision of about 1 cm: roughly 5 mm repeatability of GPS heights and 5 mm for the transformation to NAP. This endorses the need to estimate troposphere parameters.

Outline

In view of the in Chap. 1 mentioned developments and the recognized precision-limiting troposphere error for GPS height determination, this research concentrates on modeling the tropospheric delay in GPS observation models. The central question is:

How can the height component be improved by a more sophisticated troposphere modeling?

We aim at the 5-mm precision level or better, assume a classic network approach, and bear in mind the Dutch applications, although this has not really limited the covered area. By 'improving' we mean: faster (shorter sessions) or more precise than what is possible with most commercial software (applying an a-priori correction model). In fact 'faster' and 'more precise' are equivalent, since longer observation times result in more precise estimates (up to a certain level). Because of the high precision demand, faster does not necessarily mean: fast ambiguity resolution. To find possible improvements, we need to know what modeling possibilities there are.

A subdivision was made into five topics that are described in separate parts. The first three parts deal with the three aspects of troposphere modeling, namely physical, functional, and stochastic modeling. The fourth part deals with the recursive estimation/filtering of parameters (coordinates, troposphere parameters, etc.); a recursive technique allows us to analyze the precision of coordinates as function of time. The first four parts can, to a certain degree, be read separately. The notation is specific for each of the parts, but it is made consistent between parts as much as possible. Fairly all important model components are described, and the ideas and assumptions in the various model components are discussed. In Parts II and III we have worked toward a comparison of functional/stochastic models by means of simulation in Part V; in this part we look for the sensitivity of the height component to the model choices we make.

I. Physical modeling

Physical modeling of the tropospheric delay (Part I) concentrates on a description in terms of physical parameters. A-priori correction models are given, which can be used for ground-based space geodetic observation techniques such as GPS. We distinguish hydrostatic (dry) and nonhydrostatic (wet) delays. The hydrostatic delay is caused by atmospheric gases that are in hydrostatic equilibrium. This is usually the case for dry gases and part of the water vapor. This part of the delay can very well be modeled based on the surface air pressure. The wet delay, **wet delay** caused by water vapor that is not in hydrostatic equilibrium, is the source of our problems. We can derive models for this delay based on the partial pressure
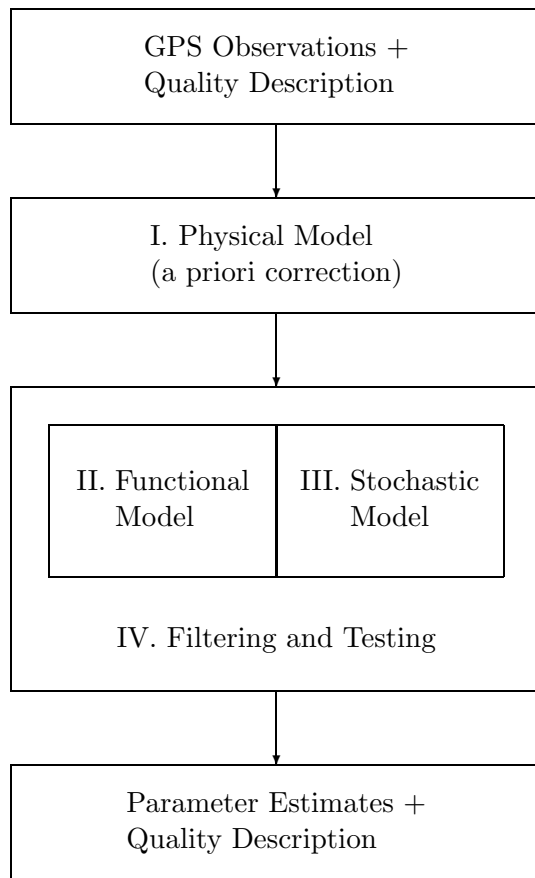
Figure 2.1   Schematic representation of the connection between the topics of Parts I–IV with respect to GPS data processing.

of water vapor or relative humidity at the surface, but these models have a low accuracy and need empirical constants that may vary widely with location and time of year.

mapping functions

Both the wet and dry delays are usually expressed as the product of the delay in the zenith direction and a mapping function, which depends on the zenith angle of the signal. The wet and dry-delay mapping functions differ because of the thickness of both parts of the troposphere. Water vapor is more concentrated near the surface than dry gases. The thicknesses may vary with time and location, and mapping functions therefore depend on physical parameters like surface temperature, total air pressure or partial pressure of water vapor, or by parameters like height, day of year, and latitude. Because physical parameters such as pressure and temperature are often not given or measured, they are usually modeled by a standard atmosphere, which gives them as a function of height under mean atmospheric conditions.

II.Functional modeling

Functional modeling of the tropospheric delay (Part II) involves the parameterization of the tropospheric delay and the functional relation with the GPS observables or additional pseudo-observables. Troposphere parameters may be zenith delays, gradients, or residual slant delays. GPS observables may be GPS phase and code observables, and pseudo-observables may be residual slant delays and time differenced zenith delays. Part II gives the functional relations between all parameters (troposphere, ionosphere, and clock parameters; ambiguities, co-

ordinates) and the (pseudo-)observables. Full-rank observation models are derived.

If we do not parameterize the tropospheric delay and just apply an a-priori correction model, we speak of a troposphere-fixed model. In this model, which is used in many less-sophisticated software packages, residual (unmodeled) tropospheric delays deteriorate the precision of the height component. In the troposphere-float model, we do parameterize the tropospheric delay. The bulk of the residual delays is then modeled by a wet zenith-delay parameter per station per time interval, using the wet mapping function as functional relation. The model is strengthened by extra pseudo-observables in the troposphere-weighted model, which is implemented in the scientific software packages and some commercial packages as well. These packages do however not always allow for any number of parameters. The observation model can be refined by parameterization of gradients or residual slant delays. Gradients have become more popular in recent years and are already implemented in, for example, the Bernese software, version 4.2 [*Beutler et al.*, 2000]. The main gain of modeling gradients however seems to be for the horizontal coordinates [*Rothacher et al.*, 1998]. Modeling of (residual) slant delays is still in a research phase.

III.Stochastic modeling

Part III deals with the stochastic model of the tropospheric pseudo-observables. If we think of the refractive index of water vapor as a spatial stochastic process, the stochastic properties of this process can be (approximately) characterized by a power-law function. The wet delay is an integral over the refractive index and can also be characterized by a power-law function. Spatial correlation can be modeled in GPS observation models by polynomial approximations.

Under the assumption of a frozen troposphere (the troposphere refractivity is fixed and moves with constant velocity), the stochastic properties of time-differenced zenith delays can also be approximated by a power-law function. No easy way to use the correlation efficiently in the data processing exists however. One needs to assume that these differences form a white-noise process, a special form of a power-law process.

We show a numerical stochastic model of the residual slant delays as derived in [*Emardson and Jarlemark*, 1999], which is based on a power-law process and the delay-defining integral over the refractivity. The covariance matrix, which is based on a polynomial approximation of a double integral, turned out to be (probably intrinsically) indefinite for all tested satellite configurations. A new positive-definite model was found in the form of a double-summation model. It could be promising because it involves stronger correlation between delays in nearby directions. The old and new model both imply a form of zenith-angle-dependent weighting when correlations are discarded.

IV.Recursive estimation

Part IV describes recursive estimation and testing procedures as can be used in GPS data processing. Similar procedures exist for VLBI [*Herring et al.*, 1990]. Both the Kalman Filter and the Square Root Information Filter are described. In the most general (troposphere-weighted) model, there are three types of parameters: global, temporal, and constrained (batch) parameters. Because the temporal parameters are of no interest for our goal, they are pre-eliminated, and thus not solved for. Two pre-elimination methods exist: one using a projector onto the orthogonal subspace of the partial design matrix of the temporal parameters, and one using the orthogonal subspace itself. The latter method is

more appealing in terms of computation time, but can only be used because this orthogonal subspace can indeed be found analytically. A recursive estimation procedure was chosen instead of a batch procedure because it was expected to give a better insight in the models. An advantage is that filter results can easily be plotted and interpreted as function of observation time.

V.Simulation   Part V gives a selection of simulation results. This part first describes the software implementation based on the theory in the preceding parts. Then some criteria are shown for judging the results: errors, precision, test values, and reliability measures. The results are first shown for a default scenario. Then several scenarios are analyzed mainly based on the precision. The troposphere-fixed model is treated separately.

contribution   Much of the given theory was already developed earlier. The main contribution herein is probably the consequent approach to obtain a consistent mathematical framework for troposphere modeling and estimation in geodetic networks using a full matrix-vector notation. This is of special interest for software developers who think in terms of procedures to implement. Because this dissertation aims at bringing the theory consistently, it is sometimes hard to distinguish between known theory and new contribution. Therefore each of the first four parts contains a concluding chapter in which also the major advances are specified. In the last part the results follow from simulation software by the author.

Bibliography

Beckers, G.W.J., H.M. de Heus, H. van der Marel, M.H.F. Martens, and H.M.E. Verhoef (1996), Gecombineerde GPS- en waterpasdata t.b.v. de bodemdalingsmetingen in Groningen, *Publ. of the Delft Geodetic Computing Centre, 15*, Delft University of Technology.

Beckers, G.W.J., H.M. de Heus, H. van der Marel, and M.H.F. Martens (1998), Onderzoek bodemdalingsmetingen Groningen '95-'96-'97, *Publ. of the Delft Geodetic Computing Centre 17*, Delft University of Technology.

Beutler, G., E. Brockmann, R. Dach, P. Fridez, W. Gurtner, U. Hugentobler, J. Johnson, L. Mervart, M. Rothacher, S. Schaer, T. Springer, and R. Weber (2000), Bernese GPS Software, Version 4.2, *software documentation*, AUIB.

Bock, Y. (1998), Reference systems, in *GPS for Geodesy*, edited by P.J.G. Teunissen and A. Kleusberg, pp. 1–41, Springer-Verlag.

Brouwer, F., J. van Buren, H. van der Marel, R. Molendijk, and J. van der Veen (1996), National report of the Netherlands, in *EUREF Symposium, Ankara*.

Bruijne, A. de, and M. Crombaghs (2002), NLGEO2002, een nieuw geoïdemodel voor Nederland, *preliminary report, md-gap-2002-33*, Meetkundige Dienst, Ministerie van Verkeer en Waterstaat.

Eisfeller, B., C. Tiberius, T. Pany, R. Biberger, and T. Schueler (2002), Real-time kinematic in the light of GPS modernization and Galileo, in *ION GPS 2001, Salt Lake City*.

Emardson, T.R., and P.O.J. Jarlemark (1999), Atmospheric modelling in GPS analysis and its effect on the estimated geodetic parameters, *Journal of Geodesy, 73*, 322–331.

Herring, T.A., J.L. Davis, and I.I. Shapiro (1990), Geodesy by Radio Interferometry: The Application of Kalman Filtering to the Analysis of Very Long Baseline Interferometry Data, *Journal of Geophysical Research (B), 95*, 12561–12581.

Kleijer, F. (2002), Time series analysis of the daily solutions of the AGRS.NL reference stations, in *IAG symposium on Vertical Reference Systems, Colombia, 2001*, edited by H. Drewes et al., pp. 60–65, Springer.

Marel, H. van der, and S. de Haan (2001), The use of ground based GNSS networks for meteorology, in *Proceedings of NaviTec 2001, 1st ESA workshop on satellite navigation user equipment technologies*, pp. 167–174.

Min, E.J. de (1996), *De geoïde voor Nederland*, Ph.D. dissertation, Delft University of Technology.

Murre, L.M. (1985), Hervereffening van de tweede, derde en vierde nauwkeurigheidswaterpassing van Nederland en vergelijking van de resultaten, *Msc. thesis*, TH Delft.

Rothacher, M., T.A. Springer, S. Schaer, and G. Beutler (1998), Processing strategies for regional GPS networks, in *Advances in Positioning and Reference Frames, IAG Scientific Assembly, Rio de Janeiro, 1997*, pp. 93–100.

Scheele, T. (2002), Adieu Niveau, het einde van 50 jaar hydrostatisch waterpassen bij Rijkswaterstaat, *Geodesia*, pp. 472–477.

Waalewijn, A. (1959), Report on hydrostatic levelling, in *Rijkswaterstaat communications*, pp. 61–86.

Waalewijn, A. (1979), *De Tweede Nauwkeurigheidswaterpassing van Nederland*, Publ. of the Netherlands Geodetic Commission, Delft.

TROPOSPHERE DELAY MODELING FOR SPACE GEODETIC MEASUREMENTS

Symbols and units in Part I

symbol	meaning	unit
a, b, c	parameters of mapping function	[-]
α	specific volume	[m^{3}kg^{-1}]
α	azimuth	[rad]
α	$\cos(z_0)$	[-]
$\beta_{d,sat}$	temperature lapse rate	[K (k)m^{-1}]
$D_{d,h,t,v,w}^{s,z}$	tropospheric delay[5,7]	[m]
$e_{0,sat}$	partial pressure of water vapor	[mbar]
ϵ	constant R_d/R_v	[-]
ϕ	latitude	[rad]
g_m	gravitational acceleration[4]	[m s^{-2}]
h_T	height above msl	[(k)m]
H$_m$	scale (or effective) height[4]	[km]
k_1, k_2, k_2'	constants	[K mbar^{-1}]
k_3, k_3'	constants	[K^{2}mbar^{-1}]
L	latent heat	[J kg^{-1}]
λ	empirical constant	[-]
$m_{d,v}$	mass[5]	[kg]
$M_{d,i,v}$	molecular mass[1,5]	[kg mol^{-1}]
$M_{h,w}(.)$	mapping function[7]	[-]
μ	constant $g_m/(R_d\beta) - 1$	[-]
n	refractive index	[-]
$N_{d,h,v,w,T}$	refractivity[5,7]	[-]
$P_{0,d}$	air pressure	[mbar]
q	specific humidity	[-]
Q	empirical constant $IPWV/D_w^z$	[-]

Table 0.1 Frequently used symbols. The symbols may have one or more subscripts or superscripts (see Table 0.3). Sometimes a subscript or superscript may have more than one meaning, as indicated with the numbers[1,...,8]. These numbers correspond with the numbers in Table 0.3.

symbol	meaning	unit
$r_{0,i,T}$	radius (distance to center of the earth)[2]	[km]
rh_0	relative humidity	[-]
R	universal gas constant	[J mol^{-1}K^{-1}]
$R_{d,i,m,v}$	gas constant[1,3,5]	[J kg^{-1}K^{-1}]
$\rho_{d,m,v,w}$	density[3,5,8]	[kg m^{-3}]
$T_{0,m,v,T}$	temperature[4,6]	[K, (°C)]
w_{sat}	mixing ratio	[-]
$z_{0,i}$	zenith angle[2]	[rad]
$Z_{d,v}$	compressibility factor[5]	[-]

Table 0.2   Frequently used symbols (continuation of Table 0.1).

index	meaning
$\cdot_0$	at base of tropospheric layer (usually at surface)
$\cdot_d$	of dry air
$\cdot_h$	of hydrostatic part of the atmosphere
$\cdot_i$	of the ith constituent[1]
$\cdot_i$	of the ith layer[2]
$\cdot_m$	of moist air[3]
$\cdot_m$	mean value for $h = 0...\infty$[4]
$\cdot_{sat}$	of saturated air
$\cdot_t$	total
$\cdot_T$	of tropopause
$\cdot_v$	of water vapor[5]
$\cdot_v$	virtual[6]
$\cdot_w$	wet (nonhydrostatic) part of atmosphere[7]
$\cdot_w$	of water[8]
$\cdot^s$	slant
$\cdot^z$	zenith

Table 0.3   Subscripts and superscripts as used for the symbols of Tables 0.1 and 0.2.

unit	composition	dimension
Joule [J]	N m	energy
millibar [mbar]	100 N m$^{-2} \equiv$ hPa	pressure
Newton [N]	kg m s^{-2}	force
degrees Celcius [°C]	K - 273.16	temperature

Table 0.4   Units.

Chapter 3

Introduction to Part I

Many space geodetic measurement techniques, like VLBI and GPS, use the propagation time of radio waves multiplied with the speed of light as a measure of range. When a signal passes the earth's atmosphere, it affects the wave in three ways: (1) it causes a propagation delay; (2) it causes a bending of the ray path; and (3) it absorbs the signal. In this 'Part I' only the propagation delay in the neutral atmosphere (or troposphere) is considered, including an indirect delay caused by the bending. Derivations and descriptions are given of several (components of) delay models that can be used to a priori correct the measured ranges.

Part I can be used as a reference work or as a first introduction to the subject. The aim is to give an overview of the most important results in modeling the propagation delay as can be found in literature. The frequently used Saastamoinen model and some of the important new(er) models are described. The intension is not to give all the existing models, since an extensive overview can be found in, for example, [*Mendes*, 1999].

Modern delay models usually distinguish a hydrostatic and a wet delay, where the hydrostatic delay can be determined more precisely by far. Slant tropospheric delays are usually modeled by the sum of products of zenith-angle-dependent mapping functions and zenith delays:

$$D_t^s = M_h(z)D_h^z + M_w(z)D_w^z, \tag{3.1}$$

D_t^s : Slant Tropospheric/Total Delay (STD) [m];
$M_h(z)$: hydrostatic mapping function [-];
D_h^z : Zenith Hydrostatic Delay (ZHD) [m];
$M_w(z)$: wet mapping function [-];
D_w^z : Zenith Wet Delay (ZWD) [m];
z : zenith angle [rad].

Zenith-delay models are given in Chap. 5; some slant-delay models are described in Chap. 6 with a focus on the mapping functions; and Chap. 7 deals with gradient delays caused by azimuthal asymmetry in the troposphere. Because all delay models are based on several physical laws and assumptions, Chap. 4 first gives some necessary theoretical background about the physics of the atmosphere. Conclusions are given in Chap. 8.

Chapter 4

Physics of the atmosphere

4.1 Introduction

This chapter deals with the physical background that is needed to derive troposphere delay models. After we have described the atmospheric layers in Sect. 4.2, in Sect. 4.3 we first give some basic physical laws. Section 4.4 then gives some water-vapor related quantities and definitions. We end with general formulas for the refractivity as function of pressure and temperature in Sect. 4.5, and describe these quantities as function of height for a standard atmosphere in Sect. 4.6.

4.2 Atmospheric layers

troposphere

The earth's atmosphere can be coarsely subdivided in several concentric layers. The characterization of these layers depends on the purpose for which the subdivision is made. Well-known characteristic features are: temperature, ionization, and propagation [*Seeber*, 1993]. Characterizing the atmosphere by the way radio waves are propagated leads to a subdivision of a troposphere and ionosphere. The ionosphere, the upper part of the atmosphere, is a dispersive medium (the propagation delay is frequency dependent), whereas the troposphere is nondispersive. The troposphere is also referred to as neutral atmosphere to distinguish with its original definition, which is actually based on the characterizing temperature profile.

msl
tropopause

In general, the temperature in the troposphere decreases almost linearly with height. At the top of the troposphere, at about 9–16 km above mean sea level (msl), the temperature stays constant for about another 10 km. This part of the neutral atmosphere is called tropopause. Above the tropopause the temperature increases again in the stratosphere up to about 50 km altitude. Between 50 and 80 km above msl the temperature drops again in the mesosphere; see Fig. 4.1.

When we speak of the troposphere, it will be clear from the context whether this is the neutral atmosphere or the layer with decreasing temperature.

4.3 Physical laws for the atmosphere

In this section the following physical laws and equations are given: the gas equation (equation of state), hydrostatic equilibrium, and Snell's law. In the sections to follow, these laws will be applied.

Altitude [km]	Temperature	Ionization	Magnetic field	Propagation	Technical
10000 —	Thermosphere	Protonosphere		Ionosphere	Upper Atmosphere
1000 —		Ionosphere	Magnetosphere		
100 —	Mesosphere				
	Stratosphere		Dynamosphere		Lower Atmosphere
10 —		Neutrosphere			
	Troposphere			Troposphere	

Figure 4.1   Possible subdivision schemes of the earth's atmosphere; after [*Seeber*, 1993].

4.3.1   Equation of state

In general, both volume and pressure increase when a gas is heated. In a closed system however, the volume or pressure may be kept constant. In [*Champion*, 1960] three laws for gases are given:

1. *Charles' constant pressure law*: "At constant pressure, for a rise in temperature of 1°C, all gases expand by a constant amount, equal to about 1/273 of their volume at 0°C."

2. *Charles' constant volume law*: "If the volume is kept constant, all gases undergo an increase in pressure equal to 1/273 of their pressure at 0°C for each degree Centigrade rise of temperature."

3. *Boyle's law*: "At constant temperature the product of pressure and volume is constant."

equation of
state

Based on these laws, the gas equation or equation of state is formulated for perfect (ideal) gases:

$$\boxed{\alpha P = R_i T,} \tag{4.1}$$

with

α : specific volume $[\text{m}^3\text{kg}^{-1}]$;
P : pressure $[\text{N m}^{-2}]$;
R_i : specific gas constant $[\text{J kg}^{-1}\text{K}^{-1}]$;
T : temperature $[\text{K}]$;
$[\text{N}]$: $[\text{kg m s}^{-2}]$;
$[\text{J}]$: $[\text{N m}]$.

specific
volume

The specific volume α is defined as:

$$\alpha \doteq \frac{1}{\rho} \doteq \frac{V}{m}, \tag{4.2}$$

with

ρ : density [kg m^{-3}];
V : volume [m^3];
m : mass [kg].

specific gas
constant

The specific gas constant R_i (for the ith gas) is related to the universal gas constant by:

$$R_i M_i \doteq R, \tag{4.3}$$

with

M_i : (mean) molecular mass [kg mol^{-1}];
R : universal gas constant [8.31434 J mol^{-1}K^{-1}].

Although there are no perfect gases, the gases in the troposphere are nearly perfect and can often be treated as such. The equation of state does not only hold for one specific gas, but also for a mixture of gases. In that case, P is the sum of the partial pressures, R_i is the specific gas constant of the mixture, and M_i is the mean molecular mass of the mixture.

4.3.2   Hydrostatic equilibrium

The atmosphere is said to be in hydrostatic equilibrium if the net vertical force on any slice of a column of air with thickness dh is equal to zero [*Haltiner and Martin*, 1957]. If one thinks of a slice with unit area, the vertical forces can be expressed in terms of pressure (force per area). The downward force is equal to the force at the top of the slice $P + dP$ plus the force due to gravity $g\,\rho\,dh$; see Fig. 4.2. The upward force is equal to the pressure at the bottom of the slice P. The hydrostatic equation therefore reads:

$$g\,\rho\,dh + (P + dP) = P. \tag{4.4}$$
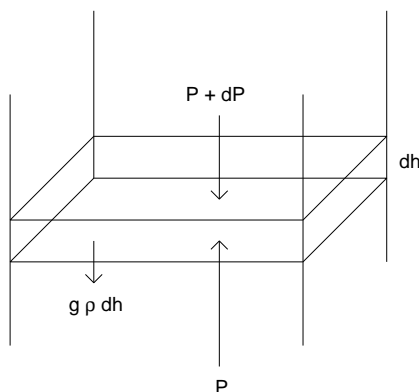
Figure 4.2   Vertical forces on a unit slice of thickness dh in hydrostatic equilibrium.

Rewriting Eq. (4.4) gives:

$$\boxed{dP = -g\,\rho\,dh \quad \text{or} \quad \frac{dP}{dh} = -g\rho,} \tag{4.5}$$

with

g : gravitational acceleration [m s^{-2}];
h : height above msl [m].

Using the definition of Eq. (4.2), this may also be written as:

$$\boxed{g\,dh = -\alpha\,dP.}$$
(4.6)

4.3.3   Snell's law

A radio signal passing through the earth's atmosphere suffers a change in direction owing to refraction. If we consider the neutral atmosphere to be horizontally stratified and neglect the ionospheric refraction, the total bending can be found by repeatedly applying Snell's law for each layer [*Smart*, 1936]. Snell's law states (see Fig. 4.3):

Snell's law

$$\boxed{n_{i+1}\sin z_{i+1} = n_i\sin z_i} = n_i\sin\psi_i,$$
(4.7)

where z_i and z_{i+1} are the zenith angles of the arriving radio signal in the layers i and $i+1$, and n_i and n_{i+1} are the corresponding refractive indexes. By applying Snell's law for each layer, one finds:

$$n_0\sin z_0 = \sin z_m,$$
(4.8)

where the index $_0$ denotes the lowest layer and the index $_m$ denotes the highest layer, where the refractive index reduces to 1. This formula holds for any refractivity profile.
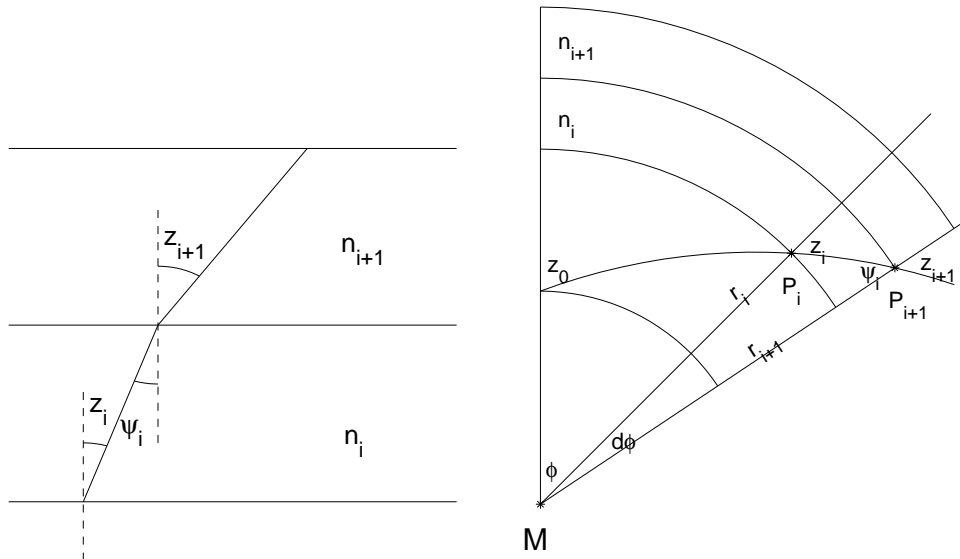
Figure 4.3   Illustration of Snell's law (left panel) and Snell's law in spherical coordinates (right panel) for horizontally stratified layers. z_i: zenith angle of layer i; n_i: refractive index of layer i.

For a spherical earth with a spherically layered atmosphere we may formulate Snell's law in spherical coordinates [*Smart*, 1936]. Application of the sine rule in the triangle MP_iP_{i+1} of Fig. 4.3 gives:

$$\frac{r_i}{\sin\psi_i} = \frac{r_{i+1}}{\sin(\pi - z_i)} = \frac{r_{i+1}}{\sin z_i},$$
(4.9)

where r_i and r_{i+1} are the distances MP_i and MP_{i+1}, with M the center of mass of the earth. Combining Eqs. (4.7) and (4.9) gives Snell's law in spherical coordinates:

$$n_{i+1}r_{i+1} \sin z_{i+1} = n_i r_i \sin z_i = n_0 r_0 \sin z_0, \tag{4.10}$$

or simply:

$$\boxed{nr \, \sin z = \text{constant.}} \tag{4.11}$$

4.4 Water vapor

The troposphere contains both dry air and water vapor. Dry air has no significant variation in composition with latitude and height [*Smith and Weintraub*, 1953]. The amount of water vapor, on the other hand, varies widely, both spatially and temporally. Most of the water vapor is contained in the boundary layer, the lowest 2 km of the troposphere.

Water also appears in the troposphere in liquid phase (fog, clouds, rain) and solid form (snow, hail, ice), and is the most important constituent in relation to weather processes, not only because of rain- and snowfall but also because large amounts of energy are released in the condensation process.

In this section, we deal with some water-vapor related measures like mixing ratio, partial pressure of water vapor, virtual temperature, and relative humidity. Most definitions follow [*Haltiner and Martin*, 1957].

4.4.1 Mixing ratio

mixing ratio

The mixture of dry air and water vapor is called moist air. A measure of moisture content is the mixing ratio, which is defined as the quotient of water-vapor mass per unit mass dry air [*Haltiner and Martin*, 1957]:

$$\boxed{w \doteq \frac{m_v}{m_d} = \frac{m_v/V}{m_d/V} = \frac{\rho_v}{\rho_d},} \tag{4.12}$$

w : mixing ratio [-];
m_v : mass of water vapor [kg];
m_d : mass of dry air [kg];
V : volume [m^3];
ρ_v : density of water vapor [kg m^{-3}];
ρ_d : density of dry air [kg m^{-3}].

equations of state

If we apply the equation of state, Eq. (4.1), for both water vapor and dry air, we obtain with use of Eq. (4.2):

$$\boxed{e = \rho_v R_v T \quad ; \quad P_d = P - e = \rho_d R_d T,} \tag{4.13}$$

e : partial pressure of water vapor [N m^{-2}];
P_d : partial pressure of dry air [N m^{-2}];
P : total pressure of (moist) air [N m^{-2}];
R_v : specific gas constant of water vapor [J kg^{-1}K^{-1}];
R_d : specific gas constant of dry air [J kg^{-1}K^{-1}].

Another expression for the mixing ratio can be obtained from Eqs. (4.12) and (4.13) as:

$$w = \frac{e/R_v T}{(P-e)/R_d T} = \epsilon \frac{e}{P-e} \approx \epsilon \frac{e}{P}, \tag{4.14}$$

$R_d = 287.06 \pm 0.01$ J kg^{-1}K^{-1};
$R_v = 461.525 \pm 0.003$ J kg^{-1}K^{-1};
$\epsilon \doteq R_d/R_v = 0.622$.

The approximation is allowed because the partial pressure of water vapor is typically about 1% of the total pressure. The constants in Eq. (4.14) are taken from [*Mendes*, 1999].

specific
humidity

An alternative way to express the humidity is the specific humidity:

$$q = \frac{\rho_v}{\rho_m} \approx w. \tag{4.15}$$

4.4.2 Virtual temperature

We can also apply the equation of state for moist air. In that case we have:

$$P = \rho_m R_m T, \tag{4.16}$$

ρ_m : density of moist air [kg m^{-3}];
R_m : specific gas constant of moist air [J kg^{-1}K^{-1}].

Since $\rho_m = \rho_v + \rho_d$, combining Eqs. (4.12), (4.13), and (4.16) gives R_m as a function of the mixing ratio:

$$R_m = \frac{R_d + wR_v}{1+w} = \frac{R_d(1+1.61w)}{1+w} \approx R_d(1+0.61w). \tag{4.17}$$

Substituting Eq. (4.17) in Eq. (4.16) gives:

$$P = \rho_m R_d T_v, \tag{4.18}$$

where

$$T_v \doteq (1+0.61w)T, \tag{4.19}$$

virtual
temperature

is the virtual temperature in kelvin. In Eq. (4.18) the fixed gas constant R_d is used instead of R_m. The presence of moisture is accounted for by the use of a fictitious temperature. The virtual temperature is the temperature that dry air at pressure P would have when its density would equal that of moist air at temperature T, pressure P, and mixing ratio w [*Haltiner and Martin*, 1957].

4.4.3 Partial pressure of saturated air

In a closed system with no air, an equilibrium will be established when equal numbers of water molecules are passing from liquid or solid to vapor, and vice versa. Under these circumstances the vapor is said to be saturated. When the vapor is mixed with air, the mixture of air and water vapor under equilibrium conditions is referred to as saturated air. When saturated air comes in contact with unsaturated air, diffusion takes place in the direction toward lower values of vapor.

saturated air

The partial pressure of saturated water vapor is a function of temperature. Larger amounts of water vapor can be contained by warmer air. By cooling saturated air, the surplus of water vapor above the saturation value at the new temperature condensates. The energy released (per unit mass) in the condensation is called latent heat. The same amount of energy is required for vaporization. We also have the latent heat of fusion, which is the amount of energy required in the change of a unit mass of ice to liquid water, and the latent heat of sublimation, which is the sum of the latent heats of vaporization and fusion.

latent heat

A relation between latent heat, partial pressure of water vapor, and temperature, is given by the Clausius–Clapeyron equation [*Haltiner and Martin*, 1957]:

Clausius–Clapeyron

$$\frac{1}{e_{sat}} \frac{de_{sat}}{dT} = \frac{L}{R_v T^2},$$ (4.20)

e_{sat} : partial pressure of saturated water vapor [N m^{-2}];
L : latent heat of fusion [$0.334 \cdot 10^6$ J kg^{-1}] or
         latent heat of vaporization [$2.500 \cdot 10^6$ J kg^{-1}] or
         latent heat of sublimation [$2.834 \cdot 10^6$ J kg^{-1}];
R_v : specific gas constant of water vapor [461.525 J kg^{-1}K^{-1}].

Integration of Eq. (4.20) gives the partial pressure of water vapor as function of temperature:

$$e_{sat} = e_{sat}(0) \exp\left[-\frac{L}{R_v}\left(\frac{1}{T} - \frac{1}{T(0)}\right)\right],$$ (4.21)

with L either the latent heat of vaporization ($T > 0°$C) or sublimation ($T < 0°$C). The integration constants are given at $0°$C:

$e_{sat}(0) = 6.11$ mbar $\equiv 611$ N m^{-2};
$T(0)$ $= 273.16$ K.

The Clausius–Clapeyron equation is a theoretical model. Slightly different values were derived by [*Baby et al.*, 1988] using a least-squares fit on laboratory tabulated e_{sat} data from [*Queney*, 1974]:

$$\boxed{e_{sat} = \exp\left[A - \frac{B}{T}\right] \quad \text{[mbar]},}$$ (4.22)

$A = 21.3195$ and $B = 5327.1157$ K for $T < T(0)$;
$A = 24.3702$ and $B = 6162.3496$ K for $T > T(0)$.

This expression is similar to Eq. (4.21) and is shown graphically in Fig. 4.4.

We may also define a saturation mixing ratio; cf. Eq. (4.14). This is given as [*Haltiner and Martin*, 1957]:

$$\boxed{w_{sat} \doteq \epsilon \frac{e_{sat}}{P - e_{sat}} \approx \epsilon \frac{e_{sat}}{P},}$$ (4.23)

w_{sat} : saturation mixing ratio [-].
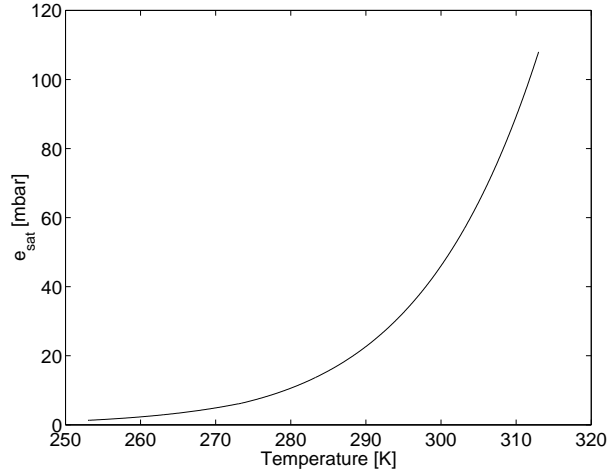
Figure 4.4 Partial pressure of saturated air according to [*Baby et al.*, 1988].

4.4.4 Relative humidity

relative
humidity

The relative humidity rh is defined as the quotient of the mixing ratio and saturation mixing ratio:

$$rh \doteq \frac{w}{w_{sat}} \approx \frac{e}{e_{sat}}, \tag{4.24}$$

rh : relative humidity [-].

The relative humidity is often multiplied by hundred to express it in percentages.

4.5 Propagation delay and refractivity

Fermat

The total delay of a radio signal caused by the neutral atmosphere depends on the refractivity along the traveled path, and the refractivity depends on pressure and temperature. The basic physical law for the propagation is Fermat's principle: Light (or any electro-magnetic wave) will follow the path between two points involving the least travel time. We define the electro-magnetic (or optical) distance between source and receiver as:

$$S \doteq \int c \, dt = \int \frac{c}{v} \, ds = \int_s n(s) \, ds, \tag{4.25}$$

S   : electro-magnetic distance [m];
s : electro-magnetic path [m];
c : speed of light in vacuum [m s^{-1}];
$v = ds/dt$: propagation speed [m s^{-1}];
$n = c/v$: refraction index [-].

In general n is considered a complex number. The imaginary part then relates to absorption whereas the real part relates to the delay and bending [*Hall et al.*, 1996]. If we denote the geometrical distance by:

$$L = \int_l dl, \tag{4.26}$$

L : geometrical distance [m];
l : geometrical path [m],

the excess path length becomes

$$D_t^s \doteq S - L = \int_s (n(s) - 1) \, ds + \{ \int_s ds - \int_l dl \}, \tag{4.27}$$

D_t^s : excess path length (Delay) in the s̲lant direction
caused by the t̲roposphere [m].

ray bending

The first term on the right-hand side is the excess path length caused by the propagation delay, whereas the second term (between braces) is the excess path length caused by ray bending. From now on we will just speak of a delay, instead of 'excess path length'. So when we speak of a delay, it is in terms of distance. In [*Mendes*, 1999] a model (called dg.v1) is given for the delay caused by the bending based on raytracing results: $a \exp(-\varepsilon/b)$, with $a = 2.256 \pm 0.0092$ m, $b = 2.072 \pm 0.0054°$, and ε the elevation angle; see Fig. 4.5.
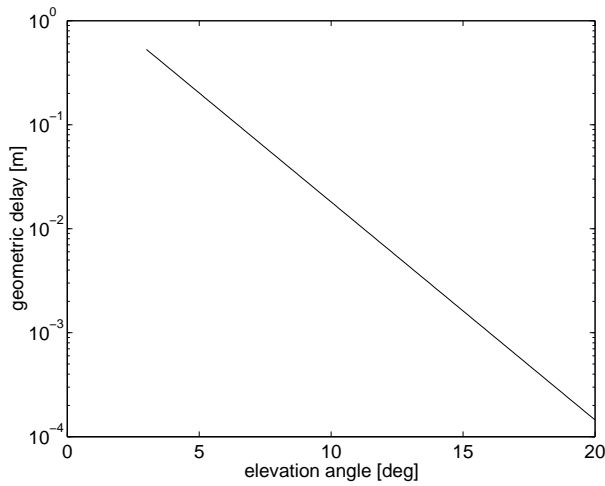
Figure 4.5   Delay caused by ray bending; model dg.v1 by [*Mendes*, 1999].

Because the value of the refractive index is close to 1, often the refractivity N (Neper) is used:

$$\boxed{N \doteq (n - 1) \cdot 10^6,} \tag{4.28}$$

N : refractivity [-].

Values for N in the atmosphere range between 0 and 300.

Since the neutral atmosphere contains both dry air and water vapor, the refractivity can also be split into a dry and a vapor part [*Smith and Weintraub*, 1953]:

$$N = N_d + N_v, \tag{4.29}$$

N_d : refractivity of dry air [-];
N_v : refractivity of water vapor [-].

This makes sense because the mixing ratio of dry-air constituents remains nearly constant in time while the water-vapor content fluctuates widely, both spatially and temporaly. According to [*Thayer*, 1974], for frequencies up to 20 GHz we may write the refractivities as a function of temperature and partial pressures:

$$\boxed{\begin{aligned} N_d &\doteq k_1 \frac{P_d}{T} Z_d^{-1}; \\ N_v &\doteq \left[k_2 \frac{e}{T} + k_3 \frac{e}{T^2} \right] Z_v^{-1}, \end{aligned}} \tag{4.30}$$

k_i : constants [K mbar^{-1} for $i = 1, 2$; K^{2}mbar^{-1} for $i = 3$];
Z_d : compressibility factor of dry air [-];
Z_v : compressibility factor of water vapor [-].

induced
molecular
polarization

permanent
dipole
moment

The constants k_i have been determined empirically. Different values are given in for example [*Bevis et al.*, 1994]; see also Table 4.1. The dry refractivity and the first term of the vapor refractivity are the result of induced molecular polarization of air and water vapor molecules respectively. The second and much larger term of the vapor refractivity represents the effects of the permanent dipole moment of the water vapor molecule [*Thayer*, 1974]; see Fig. 4.6. None of the primary constituents of dry air possess such a permanent dipole moment. The permanent dipole constitutes the main part of the delay caused by water vapor.
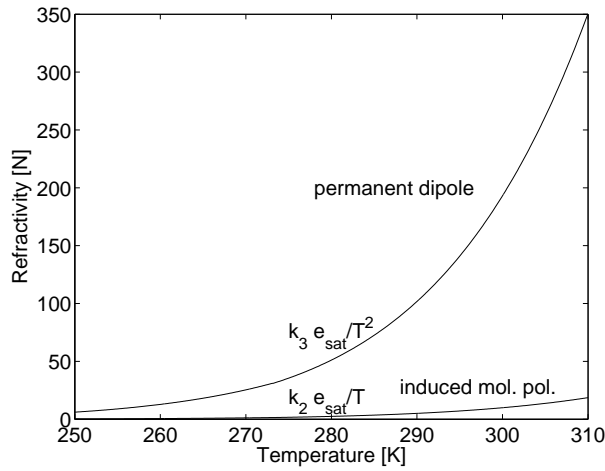
Figure 4.6   Contribution of both parts of the wet refractivity for saturation pressures; k_2 and k_3 according to [*Thayer*, 1974].

compressi-
bility

The (inverse) compressibility factors are given by the empirical formulas [*Owens*, 1967]:

$$Z_d^{-1} = 1 + P_d[57.90 \cdot 10^{-8} - 9.4581 \cdot 10^{-4} \cdot T^{-1} + 0.25844 \cdot T^{-2}];$$
$$Z_v^{-1} = 1 + e[1 + (3.7 \cdot 10^{-4})e] \cdot$$
$$[-2.37321 \cdot 10^{-3} + 2.23366 \cdot T^{-1} - 710.792 \cdot T^{-2} + 7.75141 \cdot 10^4 \cdot T^{-3}].$$

$$(4.31)$$

These factors account for the nonideal behavior of dry air and water vapor. Under normal circumstances these are close to unity.

equation of
state

With these compressibility factors the general equation of state becomes:

$$P_i = Z_i \rho_i R_i T, \tag{4.32}$$

P_i : partial pressure of the ith constituent [N m^{-2}];
ρ_i : mass density of the ith constituent [kg m^{-3}];
R_i : specific gas constant of the ith constituent [J kg^{-1}K^{-1}].

Often, dry air and water vapor are considered to be ideal gases with unit compressibility factors:

$$N_d = k_1 \frac{P_d}{T};$$
$$N_v = k_2 \frac{e}{T} + k_3 \frac{e}{T^2}. \tag{4.33}$$

hydrostatic refractivity

Instead of splitting the refractivity into a dry and a vapor part, we can also split it into a hydrostatic and a nonhydrostatic part [*Davis et al.*, 1985]. With Eq. (4.30) and the equation of state for dry air and water vapor ($i = d, v$), Eq. (4.29) becomes:

$$
\begin{aligned}
N &= k_1 R_d \rho_d + k_2 R_v \rho_v + k_3 \frac{e}{T^2} Z_v^{-1} \\
&= k_1 R_d \rho_m - k_1 R_d \rho_v + k_2 R_v \rho_v + k_3 \frac{e}{T^2} Z_v^{-1} \\
&= k_1 R_d \rho_m + (k_2 - k_1 \frac{R_d}{R_v}) R_v \rho_v + k_3 \frac{e}{T^2} Z_v^{-1},
\end{aligned}
\tag{4.34}
$$

where we used the density of moist air $\rho_m = \rho_d + \rho_v$. If we define

$$
k_2' \doteq k_2 - k_1 \frac{R_d}{R_v} = k_2 - k_1 \epsilon,
\tag{4.35}
$$

$$
k_2' : \text{constant } [\text{K mbar}^{-1}],
$$

and use the equation of state for water vapor, Eq. (4.34) becomes:

$$
\boxed{
\begin{aligned}
N &= k_1 R_d \rho_m + (k_2' \frac{e}{T} + k_3 \frac{e}{T^2}) Z_v^{-1} \\
&\doteq N_h \qquad\quad + N_w.
\end{aligned}
}
\tag{4.36}
$$

wet refractivity

The first term, N_h, is the hydrostatic refractivity. The second term, N_w, is the nonhydrostatic refractivity, but usually this is called wet refractivity, although this term is often also used for the earlier defined N_v.

Reference	k_1 [K mbar^{-1}]	k_2 [K mbar^{-1}]	k_3 $10^5[\text{K}^2\text{mbar}^{-1}]$	k_2' [K mbar^{-1}]
[Boudouris, 1963]	77.59 ± 0.08	72 ± 11	3.75 ± 0.03	24 ± 11
[Smith and Weintraub, 1953]	77.61 ± 0.01	72 ± 9	3.75 ± 0.03	24 ± 9
[Thayer, 1974]	77.60 ± 0.01	64.79 ± 0.08	3.776 ± 0.004	17 ± 10

Table 4.1 Empirical values of coefficients used in this section according to different publications; see also [*Mendes*, 1999].

4.6 Atmospheric profiles

In this section, theoretical profiles are given for the temperature, dry-air pressure, and the partial pressure of water vapor in saturated air. Based on the first two, a refractivity profile is given of dry air.

As we have seen in the previous section, the propagation delay depends on the refractivity along the ray path, and the refractivity on its turn depends on temperature and pressure. To determine the propagation delay, we therefore need information on the temperature and pressure along the ray path. Although real profiles of temperature, pressure, refractivity, and partial pressure of water vapor can only be determined by actual measurements, obtained by, for example, radiosondes, idealized or standard profiles can be given based on the temperature lapse rate and theoretical assumptions.

lapse rate

4.6.1 Temperature and pressure profile

temperature profile

Figure 4.7 shows a model temperature profile of the neutral atmosphere. The troposphere is characterized by a decreasing temperature. The measure of

decrease $\beta \doteq -dT/dh$ is called lapse rate. The lapse rate varies throughout the atmosphere, but is frequently constant in thick layers [*Haltiner and Martin*, 1957]. Some physical background on the temperature lapse rate is given in App. A.

pressure profile For the derivation of a pressure profile we need to know the temperature profile. The derivation is based on the dry-air differential equation:

$$\frac{1}{P_d}dP = -\frac{g_m}{R_d T}dh, \tag{4.37}$$

which is obtained by the equation of state Eq. (4.1), the assumption of hydrostatic equilibrium Eq. (4.5), and Eq. (4.2). We considered the gravitation to be constant

mean gravity with height and equal to a mean value

$$g_m \doteq \frac{\int_{h_0}^{\infty} \rho_m(h)g(h)dh}{\int_{h_0}^{\infty} \rho_m(h)dh}. \tag{4.38}$$

isothermal layer For isothermal layers like the tropopause, the pressure profile is found by integration of Eq. (4.37):

$$\boxed{P_d = P_{d0} \exp\left(-\frac{h - h_0}{\mathrm{H}}\right) \quad ; \quad \mathrm{H} = \frac{R_d T}{g_m},} \tag{4.39}$$

P_{d0} : pressure of dry air at the base of the layer [mbar];
h : height above msl [km];
h_0 : height above msl at the base of the layer [km];
H   : scale height [km].

scale height

In case of a polytropic layer like the troposphere and stratosphere, with the assumption of a constant lapse rate, $dh = -dT/\beta$, we integrate the right-hand side of Eq. (4.37) over dT:

$$\boxed{P_d = P_{d0}\left(\frac{T}{T_0}\right)^{\mu+1} \quad ; \quad \mu \doteq \frac{g_m}{R_d\beta} - 1,} \tag{4.40}$$

T_0 : temperature at the base of the layer [K].

standard atmosphere Troposphere delay models often use standard values for the temperature and pressure. An example of a standard atmosphere is the 1976 US Standard Atmosphere [*Stull*, 1995]:

$$
\begin{array}{lll}
\text{troposphere:} & T = 288.15 - 6.5h & 0 < h < 11 \text{ km};\\
\text{tropopause:} & T = 216.65 & 11 < h < 20 \text{ km};\\
\text{stratosphere:} & T = 216.65 + h - 20 & 20 < h < 32 \text{ km};\\
\\
\text{troposphere:} & P_d = 1013.25(288.15/T)^{-5.255877} & 0 < h < 11 \text{ km};\\
\text{tropopause:} & P_d = 226.32\exp(-0.1568(h-11)) & 11 < h < 20 \text{ km};\\
\text{stratosphere:} & P_d = 54.749(216.65/T)^{34.16319} & 20 < h < 32 \text{ km}.
\end{array}
\tag{4.41}
$$

These profiles are plotted in Fig. 4.7.
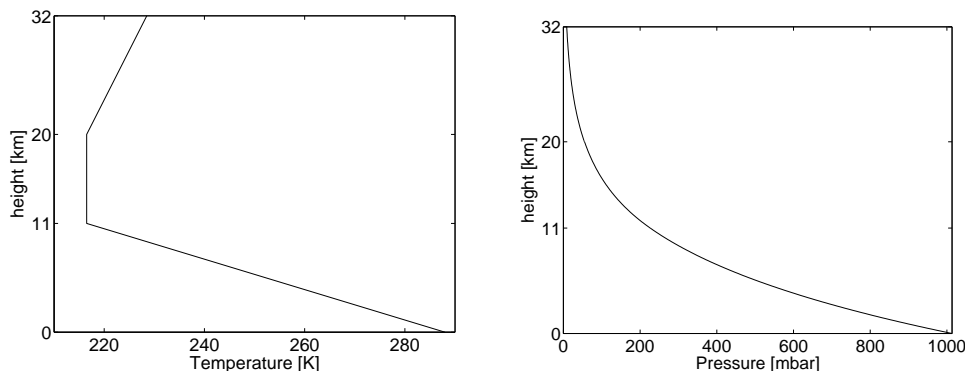
Figure 4.7   Temperature and dry-air pressure in the 1976 US Standard Atmosphere.

4.6.2   Refractivity profile of dry air

From Eqs. (4.33), (4.39), and (4.40), we can also derive the theoretical *refractivity* profiles of dry air. For polytropic layers we have:

refractivity
profile

$$\frac{N_d}{N_{d0}} = \frac{k_1 P_d/T}{k_1 P_{d0}/T_0} = \frac{P_d}{P_{d0}}\frac{T_0}{T} = \left(\frac{T}{T_0}\right)^{\mu},$$ (4.42)

N_{d0} : dry refractivity at the base of the layer [-].

In an isothermal layer $(T = T_0)$ we find:

$$\frac{N_d}{N_{d0}} = \frac{P_d}{P_{d0}} = \exp\left(-\frac{h - h_0}{\mathrm{H}}\right).$$ (4.43)

In Fig. 4.8, the refractivity profile is given for the 1976 US Standard Atmosphere (with $k_1 = 77.604$).
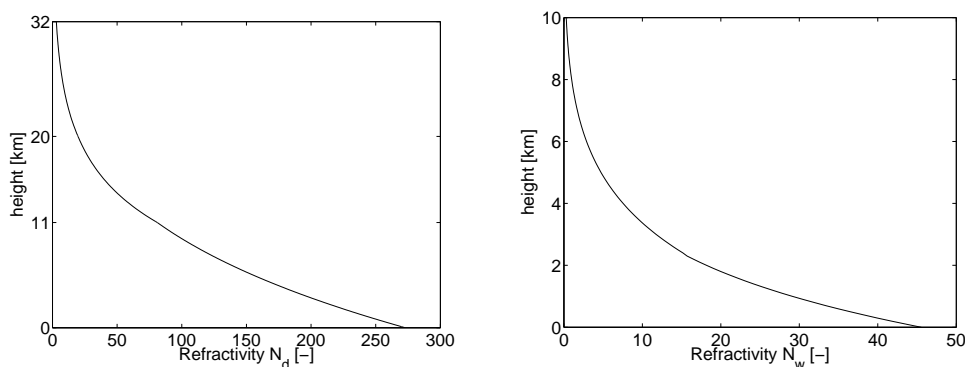
Figure 4.8   Refractivity profiles. Left: Dry-air refractivity profile of the 1976 US Standard Atmosphere. Right: Wet refractivity profile for a surface pressure of $15°$C, a constant relative humidity of 50%, a lapse rate of 6.5 K/km, and the constants of Thayer (1974).

Profiles of the hydrostatic refractivity are nearly the same as those of dry air. Instead of for the real temperature, we have to assume a profile for the lapse rate of the virtual temperature in the formulas above. The virtual temperature is only the same as the real temperature for a mixing ratio of 0.

4.6.3 Saturation pressure profile

From Model (4.22) and the assumption of a constant lapse rate, the partial pressure of saturated air e_{sat} is given as function of height by:

$$e_{sat} = \exp\left[A - \frac{B}{T_0 - \beta(h - h_0)}\right],$$
(4.44)

T_0 : surface temperature [K];
h_0 : height of the surface above msl [km].

Since hardly any water vapor is present in or above the tropopause, it is sufficient to consider only a tropospheric profile. Three different profiles are given in Fig. 4.9. Figure 4.8 shows an example of a wet refractivity profile based on Eq. (4.44).
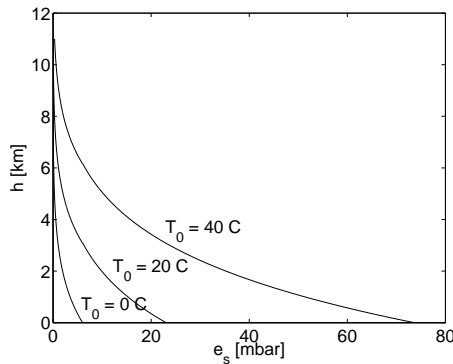
Figure 4.9   Profiles of saturated water-vapor pressure at different surface temperatures for a lapse rate of 6.5 K/km and $h_0 = 0$ km.

effective height

The thickness of the troposphere is often expressed by the effective height, a refractivity-weighted mean height, see App. C. For exponential profiles, the effective height is equal to the scale height. The effective height of the dry and wet troposphere are about 8 km and 2 km respectively. The latter value holds if the relative humidity is independent of height as in Fig. 4.8. The smaller effective height of the wet troposphere causes the wet refractivity to drop much faster with increasing altitude than the dry refractivity. The contribution of the wet part to the total refractivity is therefore at higher altitudes smaller than at the surface.

Zenith-delay models

5.1 Introduction

In this chapter we describe models for the zenith delays. In Sect. 5.2, we give an expression of the Zenith Hydrostatic Delay and Sect. 5.3 shows two models for the Zenith Wet Delay. Finally, in Sect. 5.4, the Zenith Total Delay is given as proposed by Saastamoinen. This model is still frequently used.

5.2 Zenith Hydrostatic Delay

ZHD

The Zenith Hydrostatic Delay (ZHD) is the delay of a radio signal arriving from the zenith caused by the neutral hydrostatic atmosphere. This section gives a derivation of an expression for the ZHD based on surface pressure, latitude, and height above msl.

The ZHD is defined as the integral over the hydrostatic refractivity:

$$\boxed{D_h^z \doteq 10^{-6} \int_{h_0}^{\infty} N_h(h) \, dh,} \tag{5.1}$$

D_h^z : Zenith Hydrostatic Delay [m];
h : height above msl [m];
h_0 : height of the surface above msl [m].

With Eq. (4.36), the ZHD can be written as:

$$D_h^z = 10^{-6} k_1 R_d \int_{h_0}^{\infty} \rho_m(h) dh. \tag{5.2}$$

Under the condition of hydrostatic equilibrium we can use the hydrostatic equation:

$$\frac{dP}{dh} = -\rho_m(h)g(h), \tag{5.3}$$

where $g(h)$ is the acceleration due to gravity as a function of height. Integrating Eq. (5.3) yields

$$\int_{P_0}^{0} dP = -\int_{h_0}^{\infty} \rho_m(h)g(h) dh = -P_0, \tag{5.4}$$

where P_0 is the surface value of the air pressure. Using a mean gravity as in Eq. (4.38), with Eqs. (5.4) and (5.2) the ZHD becomes:

$$\boxed{D_h^z = 10^{-6} k_1 \frac{R_d}{g_m} P_0.} \tag{5.5}$$

The mean gravity depends on height and latitude. Based on an approximation by [*Saastamoinen*, 1972], the mean gravity is given as [*Davis et al.*, 1985]:

$$g_m \doteq g_m^0 \cdot f(\phi, h), \qquad (5.6)$$

g_m : acceleration due to gravity at the center
of mass of the vertical column [m s^{-2}];
ϕ : geodetic latitude [rad],

where

$$f(\phi, h) = 1 - 0.00266 \cos 2\phi - 0.00000028\, h \approx 1 \qquad (5.7)$$

is a weak function of ϕ and h.

In [*Davis et al.*, 1985] the following constants are used:

$k_1 = 77.604$ K mbar^{-1};
$R = 8.31434$ kg m^{2}s^{-2}mol^{-1}K^{-1};
$M_d = 28.9644 \cdot 10^{-3}$ kg mol^{-1};
$g_m^0 = 9.784$ m s^{-2},

to derive the expression:

$$D_h^z = [0.0022768 \text{ m/mbar}] \cdot \frac{P_0}{f(\phi, h)}. \qquad (5.8)$$

accuracy

So, for any location on earth for which the surface pressure is given, one can compute the ZHD with Eq. (5.8). No assumptions were made about temperature or air-pressure profiles. The accuracy of the constant between brackets is given in [*Davis et al.*, 1985] as 0.5 mm/bar. The expression of [*Davis et al.*, 1985] is slightly more accurate as the one given in [*Saastamoinen*, 1972]. In [*Elgered et al.*, 1991], another slightly different constant is given, based on the value of $k_1 = 77.64 \pm 0.08$ K/mbar by [*Boudouris*, 1963]:

$$D_h^z = [0.0022779 \text{ m/mbar}] \cdot \frac{P_0}{f(\phi, h)}. \qquad (5.9)$$

They state that although this value is less precise (2.4 mm/bar), it does not rely on assumptions about the frequency dependence of the dry refractivity.

5.3   Zenith Wet Delay

ZWD

In this section, two models are given for the Zenith Wet Delay (ZWD) as derived in [*Askne and Nordius*, 1987] and [*Baby et al.*, 1988]. Both models need empirical parameters and are based on the defining integrals of Sect. 5.3.1.

5.3.1   Zenith Wet Delay and Integrated Water Vapor

The ZWD is defined as:

$$D_w^z \doteq 10^{-6} \int_{h_0}^{\infty} N_w(h)\, dh, \qquad (5.10)$$

in which N_w is (see Sect. 4.5):

$$N_w \doteq k_2' \frac{e}{T} + k_3 \frac{e}{T^2}, \tag{5.11}$$

mean temperature assuming unit compressibility. A mean temperature of water vapor can be defined as:

$$T_m \doteq \frac{\int_{h_0}^{\infty} e/T \ dh}{\int_{h_0}^{\infty} e/T^2 \ dh}. \tag{5.12}$$

Using Eqs. (5.11) and (5.12) in Eq. (5.10) then gives another expression for the ZWD:

$$\boxed{D_w^z = 10^{-6}(k_2' + k_3/T_m) \int_{h_0}^{\infty} \frac{e}{T} \ dh.} \tag{5.13}$$

One needs expressions for water-vapor pressure e and temperature T as functions of height to determine the ZWD.

IWV For meteorological applications, the ZWD is sometimes expressed in terms of Integrated Water Vapor [*Bevis et al.*, 1992]

$$IWV \doteq \int_{h_0}^{\infty} \rho_v \ dh \tag{5.14}$$

or Integrated Precipitable Water Vapor

$$IPWV \doteq IWV/\rho_w, \tag{5.15}$$

with $\rho_w = 1000$ kg m^{-3} the density of water. Using the equation of state Eq. (4.13), we have:

$$\int_{h_0}^{\infty} \frac{e}{T} \ dh = R_v \int_{h_0}^{\infty} \rho_v \ dh. \tag{5.16}$$

So, if we define a factor Q as

$$Q \doteq 10^{-6}(k_2' + k_3/T_m)R_v\rho_w, \tag{5.17}$$

the ZWD can be expressed as:

$$\boxed{D_w^z = Q \cdot IPWV.} \tag{5.18}$$

The dimensionless quantity Q is about 6.5, but varies spatially and temporally. In [*Bevis et al.*, 1992] an empirical formula is given for the mean temperature as function of the surface temperature, $T_m \approx 70.2 + 0.72 \ T_0$, with which Q can be determined to about 2% accuracy. Other empirical formulas for T_m were derived by [*Mendes et al.*, 2000]. Empirical formulas for Q are given in [*Emardson and Derks*, 1998].

5.3.2 The ZWD model of Askne and Nordius

Askne and Nordius (1987) developed a ZWD model in which they assumed a constant lapse rate. They used a model for the partial pressure of water vapor of [*Smith*, 1966]:

$$e = e_0 \left(\frac{P}{P_0} \right)^{\lambda+1}, \tag{5.19}$$

λ : constant [-].

In this model the water-vapor pressure decreases much faster with increasing height than the total pressure. This should be the case because the decreasing temperature also causes the saturation pressure to drop. The constant λ varies with season and location and typically ranges between 1 and 5. Mean values of λ for 47 stations, mainly in the northern hemisphere, are given in [*Ifadis*, 1993].

With Eqs. (5.13) and (5.19) the ZWD is expressed as:

$$D_w^z = 10^{-6}(k_2' + k_3/T_m)\frac{e_0}{P_0}\int_{h_0}^{\infty}\left(\frac{P}{P_0}\right)^{\lambda}\frac{P}{T}dh. \tag{5.20}$$

When we use the mean gravity as defined in Eq. (4.38), the differential Eq. (4.37) may be written as:

$$\frac{P}{T}dh = -\frac{R_d}{g_m}dP. \tag{5.21}$$

Substitution of Eq. (5.21) in Eq. (5.20) and evaluating the integral then gives the model for the ZWD as a function of mean temperature T_m and surface pressure e_0:

$$\boxed{D_w^z = 10^{-6}(k_2' + k_3/T_m)\frac{R_d}{(\lambda + 1)g_m}e_0.} \tag{5.22}$$

From the assumption of a constant lapse rate for the troposphere we have the relationship of Eq. (4.40), which we will write as (the surface is the base of the layer):

$$\left(\frac{P}{P_0}\right) = \left(\frac{T}{T_0}\right)^{\mu+1}. \tag{5.23}$$

Substitution of Eq. (5.23) into Eq. (5.19) gives:

$$e = e_0\left(\frac{T}{T_0}\right)^{(\lambda+1)(\mu+1)}. \tag{5.24}$$

With Eq. (5.24) substituted in Eq. (5.12) one finds an expression for the mean temperature after integration:

$$\boxed{T_m = T_0\left[1 - \frac{1}{(\lambda + 1)(\mu + 1)}\right],} \tag{5.25}$$

which can be used in Eq. (5.22) to determine the ZWD.

5.3.3 The ZWD model of Baby et al.

Another semi-empirical model was derived in [*Baby et al.*, 1988]. For tropospheric temperatures in the range of 230 to 310 K the factor $k_2' + k_3/T$ does not vary more than 0.1%. Therefore they replaced this expression by the simpler expression k_3'/T such that:

$$k_3'/T \doteq k_2' + k_3/T \quad \text{for} \quad T \doteq 273 \text{ K}, \tag{5.26}$$

$$k_3' = 3.81 \cdot 10^5 \text{ K}^2/\text{mbar}.$$

The ZWD is thus written as:

$$D_w^z = 10^{-6} k_3' \int_{h_0}^{\infty} \frac{e}{T^2} \, dh. \tag{5.27}$$

In [*Baby et al.*, 1988], a constant lapse rate, $T(h) = T_0 - \beta(h - h_0)$, and a constant relative humidity ($e = rh_0 \cdot e_{sat}$) is used. In combination with Eq. (4.22) this leads to:

$$\boxed{D_w^z = 10^{-6} k_3' \frac{rh_0}{\beta} \frac{e_{sat}(T_0)}{B_0},} \tag{5.28}$$

where B_0 is the value of B in Eq. (4.22) corresponding to the surface temperature. The model given in [*Baby et al.*, 1988] was slightly more complicated because they assumed a maximum height of 11 km to which water vapor was present, but the difference with the above given model is small. A more general form for Eq. (5.28) is:

$$D_w^z = a(T_0)rh_0, \tag{5.29}$$

where $a(T_0)$ is a slope factor that can be determined theoretically as in Eq. (5.28) or it can be derived from radio sounding profiles from all over the world and covering as long a period as possible. In [*Baby et al.*, 1988] a database of a one-year campaign carried out in 1979 was used to derive a global semi-empirical model. The general formulation for the model is:

$$\boxed{D_w^z = \nu 10^{\gamma(T_0 - 273.16)} rh_0,} \tag{5.30}$$

where ν and γ are empirical coefficients depending on various factors such as latitude, climate, time, etc. Some mean values are given in Table 5.1.

	ν [mm]	γ [K^{-1}]
global model	72.84	0.0236
continental	73.27	0.0235
oceanic	54.88	0.0291
equatorial	65.47	0.0273

Table 5.1   Empirical coefficients in the ZWD model of [*Baby et al., 1988*].

accuracy

An advantage of the semi-empirical model in [*Baby et al.*, 1988] over the model in [*Askne and Nordius*, 1987] is that no parameter fits are needed for each observing station. Both models, however, give an accuracy of only about 2 to 5 cm at zenith.

5.4   The Zenith Tropospheric Delay of Saastamoinen

ZTD

The total Zenith Tropospheric Delay (ZTD) is the sum of the ZHD and the ZWD:

$$D_t^z \doteq \int_{h_0}^{\infty} N(h) \, dh = \int_{h_0}^{\infty} N_h(h) + N_w(h) \, dh = D_h^z + D_w^z. \tag{5.31}$$

By using the Models (5.5) and (5.22), a general formulation for the ZTD is found as:

$$D_t^z = 10^{-6} k_1 \frac{R_d}{g_m} \left[P_0 + \left(\frac{k_3}{k_1(\lambda + 1 - \beta R_d/g_m)T_0} + \frac{k_2'}{k_1(\lambda + 1)} \right) e_0 \right]. \tag{5.32}$$

In [*Saastamoinen*, 1972] the following constants were used:

$k_1 = 77.624$ K mbar^{-1}; $R_d = 287.04$ m^{2}s^{-2}K^{-1};
$k_2 = 64.7$ K mbar^{-1}; $g_m = 9.784$ m s^{-2};
$k_2' = k_2 - 0.622 k_1$ K mbar^{-1}; β $= 0.0062$ K m^{-1};
$k_3 = 371900$ K^{2}mbar^{-1}; λ $= 3$.

With these values substituted in Eq. (5.32), the ZTD model of Saastamoinen reads:

$$D_t^z = 0.002277 \left[P_0 + \left(\frac{1255}{T_0} + 0.05 \right) e_0 \right], \tag{5.33}$$

where P_0 and e_0 are in [mbar], T_0 is in [K], and D_t^z is in [m]. The constants k_i, $i = 1, 2, 3$, as used by Saastamoinen, were given in [*Essen and Froome*, 1951] and were adopted by the International Association of Geodesy in 1963 [*IAG*, 1963]. Model (5.32) is a two-parameter model. Both β and λ are place and time dependent. The model given by Saastamoinen uses mean values.

Chapter 6

Slant-delay models

6.1 Introduction

The slant tropospheric delay is often modeled as the product of the delay in the zenith direction and a zenith angle dependent mapping function. Section 6.2 describes a basic integral expression for the slant delay, Sect. 6.3 describes the slant delay model by Saastamoinen, and Sect. 6.4 gives a derivation of continued fractions based mapping functions. Finally, Sect. 6.5 shows examples of these mapping functions based on atmospheric parameters, and Sect. 6.6 describes the computation of slant delays from Numerical Weather Prediction (NWP) models.

6.2 Slant Tropospheric Delay

So far, only zenith delay models have been considered. In general, however, a radio signal will not arrive from a zenith direction, but from slant directions. If we neglect the delay caused by the ray bending and assume a horizontally layered atmosphere (where the refractivity n is only a function of height), the expression for the Slant Tropospheric Delay (STD) can be found by Eqs. (4.27) and (4.28) as:

$$D_t^s = 10^{-6} \int_0^\infty N ds = 10^{-6} \int_{r_0}^\infty N \frac{ds}{dr} dr. \tag{6.1}$$

As can be seen in Fig. 6.1, for an infinitesimal thin layer we have:

$$\frac{ds}{dr} = \frac{1}{\cos z} \equiv \sec z, \tag{6.2}$$

where dr is the difference in radius (distance to the center of the earth) of the two layers and z is the zenith angle at an arbitrary layer. Because of the curvature of the atmosphere, this zenith angle changes along the ray.
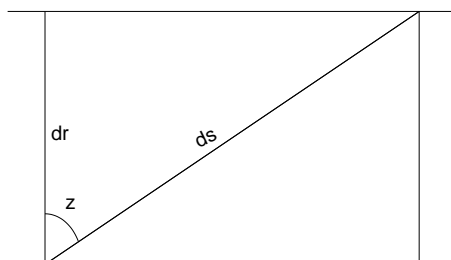
Figure 6.1 Relation between radius difference dr, distance difference ds, and zenith angle z.

Application of Snell's law in spherical coordinates (Sect. 4.3.3) gives:

$$\left.\begin{array}{l} n^2 r^2 = n^2 r^2 \cos^2 z + n^2 r^2 \sin^2 z \\ \quad\ = n^2 r^2 \cos^2 z + n_0^2 r_0^2 \sin^2 z_0 \end{array}\right\} \Rightarrow \sec z = \frac{nr}{n_0 r_0} \left[\left(\frac{nr}{n_0 r_0} \right)^2 - \sin^2 z_0 \right]^{-\frac{1}{2}}, \quad (6.3)$$

where the zero index refers to the bottom layer. Approximating $(nr)/(n_0 r_0)$ by r/r_0 and combining Eqs. (6.1–6.3), gives an integral expression for the STD:

$$D_t^s = 10^{-6} \int_{r_0}^{\infty} N(r) \frac{r}{r_0} \left[\left(\frac{r}{r_0} \right)^2 - \sin^2 z_0 \right]^{-\frac{1}{2}} dr. \tag{6.4}$$

Considering the geometry of Fig. 6.2, the same result could have been derived by:

$$s = (r^2 - r_0^2 \sin^2 z_0)^{\frac{1}{2}} - r_0 \cos z_0, \tag{6.5}$$

and

$$\frac{ds}{dr} = r(r^2 - r_0^2 \sin^2 z_0)^{-\frac{1}{2}}. \tag{6.6}$$

In general, the integral of Eq. (6.4) can be evaluated in closed form. In [*Yionoulis*, 1970] was noticed that for the refractivity profiles given in the two-quartic model of Hopfield [*Hopfield*, 1969], rounding errors were significant for high elevations. Nowadays, for any given refractivity profile, the integral of Eq. (6.4) can easily be computed by computer algebra systems. In doing so however, there result numerous large terms. Since these terms are much

numerical
instability

larger than the total integrated value, Eq. (6.4) suffers from a numerical instability.

Taylor

continued
fractions

To deal with this problem, two different approaches are used to model the STD. The first approach is to form a Taylor expansion of $\sec z(r)$ and then integrate the different terms. The other approach uses continued fractions. The Saastamoinen and some Hopfield-related models belong to the first category. The Saastamoinen models are given in [*Saastamoinen*, 1972] and in [*Saastamoinen*, 1973]. Hopfield-related models are given in [*Yionoulis*, 1970], [*Morduch*, 1978], and [*Goad and Goodman*, 1974]. Because most of the modern models use continued fractions, no extensive discussion on the latter type of models will be given here. As an illustration of the first approach, the widely used Saastamoinen model [*Saastamoinen*, 1972] is dealt with in the next section.

6.3   The Saastamoinen model

The ZTD model of Saastamoinen is given in Sect. 5.4. Here we give the STD model as shown in [*Saastamoinen*, 1972]. The derivation starts with a truncated Taylor expansion of $\sec z$:

$$\sec z = \sec z_0 + \sec z_0 \tan z_0 \Delta z. \tag{6.7}$$

Figure 6.2 shows that $\Delta z \doteq z - z_0 = -\phi$ and $\tan z = r_0 \phi / (r - r_0)$. So by approximating $\tan z \approx \tan z_0$, Eq. (6.7) becomes:

$$\sec z = \sec z_0 (1 - \tan^2 z_0 \frac{r - r_0}{r_0}). \tag{6.8}$$
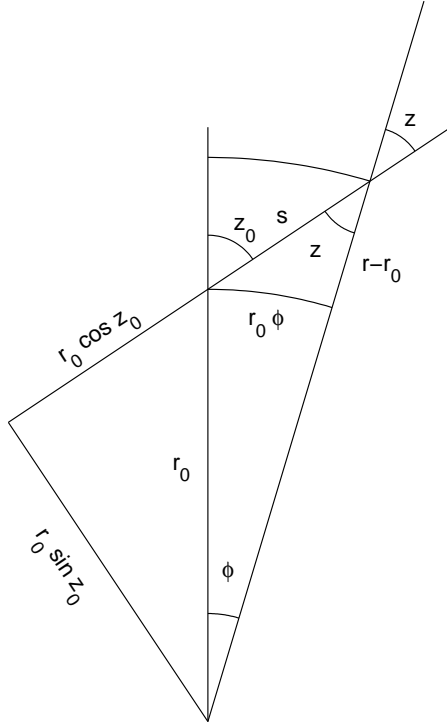
Figure 6.2   Geometry of a ray arriving through a spherical atmosphere. s: traveled distance through atmosphere; r_0: radius of the earth; z: zenith angle at the top of the atmosphere; z_0: zenith angle at the surface.

With this expression the STD reads:

$$
\begin{aligned}
D_t^s &= 10^{-6} \int_{r_0}^{\infty} N \sec z \, dr \\
&= 10^{-6} \sec z_0 \left[\int_{r_0}^{\infty} N dr - r_0^{-1} \tan^2 z_0 \int_{r_0}^{\infty} N(r - r_0) dr \right].
\end{aligned}
\tag{6.9}
$$

The first term between the brackets is the zenith delay. The second term is a correction term of which the integral part can be subdivided into three sub-integrals:

$$
\begin{aligned}
\int_{r_0}^{\infty} N \cdot (r - r_0) dr &= \int_{r_0}^{r_T} N \cdot (r - r_0) dr \\
&+ \int_{r_T}^{\infty} N \cdot (r - r_T) dr \\
&+ (r_T - r_0) \int_{r_T}^{\infty} N dr,
\end{aligned}
\tag{6.10}
$$

where r_T is the radius of the tropopause. Saastamoinen assumed the neutral atmosphere to consist of only two layers: the troposphere and the stratosphere. In this model the troposphere is a polytropic layer reaching up to r_T and the stratosphere is an isothermal layer, which for practical integration can be considered infinitely high. Each of the three integrals can be evaluated on the basis of the refractivity profiles associated with the temperature profile. Appendix B shows that the evaluation of these integrals results in:

$$
\boxed{D_t^s = 10^{-6} \underbrace{k_1 \frac{R_d}{g_m}}_{0.002277} \sec z_0 \left[P_0 + (\frac{1255}{T_0} + 0.05) e_0 - B(r) \tan^2 z_0 \right],}
\tag{6.11}
$$

where

$$
B(r) \doteq \frac{1}{r_0} \frac{g_m}{R_d} \frac{1}{k_1} \int_{r_0}^{\infty} N(r - r_0) dr.
\tag{6.12}
$$

Sometimes an additional term is used to account for the delay caused by the ray bending; tables are given in [*Saastamoinen*, 1973]. Tabular values for the correction term $B(r)$ are given in [*Saastamoinen*, 1972]. Saastamoinen does not mention the exact theoretical standard atmosphere he uses to find the tabular values of $B(r)$. However, the standard values at msl as also later used in the 1976 US Standard Atmosphere ($T_{msl} = 288.15$ K, $P_{msl} = 1013.25$ mbar), as well as the values $r_0 = 6360$ km and $h_T = 15$ km, fit quite well.

The Saastamoinen model was developed for high elevation angles. For low elevation angles the increment Δz becomes too large. The omission of higher-order terms causes an error of about 0.1 m in the range correction at a maximum zenith angle of 80° [*Saastamoinen*, 1972].

6.4   Continued fractions and mapping functions

In a parallel development to Saastamoinen, expressions for tropospheric corrections were derived and published in [*Marini*, 1972]. Marini was the first one who came up with the idea to use continued fractions. Continued fractions have the advantage over models with Taylor expansions like the Saastamoinen model that they fit for nearly the whole range of zenith angles. Starting point of the derivation is the approximation of Eq. (6.4) by

$$\boxed{D^s \approx 10^{-6} \int_{r_0}^{\infty} N(r) \left[\left(\frac{r}{r_0} \right)^2 - \sin^2 z_0 \right]^{-\frac{1}{2}} dr} \tag{6.13}$$

and the assumption of an exponential profile:

$$N(r) = N_0 \exp \left(-\frac{r - r_0}{\mathrm{H}} \right). \tag{6.14}$$

The slant delay may be either the hydrostatic or wet delay. The assumption of an exponential refractivity profile is justified for the hydrostatic refractivity. The effective height H is then close to the scale height as shown in App. C. For the wet refractivity profile this assumption is weaker.

The zenith delay corresponding to the exponential profile follows from integration as the two-parameter expression:

$$D^z = 10^{-6} N_0 \mathrm{H}. \tag{6.15}$$

By reparameterizing

$$u \doteq \frac{r - r_0}{\mathrm{H}} \quad ; \quad du = \frac{1}{\mathrm{H}} dr \quad ; \quad k \doteq 2\frac{\mathrm{H}}{r_0}, \tag{6.16}$$

and approximating

$$\left(\frac{r}{r_0} \right)^2 \approx 1 + 2\frac{r - r_0}{r_0} = 1 + ku, \tag{6.17}$$

the slant delay can be rewritten as:

$$D^s = 10^{-6} N_0 \mathrm{H} \int_0^{\infty} \exp(-u)[\cos^2 z_0 + ku]^{-\frac{1}{2}} du. \tag{6.18}$$

Now it is convenient to write the slant delay as the product of the zenith delay and a mapping function. The mapping function is thus defined as:

mapping function

$$M(z_0) \doteq \frac{D^s(z_0)}{D^z}.$$

(6.19)

Further reparameterizing

$$\alpha \doteq \cos z_0 \quad ; \quad x \doteq k \cdot \alpha^{-2},$$

(6.20)

then gives an integral expression of the mapping function:

$$M(\alpha) = \alpha^{-1} \int_0^\infty \exp(-u)[1 + xu]^{-\frac{1}{2}} du.$$

(6.21)

To evaluate this integral, a computational trick is used, as given in [*Wall*, 1948] at page 352, that starts by writing this integral as

$$M(\alpha) = \frac{\alpha^{-1}}{\Gamma(a)} \int_0^\infty \exp(-u)u^{a-1}[1 + xu]^{-b} du \quad ; \quad a = 1, b = \frac{1}{2},$$

(6.22)

where

$$\Gamma(a) \doteq \int_0^\infty \exp(-u)u^{a-1} du = (a-1) \cdot \Gamma(a-1) \stackrel{a \in \mathbb{N}^+}{=} (a-1)!$$

(6.23)

Gamma function

is the well-known Gamma function. Expanding the denominator of Eq. (6.22) and integrating each term using Eq. (6.23) leads to:

$$\begin{aligned}
M(\alpha) &= \frac{\alpha^{-1}}{\Gamma(a)} \int_0^\infty [1 - bux + b(b+1)\frac{u^2 x^2}{2!} - b(b+1)(b+2)\frac{u^3 x^3}{3!} + ...] \cdot \\
&\quad \cdot \exp(-u)u^{a-1} du \\
&= \frac{\alpha^{-1}}{\Gamma(a)}[\Gamma(a) - \Gamma(a+1)bx + \Gamma(a+2)b(b+1)\frac{x^2}{2!} - ...] \\
&= \alpha^{-1}[1 - abx + a(a+1)b(b+1)\frac{x^2}{2!} - ...] \\
&\doteq \alpha^{-1}P(a,b).
\end{aligned}$$

(6.24)

By comparing term by term, one may find the following relation:

$$P(a,b) = P(a, b+1) + axP(a+1, b+1).$$

(6.25)

This relation holds for any a and b. If we choose $a \leftarrow b$ and $b \leftarrow a$ ('$\leftarrow$' denotes assignment), we also have the relation:

$$P(b,a) = P(b, a+1) + bxP(b+1, a+1).$$

(6.26)

Here we may assign $b \leftarrow b+1$ and apply the obvious relation $P(p_1, p_2) = P(p_2, p_1)$ for all three terms:

$$P(a, b+1) = P(a+1, b+1) + (b+1)xP(a+1, b+2).$$

(6.27)

Now, after defining

$$\begin{aligned}
P_{2n} &\doteq P(a+n, b+n); \\
P_{2n+1} &\doteq P(a+n, b+n+1), n = 0, 1, 2, ...,
\end{aligned}$$

(6.28)

and assigning $a \leftarrow a + n$ and $b \leftarrow b + n$ in Eqs. (6.25) and (6.27), more general relations can be found:

$$
\begin{aligned}
P_{2n} &= P_{2n+1} + (a+n)xP_{2n+2}; \\
P_{2n+1} &= P_{2n+2} + (b+n)xP_{2n+3}.
\end{aligned}
\tag{6.29}
$$

For $n = 0$, after multiplying with α and back substitution of $k/\alpha^2 = x$, the first relation of Eq. (6.29) becomes:

$$
\alpha P_0 = \alpha P_1 + akP_2/\alpha \Rightarrow \alpha \frac{P_0}{P_1} = \alpha + \frac{ak}{\alpha P_1/P_2}.
\tag{6.30}
$$

And similarly for the second relation of Eq. (6.29) we have:

$$
\alpha \frac{P_1}{P_2} = \alpha + \frac{bk}{\alpha P_2/P_3}.
\tag{6.31}
$$

Returning to the mapping function of Eq. (6.24), with $a = 1$ and $b = 1/2$ we can apply these relations to find the continued-fractions expression:

$$
\begin{aligned}
M(\alpha) &= \alpha^{-1}P(1, \frac{1}{2}) = \alpha^{-1}P(\frac{1}{2}, 1) = \alpha^{-1}\frac{P(\frac{1}{2}, 1)}{P(\frac{1}{2}, 0)} \doteq \frac{P_1}{\alpha P_0} \mid_{a=\frac{1}{2}, b=1} \\
&= \frac{1}{\alpha P_0/P_1} = \cfrac{1}{\alpha + \cfrac{ak}{\alpha P_1/P_2}} = \cfrac{1}{\alpha + \cfrac{ak}{\alpha + \cfrac{bk}{\alpha P_2/P_3}}}.
\end{aligned}
\tag{6.32}
$$

This fraction continues in the term $\alpha P_2/P_3$ by applying the relations of Eq. (6.29) with $n = 1, 2, ...$ By back substitution of $\cos z_0 = \alpha$, we finally find a general expression for the mapping function as also found in [*Marini*, 1972]:

$$
\boxed{M(z_0) = \cfrac{1}{\cos z_0 + \cfrac{a}{\cos z_0 + \cfrac{b}{\cos z_0 + \cfrac{c}{...}}}},}
\tag{6.33}
$$

with a, b, and c newly defined constants. Theoretically these would have to be $a \leftarrow ak = \mathrm{H}/r_0, b \leftarrow bk = 2\mathrm{H}/r_0, c \leftarrow (a+1)k = 3\mathrm{H}/r_0....$ Equation (6.33) may also be written in the convenient notation:

$$
M(z_0) = \frac{1}{\cos z_0 +} \ \frac{a}{\cos z_0 +} \ \frac{b}{\cos z_0 +} \ \frac{c}{...}.
\tag{6.34}
$$

One of the disadvantages of this mapping function is that it is not equal to unity in the zenith direction. This is an immediate result of the approximation of Eq. (6.13). To overcome this problem, in [*Chao*, 1974] the following type of mapping function was suggested:

Chao mapping function

$$
M(z_0) = \frac{1}{\cos z_0 +} \ \frac{a}{\cot z_0 + b}.
\tag{6.35}
$$

The wet mapping function has constants $a = 0.00035$ and $b = 0.017$.

Davis mapping function

For elevations in the range of $20° - 60°$, $\cot z_0$ does not approach $\cos z_0$ quickly enough, even if the fraction is extended with another term as in [*Davis et al.*, 1985]:

$$
M(z_0) = \frac{1}{\cos z_0 +} \ \frac{a}{\cot z_0 +} \ \frac{b}{\cos z_0 + c}.
\tag{6.36}
$$
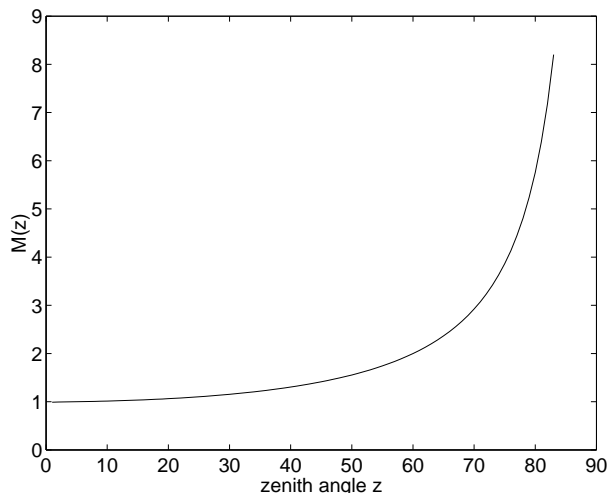
Figure 6.3   Example of a mapping function: Chao's wet mapping function.

Mapping
functions of
Ifadis,
Herring,
Niell

The most recent mapping functions are those of Ifadis (1992), Herring (1992), and Niell (1996). They use the following type of the mapping function:

$$M(z_0) = \frac{1 + a/(1 + b/(1 + c))}{\cos z_0 + \dfrac{a}{\cos z_0 + \dfrac{b}{\cos z_0 + c}}}.$$

(6.37)

Apart from different representations, the mapping functions may differ in parameterization of the coefficients a, b, and c. Since these parameters depend on the quotient of the scale height and the radius of the earth (the constant k), best fits are to be found from either standard profiles or profiles from radiosonde data. From now on the index $_0$ of the zenith angle will be omitted.

6.5   Modern troposphere mapping functions

Most of the recently developed troposphere delay models use mapping functions in the form of continued fractions. The model of [*Saastamoinen*, 1972] does not use a mapping function in the same sense as the models with continued fractions. Only after the development of models that do make use of continued fractions, the term mapping function has become current. Most of the modern models have separated mapping functions for the hydrostatic and the wet part. The total slant delay thus becomes:

total slant
delay

$$D_t^s = M_h(z)D_h^z + M_w(z)D_w^z.$$

(6.38)

In this section four models with modern mapping functions are given [*Davis et al.*, 1985], [*Ifadis*, 1992], [*Herring*, 1992], [*Niell*, 1996]. Davis' model is limited to a hydrostatic mapping function. Those of Ifadis, Herring, and Niell each include a hydrostatic and a wet mapping function. The models of Davis, Herring, and Niell are named CfA-2.2, MTT, and NMF respectively.

6.5.1   The CfA-2.2 mapping function

CfA-2.2

The CfA-2.2 mapping function [*Davis et al.*, 1985] was designed to achieve sub-centimeter accuracy at $5°$ elevation. The parameters a, b, and c in Eq. (6.36) are determined by a least-squares fit on 57 raytrace analyses for various values of a limited number of atmospheric conditions. Their raytrace algorithm was

based on a spherically symmetric, layered atmosphere. The temperature profile was taken to have a linear dependence with height up to the tropopause above which the temperature was assumed constant. Other assumptions are: hydrostatic equilibrium, constant relative humidity up to 11 km and zero above that height, and an acceleration due to gravity that is independent of height. The constants in the CfA-2.2 mapping function read:

$$
\begin{aligned}
a = {} & 0.001185[1 + 0.6071 \cdot 10^{-4}(P_0 - 1000) - 0.1471 \cdot 10^{-3}e_0 \\
& + 0.3072 \cdot 10^{-2}(T_0 - 20) + 0.1965 \cdot 10^{-1}(\beta - 6.5) \\
& - 0.5645 \cdot 10^{-2}(h_T - 11.231)]; \\
b = {} & 0.001144[1 + 0.1164 \cdot 10^{-4}(P_0 - 1000) + 0.2795 \cdot 10^{-3}e_0 \\
& + 0.3109 \cdot 10^{-2}(T_0 - 20) + 0.3038 \cdot 10^{-1}(\beta - 6.5) \\
& - 0.1217 \cdot 10^{-1}(h_T - 11.231)]; \\
c = {} & -0.0090,
\end{aligned}
\tag{6.39}
$$

with:

P_0 : total surface pressure [mbar];
e_0 : surface partial pressure of water vapor [mbar];
T_0 : surface temperature [°C];
β : temperature lapse rate [K km^{-1}];
h_T : height of the tropopause [km].

6.5.2   The Ifadis mapping functions

In [*Ifadis*, 1992] a hydrostatic mapping function is presented that can be used down to 2° elevation. Ifadis raytraced real weather profiles from various sites distributed over large areas of the world and with different climatic conditions. For most of the stations, Ifadis had a three-years database in the period 1978–1984. The global hydrostatic mapping function derived from raytracing these data has the following constants:

$$
\begin{aligned}
a = {} & 0.1237 \cdot 10^{-2} + 0.1316 \cdot 10^{-6}(P_0 - 1000) + 0.8057 \cdot 10^{-5}\sqrt{e_0} \\
& + 0.1378 \cdot 10^{-5}(T_0 - 15); \\
b = {} & 0.3333 \cdot 10^{-2} + 0.1946 \cdot 10^{-6}(P_0 - 1000) + 0.1747 \cdot 10^{-6}\sqrt{e_0} \\
& + 0.1040 \cdot 10^{-6}(T_0 - 15); \\
c = {} & 0.078.
\end{aligned}
\tag{6.40}
$$

The nonhydrostatic mapping function has the constants [*Mendes*, 1999]:

$$
\begin{aligned}
a = {} & 0.5236 \cdot 10^{-3} + 0.2471 \cdot 10^{-6}(P_0 - 1000) - 0.1328 \cdot 10^{-4}\sqrt{e_0} \\
& + 0.1724 \cdot 10^{-6}(T_0 - 15); \\
b = {} & 0.1705 \cdot 10^{-2} + 0.7384 \cdot 10^{-6}(P_0 - 1000) + 0.2147 \cdot 10^{-4}\sqrt{e_0} \\
& + 0.3767 \cdot 10^{-6}(T_0 - 15); \\
c = {} & 0.05917,
\end{aligned}
\tag{6.41}
$$

with:

P_0 : total surface pressure [mbar];
e_0 : surface partial pressure of water vapor [mbar];
T_0 : surface temperature [°C].

6.5.3   The MTT mapping functions

MTT   The MTT mapping functions [*Herring*, 1992], based on the type given by

Eq. (6.37), can be used to represent the elevation angle dependence of the tropospheric delay with an RMS of less than 0.2 mm for elevation angles larger than 3°. Herring determined expressions for the coefficients in Eq. (6.37) from raytracing radiosonde data at ten locations in the US located near VLBI stations. "The raytracing was performed at sixteen elevation angles between 3° and 90°. (...) For each raytrace the pressure was determined assuming hydrostatic equilibrium, and the path of the integration was determined from the total refractivity and its gradients at each point along the path. The delays along the path were accumulated separately for the part of the refractivity associated with the hydrostatic component of the refractivity, the residual wet component, and geometric contribution from the bending of the ray. The first ray was in the zenith direction and these zenith delays were used to determine the hydrostatic and wet mapping functions. The geometric bending terms were added to the hydrostatic delay part. After each set of ray traces was performed, the coefficients in the mapping function were determined by a least squares fitting, separately for the hydrostatic and wet components. The RMS fit of the mapping to hydrostatic delay was consistently to about 0.15 mm and to the wet delay typically less than 0.1 mm. The coefficients from each location were analyzed and deduced as a function of latitude (range used: 27° N to 65° N), height of the site (range used: 0 to 1.6 km) and surface temperature." The coefficients for the hydrostatic mapping function are:

$$
\begin{aligned}
a &= [1.2320 - 0.0139 \cos\phi - 0.0209\, h_0 + 0.00215(T_0 - 10)] \cdot 10^{-3}; \\
b &= [3.1612 - 0.1600 \cos\phi - 0.0331\, h_0 + 0.00206(T_0 - 10)] \cdot 10^{-3}; \\
c &= [71.244 - 4.293 \cos\phi - 0.149\, h_0 - 0.0021(T_0 - 10)] \cdot 10^{-3}.
\end{aligned}
\tag{6.42}
$$

For the wet mapping function the coefficients are:

$$
\begin{aligned}
a &= [0.583 - 0.011 \cos\phi - 0.052\, h_0 + 0.0014(T_0 - 10)] \cdot 10^{-3}; \\
b &= [1.402 - 0.102 \cos\phi - 0.101\, h_0 + 0.0020(T_0 - 10)] \cdot 10^{-3}; \\
c &= [45.85 - 1.91 \cos\phi - 1.29\, h_0 + 0.015(T_0 - 10)] \cdot 10^{-3},
\end{aligned}
\tag{6.43}
$$

where:

ϕ : latitude [deg];
h_0 : height of the observing station [km];
T_0 : surface temperature [°C].

6.5.4 The NMF and IMF mapping functions

Niell (1996) recognized that mapping functions like those of CfA-2.2, MTT, and Ifadis (1992) all depend on surface temperatures, which are much more variable than those at higher altitudes both diurnally and on longer time scales, resulting in an error in the mapping. Therefore, Niell aimed at developing New Mapping Functions (NMF) that are independent of surface meteorological parameters. As shown in Sect. 6.4, the coefficients of the mapping functions depend on the thickness of the troposphere (ratio of scale height and radius of the earth). This thickness varies with latitude and season. The hydrostatic mapping function (NMFh) of Niell only depends on latitude, day-of-year (DOY), and height. The wet mapping function (NMFw) varies solely with latitude because the water vapor is not in hydrostatic equilibrium and the height distribution of the water vapor is not expected to be predictable from the station height. Niell used nine temperature and relative humidity profiles of the US Standard Atmospheres, one for north latitudes of 15° for the whole year, and two for north latitudes of 30°, 45°, 60°, and 75°, for the months January and July as tabulated in

latitude, DOY, height

[*Cole et al.*, 1965]. These profiles and relative humidities accompanying them cover almost the full range of latitude north of the equator and represent averages over longitude (for North America). Mapping functions developed from them may provide a global description of the variation of path lengths with elevation angle if assumed that the southern and northern hemispheres are anti-symmetric in time. The nine profiles were raytraced at (also) nine different elevation angles between $3°$ and $90°$ to give both the hydrostatic and wet path delays. A least-squares fit to the values of each of the elevation angles was made for the three coefficients for each of the profiles with residuals not larger than 1 mm. The NMF mapping functions [*Niell*, 1996] can, like those of MTT, be used for elevation angles down to $3°$.

NMF

coeff.	$\phi = 15°$	$\phi = 30°$	$\phi = 45°$	$\phi = 60°$	$\phi = 75°$
a_{avg}	1.2769934e-3	1.2683230e-3	1.2465397e-3	1.2196049e-3	1.2045996e-3
b_{avg}	2.9153695e-3	2.9152299e-3	2.9288445e-3	2.9022565e-3	2.9024912e-3
c_{avg}	62.610505e-3	62.837393e-3	63.721774e-3	63.824265e-3	64.258455e-3
a_{amp}	0.0	1.2709626e-5	2.6523662e-5	3.4000452e-5	4.1202191e-5
b_{amp}	0.0	2.1414979e-5	3.0160779e-5	7.2562722e-5	11.723375e-5
c_{amp}	0.0	9.0128400e-5	4.3497037e-5	84.795348e-5	170.37206e-5
a_{wet}	5.8021897e-4	5.6794847e-4	5.8118019e-4	5.9727542e-4	6.1641693e-4
b_{wet}	1.4275268e-3	1.5138625e-3	1.4572752e-3	1.5007428e-3	1.7599082e-3
c_{wet}	4.3472961e-2	4.6729510e-2	4.3908931e-2	4.4626982e-2	5.4736038e-2

Table 6.1   Coefficients of the NMF hydrostatic and wet mapping functions.

The coefficient a of the hydrostatic mapping function is calculated as:

$$a_{hyd}(\phi_i, \text{DOY}) = a_{avg}(\phi_i) + a_{amp}(\phi_i) \cos\left[2\pi \frac{\text{DOY} - 28}{365.25}\right], \tag{6.44}$$

with ϕ_i the tabular latitude (Table 6.1) and DOY the time from January 0.0 in UT days. Similar expressions can be given for the other two coefficients.

sensitivity to height

The sensitivity of the hydrostatic mapping function to height above msl was determined by beginning the raytrace of each of the nine standard profiles with the values of pressure, temperature, and relative humidity at 1 km and 2 km altitude. The height correction is given as:

$$\Delta M(z) = \frac{dM(z)}{dh} h_0, \tag{6.45}$$

with

$$\frac{dM(z)}{dh} = \sec(z) - M(z, a_{ht}, b_{ht}, c_{ht}), \tag{6.46}$$

as an empirically chosen form. The coefficients $a_{ht} = 2.53e - 5$, $b_{ht} = 5.49e - 3$, and $c_{ht} = 1.14e - 3$ were determined by a least-squares fit to the height corrections at the nine elevation angles.

linear interpolation

To use the mapping function for any latitude, linear interpolation between the coefficients is required. Above $75°$ the same coefficients may be used as those at $75°$. Between $15°$ N and $15°$ S, the coefficients may be considered constant.

IMF

Recently, Niell demonstrated new mapping functions of the form of Eq. (6.37) called IMFh and IMFw (Isobaric Mapping Functions); [*Niell*, 2000], [*Niell*, 2001]. Especially IMFh showed a clear improvement. The coefficients are evaluated in $\cos(2\phi)$ terms and geopotential heights of the 200-mbar isolines, the minimum pressure level that seemed available that also showed the highest correlation with radiosonde-derived mapping functions of 28 sites at $5°$ elevation. Exact parameterizations were not yet published.

6.6 NWP-derived slant delays

Numerical Weather Prediction (NWP) models like HIRLAM (High Resolution Limited Area Model) provide an external source for determining STDs. Combining Eqs. (5.1) and (5.2) gives for the ZHD:

$$D_h^z = 10^{-6} k_1 R_d \int_0^{P_0} \frac{1}{g} \, dP, \tag{6.47}$$

and combining Eqs. (5.10), (5.11), (5.16), (5.2), and (4.15) gives for the ZWD:

$$D_w^z = 10^{-6} R_v \int_0^{P_0} (k_2' + \frac{k_3}{T}) \frac{q}{g} \, dP. \tag{6.48}$$

In NWPs at a pressure level i, the temperature T_i, specific humidity q_i, and pressure P_i are known. The gravity g_i can be computed from a standard function. The ZHD and ZWD can thus be approximated by:

$$D_h^z \approx 10^{-6} k_1 R_d \sum_i \frac{1}{g_i} \, \Delta P_i \tag{6.49}$$

and

$$D_w^z \approx 10^{-6} R_v \sum_i (k_2' + \frac{k_3}{T_i}) \frac{q_i}{g_i} \, \Delta P_i, \tag{6.50}$$

where $\Delta P_i = P_{i+1}^{1/2} - P_i^{1/2}$ is the difference of the pressures at half levels: $P_{i+1}^{1/2} = \frac{1}{2}(P_{i+1} - P_i)$. The slant delays can be derived by [*De Haan et al.*, 2001]:

$$\begin{aligned} D_h^s &\approx 10^{-6} k_1 R_d \sum_i \frac{\Delta s_i}{\Delta h_i} \frac{1}{g_i} \Delta P_i; \\ D_w^s &\approx 10^{-6} R_v \sum_i \frac{\Delta s_i}{\Delta h_i} (k_2' + \frac{k_3}{T_i}) \frac{q_i}{g_i} \Delta P_i, \end{aligned} \tag{6.51}$$

where Δh_i is the thickness of layer i and Δs_i is the slant distance through the layer.

NWPs can also be used to improve the mapping functions, especially the hydrostatic one. They can be obtained by raytracing the more realistic atmospheric profiles from the NWP. These direct mapping functions [*Rocken et al.*, 2001] are especially important when low-elevation observations are used.

Chapter 7

Azimuthal asymmetry and gradient parameters

azimuthal
asymmetry

Up till now the atmosphere was considered to be horizontally layered and azimuthal symmetric. Although for most applications this assumption is appropriate, azimuthal asymmetry may introduce significant errors in geodetic measurements where high precision is required. Some of the first publications about the effect of horizontal refractivity gradients were [*Gardner*, 1977] and [*Iyer and Bufton*, 1977]. A way to model the asymmetry is to assume a tilted atmosphere; see Fig. 7.1. In a tilted atmosphere the refractivity N as function of height h and horizontal position $\vec{x}$ reads:

$$N(\vec{x}, h) = N(0, h) + \nabla \vec{N} \cdot \vec{x}, \tag{7.1}$$

where the gradient vector is defined as:

$$\nabla \vec{N} \doteq \left. \frac{\partial N}{\partial \vec{x}} \right|_{\vec{x}=0}. \tag{7.2}$$

The dot in Eq. (7.1) denotes an inproduct. The tropospheric delay in an azimuthal asymmetric atmosphere can be given as the sum of the delay we would have in a symmetric atmosphere and a correction term owing to the azimuthal asymmetry:

$$D_{asym}(\alpha, z) = D_{sym}(z) + D_{az}(\alpha, z). \tag{7.3}$$

The correction term follows from Eq. (7.1) as:

$$D_{az} = 10^{-6} \int_0^\infty \nabla \vec{N} \cdot \vec{x} \, ds. \tag{7.4}$$

The vector $\vec{x}$ is in the projected slant direction (projection on the surface). Because we have both $|\vec{x}| \approx h \tan z$ and $ds \approx M(z)dh$, D_{az} can be approximated as:

$$\boxed{D_{az} \approx M(z) \tan z \, \vec{\gamma} \cdot \vec{e},} \tag{7.5}$$

where $\vec{e} \doteq [\cos \alpha; \sin \alpha]$ is a unit vector in the direction of $\vec{x}$, and

$$\vec{\gamma} \doteq 10^{-6} \int_{h_0}^\infty \nabla \vec{N} \, h \, dh \tag{7.6}$$

is the gradient vector in the opposite direction of the projected normal with components $\vec{\gamma} \doteq \begin{bmatrix} \gamma_N, \gamma_E \end{bmatrix}$. The first part of Eq. (7.5), $M(z) \tan z$, is a (mapping) function of the zenith angle. The second part is the azimuth (α) dependent gradient part:

$$\boxed{\gamma(\alpha) \doteq \vec{\gamma} \cdot \vec{e} = \gamma_N \cos \alpha + \gamma_E \sin \alpha.} \tag{7.7}$$

The gradient delay is thus composed of a north and an east parameter describing the azimuth dependence. The combined Model (7.5)/(7.7) was given in [*MacMillan*, 1995] and [*MacMillan and Ma*, 1997]. Using the above given definition, a slant gradient delay is a product of a gradient mapping function $M_{az}(z)$ and gradient delay $\gamma(\alpha)$:

$$\boxed{D_{az}(\alpha, z) = M_{az}(z)\,\gamma(\alpha).}\tag{7.8}$$
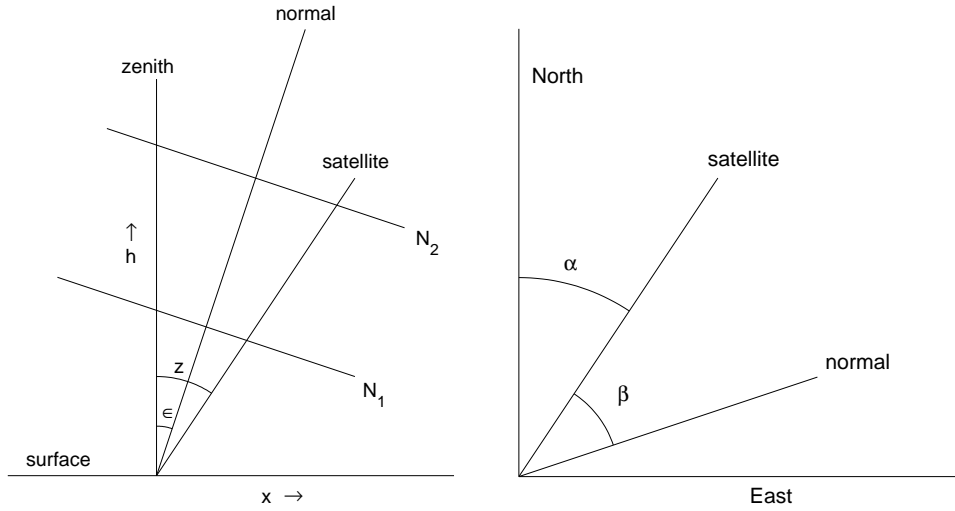
Figure 7.1   Tilted atmosphere with lines of equal refractivity N_1 and N_2.

In analogy to the symmetric case, the 'asymmetric' delay in the slant direction is also a product of a mapping function and, in this case, the delay in the direction of the normal $D^n \approx D^z$ [*De Munck*, 1991]:

$$D_{asym} = M(z - \epsilon\cos\beta)D^n \approx [M(z) - \frac{dM(z)}{dz}\epsilon\cos\beta]D^z,\tag{7.9}$$

where β is the angle between the projected normal and the projected slant direction. The approximation is allowed because $\epsilon\cos\beta$ is small. From Eqs. (7.3) and (7.9) now follows:

$$D_{az} \approx -\frac{dM(z)}{dz}\cdot D^z\epsilon\,\cos\beta.\tag{7.10}$$

So, if the radio signal is received from the same direction as the normal, the total delay is smallest, caused by a thinner atmosphere. Because the mapping function is approximately $M(z) \approx \sec z$, and so $dM(z)/dz \approx \sec z\tan z$, from Eqs. (7.5), (7.7), and (7.10) can be seen that $\vec{\gamma}\cdot\vec{e} = |\vec{\gamma}|\cdot 1\cdot\cos\beta \approx -D^z\epsilon\cos\beta$. So the
upper bound   length of the gradient vector is about $|\vec{\gamma}| \approx D^z\,|\epsilon|$, which is an upper bound of the gradient delay for a particular azimuth.

azimuthal   In [*Chen and Herring*, 1997] an azimuthal mapping function is given as:
mapping
function
$$M_{az}(z) = \frac{1}{\cos z\cot z+}\ \frac{C_1}{\cos z\cot z+}\ \frac{C_2}{\cos z\cot z + ...}\ .\tag{7.11}$$

The derivation of this mapping function is however not done in the same strict way as in Sect. 6.4. In fact, they truncated this mapping function after the first constant:

$$\boxed{M_{az}(z) = \frac{1}{\cos z\cot z + C},}\tag{7.12}$$

where $C = 0.0032$. The same type of mapping function could have been found by truncating $M(z)\tan z$ after the first constant. In [*Bar-Sever et al.*, 1998] $M(z)\tan z$ was chosen as a mapping function. They stated that essentially no different results were found in GPS Precise Point Positioning with $M(z)$ the wet or hydrostatic mapping function.
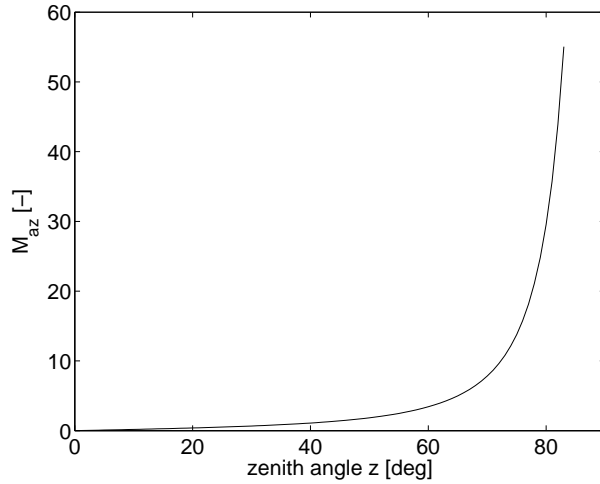
Figure 7.2   Gradient mapping function of Chen and Herring.

spatial and temporal scale

Hydrostatic gradients produced by pressure or dry temperature gradients have a large spatial scale of about 100 km and a temporal scale of days. Wet atmosphere gradients have a smaller spatial scale ($<$10 km) and can vary more rapidly on time scales of hours or less and are a function of water vapor content and temperature [*MacMillan*, 1995]. GPS or VLBI estimates of gradient delays cannot be separated into a wet and a hydrostatic component. With WVRs however, estimates of only wet gradients can be derived. In [*Davis et al.*, 1993], average values were found of approximately 1 mm for $|\vec{\gamma}|$ in Onsala. The largest value was approximately 8 mm. Bar-Sever et al. (1998) report hydrostatic gradients of about 0.6 mm from GPS estimates by comparing total gradients with wet gradients obtained by WVR-based estimates.

The model of [*Davis et al.*, 1993] was slightly more complicated. They recognized that symmetric mapping functions are evaluated at unrefracted zenith angles, where in fact $\tan z$ depends on the refracted zenith angle. They used an approximation for the bending:

$$\tau \approx -\tan z \Delta n = -\tan z \cdot 10^{-6} N_0. \tag{7.13}$$

The tangent of the refracted zenith angle then is:

$$\tan(z + \tau) = \tan z + \sec^2 z \cdot \tau = \tan z [1 - 10^{-6} N_0 \sec^2 z]. \tag{7.14}$$

Therefore, instead of the mapping function $M(z)\tan z$, they used the function $M(z)\tan z[1 - 10^{-6} N_0 \sec^2 z]$. As can be seen from Fig. 7.2, gradients become increasingly influential at lower elevation angles.

In Part II the parameterization of troposphere gradients in GPS observation models is given.

Chapter 8

Conclusions of Part I

The neutral part of the atmosphere delays radio signals like those used in space geodetic measurements. The delay is caused by the refractivity of dry-air and water-vapor molecules. Both are the result of induced molecular polarization, but water vapor also has a permanent dipole moment. Usually a distinction is made between a hydrostatic delay and a wet delay. The hydrostatic delay is caused by a mixture of dry air and water vapor that is considered to be in a hydrostatic equilibrium. The remaining delay is caused by water vapor.

We express these delays in metric units by multiplying them with the speed of light. The hydrostatic delay is about 2.20–2.40 m in the zenith direction, whereas the wet delay can range from 0–40 cm. Based on the surface pressure, the hydrostatic delay can be predicted with high accuracy. The zenith wet delay can only be determined with an accuracy of 2–5 cm based on semi-empirical models.

Air at a certain temperature can only hold a limited amount of water vapor; the surplus of water vapor condensates. At higher temperatures, the air can contain more water vapor. Since the temperature in the troposphere generally decreases with height, the saturated water-vapor pressure also decreases with height. This causes the water-vapor pressure to drop faster with increasing height than the pressure of dry air. The effective height or thickness of the wet troposphere is therefore smaller than the effective height of the dry troposphere.

The wet and hydrostatic slant tropospheric delays can be modeled as products of a zenith delay and a mapping function. In a flat horizontally layered atmosphere the mapping function is the secant of the zenith angle. In a curved atmosphere the mapping function is smaller, especially at larger zenith angles. How much smaller depends on the thickness of the atmosphere. Because the effective height of the dry troposphere is larger than the effective height of the wet troposphere, different mapping functions should be used for the wet and hydrostatic delays. Several mapping functions have been developed in the last three decades. Most of them, especially the newer ones, are described by a continued fraction. The constants in these mapping functions are functions of parameters like temperature, pressure, latitude, height, etc. The effective heights are also functions of these parameters, but are not explicitly derived.

All these mapping functions are based on the assumption of a symmetric, horizontally stratified, atmosphere. Because this assumption generally does not hold, the delay models are sometimes refined by gradients. Gradients and zenith wet delays are often parameterized in observation models used for space geodetic measurements (see Part II and Part IV for parametrization and estimation of GPS parameters).

component	model/author(s)	Eq./Table	parameters
STD	Saastamoinen	Eq. (6.11)	$P_0, T_0, e_0, B(r)$
ZHD	Davis et al.	Eq. (5.8)	P_0, h_0, ϕ
ZHD	Elgered et al.	Eq. (5.9)	P_0, h_0, ϕ
ZWD	Askne & Nordius	Eqs. (5.22) and (5.12)	T_0, e_0, λ, μ
ZWD	Baby et al.	Eq. (5.30); Table 5.1	T_0, rh_0, γ
$M_w(z)$	Ifadis	Eqs. (6.37) and (6.41)	P_0, T_0, e_0
$M_w(z)$	Herring (MTT)	Eqs. (6.37) and (6.43)	T_0, h_0, ϕ
$M_w(z)$	Niell (NMFw)	Eq. (6.37); Table 6.1	ϕ
$M_h(z)$	Ifadis	Eqs. (6.37) and (6.40)	P_0, T_0, e_0
$M_h(z)$	Herring (MTT)	Eqs. (6.37) and (6.42)	T_0, h_0, ϕ
$M_h(z)$	Niell (NMFh)	Eqs. (6.37) and (6.44–6.46); Table 6.1	ϕ, DOY, h_0
$M_h(z)$	Davis et al. (CfA2.2)	Eqs. (6.36) and (6.39)	P_0, T_0, h_T
P_0, T_0	US 1976	Eq. (4.41)	-
$e_0 = rh_0 \cdot e_{sat}$	Baby/US 1976	Eq. (4.44)	T_0, rh_0, h_0, β

Table 8.1   Discussed model components. The parameters are: P_0: surface pressure; T_0: surface temperature; e_0: surface partial pressure of water vapor; rh_0: surface relative humidity; h_0: station height above msl; h_T: height of tropopause above msl; ϕ: station latitude; DOY: Day Of Year; $B(r)$: tabular value; γ, λ, μ: model constants; β: temperature lapse rate.

state of the art

Table 8.1 shows the most important model components as given in this part. Several good mapping functions exist, but the mapping functions NMFh and NMFw of Niell (1996) are nowadays most often used and considered 'state of the art'. Their popularity is mainly due to the parameterization in latitude, height, and DOY instead of in meteorological parameters. New developments, however, show possible improvements by using actual atmospheric data. The ZHD of Elgered et al. (1991) is now used most often and can also be considered 'state of the art'. It is however important to have good (surface) pressure values available to determine this ZHD. For the ZWD several models exist, but none of them is good enough for precise applications.

contribution

Part I is mainly the result of study of literature and consists of a compilation and overview of the important model components and the retrieval of their origin. Especially the derivation of the general form of a mapping function as continued fraction is clarified with respect to the Marini (1972) paper.

Temperature lapse rate

In this appendix derivations are given of the dry-adiabatic lapse rate and the saturated-adiabatic lapse rate. The first one is a(n ideal) lapse rate in unsaturated air. The latter one is the lapse rate in saturated air. The derivation follows [*Haltiner and Martin*, 1957] and is based on the first law of thermodynamics, which is displayed here first.

A.1   First law of thermodynamics

The first law of thermodynamics states that "heat added to a system must equal the change in the internal energy of the system plus the work[1] done by the system" [*Haltiner and Martin*, 1957]. Expressed in unit mass this reads:

$$\delta H = dU + \delta W, \tag{A.1}$$

δH : infinitesimal amount of heat per unit mass [J kg^{-1}];
dU : change in internal energy per unit mass [J kg^{-1}];
δW : work done by unit mass of the system [J kg^{-1}].

A.2   Dry-adiabatic lapse rate

If we consider a closed system and allow a nonviscose gas to undergo a small expansion dV, the work done by the expansion equals (in units of mass) [*Haltiner and Martin*, 1957]:

$$\delta W = Pd\alpha. \tag{A.2}$$

By differentiating the equation of state Eq. (4.1), we find another expression for the right-hand side of Eq. (A.2):

$$Pd\alpha = R_i dT - \alpha dP. \tag{A.3}$$

Combining these two equations gives:

$$\delta W = R_i dT - \alpha dP. \tag{A.4}$$

For perfect gases, which satisfy the equation of state, the internal energy is a function of temperature alone. If the volume is kept constant while an infinitesimal amount of heat δH is added to the gas, this heat becomes internal energy. This added internal energy is:

$$dU = c_v dT \quad ; \quad c_v \doteq \left(\frac{dH}{dT}\right)_{V=constant}, \tag{A.5}$$

---

[1]force × distance.

$$c_v \text{ : specific heat at constant volume [J kg}^{-1}\text{K}^{-1}\text{]};$$
$$dT \text{ : change in temperature [K]}. \tag{A.6}$$

specific heat   The *specific heat* is a slowly varying function of the temperature, but can be considered constant for atmospheric temperatures. Substituting Eqs. (A.4) and (A.5) into Eq. (A.1) gives:

$$\delta H = (c_v + R_i)dT - \alpha dP. \tag{A.7}$$

isobaric   For an *isobaric* process we have: $dP = 0$. In this case we find from Eq. (A.7) by definition the specific heat at constant pressure:

$$c_v + R_i = \left(\frac{dH}{dT}\right)_{P=constant} \doteq c_p, \tag{A.8}$$

c_p : specific heat at constant pressure [J kg^{-1}K^{-1}].

Using Eqs. (4.6) and (A.8) in Eq. (A.7) gives:

$$\delta H = c_p dT + g dh. \tag{A.9}$$

adiabatic   If there is no heat exchange between the system and its environment ($\delta H = 0$), the process is considered *adiabatic*. In this case Eq. (A.9) becomes

$$\frac{g}{c_p} = -\frac{dT}{dh} \doteq \beta, \tag{A.10}$$

lapse rate   where β is called the temperature *lapse rate*[2]. According to Eq. (A.10), the dry-adiabatic lapse rate is equal to:

$$\boxed{\beta_d = \frac{g}{c_{pd}} \approx 9.8 \text{ K km}^{-1},} \tag{A.11}$$

β_d : dry-adiabatic lapse rate;
c_{pd} : specific heat at constant pressure for dry air [1003 J kg^{-1}K^{-1}];
g : gravitational acceleration [9.8 m s^{-2}].

A.3   Lapse rate in saturated air

Since external work is done when the gas expands, the loss in energy results in a cooling of the gas. If saturated air expands adiabatically, water vapor will condense to liquid or ice as the temperature decreases. For each gram of dry air, the saturated air contains w_{sat} gram water vapor. The amount of vapor condensed during an infinitesimal change from (T, P) to $(T + \Delta T, P + \Delta P)$ equals $-dw_{sat}$ gram. So, the amount of water vapor decreases in the condensation process. The latent heat released, equals $-Ldw_{sat}$ per $(1 + w_{sat})$ gram saturated air. The saturated-adiabatic lapse rate can now be found from (compare with Eq. (A.9)):

$$-Ldw_{sat}/(1 + w_{sat}) = c_{pm}dT + g dh, \tag{A.12}$$

with

$$c_{pm} = \frac{1}{1 + w_{sat}}c_{pd} + \frac{w_{sat}}{1 + w_{sat}}c_{pv}, \tag{A.13}$$

---

[2]Sometimes the temperature lapse rate is defined as $\beta \doteq +dT/dh$.

c_{pm} : specific heat at constant pressure for moist air [J kg^{-1}K^{-1}];

c_{pv} : specific heat at constant pressure for water vapor [1810 J kg^{-1}K^{-1}].

By rearranging Eq. (A.12) we find the lapse rate of saturated air:

$$\beta_{sat} = \frac{Lc_{pm}^{-1}}{1 + w_{sat}}\frac{dw_{sat}}{dh} + \frac{g}{c_{pm}}, \tag{A.14}$$

β_{sat} : saturated-adiabatic lapse rate [K/km].

Differentiation of the approximation Eq. (4.23) with respect to h, together with Eqs. (4.6), (4.16), and (4.20) gives:

$$\frac{dw_{sat}}{dh} = -\frac{w_{sat}L}{R_vT^2}\beta_{sat} + \frac{w_{sat}g}{R_mT}. \tag{A.15}$$

Combining Eq. (A.14) and Eq. (A.15) and using the approximations $1 + w_{sat} \approx 1$ and $R_m \approx R_d$, finally gives the following expression for the saturated-adiabatic lapse rate:

$$\boxed{\beta_{sat} = \beta_d\frac{1 + Lw_{sat}/R_dT}{1 + L^2w_{sat}/c_{pm}R_vT^2}.} \tag{A.16}$$
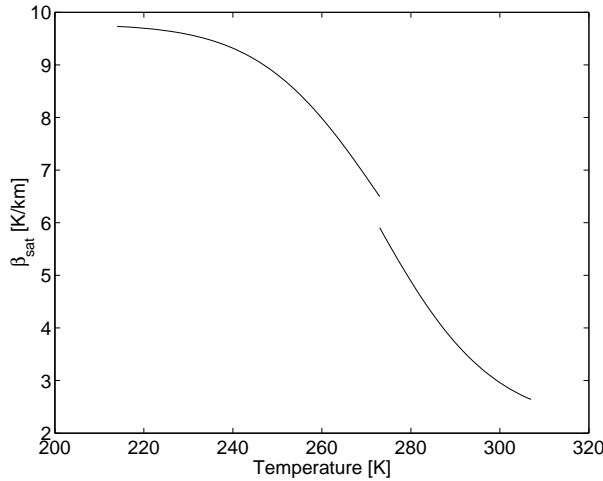
Figure A.1   Saturated-adiabatic temperature lapse rate.

In Fig. A.1, Eq. (A.16) is shown graphically for $P = 1000$ mbar. Note the jump at 0°C caused by the different values for the latent heat.

Saastamoinen integrals

Based on the three integrals of Eq. (6.11), Saastamoinen obtained his zenith-delay model.

For the *first* integral we have:

$$
N = N_0 \left(\frac{T}{T_0} \right)^\mu \quad \text{and} \quad T = T_0 - \beta(r - r_0) \Rightarrow
\begin{cases}
r - r_0 = -\left(\frac{T_0}{\beta} \right) \left(\frac{T}{T_0} - 1 \right) ; \\
dr \quad = -\frac{1}{\beta} dT.
\end{cases}
\tag{B.1}
$$

After some straightforward computations this results in:

$$
\int_{r_0}^{r_T} N(r - r_0) dr = \int_{T_0}^{T_T} N_0 \left(\frac{T}{T_0} \right)^\mu \cdot -\left(\frac{T}{\beta} \right) \left(\frac{T}{T_0} - 1 \right) \cdot \frac{1}{\beta} dT
\tag{B.2}
$$

$$
= \frac{R_d}{g_m^2 (1 - R_d \beta / g_m)} [N_0 T_0^2 - N_T T_T^2] - \frac{R_d}{g_m}(r_T - r_0) N_T T_T,
$$

where the index $_T$ stands for values at the tropopause.

The *second* integral can be evaluated by using the profile of Eq. (4.43):

$$
N = N_T \exp\left(-\frac{r - r_T}{\text{H}} \right) \quad ; \quad \text{H} = \frac{R_d T_T}{g_m},
\tag{B.3}
$$

and results in:

$$
\int_{r_T}^{\infty} N(r - r_T) dr = \frac{R_d^2}{g_m^2} N_T T_T^2.
\tag{B.4}
$$

The *third* integral is similar to Eq. (5.1) and results in:

$$
(r_T - r_0) \int_{r_T}^{\infty} N dr = (r_T - r_0) 10^{-6} k_1 \frac{R_d}{g_m} P_T = \frac{R_d}{g_m}(r_T - r_0) N_T T_T.
\tag{B.5}
$$

Summation of the three integral results gives the total integral:

$$
\int_{r_0}^{\infty} N(r - r_0) dr = \frac{R_d^2}{g_m^2} \left[\frac{N_0 T_0^2 - (R_d \beta / g_m^2) N_T T_T^2}{1 - R_d \beta / g_m} \right].
\tag{B.6}
$$

With Eqs. (5.32) and (5.33) the total Saastamoinen model then becomes:

$$
D_t^s = 10^{-6} \sec z_0 \left[\int_{r_0}^{\infty} N dr - \frac{1}{r_0} \tan^2 z_0 \cdot \int_{r_0}^{\infty} N(r - r_0) dr \right] \Rightarrow
$$

$$
\boxed{D_t^s = \underbrace{10^{-6} k_1 \frac{R_d}{g_m}}_{0.002277} \sec z_0 \left[P_0 + (\frac{1255}{T_0} + 0.05) e_0 - B(r) \tan^2 z_0 \right],}
\tag{B.7}
$$

where

$$
B(r) \doteq \frac{1}{r_0} \frac{g_m}{R_d} \frac{1}{k_1} \int_{r_0}^{\infty} N(r - r_0) dr.
\tag{B.8}
$$

Effective height

The effective height or thickness of the atmosphere is defined as:

$$\mathrm{H}_m \doteq \frac{\int_{h_0}^{\infty}(h-h_0)\, N(h)dh}{\int_{h_0}^{\infty} N(h)dh}. \qquad (\mathrm{C.1})$$

In case of an exponential refractivity profile the effective height H_m is the same as the scale height H. But for the refractivity profile of Eq. (4.42), we have with $T = T_0 - \beta h$ and the definition for μ from Eq. (4.40):

$$N(h) = N_0 \left(1 - \frac{\beta h}{T_0}\right)^{\mu} \Rightarrow \ln N(h) = \ln N_0 + \mu \ln \left(1 - \frac{\beta h}{T_0}\right). \qquad (\mathrm{C.2})$$

Since $\ln(1 - \beta h/T_0) \approx -\beta h/T_0$, this refractivity profile can be approximated as an exponential one:

$$N(h) = N_0 \exp(-\frac{\beta h}{T_0}\mu) = N_0 \exp(-\frac{h}{\overline{\mathrm{H}}}), \qquad (\mathrm{C.3})$$

where

$$\frac{1}{\overline{\mathrm{H}}} = \frac{g_m}{R_d T_0} - \frac{\beta}{T_0}. \qquad (\mathrm{C.4})$$

$\overline{\mathrm{H}}$ is a(n approximate) generalization of the scale height H as defined in Eq. (4.39): for $\beta = 0$ K km^{-1}, we have $\overline{\mathrm{H}} = \mathrm{H}$. For $g_m = 9.784$ m s^{-2}, $R_d = 287.06$ J kg^{-1}K^{-1}, and $T_0 = 288$ K, we find: $\overline{\mathrm{H}} = 8.45$ km.

Bibliography

Askne, J., and H. Nordius (1987), Estimation of tropospheric delay for microwaves from surface weather data, *Radio Science*, *22*, 379–386.

Baby, H.B., P. Golé, and J. Lavergnat (1988), A model for the tropospheric excess path length of radio waves from surface meteorological measurements, *Radio Science*, *23*, 1023–1038.

Bar-Sever, Y.E., P.M. Kroger, and J.A. Borjesson (1998), Estimating horizontal gradients of tropospheric path delay with a single GPS receiver, *Journal of Geophysical Research (B)*, *103*, 5019–5035.

Bevis, M., S. Businger, T.A. Herring, C. Rocken, R.A. Anthes, and R.H. Ware (1992), GPS meteorology: Remote sensing of atmospheric water vapor using the global positioning system, *Journal of Geophysical Research (D)*, *97*, 15787–15801.

Bevis, M., S. Businger, S. Chiswell, T.A. Herring, R. Anthes, C. Rocken, and R.H. Ware (1994), GPS meteorology: Mapping zenith wet delays onto precipitable water, *Journal of Applied Meteorology*, *33*, 379–386.

Boudouris, G. (1963), On the index of refraction of air, the absorption and dispersion of centimeter waves by gases, *J. of Res. Natl. Bur. Stand., Sect. D*, *67*, 631–684.

Champion, F.C. (1960), *University Physics*, Blackie & Son Ltd, Glasgow.

Chao, C.C. (1974), The tropospheric calibration model for mariner mars 1971, *Tech. rep., 32-1587*, JPL Pasadena.

Chen, G., and T.A. Herring (1997), Effects of atmospheric azimuthal asymmetry on the analysis of space geodetic data, *Journal of Geophysical Research (B)*, *102*, 20489–20502.

Cole, A.E., A. Court, and A.J. Cantor (1965), Model atmospheres, in *Handbook of Geophysics and Space Environments*, edited by L. Valley, pp. 19–54, McGraw-Hill Book Company.

Davis, J.L., T.A. Herring, I.I. Shapiro, A.E.E. Rogers, and G. Elgered (1985), Geodesy by radio interferometry: Effects of atmospheric modeling errors on estimates of baseline length, *Radio Science*, *20*, 1593–1607.

Davis, J.L., G. Elgered, A.E. Niell, and C.E. Kuehn (1993), Ground-based measurements of gradients in the "wet" radio refractivity of air, *Radio Science*, *28*, 1003–1018.

Elgered, G., J.L. Davis, T.A. Herring, and I.I. Shapiro (1991), Geodesy by radio interferometry: Water vapor radiometry for estimation of the wet delay, *Journal of Geophysical Research (B)*, *96*, 6541–6555.

Emardson, T.R., and H.J.P. Derks (1998), On the relation between the wet delay and the integrated precipitable water vapour in the European atmosphere, *Meteorol. Applications*, *7*, 61–68.

Essen, L., and K.D. Froome (1951), Dielectric constant and refractive index of air and its principal constituents, *Nature*, *167*, 512–513.

Gardner, C.S. (1977), Correction of laser tracking data for the effects of horizontal refractivity gradients, *Applied Optics*, *16*, 2427–2432.

Goad, C.C., and L. Goodman (1974), A modified Hopfield tropospheric refraction correction model, in *AGU Anual Fall meeting*.

Haan, S. de, H. van der Marel, and S. Barlag (2001), Use of GPS slant delay measurements for synoptic forecasting: Case study of a cold front passage.

Hall, M.P.M., L.W. Barcley, and M.T. Hewitt (ed.) (1996), *Propagation of Radiowaves*, The Institution of Electral Engineers, London, England.

Haltiner, G.J., and F.L. Martin (1957), *Dynamical and Physical Meteorology*, McGraw-Hill Book Company, New York.

Herring, T.A. (1992), Modeling atmospheric delays in the analysis of space geodetic data, in *Refraction of transatmospheric signals in geodesy*, pp. 157–164.

Hopfield, H.S. (1969), Two-quartic tropospheric refractivity profile for correcting satellite data, *Journal of Geophysical Research*, *74*, 4487–4499.

IAG (1963), Resolution 1 of the 13th general assembly, *Bulletin Géodésique*, *70*, 390–391.

Ifadis, I.M. (1992), The excess propagation path of radio waves: Study of the influence of the atmospheric parameters on its elevation dependence, *Survey Review*, *31*, 289–298.

Ifadis, I.M. (1993), Space to earth techniques: Some considerations on the zenith wet path delay parameters, *Survey Review*, *32*, 130–144.

Iyer, R.S., and J.L. Bufton (1977), Corrections for atmospheric refractivity in satellite laser ranging, *Applied Optics*, *16*, 1997–2003.

MacMillan, D.S. (1995), Atmospheric gradients from very long baseline interferometry observations, *Geophysical Research Letters*, *22*, 1041–1044.

MacMillan, D.S., and C. Ma (1997), Atmospheric gradients and the VLBI terrestrial and celestial reference frames, *Geophysical Research Letters*, *24*, 453–456.

Marini, J.W. (1972), Correction of satellite tracking data for an arbitrary tropospheric profile, *Radio Science*, *7*, 223–231.

Mendes, V.B. (1999), *Modeling the neutral-atmosphere propagation delay in radiometric space techniques*, Ph.D. dissertation, University of New Brunswick.

Mendes, V.B., G. Prates, L. Santos, and R.B. Langley (2000), An evaluation of the accuracy of models of the determination of the weighted mean temperature of the atmosphere, in *Proceedings of the 2000 National Technical Meeting, 2000: Navigation into the New Millennium*, pp. 433–438.

Morduch, G.E. (1978), A formula for computing refraction effects on range measurements, *Radio Science*, *13*, 509–510.

Munck, J.C. de (1991), Tropospheric effects on baseline measurements by using (GPS) satellites, *Tech. rep. 91-3*, DUT, Faculty of Geodetic Engineering.

Niell, A.E. (1996), Global mapping functions for the atmosphere delay at radio wavelengths, *Journal of Geophysical Research (B)*, *101*, 3227–3246.

Niell, A.E. (2000), Improved atmospheric mapping functions for VLBI and GPS, *Earth Planets Space*, *52*, 699–702.

Niell, A.E. (2001), Preliminary evaluation of atmospheric mapping functions based on numerical weather models, *Physics and Chemistry of the Earth (A)*, *26*, 475–480.

Owens, J.C. (1967), Optical refractive index of air: Dependence on pressure, temperature and composition, *Applied Optics*, *6*, 51–58.

Queney, P. (1974), *Eléments de météorologie*, Masson and Cie, Paris.

Rocken, C., S. Sokolovskiy, J.M. Johnson, and D. Hunt (2001), Improved mapping of tropospheric delays, *Journal of Atmospheric and Oceanic Technology*, *18*, 1205–1213.

Saastamoinen, J. (1972), Atmospheric correction for the troposphere and stratosphere in radio ranging of satellites, *The use of artificial satellites for geodesy, Geophys. Monogr. Ser.*, *15*, 247–251.

Saastamoinen, J. (1973), Contributions to the theory of atmospheric refraction, *Bulletin Géodésique No. 105-107*, pp. 279–298; 383–397; 13–34.

Seeber, G. (1993), *Satellite Geodesy, Foundations, Methods, and Applications*, Walter de Gruyter, Berlin.

Smart, W.M. (1936), *Text-Book on Spherical Astronomy*, Cambridge University Press, Cambridge.

Smith, E.K., and S. Weintraub (1953), The constants in the equation for atmospheric refractive index at radio frequencies, in *Proc. IRE 41*, pp. 1035–1037.

Smith, W.L. (1966), Notes on the relationship between total precipitable water and surface dew point, *Journal of Applied Meteorology*, *5*, 726–727.

Stull, R.B. (1995), *Meteorology: For Scientists and Engineers*, West Publishing, St. Paul, Minneapolis.

Thayer, G.D. (1974), An improved equation for the radio refractive index of air, *Radio Science*, *9*, 803–807.

Wall, H.S. (1948), *Analytic theory of continued fractions*, D. van Nostrand Company, Toronto.

Yionoulis, S.M. (1970), Algorithm to compute tropospheric refraction effects on range measurements, *Journal of Geophysical Research*, *75*, 7636–7637.

PARAMERIZATION OF THE TROPOSPHERIC DELAY IN GPS OBSERVATION MODELS

Symbols in Part II

symbol	meaning
$A(k)$	partial design matrix of global parameters
$\mathcal{A}$	real-valued lumped phase ambiguities
α	azimuth
$br(k)$	receiver hardware delay
$bs(k)$	satellite hardware delay
$B(k)$	partial design matrix of batch parameters
$\beta(k)$	batch parameters
c	speed of light
c_n	$n \times 1$ vector $[1, 0, \ldots, 0]'$
C_n	$n \times (n-1)$ matrix with zeros and ones
$d(k)$	differences between ZTDs
$D(k)$	STDs
$\Delta D(k)$	residual STDs
$D^z(k)$	ZTDs
$\mathcal{D}$	partial design matrix of clock errors
$\overline{\delta}(k)$	lumped (receiver and satellite) clock errors
$\delta r(k)$	receiver clock errors
$\delta s(k)$	satellite clock errors
e_n	$n \times 1$ vector with ones
ϕr	initial receiver phases
ϕs	initial satellite phases
G, G_2	partial design matrices of geometry
$\gamma(k)$	gradients
$\Gamma(k)$	partial design matrix of gradients
I_m	$m \times m$ identity matrix
$\imath(k)$	ionospheric delays
k	epoch number $(k = 1, ..., p)$

symbol	meaning
$\cdot_{,l}$	index for observation type ($l = 1, ..., 4$ means L1, L2, P1, P2)
L	2×2 matrix with zeros and wavelengths
λ_1, λ_2	GPS carrier wavelengths
Λ	4×2 matrix with zeros and wavelengths
m	number of satellites
$M(k)$	partial design matrix of ZTDs (with mapping functions)
μ	4×1 vector with quotients of wavelengths
n	number of stations
N	integer phase ambiguities
$\mathbb{N}^+$	space of natural numbers (1,2,...)
ν	3×1 vector $[-\mu_2^2,\ 1,\ \mu_2^2]'$
p	number of epochs
$\widetilde{p}$	number of epochs per batch
Q_y	covariance matrix of y
r	station coordinates
$\cdot_r$	receiver index ($r = 1, ..., n$)
ρ	receiver–satellite range
$\cdot^s$	satellite index ($s = 1, ..., m$)
$t(k)$	temporal parameters
$T(k)$	partial design matrix of temporal parameters
$T(k)^\perp$	null space of $T(k)'$
τ	travel time of GPS signal
v_k	vector of predicted residuals
x	global parameters
$\hat{x}$	estimates of x
$\overline{x}$	lumped parameters (with no significant meaning)
$x_{1..n}$	$[x_1, ..., x_n]'$
x_{12}	$x_2 - x_1$
$x_{12..n}$	$x_{2..n} - x_1$
y	observations
$\underline{y}$	observables (both GPS and pseudo-observables)
$\underline{y}_,$	GPS observables
z	zenith angle

Chapter 9

Introduction to Part II

In this part, GPS observation models are described for data processing of static geodetic networks. We focus on the functional model and discuss the stochastic model only briefly; more about the stochastic model can be found in Part III. Parameter estimation and hypothesis testing with these models is described in Part IV. Special attention is given to the inclusion of troposphere parameters. An extensive description of (undifferenced) observation equations without troposphere parameters is given in [*De Jonge*, 1998]. This part starts with a review. Like in [*De Jonge*, 1998], we distinguish global parameters, which are constant for some time span, and temporal parameters, which change every epoch.

In Chap. 10 the nonlinear and linearized GPS observation equations are derived and presented in matrix-vector notation. The system is rank deficient and therefore in Chap. 11 first a full-rank model is derived. In this chapter we assume the troposphere to be a priori corrected for (the troposphere-fixed models). We show that the redundancy is not sufficient to estimate slant tropospheric delays every epoch, unless the less precise code observables are used or additional assumptions are made.

Chapter 12 deals with several kinds of parameterization of the troposphere in the troposphere-float and weighted models. In these models the zenith delays are the primary parameters, and gradient parameters and residual slant delays may be estimated additionally. In the latter case we need extra soft-constraints, whereas in the troposphere-weighted case also differences between consecutive zenith delays are soft-constrained. Chapter 12 ends with an overview of the possible parameterizations of the tropospheric delay.

Chapter 13 describes some troposphere and geometry-related near rank deficiencies. These rank deficiencies do not occur when the satellite clocks are a priori known as in Precise Point Positioning.

Chapter 14 shows analytic expressions of a transformation with which the temporal parameters are pre-eliminated to obtain a smaller observation model. The transformation is often used in GPS data-processing software and includes double differencing and forming linear combinations to eliminate ionosphere parameters.

This part ends with conclusions in Chap. 15.

GPS observation equations

10.1   The nonlinear functional model

In precise GPS positioning, carrier phase and often also pseudo-range observations are being used. We will assume that two pseudo-range and two carrier phase observations are available per receiver-satellite combination at each epoch. Furthermore, we assume that the satellite orbits are sufficiently known, so that we do not need to parameterize them. We denote the observations and observables by

$$y_{r,l}^s(t) \text{ and } \underline{y}_{r,l}^s(t),$$

r, s, n respectively, where $r = 1, ..., n$ refers to the receiver, $s = 1, ..., m_r$ refers to
l the satellite, $l = 1, ..., 4$ refers to the observation type and t is the time of
L1, L2   measurement in GPS time. We use $l = 1$ and $l = 2$ for the L1 and L2 carrier phase observations/observables and $l = 3$ and $l = 4$ for the L1 and L2 pseudo-range or
P1, P2   code observations/observables (P1 and P2).

The basic GPS observation equation reads [*Teunissen and Kleusberg*, 1998], [*De Jonge*, 1998]:

$$
\boxed{
\begin{aligned}
y_{r,l}^s(t) - \tilde{y}_{r,l}^s(t) = &\ \rho_r^s(t, t - \tau_r^s) \\
&+ D_r^s(t) \\
&+ \iota_{r,l}^s(t) \\
&+ c \cdot \delta t_r(t) \\
&- c \cdot \delta t^s(t - \tau_r^s) \\
&+ br_{r,l}(t) \\
&- bs_{,l}^s(t - \tau_r^s) \\
&+ \lambda_{,l} \cdot N_{r,l}^s \\
&+ \lambda_{,l} \cdot \phi r_{r,l} \\
&- \lambda_{,l} \cdot \phi s_{,l}^s \\
&+ dm_{r,l}^s(t) \\
&+ \varepsilon_{r,l}^s(t),
\end{aligned}
}
\tag{10.1}
$$

where

$$
\begin{aligned}
&\tilde{y}_{r,l}^s(t) && : \text{sum of a-priori corrections [m]};\\
&\rho_r^s(t, t - \tau_r^s) && : \text{geometric range [m]};\\
&D_r^s(t) && : \text{tropospheric delay [m]};\\
&\iota_{r,l}^s(t) && : \text{ionospheric delay [m]};
\end{aligned}
$$

$c \cdot \delta t_r(t)$: receiver clock error [m];
$c \cdot \delta t^s(t - \tau_r^s)$: satellite clock error [m];
c : speed of light in vacuum [299 792 458 m/s]*;
$br_{r,l}(t)$: receiver hardware delay [m];
$bs_{,l}^s(t - \tau_r^s)$: satellite hardware delay [m];
$N_{r,l}^s$: integer-valued phase ambiguity [-];
$\lambda_{,l}$: observation-type dependent factor [m];
$\phi r_{r,l}$: initial phase at receiver [-];
$\phi s_{,l}^s$: initial phase at satellite [-];
$dm_{r,l}^s(t)$: sum of unmodeled errors [m];
$\varepsilon_{r,l}^s(t)$: observation noise [m].

      * The SI value of c was defined to be this number exactly
        at the 1983 Conference Generale des Poids et Mesures.

The a-priori correction term may include corrections for the satellite clock, phase center variations, tropospheric and ionospheric delay, phase wind-up, ocean tide loading, and solid earth tides. Unmodeled errors like multipath may occur in the term $dm_{r,l}^s(t)$. We assume that $dm_{r,l}^s(t) = 0$.

geometric range   Except for the geometric range, the tropospheric delay, and the clock errors, the parameters are observation-type dependent. The geometric range is defined as the distance between the position of a satellite (r^s) at time $t - \tau_r^s$ and the position of a receiver (r_r) at time t:

$$\rho_r^s(t, t - \tau_r^s) \doteq \parallel r^s(t - \tau_r^s) - r_r(t) \parallel, \tag{10.2}$$

where r^s and r_r are 3×1 vectors with coordinates in an Earth Centered Inertial system and τ_r^s is the travel time, which can be approximated by

$$\tau_r^s \approx \frac{\rho_r^s(t, t - \tau_r^s)}{c}, \tag{10.3}$$

if we neglect atmospheric and instrumental delays.

Like the position of the satellites, the satellite clocks and hardware delays are also evaluated at time $t - \tau_r^s$. We assume that the satellite clocks and hardware delays are sufficiently stable, so that observations of different receivers have common satellite clock errors and hardware delays, although the travel time to different receivers may differ.

$\lambda_{,l}$ The observation-type dependent factor $\lambda_{,l}$ is defined as:

$$\begin{aligned}
\lambda_{,1} &\doteq \lambda_1; \\
\lambda_{,2} &\doteq \lambda_2; \\
\lambda_{,3} &\doteq 0; \\
\lambda_{,4} &\doteq 0,
\end{aligned} \tag{10.4}$$

where $\lambda_1 \approx 0.19$ m and $\lambda_2 \approx 0.24$ m are the wavelengths of the L1 and L2 carriers with respective frequencies of $f_1 = 154 \cdot 10.23$ MHz and $f_2 = 120 \cdot 10.23$ MHz. In other words, the carrier phase ambiguities and initial phases are (of course) only present in the phase observations.

The ionospheric delay is dispersive (frequency dependent) and has a different sign for carrier phase and pseudo-range observations. In fact, the phase

ionospheric
delays

is advanced. The relation between the ionospheric delays reads in first order approximation:

$$\frac{1}{\lambda_1^2} \cdot \imath_{r,1}^s(t) = \frac{1}{\lambda_2^2} \cdot \imath_{r,2}^s(t) = -\frac{1}{\lambda_1^2} \cdot \imath_{r,3}^s(t) = -\frac{1}{\lambda_2^2} \cdot \imath_{r,4}^s(t).$$ (10.5)

10.2   Linearization

In this section the nonlinear observation equations are linearized, and a matrix-vector representation of the linearized functional model is derived. We assume that the observation noise has zero (mathematical) expectation: $\mathbf{E}\{\underline{\varepsilon}_{r,l}^s(t)\} = 0$. Linearizing Eq. (10.1) then gives the mathematical expectation of the observed-minus-computed observables:

$\mathbf{E}\{.\}$

$$
\begin{aligned}
\mathbf{E}\{\underline{\Delta y}_{r,l}^s(t)\} = &\frac{d\rho_r^s}{dr_r'}\Delta r_r & (coordinates) \\
&+ \Delta D_r^s(t) & (troposphere) \\
&+ \Delta\imath_{r,l}^s(t) & (ionosphere) \\
&+ c \cdot \Delta\delta t_r(t) & (rec.\ clock\ errors) \\
&- c \cdot \Delta\delta t^s(t - \tau_r^s) & (sat.\ clock\ errors) \\
&+ \Delta br_{r,l}(t) & (rec.\ hardware\ delays) \\
&- \Delta bs_{,l}^s(t - \tau_r^s) & (sat.\ hardware\ delays) \\
&+ \lambda_{,l} \cdot \Delta N_{r,l}^s & (ambiguities) \\
&+ \lambda_{,l} \cdot \Delta\phi r_{r,l} & (rec.\ init.\ phase) \\
&- \lambda_{,l} \cdot \Delta\phi s_{,l}^s & (sat.\ init.\ phase),
\end{aligned}
$$ (10.6)

where

$$\frac{d\rho_r^s}{dr_r'} = -\frac{r_r^s(t)'}{\parallel r_r^s(t) \parallel}.$$ (10.7)

These equations are evaluated in GPS time, whereas the actual observations are in receiver time t_r:

$$t_r = t + \delta t_r(t).$$ (10.8)

Evaluation in receiver time gives rise to several more terms in the linearization of the observation equations. For a more elaborate discussion, see [*De Jonge*, 1998]. We ignore the additional terms and assume that our observation equation is equally applicable in GPS time as in receiver time. This is the case if we have sufficient approximate values for the clock parameters (for example obtained by a code-only solution).

Before we proceed to derive a matrix-vector representation, first some notational simplifications are made. Instead of the time instant $t = t_k$, from now on we refer to the epoch number $k = 1, ..., p$. The Δ signs are dropped and the clock parameters are denoted by:

k, p

$$\delta r_r(k) \leftarrow c \cdot \Delta\delta t_r(t);$$
$$\delta s^s(k) \leftarrow c \cdot \Delta\delta t^s(t - \tau_r^s).$$

The 1×3 matrix that describes the geometry is written as:

$$G_r^s(k) \leftarrow -\frac{r_r^s(t)'}{\parallel r_r^s(t) \parallel}.$$

$\mu_{,l}$ The dispersion of the ionospheric delay will be caught in a separate factor $\mu_{,l}$:

$$\mu_{,l} \cdot \imath_r^s(k) \leftarrow \Delta\imath_{r,l}^s(t),$$

where

$$
\begin{aligned}
\mu_{,1} &\doteq \mu_1 \doteq \lambda_1/\lambda_2 = 120/154; \\
\mu_{,2} &\doteq \mu_2 \doteq \lambda_2/\lambda_1 = 154/120; \\
\mu_{,3} &\doteq -\mu_1 \doteq -\lambda_1/\lambda_2 = -120/154; \\
\mu_{,4} &\doteq -\mu_2 \doteq -\lambda_2/\lambda_1 = -154/120,
\end{aligned}
\tag{10.9}
$$

and

$$\imath_r^s(k) \doteq \mu_2 \cdot \imath_{r,1}^s(k) \doteq \mu_1 \cdot \imath_{r,2}^s(k). \tag{10.10}$$

So, we introduce an ionosphere parameter that is neither an ionospheric delay (advance) at L1 nor at L2.

Introducing all of these notational simplifications in Eq. (10.6) then gives the following linearized observation equation:

$$
\begin{aligned}
\mathbf{E}\{\underline{y}_{r,l}^s(k)\} = \ & G_r^s(k) \cdot r_r && (coordinates) \\
& + D_r^s(k) && (troposphere) \\
& + \mu_{,l} \cdot \imath_r^s(k) && (ionosphere) \\
& + \delta r_r(k) && (rec.\ clock\ errors) \\
& - \delta s^s(k) && (sat.\ clock\ errors) \\
& + br_{r,l}(k) && (rec.\ hardware\ delays) \\
& - bs_{,l}^s(k) && (sat.\ hardware\ delays) \\
& + \lambda_{,l} \cdot N_{r,l}^s && (ambiguities) \\
& + \lambda_{,l} \cdot \phi r_{r,l} && (rec.\ init.\ phase) \\
& - \lambda_{,l} \cdot \phi s_{,l}^s && (sat.\ init.\ phase).
\end{aligned}
\tag{10.11}
$$

10.3 Matrix-vector notation

Next, we derive the functional observation model displayed in matrix-vector notation. For simplicity, we assume that at all receivers, measurements are obtained to the same satellites, $m \equiv m_1 = \cdots = m_n$, and that m does not change in time m (the latter assumption is relaxed in Part IV). For the vector of all m observables from receiver r of observation type l at epoch k, the observation equations then read:

$$
\begin{aligned}
\mathbf{E}\{\underline{y}_{r,l}(k)\} = \ & G_r(k) \cdot r_r && (coordinates) \\
& + I_m \cdot D_r(k) && (troposphere) \\
& + \mu_{,l} \cdot I_m \cdot \imath_r(k) && (ionosphere) \\
& + e_m \cdot \delta r_r(k) && (rec.\ clock\ errors) \\
& - I_m \cdot \delta s(k) && (sat.\ clock\ errors) \\
& + e_m \cdot br_{r,l}(k) && (rec.\ hardware\ delays) \\
& - I_m \cdot bs_{,l}(k) && (sat.\ hardware\ delays) \\
& + \lambda_{,l} \cdot I_m \cdot N_{r,l} && (ambiguities) \\
& + \lambda_{,l} \cdot e_m \cdot \phi r_{r,l} && (rec.\ init.\ phase) \\
& - \lambda_{,l} \cdot I_m \cdot \phi s_{,l} && (sat.\ init.\ phase),
\end{aligned}
\tag{10.12}
$$

I_m, e_m where I_m is an $m \times m$ identity matrix, e_m is an $m \times 1$ vector with all ones and the

superscript ^{s} is dropped to denote all satellites. This is done for the observables

$$y_{r,l}(k) \doteq [y_{r,l}^1(k), \ldots, y_{r,l}^m(k)]' \equiv y_{r,l}^{1..m}(k), \tag{10.13}$$

as well as for the satellite-dependent parameters (tropospheric and ionospheric delays, satellite clock errors, hardware delays, initial phases, and ambiguities):

$$\begin{aligned}
D_r(k) &\equiv D_r^{1..m}(k) &; \quad \delta s(k) &\equiv \delta s^{1..m}(k) &; \quad \phi s_{,l} &\equiv \phi s_{,l}^{1..m} &; \\
\imath_r(k) &\equiv \imath_r^{1..m}(k) &; \quad bs_{,l}(k) &\equiv bs_{,l}^{1..m}(k) &; \quad N_{r,l} &\equiv N_{r,l}^{1..m} &.
\end{aligned} \tag{10.14}$$

The scalar parameters of Eq. (10.11) are replaced by these $m \times 1$ vectors; the equations are kept consistent by placing the matrices I_m and e_m. For the receiver clock errors, hardware delays, and initial phases we used the vector e_m to keep the equations consistent, since they are not satellite dependent, and the identity matrix I_m is placed in front of the satellite-dependent parameters, whereas the geometry matrix $G_r(k)$ replaces $G_r^s(k)$:

$$G_r(k) \doteq [G_r^1(k)', \ldots, G_r^m(k)']' \ (m \times 3). \tag{10.15}$$

Generalizing the observation equations for all n stations gives an $mn \times 1$ vector of observables of observation type l at epoch k, using the Kronecker product $\otimes$

⊗ Kronecker [*Rao*, 1973], see App. D:

$$\begin{aligned}
\mathbf{E}\{\underline{y}_{,l}(k)\} = \quad & G(k) \cdot r & (coordinates) \\
+ \quad & [I_n \otimes I_m] \cdot D(k) & (troposphere) \\
+ \quad & \mu_{,l} \cdot [I_n \otimes I_m] \cdot \imath(k) & (ionosphere) \\
+ \quad & [I_n \otimes e_m] \cdot \delta r(k) & (rec.\ clock\ errors) \\
+ \quad & [-e_n \otimes I_m] \cdot \delta s(k) & (sat.\ clock\ errors) \\
+ \quad & [I_n \otimes e_m] \cdot br_{,l}(k) & (rec.\ hardware\ delays) \\
+ \quad & [-e_n \otimes I_m] \cdot bs_{,l}(k) & (sat.\ hardware\ delays) \\
+ \quad & \lambda_{,l} \cdot [I_n \otimes I_m] \cdot N_{,l} & (ambiguities) \\
+ \quad & \lambda_{,l} \cdot [I_n \otimes e_m] \cdot \phi r_{,l} & (rec.\ init.\ phase) \\
+ \quad & \lambda_{,l} \cdot [-e_n \otimes I_m] \cdot \phi s_{,l} & (sat.\ init.\ phase),
\end{aligned} \tag{10.16}$$

with

$$G(k) \doteq \text{diag}[G_1(k), \ldots, G_n(k)] \ (mn \times 3n), \tag{10.17}$$

and

$$\begin{aligned}
y_{,l}(k) &\equiv y_{1..n,l}(k) \ (mn \times 1) &; \quad \delta r(k) &\equiv \delta r_{1..n}(k) & (n \times 1) &; \\
r &\equiv r_{1..n} & (3n \times 1) &; \quad br_{,l}(k) &\equiv br_{1..n,l}(k) & (n \times 1) &; \\
D(k) &\equiv D_{1..n}(k) \ (mn \times 1) &; \quad \phi r_{,l} &\equiv \phi r_{1..n,l} & (n \times 1) &; \\
\imath(k) &\equiv \imath_{1..n}(k) \ (mn \times 1) &; \quad N_{,l} &\equiv N_{1..n,l} & (mn \times 1) &.
\end{aligned} \tag{10.18}$$

The index $_r$ is dropped (to denote all receivers) for the observables, the geometry matrix, and all receiver-dependent parameters.

We kept the equations consistent by placing the vector e_n for the satellite-related parameters and the matrix I_n for the other (nongeometry) parameters.

The matrix-vector notation can be taken a step further by collecting all (four) observable types for epoch k. The observation equations then read (see

Fig. 10.1):

$$
\begin{aligned}
\mathbf{E}\{\underline{y}_{,}(k)\} = \quad & [e_4 \otimes G(k)] \cdot r \quad && (\textit{coordinates}) \\
+ \quad & [e_4 \otimes I_n \otimes I_m] \cdot D(k) && (\textit{troposphere}) \\
+ \quad & [\mu \otimes I_n \otimes I_m] \cdot \iota(k) && (\textit{ionosphere}) \\
+ \quad & [e_4 \otimes I_n \otimes e_m] \cdot \delta r(k) && (\textit{rec. clock errors}) \\
+ \quad & [e_4 \otimes -e_n \otimes I_m] \cdot \delta s(k) && (\textit{sat. clock errors}) \\
+ \quad & [I_4 \otimes I_n \otimes e_m] \cdot br(k) && (\textit{rec. hardware delays}) \\
+ \quad & [I_4 \otimes -e_n \otimes I_m] \cdot bs(k) && (\textit{sat. hardware delays}) \\
+ \quad & [\Lambda \otimes I_n \otimes I_m] \cdot N && (\textit{ambiguities}) \\
+ \quad & [\Lambda \otimes I_n \otimes e_m] \cdot \phi r && (\textit{rec. init. phase}) \\
+ \quad & [\Lambda \otimes -e_n \otimes I_m] \cdot \phi s && (\textit{sat. init. phase}),
\end{aligned}
\tag{10.19}
$$

where

$$
\begin{aligned}
y_{,}(k) &\equiv y_{,1..4}(k) && (4mn \times 1); \\
br(k) &\equiv br_{,1..4}(k) && (4n \times 1); \\
bs(k) &\equiv bs_{,1..4}(k) && (4m \times 1); \\
N &\equiv N_{,1..2} && (2mn \times 1); \\
\phi r &\equiv \phi r_{,1..2} && (2n \times 1); \\
\phi s &\equiv \phi s_{,1..2} && (2m \times 1); \\
\mu &\equiv \mu_{,1..4} && (4 \times 1); \\
\Lambda &\doteq \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} && (4 \times 2).
\end{aligned}
\tag{10.20}
$$
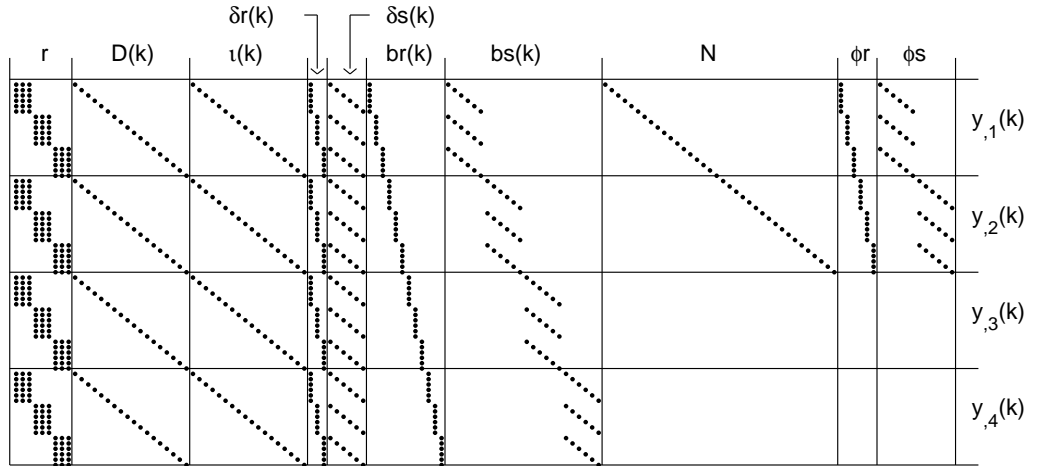
Figure 10.1   Sparsity structure of rank-deficient design matrix of Eq. (10.19) for $m = 6$ satellites and $n = 3$ stations. Corresponding parameters are shown at the top; corresponding observations are shown at the right.

$y_{,}(k)$ Note that we left the comma as index of the 'stripped' $y_{,}(k)$ vector. This notation is used for the GPS observables. We may also add pseudo-observables to our observation equations. A vector with both GPS observables and pseudo-observables for epoch k is written as $y(k)$ without comma. Since there are no pseudo-observables introduced yet, $y_{,}(k) \equiv y(k)$ for the moment.

The stochastic model of the observables at epoch k is written symbolically

as:

$$\mathbf{D}\{\underline{y}_,(k)\} = Q_{y_,}(k), \tag{10.21}$$

$\mathbf{D}\{.\}$ where $\mathbf{D}\{.\}$ denotes the mathematical dispersion.

The last step in deriving a model in matrix-vector notation is collecting all epochs. The vector of all GPS observables then reads:

$$\underline{y}_, \equiv \underline{y}_,(1..p) \quad (4mnp \times 1). \tag{10.22}$$

We assume there is no correlation between the epochs:

$$\mathbf{D}\{\underline{y}_,\} = \mathrm{diag}[Q_{y_,}(1), \ldots, Q_{y_,}(p)]. \tag{10.23}$$

global parameters

temporal parameters

We can distinguish global and local parameters, as suggested in [*De Jonge*, 1998]. Global parameters are considered constant over the complete time span whereas local parameters are estimated for each epoch. Instead of local parameters we will speak of temporal parameters, which gives a better feel for the behavior of these parameters in time. Coordinates, ambiguities, and initial phases are considered to be global parameters; clock parameters and tropospheric, ionospheric, and hardware delays are considered to be temporal parameters. With these assumptions the complete functional model for all epochs now reads:

$$
\begin{array}{rlll}
\mathbf{E}\{\underline{y}_,\} = & & G \cdot r & (coordinates) \\
& + [I_p \otimes e_4 \otimes & I_n \otimes I_m] \cdot D\,(1..p) & (troposphere) \\
& + [I_p \otimes \mu \otimes & I_n \otimes I_m] \cdot \imath\,(1..p) & (ionosphere) \\
& + [I_p \otimes e_4 \otimes & I_n \otimes e_m] \cdot \delta r\,(1..p) & (rec.\ clock\ errors) \\
& + [I_p \otimes e_4 \otimes & -e_n \otimes I_m] \cdot \delta s\,(1..p) & (sat.\ clock\ errors) \\
& + [I_p \otimes I_4 \otimes & I_n \otimes e_m] \cdot br\,(1..p) & (rec.\ hardware\ delays) \\
& + [I_p \otimes I_4 \otimes & -e_n \otimes I_m] \cdot bs\,(1..p) & (sat.\ hardware\ delays) \\
& + [e_p \otimes \Lambda \otimes & I_n \otimes I_m] \cdot N & (ambiguities) \\
& + [e_p \otimes \Lambda \otimes & I_n \otimes e_m] \cdot \phi r & (rec.\ init.\ phase) \\
& + [e_p \otimes \Lambda \otimes & -e_n \otimes I_m] \cdot \phi s & (sat.\ init.\ phase),
\end{array}
\tag{10.24}
$$

where

$$G \doteq [e_4' \otimes G(1)', \ldots, e_4' \otimes G(p)']' \quad (4mnp \times 3n). \tag{10.25}$$

For the global parameters we introduced the vector e_p, and for the temporal parameters we used the matrix I_p. A very similar model can be derived when only the phase observables are used. The model then reads:

$$
\begin{array}{rlll}
\mathbf{E}\{\underline{y}_,\} = & & G_2 \cdot r & (coordinates) \\
& + [I_p \otimes & e_2 & \otimes & I_n \otimes I_m] \cdot D\,(1..p) & (troposphere) \\
& + \left[I_p \otimes \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \otimes & I_n \otimes I_m\right] \cdot \imath\,(1..p) & (ionosphere) \\
& + [I_p \otimes & e_2 & \otimes & I_n \otimes e_m] \cdot \delta r\,(1..p) & (rec.\ clock\ errors) \\
& + [I_p \otimes & e_2 & \otimes & -e_n \otimes I_m] \cdot \delta s\,(1..p) & (sat.\ clock\ errors) \\
& + [I_p \otimes & I_2 & \otimes & I_n \otimes e_m] \cdot br\,(1..p) & (rec.\ hardware\ delays) \\
& + [I_p \otimes & I_2 & \otimes & -e_n \otimes I_m] \cdot bs\,(1..p) & (sat.\ hardware\ delays) \\
& + [e_p \otimes & L & \otimes & I_n \otimes I_m] \cdot N & (ambiguities) \\
& + [e_p \otimes & L & \otimes & I_n \otimes e_m] \cdot \phi r & (rec.\ init.\ phase) \\
& + [e_p \otimes & L & \otimes & -e_n \otimes I_m] \cdot \phi s & (sat.\ init.\ phase),
\end{array}
\tag{10.26}
$$

where:

$$
\begin{aligned}
G_2 &\doteq [e_2' \otimes G(1)', \ldots, e_2' \otimes G(p)']' & (2mnp \times 3n); \\
y, &\equiv y,(1..p) & (2mnp \times 1); \\
br(k) &\equiv br,_{1..2}(k) & (2n \times 1); \\
bs(k) &\equiv bs,_{1..2}(k) & (2m \times 1); \\
L &\doteq \mathrm{diag}[\lambda_1, \lambda_2] & (2 \times 2).
\end{aligned}
$$

$\qquad\qquad$ (10.27)

Both models contain rank deficiencies. In other words, there is more than one solution to these equations. In the next chapter, we start with a simpler model where the tropospheric delays are considered to be a priori corrected for, and derive a full-rank model with estimable parameters.

Eliminating rank deficiencies in troposphere-fixed models

11.1   Introduction

In troposphere-fixed models, the GPS observations are a priori corrected for tropospheric delays. This means that instead of parameterizing the tropospheric delays, as in a troposphere-float model, some a-priori model, like Saastamoinen's [*Saastamoinen*, 1972], is used to correct the observations (for an elaborate discussion on these kinds of models, see Part I). In Eq. (10.1), the term $\tilde{y}_{r,l}^s(t)$ includes such a correction. Although this correction term may also be used in a troposphere-float model, in troposphere-fixed models this a-priori model is assumed errorless. The troposphere-fixed models are typically applicable for small baselines, whereas troposphere-float models are suitable for long baselines.

In this chapter, full-rank troposphere-fixed models are described. In Sect. 11.2 we point out the rank-deficient parts of the design matrix and describe full-rank models. Since the troposphere-float model is an extension of the troposphere-fixed model, these rank deficiencies are common for both types of models.

The definitions for 'fixed' and 'float' can also be applied for the ionosphere [*Odijk*, 2000]. Our troposphere-fixed model for the moment will in fact be an ionosphere-float model. The ionosphere-fixed and ionosphere-weighted models are briefly discussed in Sect. 11.3. The full-rank partial design matrices are shown in Tables 11.1–11.3. In Sect. 11.4 we show the redundancy that is left to model the troposphere. Finally, in Sect. 11.5, we show how the partial design matrices can be gathered in matrix blocks for global, temporal, and batch parameters.

11.2   Eliminating basic rank deficiencies (ionosphere-float)

A_1, A_2
V

For a rank-deficient matrix $A = [A_1, A_2, A_3]$, where the columns of A_1 are a linear combination of the columns of A_2, for some $V \neq 0$ holds:

$$\boxed{A_1 = A_2 V.} \tag{11.1}$$

Eliminating the columns of A_1 leaves a design matrix $[A_2, A_3]$ with a smaller rank deficiency. The linear combination $y = A_1 \mathrm{x}_1 + A_2 \mathrm{x}_2 + A_3 \mathrm{x}_3$ then transforms into $y = A_2 \overline{\mathrm{x}} + A_3 \mathrm{x}_3$, where

$\overline{\mathrm{x}}$

$$\boxed{\overline{\mathrm{x}} \leftarrow \mathrm{x}_2 + V \mathrm{x}_1.} \tag{11.2}$$

The parameters x_3 are left unchanged and new parameters $\bar{x}$ are formed. If $[A_2, A_3]$ is of full rank, the parameters $\bar{x}$ are estimable, otherwise we need to continue reparameterizing.

We have successively discriminated columns in the design matrix that are dependent on other sets of columns, labeled these parts of the design matrix by 'A_1' and 'A_2' respectively, found a matrix V for which Eq. (11.1) holds, and gave explicit expressions of the newly formed lumped parameters $\bar{x}$. Note that we may shuffle the columns of the design matrix as we wish without altering the model, as long as the parameters are shuffled in the same way. Although the choices we make are arbitrary, other full-rank models are easier to find when we have performed this exercise once. According to Eq. (10.24) there are ten types of parameters. Because in this chapter we assume the troposphere parameters to be known and because there is no (real) rank deficiency related to the geometry part of the model, we concentrate on the remaining eight parameter groups.

The steps we need to take to determine a full-rank model with corresponding parameters are given in App. E. In Table E.3 this is done for the model with phase and code observations. Table E.4 shows a similar procedure in case only the phase observables are used. A graphical representation of the steps with the resulting parameters is given in Figs. E.1 and E.2, next to the tables. Tables E.1 and E.2 show the lumped parameters in words, rather than in mathematical notation.

In our notation, lumped parameters are usually denoted by an overbar ($\bar{x}$) suggesting that they have no significant meaning, but whenever Single Differences (SD) are formed, we use the notation $x_{12..n}$ for $x_{2..n} - x_1$ or $x^{12..m}$ for $x^{2..m} - x^1$. Undifferenced or Zero Differenced (ZD) parameters usually have no index unless it is considered useful. Double Differenced (DD) parameters (ambiguities) have both a lower and upper index.

Tables E.3 and E.4 show the steps to be taken when the satellite hardware delays:
(a) change in time and differ for the observation types, as assumed in Eqs. (10.24) and (10.26), and when we assume them to be stable;
(b) are common for the L1 and L2 phase observables as well as for the P1 and P2 code observables (which is a reasonable assumption [*Han et al.*, 2001]).
In the latter case we have for respectively the phase + code Model (10.24) and the phase-only Model (10.26):

$$
\begin{aligned}
&[e_p \otimes J \otimes -e_n \otimes I_m] \cdot \begin{bmatrix} bs_{,phase} \\ bs_{,code} \end{bmatrix} ; \\
&[e_p \otimes e_2 \otimes -e_n \otimes I_m] \cdot \quad bs_{,phase} \quad ,
\end{aligned}
\tag{11.3}
$$

J using the definition

$$
J \doteq [J_1, \ J_2] \doteq \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}.
\tag{11.4}
$$

C_n, c_n In the tables, we repeatedly use the matrices

$$C_n \atop {n \times (n-1)} \doteq \begin{bmatrix} 0 \\ 1 \\ & \ddots \\ & & 1 \end{bmatrix} \quad \text{and} \quad c_n \atop {n \times 1} \doteq \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{11.5}$$

for which the following property holds:

$$\boxed{e_n - C_n e_{n-1} = c_n.} \tag{11.6}$$

Below, we give some examples to clarify the tables, and define some matrices along the way.

Example 1:   *Receiver clocks and satellite clocks (4th step)*

In Eq. (10.24), for each epoch there are n receiver clock errors $\delta r(k) \equiv \delta r_{1..n}(k)$ and m satellite clock errors $\delta s(k) \equiv \delta s^{1..m}(k)$. According to Eq. (10.16), the part of the model that corresponds to these parameters looks like:

$$y_{,l}(k) = [I_n \otimes e_m, \; -e_n \otimes I_m] \cdot \begin{bmatrix} \delta r(k) \\ \delta s(k) \end{bmatrix} + A_3 \cdot \mathrm{x}_3, \tag{11.7}$$

with x_3 the remaining parameters. The matrix $[I_n \otimes e_m, \; -e_n \otimes I_m]$ is rank deficient and we may write Eq. (11.7) as

$$y_{,l}(k) - A_3 \cdot \mathrm{x}_3 = \underbrace{[-e_n \otimes c_m,}_{A_1} \; \underbrace{I_n \otimes e_m, -e_n \otimes C_m]}_{A_2} \cdot \begin{bmatrix} \delta s^1(k) \\ \delta r(k) \\ \delta s^{2..m}(k) \end{bmatrix} \begin{matrix} \} \mathrm{x}_1 \\ \} \mathrm{x}_2 \\ \} \end{matrix} . \tag{11.8}$$

If we use the form of Eq. (11.1), we obtain:

$$\underbrace{\left[I_n \otimes e_m, -e_n \otimes C_m \right]}_{A_2} \underbrace{\begin{bmatrix} -e_n \otimes & 1 \\ 1 & \otimes -e_{m-1} \end{bmatrix}}_{V} = \underbrace{[-e_n \otimes c_m]}_{A_1}, \tag{11.9}$$

where we made use of Eqs. (D.6) and (11.6).

The column $[-e_n \otimes c_m]$ corresponds to the satellite clock error of the first satellite. Eliminating this column from the design matrix reduces the rank
$\mathcal{D}$ deficiency. The matrix $[A_1, \; A_2]$ is now replaced by $\mathcal{D} \equiv A_2$,

$$\boxed{\mathcal{D} \doteq \left[I_n \otimes e_m, -e_n \otimes C_m \right]; \qquad (mn \times m + n - 1),} \tag{11.10}$$

to obtain a model with a lower rank deficiency (there may still be a rank deficiency in A_3). Each time a new partial design matrix is formed, in this case $[-e_n \otimes C_m]$, this is indicated in the tables.

Both receiver and satellite clock errors are lumped by this operation and
$\overline{\delta r}(k),$ the lumped clock parameters are denoted by $\overline{\delta r}(k)$ and $\delta s^{12..m}(k)$:
$\delta s^{12..m}(k)$

$$\underbrace{\begin{bmatrix} \overline{\delta r}(k) \\ \delta s^{12..m}(k) \end{bmatrix}}_{\overline{\mathrm{x}}} \leftarrow \underbrace{\begin{bmatrix} \delta r(k) \\ \delta s^{2..m}(k) \end{bmatrix}}_{\mathrm{x}_2} + \underbrace{\begin{bmatrix} -e_n \\ -e_{m-1} \end{bmatrix}}_{V} \cdot \underbrace{\delta s^1(k)}_{\mathrm{x}_1}. \tag{11.11}$$

Example 2:   *Receiver clocks and ionosphere (6th step)*

After the receiver clocks are lumped with satellite clocks as in Example 1, they are lumped with the hardware delays. Because the hardware delays differ for the observation types, the lumped receiver clocks also differ for the observation types. In the 6th step in Table E.3 we have the following model elements:

$$[I_p \otimes I_4 \otimes I_n \otimes e_m] \cdot \overline{\delta r}(1..p) \quad (rec.\ clocks);$$
$$[I_p \otimes \mu \otimes I_n \otimes I_m] \cdot \imath_{,1..4}(1..p)\ (ionosphere).$$

The rank deficiency is contained in the second matrices of these Kronecker products (I_4 and μ, respectively). We may split the identity matrix into $I_4 = [c_4,\ C_4]$, and see that:

$$\nu$$

$$\underbrace{\begin{bmatrix} 0 \\ 1 \\ & 1 \\ & & 1 \end{bmatrix}}_{C_4} \underbrace{\begin{bmatrix} -\mu_2^2 \\ 1 \\ \mu_2^2 \end{bmatrix}}_{\nu} + \mu_2 \underbrace{\begin{bmatrix} \mu_1 \\ \mu_2 \\ -\mu_1 \\ -\mu_2 \end{bmatrix}}_{\mu} = \underbrace{\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{c_4}. \tag{11.12}$$

The remaining partial design matrices in

$$A_2 = [\underbrace{I_p \otimes C_4 \otimes I_n \otimes e_m}_{\text{rec. clocks}},\ \underbrace{I_p \otimes \mu \otimes I_n \otimes I_m}_{\text{ionosphere}}]$$

do not change in the lumping steps that follow. The parameters $\overline{\delta r}_{,2..4}(1..p)$ and $\imath(1..p)$ are lumped with the L1 receiver clock errors $\overline{\delta r}_{,1}(1..p)$. Although we already used the notation with a bar for the lumped receiver clock parameters (see Example 1), we continue to use this notation for newly lumped receiver clock parameters because these parameters have no physical meaning and are considered nuisance parameters.

Example 3:   *Ambiguities and hardware delays (steps 3 and 7a)*

$$\mathcal{A}$$

In Table E.3, the integer ambiguities are first lumped with the receiver and satellite initial phases. The lumped ambiguities do not have an integer nature and are denoted by $\mathcal{A}$. By writing down this vector elementwise, we see that each element reads:

$$\mathcal{A}_{r,l}^s = N_{r,l}^s + \phi r_{r,l} - \phi s_{,l}^s. \tag{11.13}$$

A vector with double-differenced ambiguities remains once we eliminate the rank deficiency in the ambiguity part of the design matrix caused by the dependencies with receiver and satellite hardware delays[1] (case (a) in Table E.3). These double-differenced ambiguities have an integer nature, since:

double-differenced ambiguities

$$\begin{aligned} \mathcal{A}_{1r,l}^{1s} &= \mathcal{A}_{1r,l}^s - \mathcal{A}_{1r,l}^1 \\ &= (\mathcal{A}_{r,l}^s - \mathcal{A}_{1,l}^s) - (\mathcal{A}_{r,l}^1 - \mathcal{A}_{1,l}^1) \\ &= +N_{r,l}^s + \phi r_{r,l} - \phi s_{,l}^s - N_{1,l}^s - \phi r_{1,l} + \phi s_{,l}^s \\ &\quad -N_{r,l}^1 - \phi r_{r,l} + \phi s_{,l}^1 + N_{1,l}^1 + \phi r_{1,l} - \phi s_{,l}^1 \\ &= (N_{r,l}^s - N_{1,l}^s) - (N_{r,l}^1 - N_{1,l}^1) \\ &= N_{1r,l}^{1s}. \end{aligned} \tag{11.14}$$

End of Examples

---

[1] We chose the first receiver and satellite as pivots.

In each of the four cases in the tables, only four out of the eight original parameter groups remain after the final step in the full-rank model: ambiguities, receiver and satellite clocks, and ionosphere parameters. These matrices and parameters are given in Table 11.1. All parameters are lumped and the notation for the parameters is more symbolic than strict. The ionosphere parameters in case (a), for example, are not the same parameters as in case (b), because in case (b) these parameters are lumped one more time. The ambiguities in case (b) are not integer, nor are the ambiguities in the ionophere-float model with phase observations only.

11.3 Ionosphere-fixed and weighted models

11.3.1 The ionosphere-fixed model

In small networks, the ionospheric delays are nearly equal for all stations and may in fact be considered equal. In this case the ionosphere part of the model reads

$$
\begin{aligned}
&[I_p \otimes \mu \otimes e_n \otimes I_m] &\cdot \imath(1..p) \text{ and} \\
&\left[I_p \otimes \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \otimes e_n \otimes I_m \right] \cdot \imath(1..p)
\end{aligned}
$$

for the model with phase and code observables, and the model with phase observables only, respectively. This has consequences for the lumping of the parameters. The steps 6 and 9a in Table E.3 and the steps 6, 7, 8a, and 10a in Table E.4 can no longer be applied. Instead, in all four cases[2] the ionosphere part of the model can be eliminated and the satellite clocks are lumped with the ionosphere parameters. The design matrices and corresponding parameters are given in Table 11.2.

11.3.2 The ionosphere-weighted model

The ionosphere-weighted model [*Odijk*, 2002] is an extension of the ionosphere-float model. The model is suited for medium scaled networks with typical baseline lengths between 10 and 200 km. If we have a-priori information about the ionospheric delays, we may add this information in the form of pseudo-observables. The vector of observables and corresponding covariance matrix for epoch k then read

$$
\underline{y}(k) = \begin{bmatrix} \underline{y}_{,}(k) \\ \underline{\imath}(k) \end{bmatrix} \quad ; \quad Q_y(k) = \begin{bmatrix} Q_{y,}(k) & 0 \\ 0 & Q_\imath(k) \end{bmatrix}. \tag{11.15}
$$

Because of the extra constraints, the ionosphere parameters cannot be lumped with other parameters. The steps in Table E.3 and E.4 that involve the ionosphere parameters are therefore not applicable for the ionosphere-weighted model. The design matrices and corresponding parameters are given in Table 11.3.

11.4 Redundancy

Table 11.4 gives the number of observables, parameters, and the corresponding redundancy: for the three ionosphere models, for the two cases (a) and (b), and for phase + code observables and phase observables only.

The redundancies for all twelve models are also plotted in Fig. 11.1. The redundancies of the phase-only cases are shown here because the strength of the model

---

[2]Phase only (a, b), phase + code (a, b).

obs. type	(a) design matrix	(a) parameters	(b) design matrix	(b) parameters
L+P	$[e_p \otimes \Lambda \otimes C_n \otimes C_m]$	$N_{12..n}^{12..m}$	$[e_p \otimes \Lambda \otimes I_n \otimes C_m]$	$\bar{\mathcal{A}}^{12..m}$
	$[I_p \otimes C_4 \otimes I_n \otimes e_m]$	$\overline{\delta r}_{,2..4}(1..p)$	$[I_p \otimes C_4 \otimes I_n \otimes e_m]$	$\overline{\delta r}_{,2..4}(1..p)$
	$[I_p \otimes C_4 \otimes -e_n \otimes C_m]$	$\overline{\delta s}_{,2..4}^{12..m}(1..p)$	$[I_p \otimes e_4 \otimes -e_n \otimes C_m]$	$\overline{\delta s}^{12..m}(1..p)$
	$[I_p \otimes \mu \otimes I_n \otimes I_m]$	$\bar{\imath}(1..p)$	$[I_p \otimes \mu \otimes I_n \otimes I_m]$	$\bar{\imath}(1..p)$
L	$[e_p \otimes \binom{0}{\lambda_2} \otimes C_n \otimes C_m]$	$\bar{\mathcal{A}}_{12..n,2}^{12..m}$	$[e_p \otimes \binom{0}{\lambda_2} \otimes I_n \otimes C_m]$	$\bar{\mathcal{A}}_{,2}^{12..m}$
	$[I_p \otimes C_2 \otimes I_n \otimes e_m]$	$\overline{\delta r}_{,2}(1..p)$	$[I_p \otimes C_2 \otimes I_n \otimes e_m]$	$\overline{\delta r}_{,2}(1..p)$
	$[I_p \otimes C_2 \otimes -e_n \otimes C_m]$	$\overline{\delta s}_{,2}^{12..m}(1..p)$	$[I_p \otimes e_2 \otimes -e_n \otimes C_m]$	$\overline{\delta s}^{12..m}(1..p)$
	$[I_p \otimes \binom{\mu_1}{\mu_2} \otimes I_n \otimes I_m]$	$\bar{\imath}(1..p)$	$[I_p \otimes \binom{\mu_1}{\mu_2} \otimes I_n \otimes I_m]$	$\bar{\imath}(1..p)$

Table 11.1   Partial design matrices and parameters of full-rank troposphere-fixed and ionosphere-float models. (a): changing satellite hardware delays (different for each observation type) and (b): stable satellite hardware delays (common values for L1 and L2 and for P1 and P2). Top: Phase and code observables. Bottom: Phase observables only.

obs. type	(a) design matrix	(a) parameters	(b) design matrix	(b) parameters
L+P	$[e_p \otimes \Lambda \otimes C_n \otimes C_m]$	$N_{12..n}^{12..m}$	$[e_p \otimes \Lambda \otimes I_n \otimes C_m]$	$\bar{\mathcal{A}}^{12..m}$
	$[I_p \otimes I_4 \otimes I_n \otimes e_m]$	$\overline{\delta r}(1..p)$	$[I_p \otimes I_4 \otimes I_n \otimes e_m]$	$\overline{\delta r}(1..p)$
	$[I_p \otimes I_4 \otimes -e_n \otimes C_m]$	$\overline{\delta s}^{12..m}(1..p)$	$[I_p \otimes e_4 \otimes -e_n \otimes C_m]$	$\overline{\delta s}^{12..m}(1..p)$
L	$[e_p \otimes L \otimes C_n \otimes C_m]$	$N_{12..n}^{12..m}$	$[e_p \otimes L \otimes I_n \otimes C_m]$	$\mathcal{A}^{12..m}$
	$[I_p \otimes I_2 \otimes I_n \otimes e_m]$	$\overline{\delta r}(1..p)$	$[I_p \otimes I_2 \otimes I_n \otimes e_m]$	$\overline{\delta r}(1..p)$
	$[I_p \otimes I_2 \otimes -e_n \otimes C_m]$	$\overline{\delta s}^{12..m}(1..p)$	$[I_p \otimes e_2 \otimes -e_n \otimes C_m]$	$\overline{\delta s}^{12..m}(1..p)$

Table 11.2   Partial design matrix and parameters of full-rank troposphere-fixed and ionosphere-fixed models. (a): changing satellite hardware delays (different for each observation type) and (b): stable satellite hardware delays (common values for L1 and L2 and for P1 and P2). Top: Phase and code observables. Bottom: Phase observables only.

obs. type	(a) design matrix	(a) parameters	(b) design matrix	(b) parameters
L+P	$[e_p \otimes \binom{\Lambda}{0} \otimes C_n \otimes C_m]$	$N_{12..n}^{12..m}$	$[e_p \otimes \binom{\Lambda}{0} \otimes I_n \otimes C_m]$	$\bar{\mathcal{A}}^{12..m}$
	$[I_p \otimes \binom{I_4}{0} \otimes I_n \otimes e_m]$	$\overline{\delta r}(1..p)$	$[I_p \otimes \binom{I_4}{0} \otimes I_n \otimes e_m]$	$\overline{\delta r}(1..p)$
	$[I_p \otimes \binom{I_4}{0} \otimes -e_n \otimes C_m]$	$\overline{\delta s}^{12..m}(1..p)$	$[I_p \otimes \binom{e_4}{0} \otimes -e_n \otimes C_m]$	$\overline{\delta s}^{12..m}(1..p)$
	$[I_p \otimes \binom{\mu}{1} \otimes I_n \otimes I_m]$	$\imath(1..p)$	$[I_p \otimes \binom{\mu}{1} \otimes I_n \otimes I_m]$	$\imath(1..p)$
L	$[e_p \otimes \binom{L}{0} \otimes C_n \otimes C_m]$	$N_{12..n}^{12..m}$	$[e_p \otimes \binom{L}{0} \otimes I_n \otimes C_m]$	$\mathcal{A}^{12..m}$
	$[I_p \otimes \binom{I_2}{0} \otimes I_n \otimes e_m]$	$\overline{\delta r}(1..p)$	$[I_p \otimes \binom{I_2}{0} \otimes I_n \otimes e_m]$	$\overline{\delta r}(1..p)$
	$[I_p \otimes \binom{I_2}{0} \otimes -e_n \otimes C_m]$	$\overline{\delta s}^{12..m}(1..p)$	$[I_p \otimes \binom{e_2}{0} \otimes -e_n \otimes C_m]$	$\overline{\delta s}^{12..m}(1..p)$
	$[I_p \otimes \begin{pmatrix}\mu_1\\\mu_2\\1\end{pmatrix} \otimes I_n \otimes I_m]$	$\imath(1..p)$	$[I_p \otimes \begin{pmatrix}\mu_1\\\mu_2\\1\end{pmatrix} \otimes I_n \otimes I_m]$	$\imath(1..p)$

Table 11.3   Partial design matrices and parameters of full-rank troposphere-fixed and ionosphere-weighted models. (a): changing satellite hardware delays (different for each observation type) and (b): stable satellite hardware delays (common values for L1 and L2 and for P1 and P2). Top: Phase and code observables. Bottom: Phase observables only.

L/P		ionosphere-float	ionosphere-fixed	ionosphere-weighted
L+P	observables	$4mnp$	$4mnp$	$5mnp$
	coordinates	$3n$	$3n$	$3n$
	ionosphere	mnp	0	mnp
	ambiguities (a)	$2(m{-}1)(n{-}1)$	$2(m{-}1)(n{-}1)$	$2(m{-}1)(n{-}1)$
	ambiguities (b)	$2(m{-}1)n$	$2(m{-}1)n$	$2(m{-}1)n$
	clocks (a)	$3(m{+}n{-}1)p$	$4(m{+}n{-}1)p$	$4(m{+}n{-}1)p$
	clocks (b)	$(3n{+}m{-}1)p$	$(4n{+}m{-}1)p$	$(4n{+}m{-}1)p$
	redundancy (a)	$(3p{-}2)(m{-}1)(n{-}1){-}3n$	$2(2p{-}1)(m{-}1)(n{-}1){-}3n$	$2(2p{-}1)(m{-}1)(n{-}1){-}3n$
	redundancy (b)	$((3n{-}1)p{-}2n)(m{-}1){-}3n$	$((4n{-}1)p{-}2n)(m{-}1){-}3n$	$((4n{-}1)p{-}2n)(m{-}1){-}3n$
L	observables	$2mnp$	$2mnp$	$3mnp$
	coordinates	$3n$	$3n$	$3n$
	ionosphere	mnp	0	mnp
	ambiguities (a)	$(m{-}1)(n{-}1)$	$2(m{-}1)(n{-}1)$	$2(m{-}1)(n{-}1)$
	ambiguities (b)	$(m{-}1)n$	$2(m{-}1)n$	$2(m{-}1)n$
	clocks(a)	$(m{+}n{-}1)p$	$2(m{+}n{-}1)p$	$2(m{+}n{-}1)p$
	clocks(b)	$(m{+}n{-}1)p$	$(2n{+}m{-}1)p$	$(2n{+}m{-}1)p$
	redundancy (a)	$(p{-}1)(m{-}1)(n{-}1){-}3n$	$2(p{-}1)(m{-}1)(n{-}1){-}3n$	$2(p{-}1)(m{-}1)(n{-}1){-}3n$
	redundancy (b)	$((n{-}1)p{-}n)(m{-}1){-}3n$	$((2n{-}1)p{-}2n)(m{-}1){-}3n$	$((2n{-}1)p{-}2n)(m{-}1){-}3n$

Table 11.4   Redundancy count of troposphere-fixed models. L+P: phase and code observables; L: phase observables only; (a): changing satellite clocks (and different for all observables); (b): stable satellite clocks (common values for L1 and L2 and for P1 and P2).

is dominated by the phase observables, which are in the order of a hundred times more accurate (standard deviation) than the pseudo-range observables. Nearly all information that is contained in the ionosphere-float model (a) to determine the coordinates is given by $(f-1)(p-1)(m-1)(n-1)$ four-times-differenced observables (with $f=2$ the number of frequencies), that is, by differences of observables between satellites, receivers, epochs, and frequencies. If we would also fully parameterize the tropospheric delays, our model would be under-determined as shown in Fig. 11.1. We therefore need to find sufficient (soft-)constraints to model the troposphere.

11.5   Global, temporal, and batch parameters

In Sect. 10.3, global and temporal parameters were introduced. In the models of the previous section, coordinates and ambiguities are global parameters, and clock and ionosphere parameters are temporal parameters. We will continue to use this distinction, and introduce a third parameter group: the batch parameters. Batch parameters have a hybrid nature. They describe a process for which the realizations change in time like the temporal parameters, but the changes are in general not large, so that we may assume them to be constant for several (say $\widetilde{p}$) epochs. Satellite hardware delays can very well be modeled by batch parameters. In fact, up till now we have considered these delays both global (case (a)) and temporal parameters (case (b)). In both cases the delays were lumped with other parameters.

batch
parameters

$\widetilde{p}$

x

β, t

If we gather all global parameters in the vector x, all batch parameters in the vector β, and all temporal parameters in the vector t, each GPS model
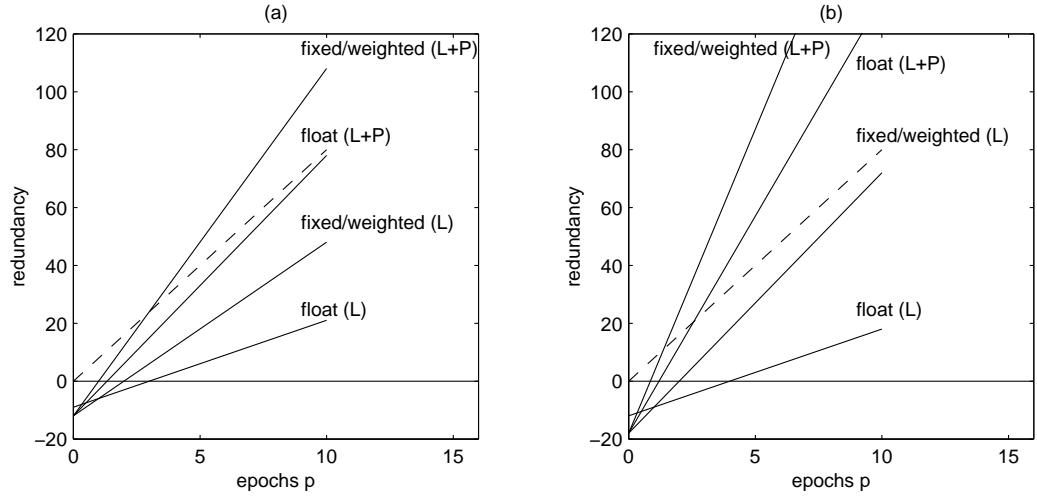
Figure 11.1   Solid lines: Redundancy of troposphere-fixed models ($m = 4$ satellites, $n = 2$ stations) for different ionosphere models (fixed, weighted, float) and phase-only (L) and phase+code observables (L+P). (a): changing satellite clocks (different for all observables); (b): stable satellite clocks (common values for L1 and L2 and for P1 and P2). Dashed: Number of tropospheric slant delays (mnp). Lines instead of points are plotted for clarity.

without pseudo-observables can be described by the general form

$$\mathbf{E}\{\underline{y}_{,}\} = Tt + B\beta + Ax \quad ; \quad \mathbf{D}\{\underline{y}_{,}\} = Q_{y,}, \tag{11.16}$$

where

$$
\begin{aligned}
A &\doteq [A(1)', \ldots, A(p)']'; \\
x &\doteq [x(1)', \ldots, x(p)']'; \\
\beta &\doteq [\beta(1)', \ldots, \beta(p/\widetilde{p})']'; \\
T &\doteq \mathrm{diag}[T(1), \ldots, T(p)]; \qquad \text{and} \quad B \doteq \begin{bmatrix} B(1) \\ \vdots \\ B(\widetilde{p}) \\ & B(\widetilde{p}+1) \\ & & \vdots \\ & & B(2\widetilde{p}) \\ & & & \ddots \end{bmatrix}. \\
t &\doteq [t(1)', \ldots, t(p)']'; \\
Q_{y,} &= \mathrm{diag}[Q_{y,}(1), \ldots, Q_{y,}(p)]; \\
p/\widetilde{p} &\in \mathbb{N}^{+},
\end{aligned}
\tag{11.17}
$$

As an example, the partial design matrices for the ionosphere-float model, case (a) $A(k), T(k)$ with phase and code observables read:

$$
\begin{aligned}
A(k) &= [\Lambda \otimes C_n \otimes C_m, \ e_4 \otimes G(k)] \text{ (global parameters)}; \\
T(k) &= [\mu \otimes I_n \otimes I_m, \ C_4 \otimes \mathcal{D}\,] \qquad \text{(temporal parameters)},
\end{aligned}
\tag{11.18}
$$

where we made use of the upper-left matrices from Table 11.1, the definition for $\mathcal{D}$ in Eq. (11.10), and the partial design matrix for the geometry as introduced in Eq. (10.19). The corresponding parameters are

$$
\begin{aligned}
x &= [N_{12..n}^{12..m'}, r']' \text{ (global parameters)}; \\
t(k) &= [\overline{\imath}(k)', \overline{\delta}(k)']' \text{ (temporal parameters)},
\end{aligned}
\tag{11.19}
$$

$\overline{\delta}(k)$ with all clock parameters in a vector:

$$\overline{\delta}(k) = [\overline{\delta r}_{,2..4}(k)', \overline{\delta s}_{,2..4}^{12..m}(k)']'. \tag{11.20}$$

Batch parameters are introduced in the next chapter, where we start to parameterize the tropospheric delay.

Troposphere-float and weighted models

12.1   Introduction

In troposphere-float models, the tropospheric delay is parameterized. Troposphere parameters may be zenith tropospheric delays, slant tropospheric delays, and troposphere gradients. These parameters may be global, temporal, or batch parameters. In the troposphere-weighted models, extra pseudo-observables for time-differenced zenith tropospheric delays are added to the troposphere-float models. In this chapter we only consider the ionosphere-float model, case (a), with phase and code observables. An extension to the other models can easily be made.

12.2   Zenith Tropospheric Delays

The total slant tropospheric delay is the sum of a hydrostatic and a wet slant delay. Both parts can be approximated by a mapping function multiplied with the delay in the zenith direction (isotropy assumption to determine the azimuthal symmetric slant delay; see Part I). So, for one slant delay between receiver r and satellite s at epoch k, we may write:

$$\boxed{\begin{aligned} D_r^s(k) &= M_h(z_r^s(k)) \cdot D_{rh}^z(k) + M_w(z_r^s(k)) \cdot D_{rw}^z(k) + \Delta D_r^s(k) \\ &\doteq D_{r,sym}^s(k) + \Delta D_r^s(k), \end{aligned}}$$

(12.1)

STD, ZHD, ZWD   where

$$\begin{aligned} D_r^s(k) \quad &: \text{Slant Total (Tropospheric) Delay (STD) [m];} \\ D_{r,sym}^s(k) \quad &: \text{azimuthal symmetric STD [m];} \\ \Delta D_r^s(k) \quad &: \text{residual STD [m];} \\ M_h(z_r^s(k)) \quad &: \text{hydrostatic mapping function [-];} \\ M_w(z_r^s(k)) \quad &: \text{wet mapping function [-];} \\ z_r^s(k) \quad &: \text{zenith angle [rad];} \\ D_{rh}^z(k) \quad &: \text{Zenith Hydrostatic Delay (ZHD) [m];} \\ D_{rw}^z(k) \quad &: \text{Zenith Wet Delay (ZWD) [m].} \end{aligned}$$

This equation holds for absolute delays, but also for the observed-minus-computed Δ quantities if we neglect the partial derivatives of the mapping functions with respect to the receiver coordinates. Since we have already derived a linearized model in absence of troposphere parameters and dropped the Δ signs, Eq. (12.1) will refer to the Δ quantities. With the definitions of Eq. (10.14) and Eq. (10.18)

we may write the slant delays in matrix-vector notation as:

$$
\begin{aligned}
D_r(k) &\doteq M_{hr}(k) \cdot D_{hr}^z(k) + M_{wr}(k) \cdot D_{wr}^z(k) + \Delta D_r(k); \\
D(k) &\doteq M_h(k) \cdot D_h^z(k) + M_w(k) \cdot D_w^z(k) + \Delta D(k),
\end{aligned}
\tag{12.2}
$$

where

$$
\begin{aligned}
D_h^z(k) &\equiv D_{h1..n}^z(k) & (n \times 1); \\
D_w^z(k) &\equiv D_{w1..n}^z(k) & (n \times 1); \\
\Delta D_r(k) &\equiv \Delta D_r^{1..m}(k) & (m \times 1); \\
\Delta D(k) &\equiv \Delta D_{1..n}(k) & (mn \times 1); \\
M_{hr}(k) &\equiv M_h(z_r^{1..m}(k)) & (m \times 1); \\
M_{wr}(k) &\equiv M_w(z_r^{1..m}(k)) & (m \times 1); \\
M_h(k) &\doteq \mathrm{diag}[M_{h1}(k), \ldots, M_{hn}(k)] & (mn \times n); \\
M_w(k) &\doteq \mathrm{diag}[M_{w1}(k), \ldots, M_{wn}(k)] & (mn \times n).
\end{aligned}
\tag{12.3}
$$

From Sect. 11.5 and Eq. (10.19), we find that extending the functional model with troposphere parameters gives:

$$
\mathbf{E}\{\underline{y}(k)\} = T(k) \cdot t(k) + [e_4 \otimes I_{mn}] \cdot D(k) + A(k) \cdot x,
\tag{12.4}
$$

for the model with phase and code observables.[1] Using Eq. (12.2), we find that the troposphere extension to our model reads:

$$
\begin{aligned}
[e_4 \otimes I_{mn}] \cdot [1 \otimes M_h(k),\ 1 \otimes M_w(k),\ 1 \otimes I_{mn}] \quad \cdot \begin{bmatrix} D_h^z(k) \\ D_w^z(k) \\ \Delta D(k) \end{bmatrix} &= \\
\left[\, e_4 \otimes M_h(k),\ e_4 \otimes M_w(k),\ e_4 \otimes I_{mn} \,\right] \cdot \begin{bmatrix} D_h^z(k) \\ D_w^z(k) \\ \Delta D(k) \end{bmatrix}&.
\end{aligned}
\tag{12.5}
$$

Since the hydrostatic and wet mapping functions do not differ much, $M_h(k) \approx M_w(k)$, the design matrix is near rank deficient:

$$
\underbrace{\left[\, e_4 \otimes M_w(k) \,\right]}_{A_2} \cdot \underbrace{\left[1 \otimes I_n \right]}_{V} \approx \underbrace{\left[\, e_4 \otimes M_h(k) \,\right]}_{A_1}.
\tag{12.6}
$$

Therefore we cannot estimate both the hydrostatic and wet delay. But if we remove, for example, the columns corresponding to the hydrostatic delay, the sum of both delays is left in the model:

$$
\underbrace{D^z(k)}_{\overline{\mathrm{x}}} \leftarrow \underbrace{D_w^z(k)}_{\mathrm{x}_2} + \underbrace{I_n}_{V} \cdot \underbrace{D_h^z(k)}_{\mathrm{x}_1}.
\tag{12.7}
$$

ZTD   This sum is called the ZTD (Zenith Total Delay or Zenith Tropospheric Delay).

From now on, we drop the index of the mapping functions and assume that we map the ZTD by the wet mapping function:

$$
\begin{aligned}
M(z_r^s(k)) & & \leftarrow M_w(z_r^s(k)) & \quad (1 \times 1); \\
M_r(k) &\equiv M(z_r^{1..m}(k)) & \leftarrow M_{wr}(k) & \quad (m \times 1); \\
M(k) &\doteq \mathrm{diag}[M_1(k), \ldots, M_n(k)] & \leftarrow M_w(k) & \quad (mn \times n).
\end{aligned}
\tag{12.8}
$$

The reason for this is that the a-priori hydrostatic delay can be modeled with much higher accuracy than the wet delay. The residual delays are therefore more

---

[1]If only phase observables are used, the vector e_4 is replaced by e_2.

likely to be (mainly) wet delays.

The functional model now reads:

$$\mathbf{E}\{\underline{y}_,(k)\} = T(k) \cdot t(k) + [e_4 \otimes M(k), \ e_4 \otimes I_{mn}] \cdot \begin{bmatrix} 1 \otimes D^z(k) \\ 1 \otimes \Delta D(k) \end{bmatrix} + A(k) \cdot x. \quad (12.9)$$

This model is still under-determined. Often the residual tropospheric delays are therefore assumed zero and left unmodeled. More about residual delays in Sect. 12.4.

We will now consider three different cases: ZTDs as temporal parameters, as global parameters, and as batch parameters.

ZTDs as temporal parameters
If we estimate ZTDs as temporal parameters, the matrix $A(k)$ remains as in Eq. (11.18), and the matrix $T(k)$ and vector $t(k)$ are extended to:

$$\begin{aligned} T(k) &= [e_4 \otimes M(k), \ \mu \otimes I_{mn}, \ C_4 \otimes \mathcal{D}\ ]; \\ t(k) &= [D^z(k)', \ \overline{\imath}(k)', \ \overline{\delta}(k)']'. \end{aligned} \quad (12.10)$$

Although this is the most obvious extension of our model, the amount of parameters is extended by one per epoch per station.

ZTDs as global parameters
To reduce the amount of parameters, one may assume that the ZTDs do not change much in time. Often, in geodetic data processing, one ZTD is estimated every few hours. If, for example, only two hours of data are available, one may consider the ZTDs to be global parameters. The temporal parameters and corresponding design matrix are then the same as in Eq. (11.18). The design matrix $A(k)$ and the vector of global parameters can be extended for the ZTDs as follows:

$$\begin{aligned} A(k) &= [\Lambda \otimes C_n \otimes C_m, \ e_4 \otimes G(k), \ e_4 \otimes M(k)]; \\ x &= \left[N_{12..n}^{12..m'}, r', D^{z'}\right]'. \end{aligned} \quad (12.11)$$

Equation (12.11) already illustrates some analogy between the geometry and troposphere. In Chap. 13 this becomes more apparent.

ZTDs as batch parameters
For long observation time spans, one ZTD parameter per epoch may extend the model by too many parameters, and one parameter for the total time span is unrealistic under changing weather conditions. In some GPS data processing software, like the Bernese software [*Beutler et al.*, 2000], batches are used. The ZTD is then considered to be constant for the duration of the batch of several **$B(k)$** (say $\widetilde{p}$) epochs. $A(k)$ and $T(k)$ are both as in Eq. (11.18), $B(k) = [e_4 \otimes M(k)]$, the vectors of global and temporal parameters are as in Eq. (11.19), and the batch parameters are formed by $\beta = [D^z(1)', \ldots, D^z(p/\widetilde{p})']'$, if $p/\widetilde{p} \in \mathbb{N}^+$.

12.3   Gradient parameters

Model (12.1) with zero residual slant delays is valid for an azimuthal symmetric and horizontally layered atmosphere. This model may be refined by the assumption of a tilted atmosphere [*MacMillan*, 1995]. The extra delay term $\Delta D_r^s(k)$ should then describe the asymmetry. This correction term can be parameterized as (see Part I):

$$\Delta D_r^s(k) = M_{az}(\alpha_r^s(k)) \cdot \gamma(\alpha_r^s(k)), \quad (12.12)$$

$M_{az}(\alpha_r^s(k))$ where $M_{az}(\alpha_r^s(k))$ is the azimuthal mapping function, as function of the azimuth α, from receiver r to satellite s, at epoch k, and

$$\gamma(\alpha_r^s(k)) = \gamma_{r;N}(k) \cdot \cos(\alpha_r^s(k)) + \gamma_{r;E}(k) \cdot \sin(\alpha_r^s(k)), \qquad (12.13)$$

with $\gamma_{r;N}(k)$ and $\gamma_{r;E}(k)$ the gradient parameters in north and east direction. In matrix-vector notation we now write:

$$\Delta D_r(k) \doteq \Gamma_r(k) \cdot \gamma_r(k) \; ; \; \Delta D(k) \doteq \Gamma(k) \cdot \gamma(k), \qquad (12.14)$$

with

$$\Gamma_r(k) \doteq \begin{bmatrix} M_{az}(z_r^1(k)) \cdot \cos(\alpha_r^1(k)) & M_{az}(z_r^1(k)) \cdot \sin(\alpha_r^1(k)) \\ \vdots & \vdots \\ M_{az}(z_r^m(k)) \cdot \cos(\alpha_r^m(k)) & M_{az}(z_r^m(k)) \cdot \sin(\alpha_r^m(k)) \end{bmatrix} \; (m \times 2), \; (12.15)$$

$\Gamma(k), \gamma(k)$ and

$$\begin{aligned} \gamma_r(k) &\doteq [\gamma_{r;N}(k)', \gamma_{r;E}(k)']' & (2 \times 1); \\ \Gamma(k) &\doteq \mathrm{diag}\,[\Gamma_1(k), \ldots, \Gamma_n(k)] & (mn \times 2n); \\ \gamma(k) &\doteq [\gamma_1(k)', \ldots, \gamma_n(k)']' & (2n \times 1). \end{aligned} \qquad (12.16)$$

If we estimate gradient parameters every epoch, the partial design matrices read (in the ionosphere-float model, case (a), phase+code observables):

$$\begin{aligned} A(k) &= [\overbrace{\Lambda \otimes C_n \otimes C_m}^{\text{ambiguities}}, \overbrace{e_4 \otimes G(k)}^{\text{coord.}}]; \\ T(k) &= [\underbrace{e_4 \otimes M(k)}_{\text{ZTDs}}, \underbrace{e_4 \otimes \Gamma(k)}_{\text{gradients}}, \underbrace{\mu \otimes I_n \otimes I_m}_{\text{ionosphere}}, \underbrace{C_4 \otimes \mathcal{D}}_{\text{clocks}}], \end{aligned} \qquad (12.17)$$

and the vector of temporal parameters reads:

$$t(k) = [D^z(k)', \gamma(k)', \bar{\imath}(k)', \overline{\delta}_{,2..4}(k)']'. \qquad (12.18)$$

Considering both ZTDs and gradients to be temporal parameters is however not very practical because the redundancy reduces by $2np$ with the introduction of gradient parameters. The gradient model is better suitable when both kinds of parameters (ZTDs, gradients) are considered global or batch, or when extra pseudo-observables are used as in the troposphere-weighted model; see Sect. 12.5. We do not describe the models where gradients and ZTDs are global or batch parameters here. Instead, these models can be found in an outline in Sect. 12.6.

gradients	type of ZTD parameter	number of extra parameters
no	temporal	np
no	global	n
no	batch	$np/\widetilde{p}$
yes	temporal	$3np$
yes	global	$3n$
yes	batch	$3np/\widetilde{p}$

Table 12.1   Number of extra troposphere parameters in the troposphere-float model with respect to the troposphere-fixed model when modeling ZTDs and gradients; n stations, p epochs, $\widetilde{p}$ epochs per batch.

12.4   Residual Slant Tropospheric Delays

Instead of extending the model of ZTDs with gradients, one can also extend the model with the original residual STDs. Modeling slant delays is only possible when they are considered temporal parameters (or maybe batch parameters for short batches) because of the changing satellite configuration. Because for each receiver-satellite combination an extra parameter is included per epoch, the amount of extra parameters is mnp. Since the redundancy for the model with residual STDs can only be sufficient for the model with code observables (see Table E.3), this model may at first not seem to be very practical. But at the same time we may introduce the same amount of pseudo-observables for the extra parameters we introduced. Since we know that the residual STDs may not deviate much from zero, these pseudo-observables are constraints on the residual STDs. We assume these constraints uncorrelated with the GPS observables. Our $5mn \times 1$ vector of observables for the kth epoch and its dispersion then read:

$$\underline{y}(k) = \begin{bmatrix} \underline{y}(k) \\ \underline{\Delta D}(k) \end{bmatrix} \quad ; \quad \mathbf{D}\{\underline{y}(k)\} = \begin{bmatrix} Q_{y,}(k) & 0 \\ 0 & Q_{\Delta D}(k) \end{bmatrix}, \tag{12.19}$$

where $\underline{\Delta D}(k) = [\underline{\Delta D_1}(k), ..., \underline{\Delta D_n}(k)]'$ is the $mn \times 1$ vector with pseudo-observables with sample values $\Delta D(k) = 0$. For a derivation of a model for $Q_{\Delta D}(k)$, see Part III. If the stations are sufficiently far separated, we can assume the slant delays of different stations to be uncorrelated: $Q_{\Delta D}(k) = \text{diag}[Q_{\Delta D_1}(k), ..., Q_{\Delta D_n}(k)]$. The design matrices of global and temporal parameters read:

$$A(k) = \begin{bmatrix} \Lambda \otimes C_n \otimes C_m \ e_4 \otimes G(k) \\ 0 & 0 \end{bmatrix};$$

$$T(k) = \begin{bmatrix} e_4 \otimes I_n \otimes I_m \ \mu \otimes I_n \otimes I_m \ C_4 \otimes \mathcal{D} \\ I_n \otimes I_m & 0 & 0 \end{bmatrix}, \tag{12.20}$$

where one of these matrices may be extended with a partial design matrix that takes into account the ZTDs, depending on whether we estimate the ZTDs as global or temporal parameters.

12.5   The troposphere-weighted model

The equations for the troposphere-float model follow from Eqs. (12.9) as:

$$\mathbf{E}\{\underline{y}(k)\} = \begin{bmatrix} T(k), \ B(k), \ A(k) \end{bmatrix} \begin{bmatrix} t(k) \\ \beta(k) \\ x \end{bmatrix}, \tag{12.21}$$

where $B(k) = [e_4 \otimes M(k)]$ and $\beta(k) = D^z(k)$. In a(n absolute) troposphere-weighted model, extra observation equations of the form $\mathbf{E}\{\underline{D}^z(k)\} = D^z(k)$ result in the augmented model:

$$\mathbf{E}\{\begin{bmatrix} \underline{y}(k) \\ \underline{D}^z(k) \end{bmatrix}\} = \begin{bmatrix} T(k) \ B(k) \ A(k) \\ 0 & I_n & 0 \end{bmatrix} \begin{bmatrix} t(k) \\ D^z(k) \\ x \end{bmatrix}. \tag{12.22}$$

This would imply that we have knowledge about the absolute ZTDs. Unless we have some external source, like Water Vapor Radiometer data, this will generally

not be true. But what we do know is that the ZTDs only change in time within
bounds. Therefore, except for the first epoch, for each epoch an $n \times 1$ vector of
pseudo-observables $\underline{d}(k)$ is formed for which we have the observation equations
[*Rothacher*, 1992]:

$$\mathbf{E}\{\underline{d}(k)\} = D^z(k) - D^z(k-1).$$ (12.23)

As sample values we use $d(k) = 0$. Instead of Model (12.22) we rather use the
following relative troposphere-weighted model:

$$\mathbf{E}\{ \begin{bmatrix} \underline{y}(k-1) \\ \underline{d}(k) \\ \underline{y}(k) \end{bmatrix} \} = \begin{bmatrix} T(k-1) & B(k-1) & 0 & 0 & A(k-1) \\ 0 & -I_n & 0 & I_n & 0 \\ 0 & 0 & T(k) & B(k) & A(k) \end{bmatrix} \begin{bmatrix} t(k-1) \\ \beta(k-1) \\ t(k) \\ \beta(k) \\ x \end{bmatrix},$$ (12.24)

in which, for notational convenience, the parameters relating to epochs other than
k and $k-1$ are ignored. We assume that the pseudo-observables $\underline{d}(k)$ are uncor-
related with other (pseudo-)observables:

$$\mathbf{D}\{ \begin{bmatrix} \vdots \\ \underline{d}(k-1) \\ \underline{y}(k-1) \\ \underline{d}(k) \\ \underline{y}(k) \\ \vdots \end{bmatrix} \} = \begin{bmatrix} \ddots & & & & \\ & Q_d(k-1) & & & 0 \\ & & Q_y(k-1) & & \\ & & & Q_d(k) & \\ & 0 & & & Q_y(k) \\ & & & & & \ddots \end{bmatrix}.$$ (12.25)

A more explicit derivation of the structure of the $Q_d(k)$ can be found in Part III.

12.6   Observation scenarios

As discussed in the previous chapter we can distinguish troposphere-fixed, float,
and weighted models. In both the troposphere-float and weighted model we
may estimate zenith delays as well as gradients or slant delays. The slant delays
are temporal parameters, but the zenith delays and gradients may be temporal,
batch, or global parameters. Because in the troposphere-weighted model relative
constraints are formed between batches or epochs (batches of one epoch), the
troposphere-weighted model does not make sense in the case where ZTDs are
global parameters. This model will be referred to as the troposphere-constant
model.

troposphere-
constant

An overview of possible models is given in tables. Table 12.2 shows the
partial design matrices and corresponding parameters, Table 12.3 gives a sum-
mary of all different models, and Table 12.4 gives the corresponding design
matrices for the models of Table 12.3 with numbers 1–10.

The batch cases 11–16 of Table 12.3 are mixed models, where in the first $\widetilde{p}$
epochs the ZTD parameters are considered to be constant. For these epochs the
models are in fact the same as the model where ZTDs are global parameters.
Then, in the $(\widetilde{p}+1)$th epoch, new ZTD parameters are introduced. This is the

type of parameters	global/batch temporal	parameters	partial design matrices	number of parameters
ambiguities	global	$N_{12..n}^{12..m}$	$\Lambda \otimes C_n \otimes C_m$	$2(m-1)(n-1)$
geometry	global	r	$e_4 \otimes G(k)$	$3n$
ZTDs	batch	$D^z, D^z(k)$	$e_4 \otimes M(k)$	n
gradients	batch	$\gamma, \gamma(k)$	$e_4 \otimes \Gamma(k)$	$2n$
residual STDs	temporal	$\Delta D(k)$	$e_4 \otimes I_{mn}, I_{mn}$	mn
ionospheric delays	temporal	$\iota(k)$	$\mu \otimes I_{mn}$	mn
clocks	temporal	$\overline{\delta}_{,2..4}(k)$	$C_4 \otimes \mathcal{D}$	$3(m+n-1)$

Table 12.2   Partial design matrices and corresponding parameters. Recall the definitions of Λ in Eq. (10.20), C_n in Eq. (11.5), $G(k)$ in Eqs. (10.17) and (10.15), $M(k)$ in Eq. (12.8), $\Gamma(k)$ in Eqs. (12.16) and (12.15), μ in Eqs. (10.9) and (10.20), and $\mathcal{D}$ in Eq. (11.10). Note that, because of near rank deficiencies, the matrices $G(k)$, $M(k)$, and $\Gamma(k)$ are redefined in Eq. (13.7).

same as in the model where ZTDs are temporal parameters. The model matrices keep alternating in this way for the troposphere-float models. The batch cases refer to the most general models; the ZTDs as temporal or global parameters are special cases of the model with the ZTDs as batch parameter.

$$
\begin{array}{ccccc}
\text{temporal} & \leftarrow & \text{batch} & \rightarrow & \text{global} \\
\widetilde{p} = 1 & & 1 \leq \widetilde{p} \leq p & & \widetilde{p} = p \\
\\
& & \text{troposphere} & & \\
\text{-float} & \leftarrow & \text{-weighted} & \rightarrow & \text{-constant} \\
Q_d(k) = \infty & & \infty > Q_d(k) > 0 & & Q_d(k) = 0
\end{array}
$$

In much the same way, we can consider the troposphere-constant and float models special cases of the troposphere-weighted model. In the troposphere-constant case we use hard constraints $Q_d(k) = 0$, and in the troposphere-float case we use very loose constraints: $Q_d(k) = \infty$.

nr.	model	delay parameters	ZTD parameter
1.	troposphere-fixed	none	none
2.	troposphere-constant	zenith	global
3.	troposphere-constant	zenith and gradient	global
4.	troposphere-constant	zenith and slant	global
5.	troposphere-float	zenith	temporal
6.	troposphere-float	zenith and gradient	temporal
7.	troposphere-float	zenith and slant	temporal
8.	troposphere-weighted	zenith	temporal
9.	troposphere-weighted	zenith and gradient	temporal
10.	troposphere-weighted	zenith and slant	temporal
11.	troposphere-float	zenith	batch
12.	troposphere-float	zenith and gradient	batch
13.	troposphere-float	zenith and slant	batch
14.	troposphere-weighted	zenith	batch
15.	troposphere-weighted	zenith and gradient	batch
16.	troposphere-weighted	zenith and slant	batch

Table 12.3   Parameterization of the tropospheric delay. The numbers correspond with Table 12.4.

nr.	$A(k)$	$B(k)$	$T(k)$
1.	$\begin{bmatrix} \Lambda \otimes C_n \otimes C_m & e_4 \otimes G(k) \end{bmatrix}$	-	$\begin{bmatrix} \mu \otimes I_{mn} & C_4 \otimes \mathcal{D} \end{bmatrix}$
2.	$\begin{bmatrix} \Lambda \otimes C_n \otimes C_m & e_4 \otimes G(k) & e_4 \otimes M(k) \end{bmatrix}$	-	$\begin{bmatrix} \mu \otimes I_{mn} & C_4 \otimes \mathcal{D} \end{bmatrix}$
3.	$\begin{bmatrix} \Lambda \otimes C_n \otimes C_m & e_4 \otimes G(k) & e_4 \otimes M(k) & e_4 \otimes \Gamma(k) \end{bmatrix}$	-	$\begin{bmatrix} \mu \otimes I_{mn} & C_4 \otimes \mathcal{D} \end{bmatrix}$
4.	$\begin{bmatrix} \Lambda \otimes C_n \otimes C_m & e_4 \otimes G(k) & e_4 \otimes M(k) \\ 0 & 0 & 0 \end{bmatrix}$	-	$\begin{bmatrix} e_4 \otimes I_{mn} & \mu \otimes I_{mn} & C_4 \otimes \mathcal{D} \\ I_{mn} & 0 & 0 \end{bmatrix}$
5.	$\begin{bmatrix} \Lambda \otimes C_n \otimes C_m & e_4 \otimes G(k) \end{bmatrix}$	-	$\begin{bmatrix} e_4 \otimes M(k) & \mu \otimes I_{mn} & C_4 \otimes \mathcal{D} \end{bmatrix}$
6.	$\begin{bmatrix} \Lambda \otimes C_n \otimes C_m & e_4 \otimes G(k) \end{bmatrix}$	-	$\begin{bmatrix} e_4 \otimes M(k) & e_4 \otimes \Gamma(k) & \mu \otimes I_{mn} & C_4 \otimes \mathcal{D} \end{bmatrix}$
7.	$\begin{bmatrix} \Lambda \otimes C_n \otimes C_m & e_4 \otimes G(k) \\ 0 & 0 \end{bmatrix}$	-	$\begin{bmatrix} e_4 \otimes M(k) & e_4 \otimes I_{mn} & \mu \otimes I_{mn} & C_4 \otimes \mathcal{D} \\ 0 & I_{mn} & 0 & 0 \end{bmatrix}$
8.	$\begin{bmatrix} \Lambda \otimes C_n \otimes C_m & e_4 \otimes G(k) \end{bmatrix}$	$\begin{bmatrix} e_4 \otimes M(k) \end{bmatrix}$	$\begin{bmatrix} \mu \otimes I_{mn} & C_4 \otimes \mathcal{D} \end{bmatrix}$
9.	$\begin{bmatrix} \Lambda \otimes C_n \otimes C_m & e_4 \otimes G(k) \end{bmatrix}$	$\begin{bmatrix} e_4 \otimes M(k) & e_4 \otimes \Gamma(k) \end{bmatrix}$	$\begin{bmatrix} \mu \otimes I_{mn} & C_4 \otimes \mathcal{D} \end{bmatrix}$
10.	$\begin{bmatrix} \Lambda \otimes C_n \otimes C_m & e_4 \otimes G(k) \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} e_4 \otimes M(k) \\ 0 \end{bmatrix}$	$\begin{bmatrix} e_4 \otimes I_{mn} & \mu \otimes I_{mn} & C_4 \otimes \mathcal{D} \\ I_{mn} & 0 & 0 \end{bmatrix}$

Table 12.4   Partial design matrices in different models. The numbers correspond with Table 12.3.

Chapter 13

Near rank deficiencies

13.1 Introduction

In the preceding chapters, full-rank models were derived by lumping as many parameters as needed. The models that were derived in this way, may however still be *near* rank deficient. An example of a near rank deficiency was shown in Sect. 12.2 for the ZHD and ZWD, which was solved by lumping the delays the same way as in case of a real rank deficiency. A near rank deficiency may cause numerical instability and a poor precision of the (estimated) parameters. Lumping parameters in case of a near rank deficiency on the other hand, may cause biases in the estimated parameters. The near rank deficiencies that are still present in our models concern the geometry and troposphere parameters. This chapter shows in Sect. 13.2 how they appear in 'standard' network processing and, in Sect. 13.3, why they do not appear when the concept of Precise Point Positioning is used.

13.2 Revealing the near rank deficiencies

In small networks the zenith angles and azimuths of the satellites are approximately equal for all stations. This means that for the matrices with values of mapping functions holds: $M_1(k) \approx M_2(k) \approx \cdots \approx M_n(k)$ or $M(k) \approx [I_n \otimes M_r(k)]$, for any receiver r. But also for the geometry holds: $G_1(k) \approx G_2(k) \approx \cdots \approx G_n(k)$ or $G(k) \approx [I_n \otimes G_r(k)]$; and for the gradient matrices: $\Gamma_1(k) \approx \Gamma_2(k) \approx \cdots \approx \Gamma_n(k)$ or $\Gamma(k) \approx [I_n \otimes \Gamma_r(k)]$. These near equalities, in combination with the partial design matrices of the ionosphere and receiver clock errors, cause near rank deficiencies.

In the ionosphere-float model with four observation types, case (a), the design matrix of the temporal parameters $T(k)$ reads (see Table 11.1):

$$T(k) = [\underbrace{\mu \otimes I_n \otimes I_m}_{\text{ionosphere}}, \underbrace{C_4 \otimes I_n \otimes e_m}_{\text{rec. clocks}}, \underbrace{C_4 \otimes -e_n \otimes C_m}_{\text{sat. clocks}}], \tag{13.1}$$

which is in fact not time dependent as long as the number of satellites or stations does not change. If we now define

$$v \doteq \begin{bmatrix} \mu_2 & \otimes\ e_n\ \otimes & M_r(k) \\ (\nu + e_3) \otimes\ e_n\ \otimes & c'_m \cdot M_r(k) \\ (\nu + e_3) \otimes\ -1\ \otimes (-e_{m-1},\ I_{m-1}) \cdot M_r(k) \end{bmatrix} \quad (mn+3n+3(m-1)\times 1), \tag{13.2}$$

with ν as defined in Eq. (11.12), we can verify that:

$$T(k) \cdot v = [e_4 \otimes e_n \otimes M_r(k)]. \tag{13.3}$$

troposphere-float model

Using Eq. (13.3), in the troposphere-float model with ZTDs as global parameters, one (near) rank deficiency can be shown as:

$$\underbrace{[e_p \otimes e_4 \otimes C_n \otimes M_r(k), \ I_p \otimes T(k)]}_{A_2} \underbrace{\begin{bmatrix} 1 \otimes 1 \otimes -e_{n-1} \otimes 1 \\ e_p \otimes v \end{bmatrix}}_{V} =$$

$$\underbrace{[e_p \otimes e_4 \otimes c_n \otimes M_r(k)]}_{A_1}. \quad (13.4)$$

And the lumped parameters read:

$$\underbrace{\begin{bmatrix} D_{12..n}^z \\ \overline{t}(1..p) \end{bmatrix}}_{\overline{x}} \leftarrow \underbrace{\begin{bmatrix} D_{2..n}^z \\ t(1..p) \end{bmatrix}}_{x_2} + \underbrace{\begin{bmatrix} -e_{n-1} \\ e_p \otimes v \end{bmatrix}}_{V} \cdot \underbrace{D_1^z}_{x_1}, \quad (13.5)$$

where $t(1..p)$ stands for the vector of all temporal parameters in all epochs. We can thus solve the near rank deficiency by lumping the ZTDs into between-station single-differenced ZTD parameters. When the ZTDs are temporal parameters, the same rank deficiency is present, since this is a special case in which $p = 1$. More generally, the number of rank deficiencies for the ZTDs is one per batch.

alternatively shown rank deficiencies

The near rank deficiency as shown above may look a bit complicated because of the complicated partial design matrices for the ionosphere and receiver clock parameters. An alternative way to show the near rank deficiency starts with the still rank-deficient design matrix of Eq. (10.24). First we may lump the satellite clock errors and satellite hardware delays into $[I_4 \otimes -e_n \otimes I_m] \cdot \overline{\delta s}(k)$ for each epoch (this may be done for both case (a) and (b)). Next, as an example, the rank deficiency in the troposphere-float model with ZTDs as temporal parameters can then be shown, using Eq. (11.6):

$$\underbrace{\left[I_4 \otimes -e_n \otimes I_m, e_4 \otimes C_n \otimes M_r(k) \right]}_{A_2} \underbrace{\begin{bmatrix} e_4 \otimes -1 \otimes M_r(k) \\ 1 \otimes -e_{n-1} \otimes 1 \end{bmatrix}}_{V} = \underbrace{[e_4 \otimes c_n \otimes M_r(k)]}_{A_1}.$$

$$(13.6)$$

Only because we started with another sequence of lumping parameters, we ended up with a (near) rank deficiency that was also related to the ionosphere and receiver clock parameters, since these parameters were lumped with the satellite clock errors (steps 4 and 6 in Figs. E.1 and E.2). If the (lumped) satellite clocks would be know beforehand, the near rank deficiencies would not have been present. In fact we see the same (near) rank deficiency occur over and over again in the receiver dependent parameters. Completely analogue to the shown rank deficiencies in the ZTDs, we could show rank deficiencies for the geometry (3) and the gradients (2/batch). Eliminating the near rank deficiencies by forming inter-station single differences, results in the following partial design matrices:

$$\begin{aligned} G(k) &= \begin{bmatrix} 0 & & & \\ & G_2(k) & & \\ & & \ddots & \\ & & & G_n(k) \end{bmatrix}; \\ M(k) &= \begin{bmatrix} 0 & & & \\ & M_2(k) & & \\ & & \ddots & \\ & & & M_n(k) \end{bmatrix}; \Gamma(k) = \begin{bmatrix} 0 & & & \\ & \Gamma_2(k) & & \\ & & \ddots & \\ & & & \Gamma_n(k) \end{bmatrix}. \end{aligned} \quad (13.7)$$

parameters	solved near rank deficiencies	partial design matrices
coordinates	3	$G(k)$
ZTDs	1/batch	$M(k)$
gradients	2/batch	$\Gamma(k)$

Table 13.1   Number of solved near rank deficiencies.

troposphere-weighted model

In the troposphere-weighted model, the near rank deficiency that appears in the troposphere-float model seems to be resolved by the partial design matrix of the constraints. Considering the third and fourth column of the design matrix of Eq. (12.24) and only the lumped satellite clocks as in Eq. (13.6), we have an apparently full-rank design matrix

$$
A \doteq \underbrace{\begin{bmatrix} 0 & 1 \otimes I_n \otimes 1 \\ I_4 \otimes -e_n \otimes I_m & e_4 \otimes I_n \otimes M_r(k) \end{bmatrix}}_{\substack{\text{sat. clocks} \qquad \text{ZTDs}}}
\tag{13.8}
$$

for the satellite clocks and ZTDs at epoch k. Matrix A does not have a real rank deficiency because

$$
\underbrace{\begin{bmatrix} 0 & 1 \otimes C_n \otimes 1 \\ I_4 \otimes -e_n \otimes I_m & e_4 \otimes C_n \otimes M_r(k) \end{bmatrix}}_{A_2} \underbrace{\begin{bmatrix} e_4 \otimes -1 \otimes M_r(k) \\ 1 \otimes -e_{n-1} \otimes 1 \end{bmatrix}}_{V} = \underbrace{\begin{bmatrix} -C_n e_{n-1} \\ e_4 \otimes c_n \otimes M_r(k) \end{bmatrix}}_{\bar{A}_1}
$$

$$
\neq \underbrace{\begin{bmatrix} -c_n \\ e_4 \otimes c_n \otimes M_r(k) \end{bmatrix}}_{A_1},
\tag{13.9}
$$

where $\bar{A}_1$ is not a column of A. A rank deficiency would have been shown if $A_2 V = A_1$. A near rank deficiency is still present because A_1 and $\bar{A}_1$ make a small angle γ:

$$
\cos(\gamma) = \frac{A_1' \bar{A}_1}{\parallel A_1 \parallel \cdot \parallel \bar{A}_1 \parallel} = \frac{4 M_r(k)' M_r(k)}{(\{n - 1 + 4 M_r(k)' M_r(k)\}\{1 + 4 M_r(k)' M_r(k)\})^{\frac{1}{2}}}.
\tag{13.10}
$$

The larger the product $M_r(k)' M_r(k)$, the smaller γ. In other words, a closer near rank deficiency occurs if we are observing too many satellites or satellites with too low elevations (large mapping values).

correlation of ZTDs

This near rank deficiency translates into a strong correlation of estimated ZTDs. This is shown below for a simple case. Consider the short baseline case $n = 2$. Furthermore, assume that we have only one observation type[1] and the following covariance matrices:

$$
\begin{aligned}
Q_d(k) &= \sigma_d^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}; \\
Q_y(k) &= \sigma_y^2 (I_2 \otimes I_m); \\
Q(k) &= \mathrm{diag}[Q_d(k), Q_y(k)].
\end{aligned}
\tag{13.11}
$$

If the satellite clocks and ZTDs of epoch k were the only parameters, the partitioned normal matrix $N = A' Q(k)^{-1} A$ reads:

$$
\begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix} = \sigma_y^{-2} \begin{bmatrix} 2 \otimes I_m & -e_2' \otimes M_r(k) \\ -e_2 \otimes M_r(k)' & I_2 \otimes M_r(k)' M_r(k) \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & Q_d(k)^{-1} \end{bmatrix}.
\tag{13.12}
$$

[1]The near rank deficiency generalizes to more observation types.

The covariance matrix of the least-squares estimates of the ZTDs can now be obtained from:

$$Q_{\hat{D}^z} \doteq \begin{bmatrix} \sigma^2_{\hat{D}^z_1} & \sigma_{\hat{D}^z_1 \hat{D}^z_2} \\ \sigma_{\hat{D}^z_1 \hat{D}^z_2} & \sigma^2_{\hat{D}^z_2} \end{bmatrix} =$$
$$(N^{-1})_{22} = (N_{22} - N_{21} N_{11}^{-1} N_{12})^{-1} = Det^{-1} \begin{bmatrix} \alpha + \sigma_d^{-2} & \alpha - \sigma_d^{-2} \rho \\ \alpha - \sigma_d^{-2} \rho & \alpha + \sigma_d^{-2} \end{bmatrix},$$

(13.13)

where $\alpha \doteq \frac{1}{2} \sigma_y^{-2} M_r(k)' M_r(k)$ and $Det \doteq \sigma_d^{-4}(1 - \rho^2) + 2\alpha \sigma_d^{-2}(1 + \rho)$. From the covariance matrix of Eq. (13.13) we find the following correlation coefficient between the ZTDs of the stations:

$$\boxed{\frac{\sigma_{\hat{D}^z_1 \hat{D}^z_2}}{\sigma_{\hat{D}^z_1} \sigma_{\hat{D}^z_2}} = \frac{\alpha - \sigma_d^{-2} \rho}{\alpha + \sigma_d^{-2}}.}$$

(13.14)

The correlation is large if $\sigma_d^{-2} \ll \alpha$, which will generally be the case when σ_y^2 and σ_d^2 are of the same order of magnitude.

precision
relative
ZTDs

From Eq. (13.13) we can also derive the precision of the relative ZTD:

$$\boxed{\sigma^2_{\hat{D}^z_{12}} = \sigma^2_{\hat{D}^z_1} + \sigma^2_{\hat{D}^z_2} - 2\sigma_{\hat{D}^z_1 \hat{D}^z_2} = (\alpha + \frac{1}{2} \sigma_d^{-2}(1 - \rho))^{-1},}$$

(13.15)

which shows a better accuracy if α is large, that is, when more satellites and when low-elevation satellites are used.

Numerically more stable estimates are obtained by fixing one of the ZTDs, or estimating single-differenced ZTDs instead of one ZTD per station. As a consequence of eliminating columns of the design matrix, not only the matrix $M(k)$ (or $G(k)$, $\Gamma(k)$) are altered in Eq. (13.7), but also the constraints change to an alternative form of Eq. (12.23):

$$\boxed{\mathbf{E}\{\underline{d}(k)\} = C_n \beta(k) - C_n \beta(k-1) \quad ; \quad \mathbf{D}\{\underline{d}(k)\} = Q_d(k).}$$

(13.16)

Although the first constraint does not have a functional relation with any of the parameters, it may still be correlated with the other constraints and should therefore remain in the observation model. One should however realize that in the original Model (12.24) we have $\beta(k) = [D^z_1(k), ..., D^z_n(k)]'$, and eliminating the first column of $B(k)$ is equivalent with lumping the parameters to $\beta(k) = [D^z_{12}(k), ..., D^z_{1n}(k)]'$ as in Eq. (13.16), cf. Eq. (13.5).

13.3   **Precise Point Positioning**

PPP

IGS

A method that has become more and more popular in recent years is 'Precise Point Positioning' (PPP). The method as described in [*Zumberge et al.*, 1997] makes it possible to (post)process large networks and is also useful for GPS users with a single receiver. Two steps can be distinguished in the PPP method. First, precise GPS satellite positions and clock corrections are determined from a globally distributed network of receivers. These satellite positions and clock errors are made available by the International GPS Service (IGS) with a certain interval (for example, every 5 minutes). For a higher sampling rate, these positions and clock errors need to be interpolated. Since Selective Availability (SA) was turned off (May, 2000), the satellite clock errors have a smooth behavior and can be

interpolated with a higher precision [*Han et al.*, 2001]. Second, once the satellite clock errors are known or assumed known, the GPS data of a local network or single station can be processed in a second step.

An exact description of the PPP method will not be given here. But if the satellite hardware delays are assumed constant and equal for L1 and L2 and for P1 and P2 (case(b), see Table E.3) in the first step, then the satellite clocks are lumped with the hardware delays and only relative clock errors can be estimated; see Figs. E.1 and E.2. Absolute clock corrections can be determined with the aid of one or more stable external clocks. For a more elaborate description, see [*Kouba et al.*, 1995], [*Kouba and Springer*, 2001].

With absolute satellite clocks errors, we can process the data of a single station or a (local) network. As we can see in Table 11.1, case (b), the remaining parameters have partial design matrices for which the Kronecker product has an identity matrix (I_n) in place for the stations, while for the geometry and troposphere we have block-diagonal partial design matrices. As long as there is no stochastic correlation between the stations, we may therefore process the data one station at a time.

Chapter 14

Pre-elimination transformations

By a transformation of Eq. (12.21), an alternative model can be derived in which there are no temporal parameters. The temporal parameters are thus pre-eliminated. This has the advantage that a smaller system has to be solved to obtain estimates of the nontemporal parameters. Analytic expressions of the transformation are given below.

The pre-elimination transformation is of the form:

$$
T(k)^{\perp'}\mathbf{E}\{\underline{y}(k)\} = T(k)^{\perp'}\left[\, B(k),\, A(k)\,\right]\begin{bmatrix}\beta(k)\\x\end{bmatrix}, \tag{14.1}
$$

$T(k)^{\perp}$ where $T(k)^{\perp'}T(k) \doteq 0$. No information about the batch or global parameters is lost when $T(k)^{\perp}$ spans the null space of $T(k)'$. In the most general case of the troposphere-weighted model, *two different cases* can be distinguished: the model with ZTDs (+ gradients) and the model with ZTDs and residual STDs, respectively corresponding with the cases 8/9 and 10 in Table 12.3. One can easily verify that in the *first case*[1] (we assume the ionosphere-float model):

$$
T(k)^{\perp} = \begin{bmatrix}\nu'\\I_3\end{bmatrix}\otimes\mathcal{D}^{\perp} \qquad (4mn\times 3(m-1)(n-1))\,, \tag{14.2}
$$

and in the *second case*:

$$
T(k)^{\perp} = \begin{bmatrix}I_{4mn}\\-e_4'\otimes I_{mn}\end{bmatrix}\cdot(\begin{bmatrix}\nu'\\I_3\end{bmatrix}\otimes\mathcal{D}^{\perp})\;\;(5mn\times 3(m-1)(n-1))\,, \tag{14.3}
$$

$\mathcal{D}^{\perp'}$ with $\mathcal{D}^{\perp'}$ the double-difference operator $(\mathcal{D}^{\perp'}\mathcal{D} = 0)$:

$$
\mathcal{D}^{\perp'} = [-e_{n-1},\, I_{n-1}]\otimes[-e_{m-1},\, I_{m-1}]\quad\;\;((m-1)(n-1)\times mn)\,. \tag{14.4}
$$

transfor-
mation of
observables   Premultiplying the observables with the $T(k)^{\perp'}$ of Eq. (14.3) then gives the transformed observables:

$$
\begin{aligned}
T(k)^{\perp'}\underline{y}(k) &= ([\nu,\, I_3]\otimes\mathcal{D}^{\perp'})\cdot[I_{4mn},\, -e_4\otimes I_{mn}]\begin{bmatrix}\underline{y}(k)\\\underline{\Delta D}(k)\end{bmatrix}\\
&\doteq ([\nu,\, I_3]\otimes\mathcal{D}^{\perp'})\cdot\underline{y}^{STD-free}(k).
\end{aligned} \tag{14.5}
$$

Now we can distinguish two multiplications. The first multiplication results in residual STD-free observables, and the second multiplication results in DD (clock error-free) observables $\underline{y}_{12..n,l}^{12..m}$. The latter multiplication is combined with forming

---

[1]Recall the definitions of $\mathcal{D}$ and ν from Eqs. (11.10) and (11.12).

ionosphere-free observables.

The residual STD-free observa*bles* read:

$$\underline{y}^{STD-free}(k) \doteq \underline{y}_{,}(k) - e_4 \otimes \underline{\Delta D}(k). \tag{14.6}$$

But because we assume $\Delta D(k) = 0$, this product does not actually have to be computed to obtain the STD-free observa*tions*. In both cases the transformed observations are:

$$\boxed{T(k)^{\perp'}y(k) = ([\nu,\ I_3] \otimes \mathcal{D}^{\perp'}) \cdot y_{,}(k).} \tag{14.7}$$

In the second case however, the covariance matrix of the STD-free observables has an additional term:

$$\boxed{Q_y^{STD-free}(k) = Q_{y,}(k) + e_4 e_4' \otimes Q_{\Delta D}(k).} \tag{14.8}$$

<div style="float:left">analytical design-matrix transforma-tions</div>

The transformations on the right-hand side of Eq. (14.1), $T(k)^{\perp'}A(k)$ and $T(k)^{\perp'}B(k)$, can be done analytically. If the residual STDs are modeled as in case 10 of Table 12.4, the first part of the transformation simply reads:

$$\begin{aligned}
[I_{4mn},\ -e_4 \otimes I_{mn}] \cdot A(k) &= [\Lambda \otimes C_n \otimes C_m,\ e_4 \otimes G(k)]\,; \\
[I_{4mn},\ -e_4 \otimes I_{mn}] \cdot B(k) &= [e_4 \otimes M(k)]\,.
\end{aligned} \tag{14.9}$$

In other words, the first part of this design-matrix transformation reduces the design matrices to those of the case where no residual STDs are modeled (case 8). Continuing with the second part of the transformation, we find the following useful analytical expressions:

$$\begin{aligned}
[\nu,\ I_3] \cdot e_4 &= \nu + e_3\,; \\
[\nu,\ I_3] \cdot \Lambda &= [\lambda_1 \nu,\ \lambda_2 c_3]\,; \\
\mathcal{D}^{\perp'} \cdot (C_n \otimes C_m) &= I_{(n-1)(m-1)},
\end{aligned} \tag{14.10}$$

with which the analytical design-matrix transformations become:

$$\boxed{\begin{aligned}
T(k)^{\perp'}A(k) &= \left[[\lambda_1\nu,\ \lambda_2 c_3] \otimes I_{(n-1)(m-1)},\ (\nu + e_3) \otimes \mathcal{D}^{\perp'}G(k)\right]; \\
T(k)^{\perp'}B(k) &= (\nu + e_3) \otimes \mathcal{D}^{\perp'}M(k).
\end{aligned}} \tag{14.11}$$

<div style="float:left">analytical covariance-matrix transforma-tion</div>

The covariance matrix of $T(k)^{\perp'}\ \underline{y}(k)$ can also be computed analytically, if we have a simple model for $Q_{y,}(k)$, like $\underline{Q}_{y,}(k) = \mathrm{diag}[\sigma_\phi^2 I_2,\ \sigma_p^2 I_2] \otimes I_{mn}$, with σ_ϕ^2 and σ_p^2 the variances of the phase and code observables respectively. In that case we have, with $T(k)^\perp$ as in Eq. (14.2):

$$\boxed{\begin{aligned}
&T(k)^{\perp'}\ Q_y^{STD-free}(k)\ T(k)^\perp\ = \\
&T(k)^{\perp'}\ Q_{y,}(k)\ T(k)^\perp\ +\ T(k)^{\perp'}\ (e_4 e_4' \otimes Q_{\Delta D}(k))\ T(k)^\perp\ = \\
&(\sigma_\phi^2 \nu\nu' + \mathrm{diag}[\sigma_\phi^2, \sigma_p^2, \sigma_p^2]) \otimes ([e_{n-1}e_{n-1}' + I_{n-1}] \otimes [e_{m-1}e_{m-1}' + I_{m-1}]) \\
&+ (\nu + e_3)(\nu + e_3)' \otimes \mathcal{D}^{\perp'}Q_{\Delta D}(k)\mathcal{D}^\perp,
\end{aligned}} \tag{14.12}$$

where the first term refers to the model where no residual slant delays are modeled.

Chapter 15

Conclusions of Part II

To facilitate GPS data processing, the observations have to be expressed in terms of (unknown) parameters. Several observation models are derived in this part. Because the models with all parameters present are rank deficient, full-rank models with fewer parameters are derived by reparameterizations. We restricted ourselves to the functional model and give only a very general description of the stochastic model. An overview is given of models that can be used to parameterize the tropospheric delay.

By extensive use of the Kronecker product, partial design matrices are obtained in a very compact notation. Each partial design matrix corresponds with a parameter group. We use these matrices as building blocks and distinguish global, batch, and temporal parameters. Global parameters remain constant over the complete time span, whereas batch parameters change from batch (several epochs) to batch, and temporal parameters change from epoch to epoch. For static geodetic networks the coordinates are global parameters, at least for the duration of a run of one day. Batch and temporal parameters are considered nuisance parameters.

Three types of troposphere parameters are introduced: residual Slant Tropospheric Delays (STDs), Zenith Tropospheric Delays (ZTDs), and gradients. Residual STDs are temporal parameters, whereas ZTDs and gradient parameters can be considered batch parameters. This gives the possibility to use batches of just one epoch (temporal parameters), all epochs (global parameters) or anything in between. If neither ZTDs nor gradients are parameterized and only an a-priori model is used, we speak of a troposphere-fixed model. In the troposphere-float and weighted models, ZTDs are estimated with a possible addition of either gradients or residual STDs. In the latter case we also introduce additional constraints for the residual STDs. The troposphere-weighted model is an extension of the troposphere-float model with soft constraints on ZTD differences between batches. Table 15.1 gives an overview of the modeling options that are dealt with in this part.

In principle one ZTD per station per batch can be parameterized. But because of a near rank deficiency, in small networks it may be advisable to constrain one ZTD per batch (or model only single-differenced ZTDs) to obtain estimates with higher precision. Similar near rank deficiencies occur for the geometry (baseline coordinates) and gradients (single-differenced gradients). Solving for the near rank deficiencies by fixing parameters (one ZTD, three coordinates, two gradients) has the disadvantage that biases are introduced when these parameters are fixed to a wrong value.

observables:	*phase + code*; phase only
ionosphere:	fixed; *float*; weighted
satellite clocks:	*changing*; constant
troposphere parameters:	none; ZTDs; ZTDs+gradients; ZTD+residual STDs
troposphere:	fixed; float (constant); weighted
ZTDs:	global; batch; temporal

Table 15.1   Described modeling options. The top three rows are considered in the troposphere-fixed model. Only the italicized options are assumed in the troposphere-float and weighted models.

By means of transformations, smaller GPS observation models can be derived in which the temporal parameters are pre-eliminated and therefore not solved for. Analytical expressions can be found for these transformations (double differencing, ionosphere-free linear combinations). If residual STDs are modeled, pre-elimination of these parameters boils down to using the same observables and design matrices as the model with no residual STDs, however, the covariance matrix of the observables will have an additional term.

state of the art and contribution
The troposphere-fixed model is widely used, mainly in commercial software; the troposphere-constant, weighted, and float models are implemented in most scientific software. Although these models are not new, the terminology is by the author. Estimation of troposphere gradients is possible with recent versions of scientific software. The modeling of residual STDs is introduced in this part. Undifferenced observation models and the presence of several rank deficiencies in these models were earlier shown in [*De Jonge*, 1998]. The systematics in deriving rank deficiencies in the troposphere-fixed model and the derivation of partial design matrices is by the author in cooperation with Odijk (2002). The classification of models in Tables 12.3 and 12.4 is by the author. The near rank deficiencies shown in Chap. 13 were generally known, but the derivation is by the author. The known method of double differencing and inter-frequency linear combinations are formalized as a pre-elimination transformation in this part.

Possible improvements
Based on the theory given in this chapter, several questions can be asked to which we come back in Part V: What is the impact of troposphere weighting; should we use a troposphere-weighted model or the troposphere-constant or float model? What is the impact of the batch size? Over what distances is the troposphere-fixed model preferred? What improvements can be reached by modeling residual STDs? How 'bad' is the near rank deficiency of Chap. 13? A related question is: How do model choices compare as to the known precision-improving ambiguity fixing and increasing the zenith cut-off angle? Before we can answer these questions, the stochastic model of the troposphere-weighted model and the covariance matrix of the residual STDs needs to be specified. We do so in Part III.

<div align="right">Appendix D</div>

The Kronecker product

The following definition and the results, Eqs. (D.2–D.10), are taken from [*Rao*, 1973].

DEFINITION. Let $A = (a_{ij})$ and $B = (b_{ij})$ be $m \times n$ and $p \times q$ matrices respectively. Then the Kronecker product

$$A \otimes B \doteq (a_{ij}B) \tag{D.1}$$

is an $mp \times nq$ matrix expressible as a partitioned matrix with $a_{ij}B$ as the (i,j)th partition, $i = 1, \ldots, m$ and $j = 1, \ldots, n$. ∎

The following results are consequences of Definition (D.1):

$$0 \otimes A = A \otimes 0 = 0; \tag{D.2}$$
$$(A_1 + A_2) \otimes B = (A_1 \otimes B) + (A_2 \otimes B); \tag{D.3}$$
$$A \otimes (B_1 + B_2) = (A \otimes B_1) + (A \otimes B_2); \tag{D.4}$$
$$aA \otimes bB = abA \otimes B; \tag{D.5}$$
$$A_1 A_2 \otimes B_1 B_2 = (A_1 \otimes B_1)(A_2 \otimes B_2); \tag{D.6}$$
$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}, \text{ if the inverses exist;} \tag{D.7}$$
$$(A \otimes B)^- = A^- \otimes B^- \text{ using any generalized inverses;} \tag{D.8}$$
$$(A \otimes B)' = A' \otimes B'; \tag{D.9}$$
$$(A \otimes B)(A^{-1} \otimes B^{-1}) = I. \tag{D.10}$$

Repeated application of Eq. (D.6) gives:

$$A_1 A_2 \cdots A_k \otimes B_1 B_2 \cdots B_k = (A_1 \otimes B_1)(A_2 \otimes B_2) \cdots (A_k \otimes B_k). \tag{D.11}$$

This rule can of course only be applied if the dimensions of the matrices are such that the matrix multiplications on the left-hand side are allowed.

Appendix E

Tables on elimination of rank deficiencies

step	parameters	lumped parameters
1.	initial phases and integer ZD ambiguities	real-valued ZD ambiguities
2.	receiver and satellite hardware delays	receiver hardware delays and SD satellite hardware delays
3.	real-valued ZD ambiguities and receiver hardware delays	SD ambiguities and receiver hardware delays
4.	receiver and satellite clocks	receiver clocks and SD satellite clocks
5.	receiver hardware delays and receiver clocks	receiver clocks
6.	receiver clocks (4 observation types) and ionosphere	receiver clocks (3 observation types) and ionosphere
7a.	SD ambiguities and satellite hardware delays	DD ambiguities and satellite hardware delays
8a.	satellite hardware delays and SD satellite clocks	SD satellite clocks
9a.	SD satellite clocks (4 observation types) and ionosphere	SD satellite clocks (3 observation types) and ionosphere
7b.	satellite hardware delays (phase and code) and SD satellite clocks	satellite hardware delays (phase) and SD satellite clocks
8b.	SD ambiguities and satellite hardware delays (phase)	SD ambiguities

Table E.1   Lumping steps of Fig. E.1.

step	parameters	lumped parameters
1.	initial phases and integer ZD ambiguities	real-valued ZD ambiguities
2.	receiver hardware delays and satellite hardware delays (phase)	receiver hardware delays and SD satellite hardware delays (phase)
3.	real-valued ZD ambiguities and receiver hardware delays	SD ambiguities and receiver hardware delays
4.	receiver and satellite clocks	receiver clocks and SD satellite clocks
5.	receiver hardware delays and clocks	receiver clocks
6.	receiver clocks (2 observation types) and ionosphere	receiver clocks (1 observation type) and ionosphere
7.	SD ambiguities (2 frequencies) and ionosphere	SD ambiguities (1 frequency) and ionosphere
8a.	SD ambiguities and satellite hardware delays	DD ambiguities and satellite hardware delays
9a.	satellite hardware delays and SD satellite clocks	SD satellite clocks
10a.	SD satellite clocks (2 observation types) and ionosphere	SD satellite clocks (1 observation type) and ionosphere
8b.	satellite hardware delays and SD satellite clocks	SD satellite clocks

Table E.2   Lumping steps of Fig. E.2.

step	A_2	x_2	$\overline{x}$	V	A_1	x_1
1.	$[e_p \otimes \Lambda \otimes I_n \otimes I_m]$	N	$\mathcal{A}$	$[1 \otimes I_2 \otimes I_n \otimes e_m,\ 1 \otimes I_2 \otimes -e_n \otimes I_m]$	$[e_p \otimes \Lambda \otimes I_n \otimes e_m,\ e_p \otimes \Lambda \otimes -e_n \otimes I_m]$	$\begin{bmatrix} \phi r \\ \phi s \end{bmatrix}$
2a.	$[\underbrace{I_p \otimes I_4 \otimes -e_n \otimes C_m}_{\text{new}},\ I_p \otimes I_4 \otimes I_n \otimes e_m]$	$\begin{bmatrix} bs^{2..m}(1..p) \\ br(1..p) \end{bmatrix}$	$\begin{bmatrix} bs^{12..m}(1..p) \\ \overline{br}(1..p) \end{bmatrix}$	$\begin{bmatrix} I_p \otimes I_4 \otimes 1 \otimes -e_{m-1} \\ I_p \otimes I_4 \otimes -e_n \otimes 1 \end{bmatrix}$	$[I_p \otimes I_4 \otimes -e_n \otimes c_m]$	$bs^1(1..p)$
2b.	$[\underbrace{e_p \otimes J \otimes -e_n \otimes C_m}_{\text{new}},\ I_p \otimes I_4 \otimes I_n \otimes e_m]$	$\begin{bmatrix} bs^{2..m} \\ br(1..p) \end{bmatrix}$	$\begin{bmatrix} bs^{12..m} \\ \overline{br}(1..p) \end{bmatrix}$	$\begin{bmatrix} 1 \otimes I_2 \otimes 1 \otimes -e_{m-1} \\ e_p \otimes J \otimes -e_n \otimes 1 \end{bmatrix}$	$[e_p \otimes J \otimes -e_n \otimes c_m]$	bs^1
3.	$[\underbrace{e_p \otimes \Lambda \otimes I_n \otimes C_m}_{\text{new}},\ I_p \otimes I_4 \otimes I_n \otimes e_m]$	$\begin{bmatrix} \mathcal{A}^{2..m} \\ \overline{br}(1..p) \end{bmatrix}$	$\begin{bmatrix} \mathcal{A}^{12..m} \\ \overline{br}(1..p) \end{bmatrix}$	$\begin{bmatrix} 1 \otimes I_2 \otimes I_n \otimes -e_{m-1} \\ e_p \otimes \Lambda \otimes I_n \otimes 1 \end{bmatrix}$	$[e_p \otimes \Lambda \otimes I_n \otimes c_m]$	$\mathcal{A}^1$
4.	$[\underbrace{I_p \otimes e_4 \otimes -e_n \otimes C_m}_{\text{new}},\ I_p \otimes e_4 \otimes I_n \otimes e_m]$	$\begin{bmatrix} \delta s^{2..m}(1..p) \\ \delta r(1..p) \end{bmatrix}$	$\begin{bmatrix} \delta s^{12..m}(1..p) \\ \overline{\delta r}(1..p) \end{bmatrix}$	$\begin{bmatrix} I_p \otimes 1 \otimes 1 \otimes -e_{m-1} \\ I_p \otimes 1 \otimes -e_n \otimes 1 \end{bmatrix}$	$[I_p \otimes e_4 \otimes -e_n \otimes c_m]$	$\delta s^1(1..p)$
5.	$[\underbrace{I_p \otimes I_4 \otimes I_n \otimes e_m}_{\text{new}}]$	$\overline{br}(1..p)$	$\overline{\delta r},.1..4(1..p)$	$[I_p \otimes e_4 \otimes I_n \otimes 1]$	$[I_p \otimes e_4 \otimes I_n \otimes e_m]$	$\overline{\delta r}(1..p)$
6.	$[\underbrace{I_p \otimes C_4 \otimes I_n \otimes e_m,\ I_p \otimes \mu \otimes I_n \otimes I_m}_{\text{new, final (a,b)}}]$	$\begin{bmatrix} \overline{\delta r},2..4(1..p) \\ \iota(1..p) \end{bmatrix}$	$\begin{bmatrix} \overline{\delta r},2..4(1..p) \\ \overline{\iota}(1..p) \end{bmatrix}$	$\begin{bmatrix} I_p \otimes \nu \otimes I_n \otimes 1 \\ I_p \otimes \mu_2 \otimes I_n \otimes e_m \end{bmatrix}$	$[I_p \otimes c_4 \otimes I_n \otimes e_m]$	$\overline{\delta r},1(1..p)$
7a.	$[\underbrace{e_p \otimes \Lambda \otimes C_n \otimes C_m,\ I_p \otimes I_4 \otimes -e_n \otimes C_m}_{\text{new, final}}]$	$\begin{bmatrix} \mathcal{A}^{12..m}_{2..n}(1..p) \\ bs^{12..m}(1..p) \end{bmatrix}$	$\begin{bmatrix} \mathcal{A}^{12..m}_{12..n}(1..p) \\ \overline{bs}^{12..m}(1..p) \end{bmatrix}$	$\begin{bmatrix} 1 \otimes I_2 \otimes -e_{n-1} \otimes I_{m-1} \\ e_p \otimes \Lambda \otimes -1 \otimes I_{m-1} \end{bmatrix}$	$[e_p \otimes \Lambda \otimes c_n \otimes C_m]$	$\mathcal{A}^{12..m}_1$
8a.	$[\underbrace{I_p \otimes I_4 \otimes -e_n \otimes C_m}_{\text{new, final}}]$	$\overline{bs}^{12..m}(1..p)$	$\overline{\delta s}^{12..m}(1..p)$	$[I_p \otimes e_4 \otimes 1 \otimes I_{m-1}]$	$[I_p \otimes e_4 \otimes -e_n \otimes C_m]$	$\delta s^{12..m}(1..p)$
9a.	$[\underbrace{I_p \otimes C_4 \otimes -e_n \otimes C_m,\ I_p \otimes \mu \otimes I_n \otimes I_m}_{\text{new, final}}]$	$\begin{bmatrix} \overline{\delta s},2..4(1..p) \\ \iota(1..p) \end{bmatrix}$	$\begin{bmatrix} \overline{\delta s},2..4 \\ \overline{\iota}(1..p) \end{bmatrix}$	$\begin{bmatrix} I_p \otimes \nu \otimes 1 \otimes I_{m-1} \\ I_p \otimes \mu_2 \otimes -e_n \otimes C_m \end{bmatrix}$	$[I_p \otimes c_4 \otimes -e_n \otimes C_m]$	$\overline{\delta s}^{12..m}_1(1..p)$
7b.	$[\underbrace{I_p \otimes e_4 \otimes -e_n \otimes C_m,\ e_p \otimes J_1 \otimes -e_n \otimes C_m}_{\text{final}}]$	$\begin{bmatrix} \delta s^{12..m}(1..p) \\ bs^{12..m}_{,phase} \end{bmatrix}$	$\begin{bmatrix} \overline{\delta s}^{12..m}(1..p) \\ \overline{bs}^{12..m}_{,phase} \end{bmatrix}$	$\begin{bmatrix} e_p \otimes 1 \otimes 1 \otimes I_{m-1} \\ 1 \otimes -1 \otimes 1 \otimes I_{m-1} \end{bmatrix}$	$[e_p \otimes J_2 \otimes -e_n \otimes C_m]$	$bs^{12..m}_{,code}$
8b.	$[\underbrace{e_p \otimes \Lambda \otimes I_n \otimes C_m}_{\text{final}}]$	$\mathcal{A}^{12..m}$	$\overline{\mathcal{A}}^{12..m}$	$[1 \otimes \begin{pmatrix} -\lambda_1^{-1} \\ -\lambda_2^{-1} \end{pmatrix} \otimes -e_n \otimes I_{m-1}]$	$[e_p \otimes J_1 \otimes -e_n \otimes C_m]$	$bs^{12..m}_{,phase}$

Table E.3 Different lumping steps for model with phase + code observables and two model assumptions: (a) changing satellite hardware delays; (b) stable satellite hardware delays (common delays for phase observables and code observables). The rank-deficient systems $A_1 x_1 + A_2 x_2$ for which $A_2 V = A_1$ are transformed into the systems $A_2 \overline{x}$, where $\overline{x} \leftarrow x_2 + V x_1$. 'new': newly formed partial design matrix. 'final': partial design matrix that is no longer changed (part of full-rank model). This table corresponds with Fig. E.1.
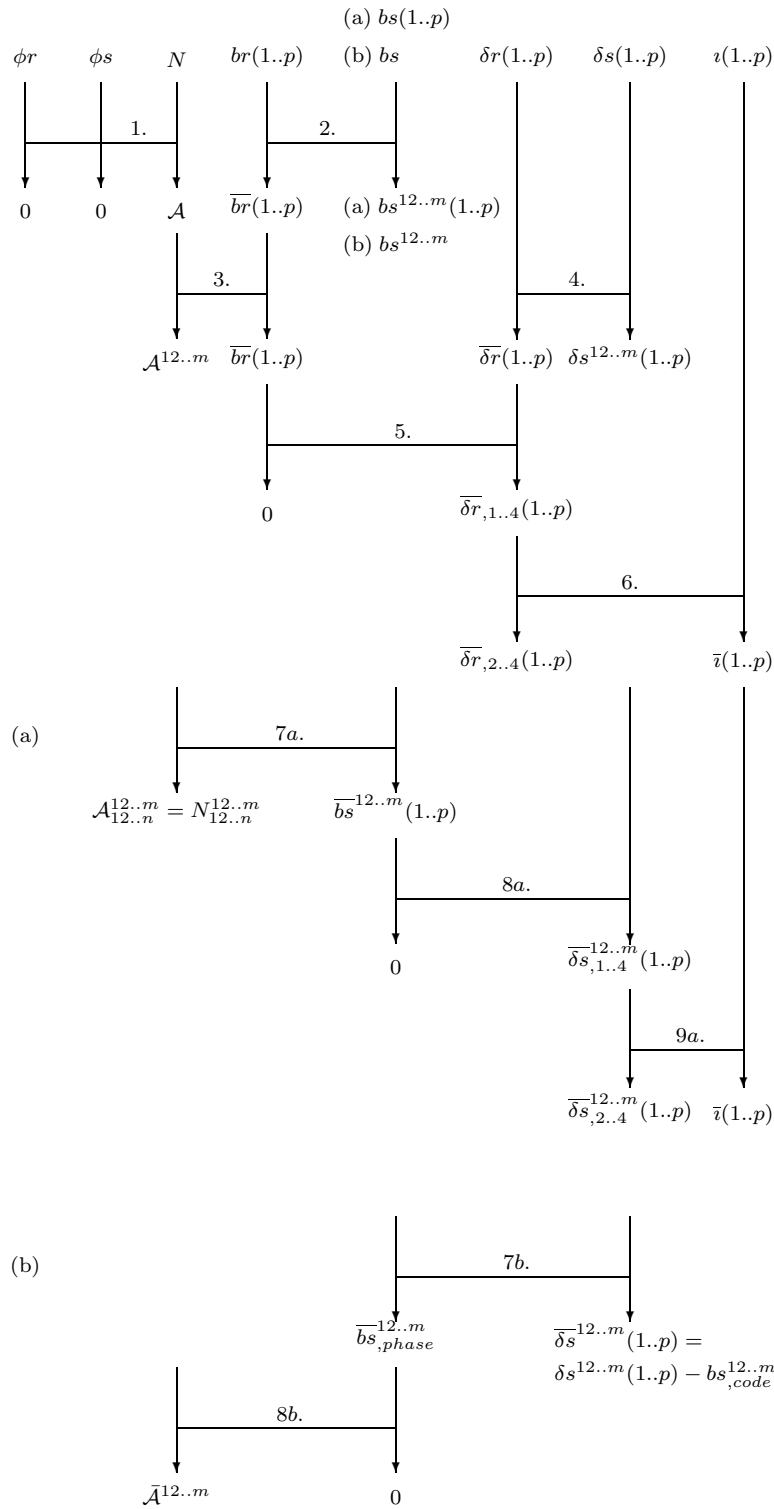
Figure E.1   Lumping of the parameters. <u>Phase + code observables.</u> Top: General steps. Bottom: (a) changing satellite hardware delays (different for all observation types); (b) constant satellite hardware delays (common values for phase and code). Zeros indicate parameters that exist no more after lumping. This figure corresponds with Table E.3.

step	A_2	x_2	$\overline{x}$	V	A_1	x_1
1.	$[e_p \otimes L \otimes I_n \otimes I_m]$	N	$\mathcal{A}$	$[1 \otimes I_2 \otimes I_n \otimes e_m, \ 1 \otimes I_2 \otimes -e_n \otimes I_m]$	$[e_p \otimes L \otimes I_n \otimes e_m, \ e_p \otimes L \otimes -e_n \otimes I_m]$	$\begin{bmatrix} \phi r \\ \phi s \end{bmatrix}$
2a.	$\underbrace{[I_p \otimes I_2 \otimes -e_n \otimes C_m,}_{\text{new}} \ I_p \otimes I_2 \otimes I_n \otimes e_m]$	$\begin{bmatrix} bs^{2..m}_{,phase}(1..p) \\ br(1..p) \end{bmatrix}$	$\begin{bmatrix} bs^{12..m}_{,phase}(1..p) \\ br(1..p) \end{bmatrix}$	$\begin{bmatrix} I_p \otimes I_2 \otimes 1 \otimes -e_{m-1} \\ I_p \otimes I_2 \otimes -e_n \otimes 1 \end{bmatrix}$	$[I_p \otimes I_2 \otimes -e_n \otimes c_m]$	$bs^1(1..p)$
2b.	$\underbrace{[e_p \otimes e_2 \otimes -e_n \otimes C_m,}_{\text{new}} \ I_p \otimes I_2 \otimes I_n \otimes e_m]$	$\begin{bmatrix} bs^{2..m}_{,phase} \\ br(1..p) \end{bmatrix}$	$\begin{bmatrix} bs^{12..m}_{,phase} \\ br(1..p) \end{bmatrix}$	$\begin{bmatrix} 1 \otimes 1 \otimes 1 \otimes -e_{m-1} \\ e_p \otimes e_2 \otimes -e_n \otimes 1 \end{bmatrix}$	$[e_p \otimes e_2 \otimes -e_n \otimes c_m]$	bs^1
3.	$\underbrace{[e_p \otimes L \otimes I_n \otimes C_m,}_{\text{new}} \ I_p \otimes I_2 \otimes I_n \otimes e_m]$	$\begin{bmatrix} \mathcal{A}^{2..m} \\ br(1..p) \end{bmatrix}$	$\begin{bmatrix} \mathcal{A}^{12..m} \\ br(1..p) \end{bmatrix}$	$\begin{bmatrix} 1 \otimes I_2 \otimes I_n \otimes -e_{m-1} \\ e_p \otimes L \otimes I_n \otimes 1 \end{bmatrix}$	$[e_p \otimes L \otimes I_n \otimes c_m]$	$\mathcal{A}^1$
4.	$\underbrace{[I_p \otimes e_2 \otimes -e_n \otimes C_m,}_{\text{new}} \ I_p \otimes e_2 \otimes I_n \otimes e_m]$	$\begin{bmatrix} \delta s^{2..m}(1..p) \\ \delta r(1..p) \end{bmatrix}$	$\begin{bmatrix} \delta s^{12..m}(1..p) \\ \delta r(1..p) \end{bmatrix}$	$\begin{bmatrix} I_p \otimes 1 \otimes 1 \otimes -e_{m-1} \\ I_p \otimes 1 \otimes -e_n \otimes 1 \end{bmatrix}$	$[I_p \otimes e_2 \otimes -e_n \otimes c_m]$	$\delta s^1(1..p)$
5.	$\underbrace{[I_p \otimes I_2 \otimes I_n \otimes e_m]}_{\text{new}}$	$\overline{br}(1..p)$	$\overline{\delta r}_{,1..2}(1..p)$	$[I_p \otimes e_2 \otimes I_n \otimes 1]$	$[I_p \otimes e_2 \otimes I_n \otimes e_m]$	$\overline{\delta r}(1..p)$
6.	$\underbrace{[I_p \otimes C_2 \otimes I_n \otimes e_m, \ I_p \otimes \begin{pmatrix}\mu_1\\\mu_2\end{pmatrix} \otimes I_n \otimes I_m]}_{\text{new, final (a,b)}}$	$\begin{bmatrix} \overline{\delta r}_{,2}(1..p) \\ \iota(1..p) \end{bmatrix}$	$\begin{bmatrix} \overline{\delta r}_{,2}(1..p) \\ \overline{\iota}(1..p) \end{bmatrix}$	$\begin{bmatrix} I_p \otimes -\mu_2\mu_1^{-1} \otimes I_n \otimes 1 \\ I_p \otimes \mu_1^{-1} \otimes I_n \otimes e_m \end{bmatrix}$	$[I_p \otimes c_2 \otimes I_n \otimes e_m]$	$\overline{\delta r}_{,1}(1..p)$
7.	$\underbrace{[e_p \otimes \begin{pmatrix}0\\\lambda_2\end{pmatrix} \otimes I_n \otimes C_m, \ I_p \otimes \begin{pmatrix}\mu_1\\\mu_2\end{pmatrix} \otimes I_n \otimes I_m]}_{\text{new, final (b)}}$	$\begin{bmatrix} \mathcal{A}^{12..m}_{,2} \\ \overline{\iota}(1..p) \end{bmatrix}$	$\begin{bmatrix} \overline{\mathcal{A}}^{12..m}_{,2} \\ \overline{\iota}(1..p) \end{bmatrix}$	$\begin{bmatrix} 1 \otimes -\lambda_2\lambda_1^{-1} \otimes I_n \otimes I_{m-1} \\ e_p \otimes \lambda_2 \otimes I_n \otimes C_m \end{bmatrix}$	$[e_p \otimes \begin{pmatrix}\lambda_1\\0\end{pmatrix} \otimes I_n \otimes C_m]$	$\mathcal{A}^{12..m}_{,1}$
8a.	$\underbrace{[e_p \otimes \begin{pmatrix}0\\\lambda_2\end{pmatrix} \otimes C_n \otimes C_m, \ I_p \otimes I_2 \otimes -e_n \otimes C_m]}_{\text{new, final}}$	$\begin{bmatrix} \mathcal{A}^{12..m}_{2..n,2} \\ bs^{12..m}_{,phase}(1..p) \end{bmatrix}$	$\begin{bmatrix} \overline{\mathcal{A}}^{12..m}_{12..n,2} \\ \overline{bs}^{12..m}_{,phase}(1..p) \end{bmatrix}$	$\begin{bmatrix} 1 \otimes 1 \otimes -e_{n-1} \otimes I_{m-1} \\ e_p \otimes \begin{pmatrix}0\\\lambda_2\end{pmatrix} \otimes -1 \otimes I_{m-1} \end{bmatrix}$	$[e_p \otimes \begin{pmatrix}0\\\lambda_2\end{pmatrix} \otimes c_n \otimes C_m]$	$\mathcal{A}^{12..m}_{1,2}$
9a.	$\underbrace{[I_p \otimes I_2 \otimes -e_n \otimes C_m]}_{\text{new, final}}$	$\overline{bs}^{12..m}_{,phase}(1..p)$	$\overline{\delta s}^{12..m}_{,1..2}(1..p)$	$[I_p \otimes e_2 \otimes 1 \otimes I_{m-1}]$	$[I_p \otimes e_2 \otimes -e_n \otimes C_m]$	$\delta s^{12..m}(1..p)$
10a.	$\underbrace{[I_p \otimes C_2 \otimes -e_n \otimes C_m, \ I_p \otimes \begin{pmatrix}\mu_1\\\mu_2\end{pmatrix} \otimes I_n \otimes I_m]}_{\text{new, final}}$	$\begin{bmatrix} \overline{\delta s}^{12..m}_{,2}(1..p) \\ \overline{\iota}(1..p) \end{bmatrix}$	$\begin{bmatrix} \overline{\delta s}^{12..m}_{,2}(1..p) \\ \overline{\iota}(1..p) \end{bmatrix}$	$\begin{bmatrix} I_p \otimes -\mu_2\mu_1^{-1} \otimes 1 \otimes I_{m-1} \\ I_p \otimes \mu_1^{-1} \otimes -e_n \otimes C_m \end{bmatrix}$	$[I_p \otimes c_2 \otimes -e_n \otimes C_m]$	$\overline{\delta s}^{12..m}_{,1}(1..p)$
8b.	$\underbrace{[I_p \otimes e_2 \otimes -e_n \otimes C_m]}_{\text{final}}$	$\delta s^{12..m}(1..p)$	$\overline{\delta s}^{12..m}(1..p)$	$[e_p \otimes 1 \otimes 1 \otimes I_{m-1}]$	$[e_p \otimes e_2 \otimes -e_n \otimes C_m]$	$bs^{12..m}_{,phase}$

Table E.4   Different lumping steps for model with phase observables only and two model assumptions: (a) changing satellite hardware delays; (b) stable satellite hardware delays (common delays for phase observables). The rank-deficient systems $A_1 x_1 + A_2 x_2$ for which $A_2 V = A_1$ are transformed into the systems $A_2 \overline{x}$, where $\overline{x} \leftarrow x_2 + V x_1$. 'new': newly formed partial design matrix. 'final': partial design matrix that is no longer changed (part of full-rank model). This table corresponds with Fig. E.2.
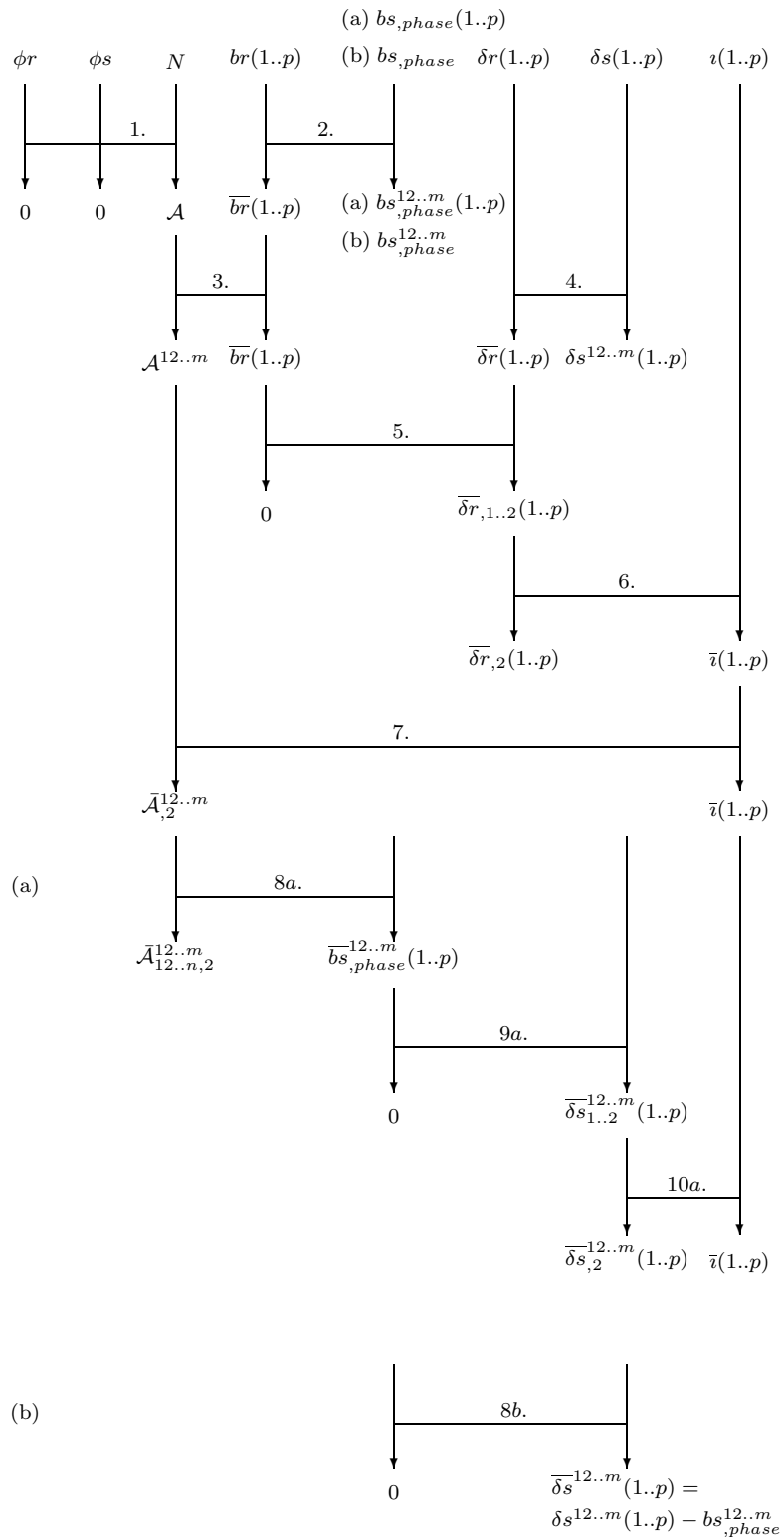
Figure E.2   Lumping of the parameters. Phase observables only. Top: General steps. Bottom: (a) changing satellite hardware delays (different for all observation types); (b) constant satellite hardware delays (common delays for phase observables). Zeros indicate parameters that exist no more after lumping. This figure corresponds with Table E.4.

Bibliography

Beutler, G., E. Brockmann, R. Dach, P. Fridez, W. Gurtner, U. Hugentobler, J. Johnson, L. Mervart, M. Rothacher, S. Schaer, T. Springer, and R. Weber (2000), Bernese GPS Software, Version 4.2, *software documentation*, AUIB.

Han, S.-C., J.H. Kwon, and C. Jekeli (2001), Accurate absolute positioning through satellite clock error determination, *Journal of Geodesy, 75*, 33–43.

Jonge, P.J. de (1998), *A processing strategy for the application of the GPS in networks*, Ph.D. dissertation, Delft University of Technology.

Kouba, J., and T. Springer (2001), New IGS station and satellite clock combination, *GPS Solutions, 4*, 31–36.

Kouba, J., Y. Mireault, and F. Lahaye (1995), 1994 IGS orbit/clock combination and evaluation, in *IGS 1994 Annual Report*, pp. 70–94.

MacMillan, D.S. (1995), Atmospheric gradients from very long baseline interferometry observations, *Geophysical Research Letters, 22*, 1041–1044.

Odijk, D. (2000), Weighting ionosphere corrections to improve fast GPS positioning over medium distances, in *ION GPS 2000, Salt Lake City*, pp. 1113–1123.

Odijk, D. (2002), *Fast precise GPS positioning in the presence of ionospheric delays*, Ph.D. dissertation, Delft University of Technology.

Rao, C.R. (1973), *Linear Statistical Inference and Its Applications*, John Wiley & Sons, New York.

Rothacher, M. (1992), Orbits of satellite systems in space geodesy, *Geodaetisch-geophysikalische Arbeiten in der Schweiz 46*, Schweizerischen Geodaetischen Kommission.

Saastamoinen, J. (1972), Atmospheric correction for the troposphere and stratosphere in radio ranging of satellites, *The use of artificial satellites for geodesy, Geophys. Monogr. Ser., 15*, 247–251.

Teunissen, P.J.G., and A. Kleusberg (1998), GPS observation equations and positioning concepts, in *GPS for Geodesy*, edited by P.J.G. Teunissen and A. Kleusberg, pp. 187–229, Springer-Verlag.

Zumberge, J.F., M.B. Heflin, D.C. Jefferson, M.M. Watkins, and F.H. Webb (1997), Precise point positioning for the efficient and robust analysis of GPS data from large networks, *Journal of Geophysical Research (B), 102*, 5005–5017.

STOCHASTIC MODELING OF (SLANT) TROPOSPHERIC DELAYS OBSERVED BY GPS

Symbols in Part III

symbol	meaning
$\forall$	for all
α	azimuth
C_X^2	constant (scaling of power-law structure function of $\underline{X}(r)$)
$\underline{d}(k)$	constraints on $\underline{D}^z(k) - \underline{D}^z(k-1)$
$\mathbf{D}_X(r)$	structure function of $\underline{X}(r)$
$D^z(k)$	zenith delay(s) on epoch k
$D(\alpha, z)$	Slant Tropospheric Delay
$\Delta D(\alpha, z)$	(scalar) residual STD as function of α, z
$\Delta D(k)$	(vector) of residual STDs at epoch k
$\Gamma(a)$	Gamma function of a
h	height
H	effective height
j	imaginary unit: $\sqrt{-1}$
k	kth epoch
κ	angular frequencies [rad m^{-1}]
$M(z)$	mapping function
N	refractivity
p	power of power-law structure function
p	number of epochs
$Q_X(t_1, t_2)$	autocovariance function of $\underline{X}(t_1)$ and $\underline{X}(t_2)$
$Q_{XY}(t_1, t_2)$	cross-covariance function of $\underline{X}(t_1)$ and $\underline{Y}(t_2)$
r	distance $|r_2 - r_1|$
$R_X(t_1, t_2)$	autocorrelation of $\underline{X}(t)$
$R_X(\tau)$	autocorrelation of (stationary) $\underline{X}(\tau)$
$R_{XY}(t_1, t_2)$	cross-correlation of $\underline{X}(t)$ and $\underline{Y}(t)$
$S_X(\omega)$	Power Spectral Density of $\underline{X}(t)$
t	time
τ	time lag $|t_2 - t_1|$
$x(t)$	representation of stochastic process $\underline{X}(t)$
$X(\omega)$	spectral representation of process $\underline{X}(t)$
$\underline{X}_{ij}$	difference $\underline{X}(r_j) - \underline{X}(r_i)$
$\underline{y}_k$	observables at epoch k
z	zenith angle
ω	angular frequencies [rad s^{-1}]

Chapter 16

Introduction to Part III

Besides the actual GPS observables, GPS observation models may also contain pseudo-observables to improve the quality of the parameters of interest. A good stochastic model for both types of observables is needed for weighting the observables and for hypothesis testing. Part III describes stochastic models of tropospheric pseudo-observables, based on stochastic characteristics of the tropospheric refractivity. Other types of pseudo-observables are beyond our scope.

We consider an observation model with two types of pseudo-observables, cf. Part II:

1. spatiotemporal constraints[1] on Zenith Tropospheric Delays (ZTDs);

2. constraints on residual Slant Tropospheric Delays (STDs): differences between actual STDs and mapped ZTDs.

For both types of constraints the mathematical expectation is assumed zero and in the actual processing of the GPS data, zero sample values are used. It is therefore sufficient to find auto- and cross-correlation functions to describe the covariance matrices of these constraints.

The derivations of the covariance matrices of the constraints are given in Chap. 18. First a model is derived for inter-station differenced tropospheric delays. Then, based on this model and the assumption of a 'frozen troposphere', a model is given for the inter-epoch single differences. And finally, a model is given of the residual STDs. Because these models rely on several assumptions and definitions, in Chap. 17 first a review is given on definitions and terminology concerning stochastic processes. Three distribution functions are described: the autocorrelation function, the power spectral density function, and the structure function. Our special interest goes to power-law processes, more specifically the fractional Brownian motions, because they are believed to describe the stochastic behavior of the tropospheric refractivity. The distribution functions for power-law processes are given in Chap. 17 and applied in Chap. 18. Chapter 19 gives the conclusions of this part.

---

[1]Models using these constraints are referred to as troposphere-weighted models.

<div align="right">

Chapter 17

Power-law processes

</div>

17.1 Definitions and terminology

In this section we introduce some basic statistical concepts. The most important definitions are given and three important representations are shown of wide-sense stationary stochastic processes: the power spectral density function, the autocorrelation function, and the structure function.

A signal can be presented in either the time domain or frequency domain (spectrum). Their representations are written as $x(t)$ and $X(\omega)$ respectively, where t is the time $[s]$, $\omega \doteq 2\pi f$ the angular frequency $[\text{rad s}^{-1}]$, and f the frequency $[s^{-1}]$.

Fourier transform

The *Fourier transform* of a signal $x(t)$ is given by [*Peebles*, 1987]:

$$\boxed{X(\omega) \doteq \int_{-\infty}^{\infty} x(t)\exp(-j\omega t)dt,}$$

(17.1)

and the *inverse Fourier transform* is given by:

$$\boxed{x(t) \doteq \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)\exp(j\omega t)d\omega,}$$

(17.2)

with $j \doteq \sqrt{-1}$.

The factor $1/2\pi$ is sometimes included in the first equation and sometimes equally distributed over both equations. We continue with the definition as above because it seems to be most frequently used. Sufficient but not necessary conditions for the existence of a Fourier pair are [*Peebles*, 1987]:

1. $x(t)$ is bounded with at most a finite number of maxima and minima and a finite number of discontinuities in any finite time interval, and

2. $\int_{-\infty}^{\infty} |x(t)|dt < \infty.$

Equations (17.1) and (17.2) can also be written as:

$$\begin{aligned} X(\omega) &= \mathcal{F}\{x(t)\}; \\ x(t) &= \mathcal{F}^{-1}\{X(\omega)\}, \end{aligned}$$

(17.3)

$\mathcal{F}\{.\}$

where $\mathcal{F}\{\cdot\}$ and $\mathcal{F}^{-1}\{\cdot\}$ are symbolic notations for the Fourier and inverse Fourier transform. Note that if $\mathcal{F}\{x(t)\} = X(\omega)$, then $\mathcal{F}\{x(-t)\} = X(-\omega)$ follows from the definition.

The *mathematical expectation* of a continuous, random (stochastic) process
E{.} $\underline{X}(t)$ is defined as [*Peebles*, 1987]:

$$\mathbf{E}\{\underline{X}(t)\} \doteq \int_{-\infty}^{\infty} x \cdot p_X(x;t)dx, \tag{17.4}$$

auto-   where $p_X(x;t)$ is the *probability density function* of $\underline{X}(t)$, and $x(t)$ denotes a rea-
correlation lization of $\underline{X}(t)$. The *autocorrelation function* of $\underline{X}(t)$ is defined as [*Peebles*, 1987]:

$$R_X(t_1, t_2) \doteq \mathbf{E}\{\underline{X}(t_1)\underline{X}(t_2)^*\}, \tag{17.5}$$

auto-   where * denotes the complex conjugate. The *autocovariance function* of $\underline{X}(t)$ is
covariance defined as

$$Q_X(t_1, t_2) \doteq \mathbf{E}\{(\underline{X}(t_1) - \mathbf{E}\{\underline{X}(t_1)\})(\underline{X}(t_2) - \mathbf{E}\{\underline{X}(t_2)\})^*\}, \tag{17.6}$$

cross-   and the *cross-correlation* and *cross-covariance function* between two processes $\underline{X}(t)$
correlation and $\underline{Y}(t)$ are respectively:
cross-
covariance
$$R_{XY}(t_1, t_2) \doteq \mathbf{E}\{\underline{X}(t_1)\underline{Y}(t_2)^*\};$$
$$Q_{XY}(t_1, t_2) \doteq \mathbf{E}\{(\underline{X}(t_1) - \mathbf{E}\{\underline{X}(t_1)\})(\underline{Y}(t_2) - \mathbf{E}\{\underline{Y}(t_2)\})^*\}. \tag{17.7}$$

For a *real process* $\underline{X}(t) = \underline{X}(t)^*$, and the autocorrelation, cross-correlation, auto-
covariance, and cross-covariance function read respectively:

$$\begin{aligned}
R_X(t_1, t_2) &= \mathbf{E}\{\underline{X}(t_1)\underline{X}(t_2)\}; \\
R_{XY}(t_1, t_2) &= \mathbf{E}\{\underline{X}(t_1)\underline{Y}(t_2)\}; \\
Q_X(t_1, t_2) &= \mathbf{E}\{(\underline{X}(t_1) - \mathbf{E}\{\underline{X}(t_1)\})(\underline{X}(t_2) - \mathbf{E}\{\underline{X}(t_2)\})\}; \\
Q_{XY}(t_1, t_2) &= \mathbf{E}\{(\underline{X}(t_1) - \mathbf{E}\{\underline{X}(t_1)\})(\underline{Y}(t_2) - \mathbf{E}\{\underline{Y}(t_2)\})\}.
\end{aligned} \tag{17.8}$$

wide-sense $\underline{X}(t)$ is called *wide-sense stationary* if [*Peebles*, 1987]:
stationary
$$\begin{aligned}
&1. \ \mathbf{E}\{\underline{X}(t)\} = \mu = \text{constant (not a function of time)}; \\
&2. \ R_X(t_1, t_2) = R_X(\tau), \ \tau \doteq |t_2 - t_1|.
\end{aligned} \tag{17.9}$$

The autocovariance function of a real wide-sense stationary process $\underline{X}(t)$ is:

$$\begin{aligned}
Q_X(t_1, t_2) &= \mathbf{E}\{(\underline{X}(t_1) - \mu)(\underline{X}(t_2) - \mu)\} \\
&= R_X(t_1, t_2) - \mu^2 \\
&= R_X(\tau) - \mu^2 \\
&\equiv Q_X(\tau).
\end{aligned} \tag{17.10}$$

The *variance* of this process is defined as $Q_X(0) = R_X(0) - \mu^2$.

ergodic   Often a process is considered to be *ergodic*, in which case [*Peebles*, 1987]:

$$\begin{aligned}
\mathbf{E}\{\underline{X}(t)\} &= <x(t)>; \\
\mathbf{E}\{\underline{X}(t)\underline{X}(t+\tau)\} &= <x(t)x(t+\tau)>,
\end{aligned} \tag{17.11}$$

using the *time average* notation

$$\boxed{<\cdot> \doteq \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \cdot \, dt.} \tag{17.12}$$

ensemble   Generally, when integrating over any dimension (time, distance), we speak of an
average   *ensemble average*.

Now, let $x_T(t)$ be the following function:

$$x_T(t) \doteq \begin{cases} x(t) & \text{for } -T < t < T; \\ 0 & \text{elsewhere.} \end{cases}$$

The *power spectral density function* or power spectrum of $\underline{X}(t)$ is then defined as [*Peebles*, 1987]:

$$S_X(\omega) \doteq \lim_{T \to \infty} \frac{1}{2T} \mathbf{E}\{|X_T(\omega)|^2\}. \tag{17.13}$$

Since $x(t)$ is limited, there exists a Fourier transform $X(\omega)$ and

$$\begin{aligned} S_X(\omega) &= \lim_{T \to \infty} \frac{1}{2T} \mathbf{E}\{\int_{-T}^{T} \underline{X}(t_1) \exp(j\omega t_1) \, dt_1 \int_{-T}^{T} \underline{X}(t_2) \exp(-j\omega t_2) \, dt_2\} \\ &= \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T}\!\!\int_{-T}^{T} \mathbf{E}\{\underline{X}(t_1)\underline{X}(t_2)\} \exp(-j\omega(t_2 - t_1)) \, dt_2 \, dt_1 \\ &= \lim_{T \to \infty} \frac{1}{2T} \int_{-T-t}^{T-t}\!\!\int_{-T}^{T} R_X(t, t+\tau) \, dt \, \exp(-j\omega\tau) \, d\tau \\ &= \int_{-\infty}^{\infty} \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} R_X(t, t+\tau) \, dt \, \exp(-j\omega\tau) \, d\tau \\ &= \int_{-\infty}^{\infty} < R_X(t, t+\tau) > \exp(-j\omega\tau) \, d\tau, \end{aligned}$$
$$\tag{17.14}$$

using Eqs. (17.13), (17.1), and (17.8) as well as the change of variables $t = t_1$, $\tau = t_2 - t_1$. If $\underline{X}(t)$ is a wide-sense stationary and ergodic process, the power spectral density function reads:

$$\boxed{S_X(\omega) = \mathcal{F}\{R_X(\tau)\} \equiv \int_{-\infty}^{\infty} R_X(\tau) \exp(-j\omega\tau) d\tau,} \tag{17.15}$$

and from the inverse Fourier transform follows:

$$\boxed{R_X(\tau) = \mathcal{F}^{-1}\{S_X(\omega)\} \equiv \frac{1}{2\pi} \int_{-\infty}^{\infty} S_X(\omega) \exp(j\omega\tau) d\omega.} \tag{17.16}$$

If $\underline{X}(t)$ is real, Eq. (17.16) can also be written as

$$\begin{aligned} R_X(\tau) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} S_X(\omega) \cos(\omega\tau) d\omega \\ &= \frac{1}{\pi} \int_{0}^{\infty} S_X(\omega) \cos(\omega\tau) d\omega, \end{aligned} \tag{17.17}$$

since $R_X(\tau)$ and $S_X(\omega)$ are even.

The *average power* of a wide-sense stationary and ergodic process is:

$$R_X(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_X(\omega) d\omega, \tag{17.18}$$

which is the same as the variance for a zero-mean process.

The *time derivative* of $x(t)$ is:

$$\dot{x}(t) = \int_{-\infty}^{\infty} X(\omega) j\omega \exp(j\omega t) d\omega. \tag{17.19}$$

Therefore, $\dot{x}(t)$ and $j\omega X(\omega)$ form a Fourier pair, and from Eq. (17.14) follows that:

$$S_{\dot{X}}(\omega) = \omega^2 S_X(\omega). \tag{17.20}$$

power-law This is of special interest for processes with a power-law behavior: $S_X(\omega) = S_0 \cdot |\omega|^{-1-p}$, with S_0 some scaling factor. The spectrum of the time derivative is then shifted by a power 2: $S_{\dot{X}}(\omega) = S_0 \cdot |\omega|^{1-p}$. The case $p = 1$ **random walk** corresponds to a random-walk process and is also called Brownian motion. The time derivative of this process has a spectral power of zero, which corresponds to white noise, and the Power Spectral Density (PSD) is a constant (S_0). The total **fractional** range $0 < p < 2$ is referred to as 'fractional Brownian motion' [*Agnew*, 1992].
Brownian
motion

We may also think of stochastic processes that are functions of distance instead of time. The above given definitions are then equally applicable if we replace t_1 and t_2 by r_1 and r_2, and τ by $r = |r_2 - r_1|$; where r_1, r_2, and r are in distance units (we use [m]). The spectrum of a wide-sense stationary process is then given in terms of angular frequency κ [rad/m]:

$$S_X(\kappa) = \int_{-\infty}^{\infty} R_X(r) \exp(-j\kappa r) dr. \tag{17.21}$$

random Random functions of three variables are called *random fields*. The concept of
fields stationarity then generalizes to the concept of *homogeneity* [*Tatarski*, 1961]. A
homogeneity random field is called homogeneous if:

1. $\mathbf{E}\{\underline{X}(\vec{r})\} = \mu = $ constant;
2. $R_X(\vec{r}_1, \vec{r}_2) = R_X(\vec{r}_1 + \vec{r}_0, \vec{r}_2 + \vec{r}_0).$ (17.22)

By choosing $\vec{r}_0 = -\vec{r}_1$, one can see that $R_X(\vec{r}_1, \vec{r}_2) = R_X(\vec{r}_2 - \vec{r}_1) \doteq R_X(\vec{r})$, where
isotropic $\vec{r} \doteq \vec{r}_2 - \vec{r}_1$. A homogeneous random field is called *isotropic* if $R_X(\vec{r}) = R_X(|\vec{r}|)$. A random field is *locally homogeneous* in a region G if the conditions of Eq. (17.22) hold for any $\vec{r}_1$ and $\vec{r}_2$ in G, and *locally isotropic* in the region G if $R_X(\vec{r}) = R_X(|\vec{r}|)$ for any $\vec{r}$ in G [*Tatarski*, 1961].

structure The *structure function* or variogram of $\underline{X}(\vec{r})$ is defined as [*Tatarski*, 1961]:
function
$$\begin{aligned} \mathbf{D}_X(\vec{r}) &\doteq < [x(\vec{r}_2) - x(\vec{r}_1)]^2 > \\ &= < x(\vec{r}_2)x(\vec{r}_2) > + < x(\vec{r}_1)x(\vec{r}_1) > - \\ &\quad < x(\vec{r}_1)x(\vec{r}_2) > - < x(\vec{r}_2)x(\vec{r}_1) > . \end{aligned} \tag{17.23}$$

For a (locally) homogeneous and ergodic random field the structure function reduces to:

$$\mathbf{D}_X(\vec{r}) = 2[R_X(0) - R_X(\vec{r})], \tag{17.24}$$

and for a(n) (locally) isotropic and ergodic random field the structure function reads:

$$\mathbf{D}_X(r) = 2[R_X(0) - R_X(r)], \text{ with } r = |\vec{r}|. \tag{17.25}$$

If $R_X(\infty) = 0$, which is reasonable for many physical processes, then

$$\mathbf{D}_X(\infty) = 2R_X(0). \tag{17.26}$$

In other words, the structure function approaches some constant (sill) value as it approaches infinity. From Eqs. (17.25) and (17.26) follows that the normalized autocorrelation function then reads:

$$\frac{R_X(r)}{R_X(0)} = 1 - \frac{\mathbf{D}_X(r)}{\mathbf{D}_X(\infty)}. \tag{17.27}$$

17.2   Power-law distribution functions

A special class of stochastic processes is formed by processes with power-law distribution functions. Three distribution functions were introduced in the previous section: the structure function, the autocorrelation function, and the power spectral density function. We start with the power-law formulation of the structure function and derive the other two.

Suppose we have a real, isotropic, and ergodic random process $\underline{X}(r)$ (or random field $\underline{X}(\vec{r})$ if $r = |\vec{r}|$) with a power-law *structure function*

$$\boxed{\mathbf{D}_X(r) = C_X^2 r^p,} \tag{17.28}$$

where $0 < p < 2$ and C_X^2 is a positive constant [m^{2-p}]. From Eq. (17.25) then follows that the *autocorrelation function* reads:

$$R_X(r) = R_X(0) - \frac{1}{2}\mathbf{D}_X(r), \tag{17.29}$$

and the *power spectral density function* reads (see App. F):

$$
\begin{aligned}
S_X(\kappa) &= \mathcal{F}\{R_X(r)\} = \mathcal{F}\{R_X(0) - \tfrac{1}{2}\mathbf{D}_X(r)\} \\
&= \mathcal{F}\{R_X(0) - \tfrac{1}{2}C_X^2|r|^p\} \\
&= 2\pi R_X(0)\delta(\kappa) + C_X^2\Gamma(p+1)\sin(p\pi/2)|\kappa|^{-(p+1)},
\end{aligned} \tag{17.30}
$$

with $\delta(.)$ the Dirac function[1]. So this random field has infinite spectral density at $\kappa = 0$ and infinite average power, and is in fact not realizable.

The random field can however be realizable if we assume that the spectrum is limited to frequencies larger than some (positive low) frequency $\kappa_{\min}$ [rad m^{-1}]. Based on this assumption, we approximate the autocorrelation function and the structure function. Since we always work with finite length processes, this means we make an assumption about the behavior of the autocorrelation function beyond the length $r_{\max}$ of the process, if $\kappa_{\min} \leq 2\pi/r_{\max}$. Based on this assumption we can start with the power-law formulation of the Power Spectrum and again derive the other two distribution functions.

Assume we have the following power spectrum of a real, wide-sense stationary, and ergodic stochastic process $\underline{X}(r)$ [m]:

$$
S_X(\kappa) = \begin{cases} S_0 \cdot |\kappa|^{-(p+1)} & \text{for } |\kappa| \geq \kappa_{\min}; \\ S_0 \cdot \kappa_{\min}^{-(p+1)} & \text{for } |\kappa| \leq \kappa_{\min}, \end{cases} \tag{17.31}
$$

for which $0 < p < 2$, $S_0 \doteq \Gamma(p+1)\sin(p\pi/2)C_X^2$ a positive constant, and $\kappa_{\min}r \ll 1$ for $0 < r < r_{\max}$. The autocorrelation function of $\underline{X}(r)$ then reads (see App. G):

$$R_X(r) \approx \frac{(p+1)S_0}{p\pi\kappa_{\min}^p} - \frac{S_0}{2\Gamma(p+1)\sin(p\pi/2)}r^p. \tag{17.32}$$

This autocorrelation function is of the form

$$\boxed{R_X(r) = c_0 - c_1 r^p,} \tag{17.33}$$

---

[1] $\int_{-\infty}^{\infty} f(t-x)\delta(x)dx \doteq f(t);\ \delta(x) = 0,\ \forall x \neq 0;\ \int_{-\infty}^{\infty} \delta(x)dx = 1.$
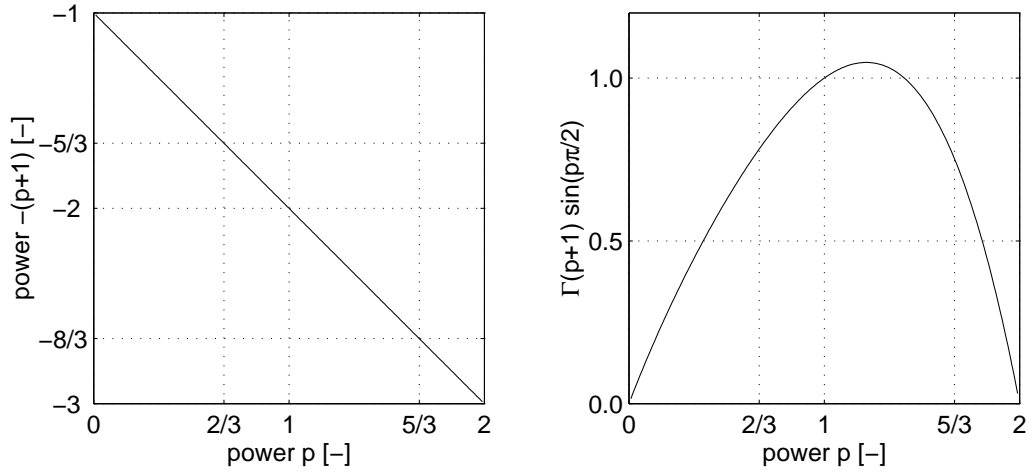
Figure 17.1   Relation between power p of the structure function (and autocorrelation function) and the spectral power $-(p+1)$ (left panel) and constant $\Gamma(p+1)\sin(p\pi/2)$ (right panel) in the range $0 < p < 2$. The case $p = 1$ corresponds to a random-walk process. $p = 2/3$ and $p = 5/3$ correspond to the extreme cases of the structure function of Sect. 18.3.

with c_0 and c_1 positive constants:

$$c_0 = R_X(0) = \pi^{-1}(p+1)\Gamma(p)\sin(\frac{p\pi}{2})C_X^2\kappa_{\min}^{-p} \quad ; \quad c_1 = \frac{1}{2}C_X^2. \qquad (17.34)$$

This autocorrelation function has a finite variance that strongly depends on $\kappa_{\min}$: when $\kappa_{\min} \to 0$ then $c_0 \to \infty$. For $p = 1$ we have the well-known case of the autocorrelation function of Baarda's criterion matrix [$Baarda$, 1973] for the description of the coordinate precision in geodetic networks. In that case $c_0 = 2C_X^2/\pi\kappa_{\min}$. Clearly, for a small $\kappa_{\min}$, c_0 becomes large, but it is not in the structure function we started with[2]:

$$\begin{aligned}
\mathbf{D}_X(r) &= 2[R_X(0) - R_X(r)] \\
&= 2[c_0 - (c_0 - c_1 r^p)] \\
&= 2c_1 r^p \\
&= C_X^2 r^p.
\end{aligned} \qquad (17.35)$$

In other words, contrary to the autocorrelation function, the structure function is not sensitive for the value of $\kappa_{\min}$.

17.3   Power-law cross-correlation of differences

Suppose we have a real, wide-sense stationary, and ergodic process $\underline{X}(r)$ with the power-law structure function of Eq. (17.35). We form differences $\underline{X}_{ij} \doteq \underline{X}(r_j) - \underline{X}(r_i)$ and $\underline{X}_{kl} \doteq \underline{X}(r_l) - \underline{X}(r_k)$. Using the equality

$$(a - b)(c - d) = \frac{1}{2}[(a - d)^2 + (b - c)^2 - (a - c)^2 - (b - d)^2], \qquad (17.36)$$

---

[2]Baarda noticed that the S-transformed (Similarity transform to obtain a nonstochastic basis) criterion matrix did not depend on c_0 either.

the cross-correlation of $\underline{X}_{ij}$ and $\underline{X}_{kl}$ can be found as:

$$
\begin{aligned}
R_{X_{ij}X_{kl}} &= \mathbf{E}\{[\underline{X}(r_j) - \underline{X}(r_i)][\underline{X}(r_l) - \underline{X}(r_k)]\} \\
&= <[x(r_j) - x(r_i)][x(r_l) - x(r_k)]> \\
&= \tfrac{1}{2} <[x(r_j) - x(r_k)]^2 + [x(r_i) - x(r_l)]^2 - \\
&\qquad [x(r_j) - x(r_l)]^2 - [x(r_i) - x(r_k)]^2 > \\
&= \tfrac{1}{2}[\mathbf{D}_X(r_{jk}) + \mathbf{D}_X(r_{il}) - \mathbf{D}_X(r_{jl}) - \mathbf{D}_X(r_{ik})] \\
&= \tfrac{1}{2}C_X^2[r_{jk}^p + r_{il}^p - r_{jl}^p - r_{ik}^p],
\end{aligned}
\tag{17.37}
$$

with $r_{ij} \doteq |r_j - r_i|$.

There are several special cases. The case $k = i \neq j \neq l$ corresponds to the S-transform where all differences are with respect to $\underline{X}(r_i)$, like $\underline{X}_{12}, \underline{X}_{13}, \underline{X}_{14}$, etc. The case $i < j = k < l$ corresponds to a transform to obtain a sequence of differences: $\underline{X}_{12}, \underline{X}_{23}, \underline{X}_{34}$, etc. And in case $k = i \neq l = j$, we have the auto-correlation of a difference, which is by definition equal to the structure function. Another special case is the random-walk process ($p = 1$). The cross-correlation then becomes zero when $i < j \leq k < l$. Table 17.1 gives eight different cases; the four cases at the right-hand side are special cases of the four at the left-hand side.

case ($0 < p < 2$)	cross-correlation	case ($p = 1$)	cross-corr.
$i \neq j \neq k \neq l$	$\tfrac{1}{2}C_X^2[r_{jk}^p + r_{il}^p - r_{jl}^p - r_{ik}^p]$	$i < j < k < l$	0
$i \neq j = k \neq l$	$\tfrac{1}{2}C_X^2[r_{il}^p - r_{jl}^p - r_{ij}^p]$	$i < j = k < l$	0
$k = i \neq j \neq l$	$\tfrac{1}{2}C_X^2[r_{il}^p + r_{ij}^p - r_{jl}^p]$	$k = i < j,l$	$C_X^2 \min(r_{ij}, r_{kl})$
	autocorrelation		autocorr.
$k = i \neq l = j$	$C_X^2 r_{ij}^p$	$k = i \neq l = j$	$C_X^2 r_{ij}$

Table 17.1   Cross-correlation and autocorrelation of differences $\underline{X}_{ij} = \underline{X}(r_j) - \underline{X}(r_i)$ and $\underline{X}_{kl} = \underline{X}(r_l) - \underline{X}(r_k)$ for a process $\underline{X}(r)$ with structure function $\mathbf{D}_X(r) = C_X^2\, r^p$.

Stochastic modeling of troposphere constraints

18.1   Troposphere constraints in the GPS observation model

Based on stochastic models for differences of tropospheric delays, we can compile a stochastic model for pseudo-observables (constraints) in a GPS observation model. In GPS observation models we can distinguish global, temporal, and batch parameters. Global parameters are constant for the complete time span, temporal parameters change on an epoch-to-epoch basis, and batch parameters change slowly from epoch to epoch and may therefore be considered constant for a batch of several epochs. Batch parameters may also be estimated every epoch, but they can still be considered a different parameter group than the temporal parameters if we use relative constraints (this is of no use for the quickly changing temporal parameters). This is what is applied in a troposphere-weighted model. For a more elaborated discussion on different types of parameters, see Part II.

We assume that the ZTDs are the only batch parameters. In static GPS data processing of GPS observations and pseudo-observations the following functional model can be given for three consecutive epochs[1] $k-1$, k, and $k+1$:

$$
\mathbf{E}\left\{
\begin{bmatrix}
\underline{y}(k-1) \\
\underline{d}(k) \\
\underline{y}(k) \\
\underline{d}(k+1) \\
\underline{y}(k+1)
\end{bmatrix}
\right\} =
\begin{bmatrix}
T_{k-1} & B_{k-1} & 0 & 0 & 0 & 0 & A_{k-1} \\
0 & -I_n & 0 & I_n & 0 & 0 & 0 \\
0 & 0 & T_k & B_k & 0 & 0 & A_k \\
0 & 0 & 0 & -I_n & 0 & I_n & 0 \\
0 & 0 & 0 & 0 & T_{k+1} & B_{k+1} & A_{k+1}
\end{bmatrix}
\begin{bmatrix}
t(k-1) \\
\beta(k-1) \\
t(k) \\
\beta(k) \\
t(k+1) \\
\beta(k+1) \\
x
\end{bmatrix}. \quad (18.1)
$$

This model includes the following parameters:

x : global parameters: ambiguities, coordinates;
$\beta(k)$: batch parameters of epoch k: (absolute) ZTDs;
$t(k)$: temporal parameters of epoch k: ionospheric,
       residual STD, and clock parameters.

The vector of observables $\underline{y}(k)$ is a concatenated vector: $\underline{y}(k) = [\underline{y}_{,}(k)', \ \underline{\Delta D}(k)']'$. So there are in total three different types of (pseudo-)observables:

$\underline{y}_{,}(k)$: GPS code and phase observables;
$\underline{\Delta D}(k)$: pseudo-observables of residual STDs;
$\underline{d}(k)$: pseudo-observables of differences of ZTDs.

---

[1]This model generalizes for more epochs but only three epochs are shown for clarity.

Model (18.1) is a linearized troposphere-weighted model with 'observed-minus-computed' observables. The GPS observables are a priori corrected for STDs based on a model like the one in [*Saastamoinen*, 1972], see also Eq. (6.11) in Part I. Zero sample values are taken as pseudo-observations when estimating the parameters; $\beta(k) = [D_1^z(k), ..., D_n^z(k)]'$ is an $n \times 1$ vector of undifferenced ZTDs (with n is the number of stations); T_k, B_k, and A_k in Eq. (18.1) are the partial design matrices for epoch k of the temporal, batch, and global parameters respectively; and I_n is an $n \times n$ identity matrix.

The following stochastic model of untransformed observables is assumed as in Part II:

$$
\mathbf{D}\left\{ \begin{bmatrix} \underline{y}(k-1) \\ \underline{d}(k) \\ \underline{y}(k) \\ \underline{d}(k+1) \\ \underline{y}(k+1) \end{bmatrix} \right\} = \begin{bmatrix} Q_y(k-1) & & & & \\ & Q_d(k) & & & \\ & & Q_y(k) & & \\ & & & Q_d(k+1) & \\ & & & & Q_y(k+1) \end{bmatrix} ;
$$

$$
Q_y(k) \quad = \begin{bmatrix} Q_{y,}(k) & \\ & Q_{\Delta D}(k) \end{bmatrix}.
$$

(18.2)

In other words, the observation types are uncorrelated and there is no correlation between the observables of different epochs. The parameters in this model can be solved in a batch processing or in a recursive filtering procedure. In the first case, time correlation makes the parameter estimation very time consuming and therefore impractical. In the second case the absence of time correlations is a requisite. Generally, the observables are correlated and the block diagonal structure is a simplifying assumption. Still, autocovariance functions are needed to describe the matrices $Q_d(k)$ and $Q_{\Delta D}(k)$ (we assume $Q_{y,}(k)$ is known). These are derived in this chapter.

18.2   Atmospheric turbulence

To derive the stochastic properties of a tropospheric delay, we need the stochastic properties of the refractivity first. The STD, $D(\alpha, z)$, of a radio signal from a satellite that arrives at a ground receiver is a function of the refractivity along the ray path:

$$
D(\alpha, z) = 10^{-6} M(z) \int_0^{h_{max}} N(\vec{r}(h)) \, dh,
$$

(18.3)

where $\vec{r}(h)$ is a vector with coordinates of a point along the ray path as function of height; h_{max} is the top of the (wet) troposphere, α and z are the azimuth and zenith angle of the ray, and $M(z) \approx \sec(z)$ is the mapping function (we assume a horizontally stratified atmosphere).

Tatarski (1961) showed that based on atmospheric turbulence theory, the structure function of the refractivity has a power-law behavior:

$$
\mathbf{D}_N(r) = \mathbf{D}_N(\vec{r}_1, \vec{r}_2) \doteq < [N(\vec{r}_2) - N(\vec{r}_1)]^2 > = C_N^2 r^{\frac{2}{3}} , \ l_0 \le r \le L_0,
$$

(18.4)

where $r = |\vec{r}_2 - \vec{r}_1|$, $C_N[\mathrm{m}^{-\frac{1}{3}}]$ some constant under certain atmospheric conditions, and l_0 and L_0 the inner and outer scale of the turbulence. The refractivity

is thus considered a locally isotropic random field. Equation (18.4) was first obtained (in a slightly different form) in [*Kolmogorov*, 1941] and [*Obukhov*, 1941], and is known as the two-thirds law. Although this function grows with larger values of r, there should be a saturation point where the function becomes constant.[2] The structure function is therefore only valid between the lower and upper bound. A probably better approximation for the complete range of r is given in [*Treuhaft and Lanyi*, 1987]:

$$\mathbf{D}_N(r) = \frac{C_N^2 r^{\frac{2}{3}}}{1 + (r/L)^{\frac{2}{3}}} \; ; \; L = 3000 \text{ km} \rightarrow \mathbf{D}_N(\infty) = C_N^2 L^{\frac{2}{3}}. \tag{18.5}$$

Both functions, Eq. (18.4) and Eq. (18.5), are illustrated in Fig. 18.1. In [*Gradinarsky*, 2002] a more refined model was shown where the expectation and structure function of the refractivity have a different dependence of distance in vertical and horizontal direction. Although this is a more reasonable assumption, it contradicts the assumption of isotropy.
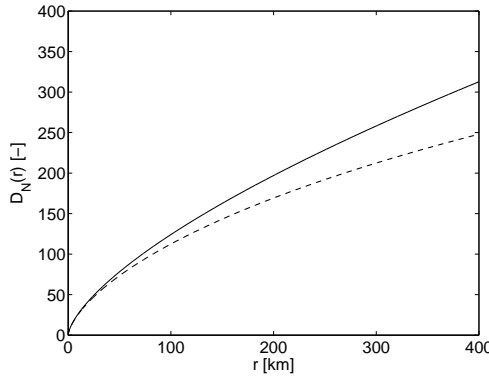
Figure 18.1   Structure function according to Kolmogorov (solid line) and Treuhaft and Lanyi (dashed line) as function of distance. $C_N = 0.24 \text{ m}^{-\frac{1}{3}}$ was assumed as by Treuhaft and Lanyi.

Based on Eq. (18.4) or Eq. (18.5) we can develop stochastic models for single differenced tropospheric delays as described in the sections that follow.

18.3   Inter-station troposphere single difference

For two stations (say 1 and 2) sufficiently close to each other, the zenith angles are approximately equal and the difference of the ZTDs, $D_{12}^z \doteq D_2^z - D_1^z$, then reads:

$$D_{12}^z = 10^{-6} M(z) \int_0^{h_{max}} [N(\vec{r}_2(h)) - N(\vec{r}_1(h))] \, dh. \tag{18.6}$$

From Eqs. (18.6) and (17.37) then follows that (see Fig. 18.2):

$$< [D_2^z - D_1^z]^2 >=$$
$$10^{-12} M^2(z) \int_0^{h_{max}} \int_0^{h_{max}} < [N(\vec{r}_2(h_1)) - N(\vec{r}_1(h_1))] \cdot$$
$$[N(\vec{r}_2(h_2)) - N(\vec{r}_1(h_2))] > dh_1 dh_2 =$$
$$10^{-12} M^2(z) \int_0^{h_{max}} \int_0^{h_{max}} \tfrac{1}{2} \mathbf{D}_N(|\vec{r}_2(h_1) - \vec{r}_1(h_2)|) \; + \; \tfrac{1}{2} \mathbf{D}_N(|\vec{r}_1(h_1) - \vec{r}_2(h_2)|)$$
$$- \tfrac{1}{2} \mathbf{D}_N(|\vec{r}_2(h_1) - \vec{r}_2(h_2)|) \; - \; \tfrac{1}{2} \mathbf{D}_N(|\vec{r}_1(h_1) - \vec{r}_1(h_2)|) \, dh_1 dh_2.$$

---

[2]This would be physically reasonable to obtain zero autocovariance at infinity. And zero autocovariance is sufficient to prove the ergodic property [*Yaglom*, 1962].

$$(18.7)$$
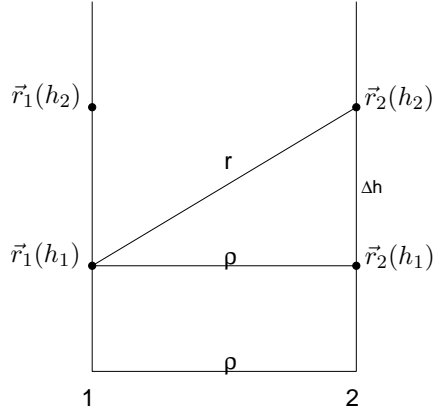
Figure 18.2   Geometry of two parallel zenith rays of a radio signal arriving from one satellite at two receiving stations (1, 2). The refractivity structure function is evaluated at a distance r between two points along the rays and at a distance Δh along one of the rays.

From Fig. 18.2 we can also see that:

$$
\begin{aligned}
|\vec{r}_2(h_2) - \vec{r}_1(h_1)| &= |\vec{r}_1(h_2) - \vec{r}_2(h_1)| = r; \\
|\vec{r}_2(h_2) - \vec{r}_2(h_1)| &= |\vec{r}_1(h_2) - \vec{r}_1(h_1)| = \Delta h,
\end{aligned}
\tag{18.8}
$$

with $\Delta h \doteq h_2 - h_1$ and $r = [\Delta h^2 + \rho^2]^{\frac{1}{2}}$.

The structure function of the delays at two receivers, as function of their distance ρ over the surface, therefore reads:

$$
\begin{aligned}
\mathbf{D}_{D^z}(\rho) &\doteq < [D_2^z - D_1^z]^2 > = \\
&10^{-12} M^2(z) \int_0^{h_{max}} \int_0^{h_{max}} \mathbf{D}_N([\Delta h^2 + \rho^2]^{\frac{1}{2}}) - \mathbf{D}_N([\Delta h^2]^{\frac{1}{2}}) \, dh_1 dh_2.
\end{aligned}
\tag{18.9}
$$

This integral was given in [*Treuhaft and Lanyi*, 1987]; even a more general integral was given for arbitrary slant directions. They replaced h_{max} by the effective height H. Although only about $(1 - \exp(-1)) \times 100\% \approx 63\%$ of the delay is caused by the atmosphere up to the effective height in case of an exponential refractivity profile, this may be reasonable if one only considers the stochastic properties; the Kolmogorov turbulence may not be valid at large altitudes. In fact, Gradinarski (2002) experimentally showed a height dependence of the refractivity structure function.

Using the structure function of Eq. (18.4) and Eq. (18.5), we obtain:

$$
\mathbf{D}_{D^z}(\rho) = 10^{-12} C_N^2 \int_0^{H} \int_0^{H} [\Delta h^2 + \rho^2]^{\frac{1}{3}} - [\Delta h^2]^{\frac{1}{3}} \, dh_1 dh_2
\tag{18.10}
$$

and

$$
\mathbf{D}_{D^z}(\rho) = 10^{-12} C_N^2 \int_0^{H} \int_0^{H} \frac{[\Delta h^2 + \rho^2]^{\frac{1}{3}} - [\Delta h^2]^{\frac{1}{3}}}{1 + L^{-\frac{2}{3}}[\Delta h^2 + \rho^2]^{\frac{1}{3}}} \, dh_1 dh_2,
\tag{18.11}
$$

$u \doteq \rho/\mathsf{H}$ respectively. By changing to the dimensionless variables $x_1 \doteq h_1/\mathrm{H}$, $x_2 \doteq h_2/\mathrm{H}$, and $u \doteq \rho/\mathrm{H}$, Eqs. (18.10) and (18.11) may be replaced by:

$$\boxed{\begin{aligned} \mathbf{D}_{D^z}(\rho) &= \underbrace{10^{-12}C_N^2 \mathrm{H}^{\frac{8}{3}}}_{C_D^2} \overline{\mathbf{D}}_{D^z}(u); \\ &\doteq \quad C_D^2 \quad \overline{\mathbf{D}}_{D^z}(u), \end{aligned}} \tag{18.12}$$

normalized structure functions

where C_D^2 is in $[\mathrm{m}^2]$ and the (dimensionless) normalized structure functions are given as:

$$\boxed{\overline{\mathbf{D}}_{D^z}(u) \doteq \int_0^1 \int_0^1 [\Delta x^2 + u^2]^{\frac{1}{3}} - [\Delta x^2]^{\frac{1}{3}} \, dx_1 dx_2} \tag{18.13}$$

and

$$\boxed{\overline{\mathbf{D}}_{D^z}(u) \doteq \int_0^1 \int_0^1 \frac{[\Delta x^2 + u^2]^{\frac{1}{3}} - [\Delta x^2]^{\frac{1}{3}}}{1 + (\mathrm{H}/L)^{\frac{2}{3}}[\Delta x^2 + u^2]^{\frac{1}{3}}} \, dx_1 dx_2,} \tag{18.14}$$

respectively. Since for any even function holds [*Tatarski*, 1961]:

$$\int_0^b \int_0^b f(x_2 - x_1) \, dx_1 dx_2 = 2 \int_0^b (b - x) f(x) \, dx, \tag{18.15}$$

we may replace the double integral of Eq. (18.13) by a single integral:

$$\overline{\mathbf{D}}_{D^z}(u) = 2 \int_0^1 (1 - x) \cdot \{ [x^2 + u^2]^{\frac{1}{3}} - [x^2]^{\frac{1}{3}} \} \, dx. \tag{18.16}$$

This integral and the integral of Eq. (18.13) do not give simple analytical expressions. The result was first presented in [*Treuhaft and Lanyi*, 1987] and is shown in Fig. 18.3. This figure also shows the integral of Eq. (18.14). Treuhaft and Lanyi (1987) also gave a polynomial approximation of Eq. (18.13). A similar polynomial can be given for Eq. (18.14), and is of the same form:

$$\log(\overline{\mathbf{D}}_{D^z}(u)) = \sum_{i=0}^{10} a_i (\log(u))^i, \tag{18.17}$$

TL2

where the constants a_i are as in Table 18.1; see also Fig. 18.5. This model will be referred to as the TL2 model; see Part V. Contrary to the polynomial of [*Treuhaft and Lanyi*, 1987] the constants depend on H.

i	$a(i)$	i	$a(i)$
0	-0.24642825974092		
1	0.99652032780210	6	-0.00169423152584
2	-0.23032140825582	7	0.00079874140271
3	0.02014514419047	8	0.00002208367722
4	0.01957239963592	9	-0.00003423986001
5	-0.00672815490738	10	0.00000318042977

Table 18.1   Constants $a(i)$ corresponding to Model (18.17) for H = 2 km (TL2 model).

short distance regime

For $u \ll 1$, the function $\overline{\mathbf{D}}_{D^z}(u) = u^p$ shows a power-law behavior with a power of $p = 5/3$ (right panel of Fig. 18.3), which was shown earlier in [*Tatarski*, 1961] for a horizontal planar wave. The integral of Eq. (18.16) can be computed analytically by a computer algebra system. With the symbolic toolbox of MATLAB™ [*Moler and Costa*, 1997] the following expression was found:

$$\frac{6}{5} F(-\frac{1}{3}, -\frac{5}{6}; \frac{1}{6}; -u^2) - \frac{9}{20} - \frac{3}{4}(1 + u^2)^{\frac{4}{3}} + \frac{3}{4} u^{\frac{8}{3}} + \frac{\frac{4}{5}\pi^{\frac{3}{2}}}{\Gamma(\frac{5}{6})\Gamma(\frac{2}{3})} u^{\frac{5}{3}}, \tag{18.18}$$
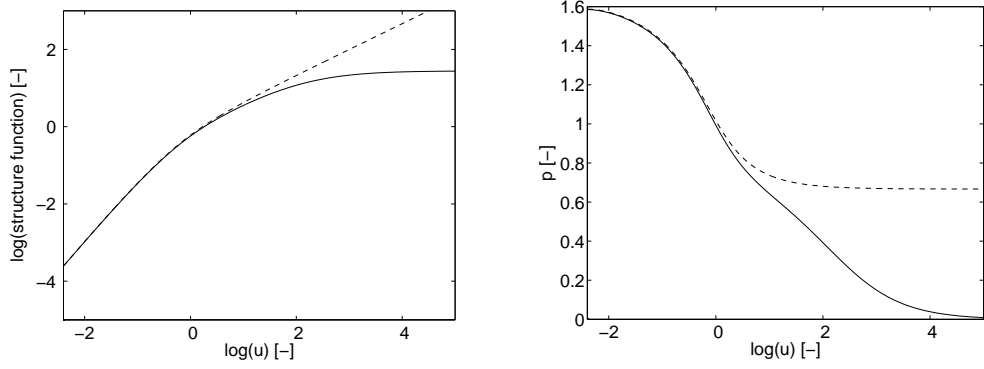
Figure 18.3　Left panel: Normalized structure function $\overline{\mathbf{D}}_{D^z}(u)$. Right panel: Power p of $\overline{\mathbf{D}}_{D^z}(u) = u^p$. Dashed lines: based on Eq. (18.13). Solid lines: based on Eq. (18.14) with H = 2 km.
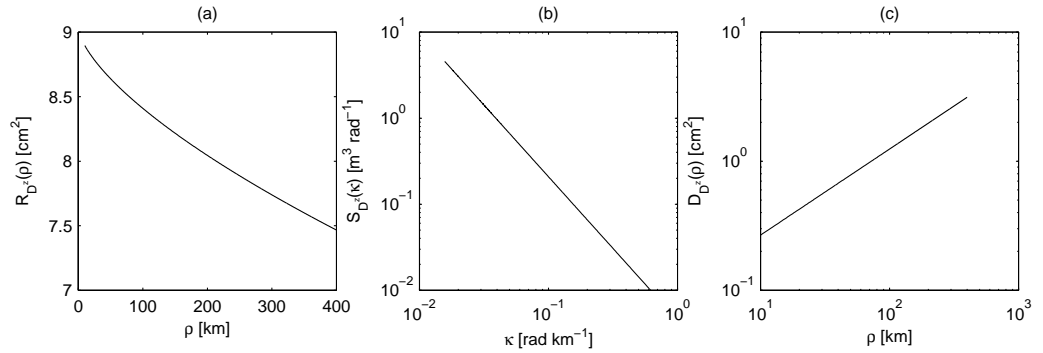
Figure 18.4　Three distribution functions of the ZTD. The range 10 km $< r <$ 400 km is shown where $p \approx \frac{2}{3}$ ($p = \frac{2}{3}$ was assumed) and $\kappa_{\mathsf{min}} = 0.1/400$ rad km^{-1} is used. $C_N = 0.24$ m$^{-\frac{1}{3}}$, H = 1 km. (a) Autocorrelation function $R_{D^z}(\rho) = c_0 - c_1\rho^p$; (b) Power spectral density function $S_{D^z}(\kappa) = S_0\kappa^{-(p+1)}$; (c) Structure function $\mathbf{D}_{D^z}(\rho) = 10^{-12}C_N^2\mathsf{H}^2\rho^p$.

where $\Gamma(.)$ is the Gamma function[3] and

$$F(a, b; c; f(u)) \doteq \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{n=0}^{\infty} \frac{\Gamma(a+n)\Gamma(b+n)}{\Gamma(c+n)} \frac{f(u)^n}{n!} \tag{18.19}$$

is a Gauss hypergeometric series [*Abramowitz and Stegun*, 1970]. For $f(u) = -u^2 \to 0$ this series approaches 1, and the expression of Eq. (18.18) is dominated by the last term, which confirms the 5/3th-power behavior of the structure function for $u \ll 1$.

large distance regime

For $u \gg 1 \geq x$, the integral of Eq. (18.16) must approach $\frac{1}{2}u^{\frac{2}{3}}$ because the expression between braces is approximately $u^{\frac{2}{3}}$. In other words, for large u the normalized structure function becomes $\overline{\mathbf{D}}_{D^z}(u) = u^{\frac{2}{3}}$. This was confirmed numerically.

Similar results, with a transition of power from 5/3 to 2/3, can numerically be found for other α and z than the zenith direction [*Treuhaft and Lanyi*, 1987]. Note that this type of power-law behavior was confirmed in [*Hanssen*, 2001] based on InSAR data. A power of $p = 1$ ($\log(u) \approx 0$) is reached when $\rho \approx$ H; see Fig. 18.3.

Figure 18.5 shows the correlation between two delays as function of the

---

[3]$\Gamma(a) \doteq \int_0^\infty \exp(-x)x^{a-1}dx.$

distance based on the TL2 model and follows from Eq. (17.27) as:

$$\overline{Q}_{D^z} \doteq \frac{R_{D^z}(\rho)}{R_{D^z}(0)} = 1 - \frac{\mathbf{D}_{D^z}(\rho)}{\mathbf{D}_{D^z}(\infty)} = 1 - \frac{\overline{\mathbf{D}}_{D^z}(u)}{\overline{\mathbf{D}}_{D^z}(\infty)}, \tag{18.20}$$

where $\mathbf{D}_{D^z}(\infty)/2 = R_{D^z}(0) = 5$ cm^2. In fact, the value of $L = 3000$ km was chosen in [*Treuhaft and Lanyi*, 1987] such that for H $= 2$ km and $C_N = 0.24$ m$^{-\frac{1}{3}}$ this value was consistent with annual fluctuations of the tropospheric delay at mid-latitudes.
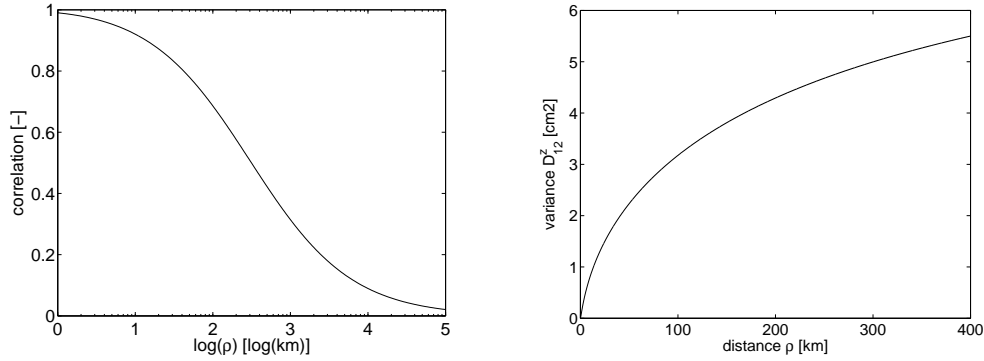
Figure 18.5   Left panel: Correlation between two zenith delays as function of the distance ρ for an assumed effective height of H $= 2$ km. Right panel: Structure function of zenith delay as function of distance according to the TL2 model of Eq. (18.17) and Table 18.1.

18.4   Inter-epoch troposphere single difference

The structure function of inter-epoch troposphere single differenced delays can be derived by assuming a 'frozen turbulence' [*Taylor*, 1938], [*Treuhaft and Lanyi*, 1987]. In this case we may think of the troposphere as passing over a station with a constant wind velocity $v = \rho/\tau$, where τ is the time lag. With $u = v\tau/$H, the structure function is then in terms of τ:

$$\mathbf{D}_{D^z}(\tau) = C_D^2 \frac{v^p}{\mathrm{H}^p} \tau^p. \tag{18.21}$$

For typical values of H $= 2$ km and $v = 8$ m/s, a power of $p = 1$ ($u \approx 1$) is reached when $\tau = 250$ s, or about 4 minutes. The corresponding value of the PSD $= C_D^2 v/$H is then about $1.5 \cdot 10^{-7}$m^2/s.

PSD

If we form a sequence of differences in time, $\underline{d}(k) \doteq \underline{D}^z(k) - \underline{D}^z(k-1)$, $\underline{d}(k+1) \doteq \underline{D}^z(k+1) - \underline{D}^z(k)$, etc., the cross-*correlation* function between any $\underline{d}(k)$ and $\underline{d}(l)$ reads:

$$R_{d(k)d(l)} = \frac{1}{2} \frac{v^p}{\mathrm{H}^p} C_D^2 [\tau_{k+1\,l}^p + \tau_{k\,l+1}^p - \tau_{k+1\,l+1}^p - \tau_{kl}^p], \tag{18.22}$$

where $\tau_{kl} = |t_l - t_k|$; cf. Eq. (17.37). This function is in general not equal to zero, which frustrates a recursive filtering approach of GPS data processing. Only in case of a random-walk process, there is no correlation between any $\underline{d}(k)$ and $\underline{d}(l \neq k)$; cf. Table 17.1. But, as shown in Fig. 18.6, the correlation is small for $p = 2/3$. Since $\mathbf{E}\{\underline{d}(k)\} = 0$ for all k, Eq. (18.22) also describes the cross-*covariance* function. In the random-walk case we therefore have the following
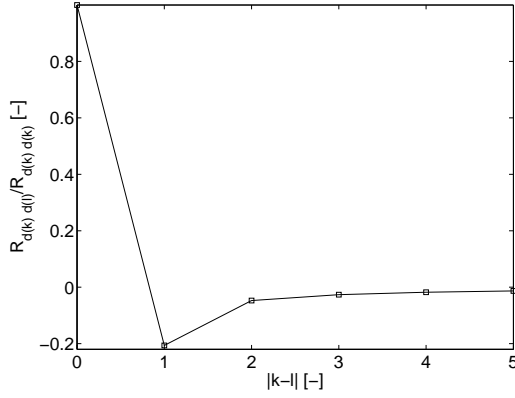
Figure 18.6   Normalized cross-correlation between troposphere differences $d(k)$ and $d(l)$ for power $p = 2/3$. Lines between values are shown for clarity.

stochastic model for constraints $\underline{d}(k)$:

$$\mathbf{D}\left\{\begin{bmatrix} \underline{d}(1) \\ \vdots \\ \underline{d}(p) \end{bmatrix}\right\} = \text{PSD} \cdot \text{diag}[\tau_{01}, ..., \tau_{p-1p}]. \tag{18.23}$$

If, like is assumed in the previous section, ZTDs are also spatially correlated and $\underline{d}(k)' = [\underline{D}_1^z(k), ..., \underline{D}_n^z(k)] - [\underline{D}_1^z(k-1), ..., \underline{D}_n^z(k-1)]$, the stochastic model reads

$$\mathbf{D}\left\{\begin{bmatrix} \underline{d}(1) \\ \vdots \\ \underline{d}(p) \end{bmatrix}\right\} \doteq \text{diag}[Q_d(1), ..., Q_d(p)] = \text{PSD} \cdot \text{diag}[\tau_{01}\overline{Q}_{D^z}, ..., \tau_{p-1p}\overline{Q}_{D^z}]. \tag{18.24}$$

with $\overline{Q}_{D^z}$ obtained from Eq. (18.20).

18.5   Residual Slant Tropospheric Delays

This section describes the development of a stochastic model of residual STDs. A residual STD, $\Delta D(\alpha, z)$, is the difference between an actual slant delay $D(\alpha, z)$ and a mapped ZTD, D^z:

$$\Delta D(\alpha, z) \doteq D(\alpha, z) - M(z)D^z, \tag{18.25}$$

where $D(\alpha, z)$ and D^z are in fact already residuals when we subtract an a-priori model like the one by Saastamoinen (1972). Because we assume $\mathbf{E}\{\underline{\Delta D}(\alpha, z)\} = 0$, the cross-*covariance* function of two residual STDs of the same station, which we want to obtain, equals the cross-*correlation* function and reads [*Emardson and Jarlemark*, 1999]:

$$\begin{aligned}
Q_{\Delta D}([\alpha_1, z_1], [\alpha_2, z_2]) &\doteq\; <\Delta D(\alpha_1, z_1) \cdot \Delta D(\alpha_2, z_2) >= \\
&< [D(\alpha_1, z_1) - M(z_1) \cdot D^z][D(\alpha_2, z_2) - M(z_2) \cdot D^z] >= \\
&M(z_1)M(z_2) < [M(z_1)^{-1}D(\alpha_1, z_1) - D^z][M(z_2)^{-1}D(\alpha_2, z_2) - D^z] >= \\
&\tfrac{1}{2}M(z_1)M(z_2) \left(< [M(z_1)^{-1}D(\alpha_1, z_1) - D^z]^2 > + \right. \\
&\qquad\qquad < [M(z_2)^{-1}D(\alpha_2, z_2) - D^z]^2 > - \\
&\qquad\qquad \left. < [M(z_1)^{-1}D(\alpha_1, z_1) - M(z_2)^{-1}D(\alpha_2, z_2)]^2 > \right),
\end{aligned} \tag{18.26}$$
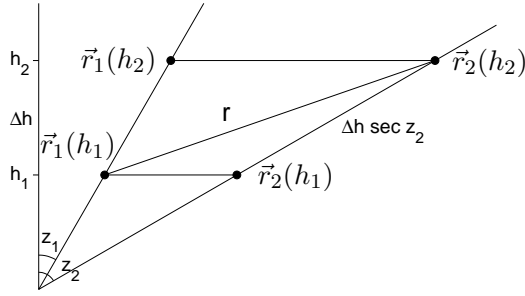
Figure 18.7   Geometry of two slant rays in the same vertical plane.

where we made use of Eq. (17.36); see also the third case of Table 17.1. This autocorrelation function is given as function of:

$$
< [M(z_1)^{-1}D(\alpha_1, z_1) - M(z_2)^{-1}D(\alpha_2, z_2)]^2 >=
$$
$$
10^{-12} \int_0^H \int_0^H \tfrac{1}{2}\mathbf{D}_N(|\vec{r}_2(h_1) - \vec{r}_1(h_2)|) + \tfrac{1}{2}\mathbf{D}_N(|\vec{r}_1(h_1) - \vec{r}_2(h_2)|)
$$
$$
- \tfrac{1}{2}\mathbf{D}_N(|\vec{r}_1(h_1) - \vec{r}_1(h_2)|) - \tfrac{1}{2}\mathbf{D}_N(|\vec{r}_2(h_1) - \vec{r}_2(h_2)|) \, dh_1 dh_2 =
$$
$$
10^{-12} \int_0^H \int_0^H \mathbf{D}_N([h_1^2 \sec^2(z_1) + h_2^2 \sec^2(z_2) - 2h_1 h_2 \sec(z_1)\sec(z_2)\cos(\theta)]^{\frac{1}{2}})
$$
$$
- \tfrac{1}{2}\mathbf{D}_N([\Delta h^2 \sec^2(z_1)]^{\frac{1}{2}}) - \tfrac{1}{2}\mathbf{D}_N([\Delta h^2 \sec^2(z_2)]^{\frac{1}{2}}) \, dh_1 dh_2 =
$$
$$
10^{-12} C_N^2 H^{\frac{8}{3}} \int_0^1 \int_0^1 [x_1^2 \sec^2(z_1) + x_2^2 \sec^2(z_2) - 2x_1 x_2 \sec(z_1)\sec(z_2)\cos(\theta)]^{\frac{1}{3}}
$$
$$
- \tfrac{1}{2}[\Delta x^2 \sec^2(z_1)]^{\frac{1}{3}} - \tfrac{1}{2}[\Delta x^2 \sec^2(z_2)]^{\frac{1}{3}} \, dx_1 dx_2,
$$

$$(18.27)$$

with θ the angle between both slant directions and $x_1 \doteq h_1/H$, $x_2 \doteq h_2/H$. The geometry of this derivation is shown in Fig. 18.7, where both rays are in the same vertical plane, $\theta = |z_1 - z_2|$, which is in general not the case. In both [*Treuhaft and Lanyi*, 1987] and [*Emardson and Jarlemark*, 1999] the constants $C_N = 0.24$ m$^{-\frac{1}{3}}$ and H = 1 km were used.
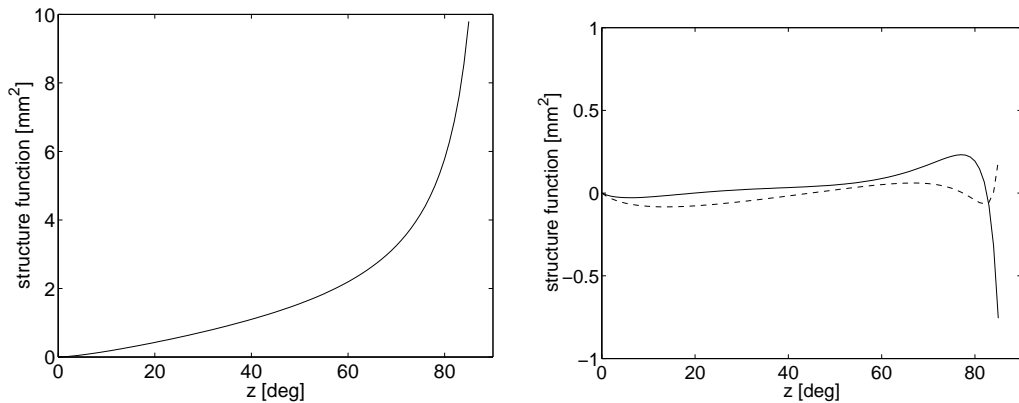
Figure 18.8   Left panel: $< [M(z)^{-1} \cdot D(\alpha, z) - D^z]^2 >$ as function of z for $C_N = 0.24$ m$^{-\frac{1}{3}}$, H = 1 km, and α fixed. Right panel: Error in approximation (real minus model) of same function by [*Emardson and Jarlemark*, 1999] (solid line) and as in Eq. (18.31) (dashed line).

In Eq. (18.27), we have obtained an integral that cannot be solved analytically. Numerical integration with $z_2 = 0$ and $\theta = z_1$ gives the result of Fig. 18.8. This is a special case where $M(z_2)^{-1}D(\alpha_2, z_2) = D^z$, and

was presented in [*Treuhaft and Lanyi*, 1987]. A Taylor expansion was given in [*Emardson and Jarlemark*, 1999]:

$$< \left[M(z_1)^{-1}D(\alpha_1, z_1) - M(z_2)^{-1}D(\alpha_2, z_2) \right]^2 > = 10^{-7} \sum_{i=1}^{12} a(i)k_i \ [\text{m}^2]. \quad (18.28)$$

The $a(i)$ and k_i are given in Table 18.2. An error plot of the above given model is given in Fig. 18.8. Emardson and Jarlemark (1999) gave an expansion in the mapping functions $M(z_1)$ and $M(z_2)$. The Lanyi (wet) mapping function [*Lanyi*, 1984] was used to derive the constants of Eq. (18.28) [*Jarlemark*, 2002], but the choice of mapping function is not critical. The model is based on 480 double integrals of combinations (α_1, z_1) and (α_2, z_2) [*ibid*]. In degrees they are:

$$
\begin{aligned}
\alpha_1 &= e_{480} \otimes 0; \\
\alpha_2 &= e_{48} \otimes [0\ 20\ 40\ 60\ 80\ 100\ 120\ 140\ 160\ 180]'; \\
z_1 &= [83\ 82\ 80\ 74\ 58\ 19\ ]' \ \otimes e_{80}; \\
z_2 &= e_6 \otimes [85\ 80\ 75\ 70\ 60\ 40\ 20\ 0]' \otimes e_{10},
\end{aligned}
\qquad (18.29)
$$

where $\otimes$ denotes the Kronecker product (see Part II) and e_n is an $n \times 1$ vector with ones.

i	$a(i)$	i	$a(i)$	i	$a(i)$
1	-0.2463797631	5	0.0057411608	9	-0.0000245648
2	0.2306443703	6	-0.0021812591	10	0.0000070851
3	-0.1178244362	7	0.0017793703	11	-0.0000069178
4	0.0179544380	8	-0.0003054072	12	0.0000012543

i	k_i	i	k_i	i	k_i
1	θ	5	θ^2	9	θ^3
2	$\theta[M(z_1) + M(z_2)]$	6	$\theta^2[M(z_1) + M(z_2)]$	10	$\theta^3[M(z_1) + M(z_2)]$
3	$\theta[M(z_1) \cdot M(z_2)]$	7	$\theta^2[M(z_1) \cdot M(z_2)]$	11	$\theta^3[M(z_1) \cdot M(z_2)]$
4	$\theta[M(z_1) + M(z_2)]^2$	8	$\theta^2[M(z_1) + M(z_2)]^2$	12	$\theta^3[M(z_1) + M(z_2)]^2$

Table 18.2   Constants $a(i)$ and function k_i as in the model of [Emardson and Jarlemark, 1999], Eq. (18.28); θ in degrees.

Unfortunately, tests with the model of [*Emardson and Jarlemark*, 1999] for actual satellite configurations all over the world resulted in covariance matrices that were indefinite. Probably this is caused by the crude polynomial approximation. This may not be seen from Fig. 18.8, but for arbitrary θ the errors become larger. One way to overcome this problem is neglecting the correlations. In fact, this is what is usually done in GPS data processing. The variances can then be described by some zenith angle dependent function. Using Eq. (18.26), we find for $z \equiv z_1 = z_2$ and $\alpha \equiv \alpha_1 = \alpha_2$:

$$\sigma^2_{\Delta D} = < [\Delta D(\alpha, z)]^2 > = M(z)^2 \cdot < [M(z)^{-1}D(\alpha, z) - D^z]^2 >, \qquad (18.30)$$

as shown in Fig. 18.9. We do not need a very sophisticated polynomial to approximate this structure function, as in this case: $\theta = z$. Figure 18.8 shows the error of the following model, with z in degrees:

$$< [\Delta D(\alpha, z)]^2 > = 0.0128 \cdot z \cdot (1 + \sec(z)) - 0.0003 \cdot z \cdot (1 + \sec(z))^2, \quad (18.31)$$

which seems to be a good (theoretical) approximation.

In GPS data processing, residual slant delays are usually not modeled explicitly. They are considered to be one of the noise terms of the observations. If the variance of a phase observable $\underline{\phi}$ is σ_ϕ^2, then $\underline{\phi} - \underline{\Delta D}(\alpha, z)$ has variance $\sigma_\phi^2 + \sigma_{\Delta D}^2$, which gives an apparent zenith angle dependent variance of the phase observation. Other zenith angle dependent noise terms, like multipath or antenna phase center variations, may be present.
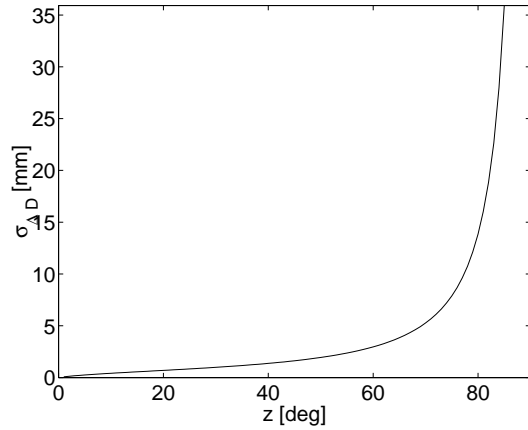
Figure 18.9   Standard deviation of residual slant delay as function of zenith angle; $M(z) = \sec(z)$ used.

A better way to prevent indefiniteness of the covariance matrix is computing the integrals explicitly. In actual data processing this would cost a lot of computation time, which makes it impractical. This was the reason for the polynomial approximation in [*Emardson and Jarlemark*, 1999] in the first place. The integral terms do however not need to be determined exactly; the integrals may be approximated by summations. Combining Eqs. (18.26) and (18.27) then gives the model for station r:

$$
\boxed{
\begin{aligned}
Q_{\Delta D_r}([\alpha_1, z_1], [\alpha_2, z_2]) &= \tfrac{1}{2} \cdot 10^{-12} C_N^2 \mathrm{H}^{\frac{8}{3}} M(z_1) M(z_2) \cdot \{ n^{-2} \textstyle\sum_{x_1} \sum_{x_2} \\
&\quad [x_1^2 \sec^2(z_1) + x_2^2 - 2x_1 x_2]^{\frac{1}{3}} + [x_1^2 + x_2^2 \sec^2(z_2) - 2x_1 x_2]^{\frac{1}{3}} \\
&\quad - [x_1^2 + x_2^2 - 2x_1 x_2]^{\frac{1}{3}} \\
&\quad - [x_1^2 \sec^2(z_1) + x_2^2 \sec^2(z_2) - 2x_1 x_2 \sec(z_1) \sec(z_2) \cos(\theta)]^{\frac{1}{3}} \},
\end{aligned}
}
\tag{18.32}
$$

where $x_1, x_2 = 1/2n, 3/2n, ..., (2n-1)/2n$. Tests with $n = 5$ and $n = 10$ at several locations around the world showed it to be positive definite every 30 s for a complete day. Random checking of Model (18.32) showed better approximations than the model from [*Emardson and Jarlemark*, 1999], even for $n = 5$. Model (18.32) is expected to be a reasonable model to be used in GPS data processing as long as a correct scaling ($C_D^2 = 10^{-12} C_N^2 \mathrm{H}^{\frac{8}{3}}$) can be assumed.

Chapter 19

Conclusions of Part III

The tropospheric (wet) refractive index can be considered a locally isotropic random field[1], with power-law distribution functions. The structure function, or ensemble autocorrelation function of differences, follows a two-thirds power-law behavior: $C_N^2 r^p$, with C_N^2 some constant, r the distance between two points in the troposphere, and $p = 2/3$ the power. This is a special case of the fractional Brownian motion where the power is between 0 and 2. Although, strictly speaking, these kinds of stochastic processes are not wide-sense stationary (or isotropic), because of the local validity of this structure function and assumptions on the power spectral density for low frequencies, we may treat them as such. Unlike the autocorrelation function, the structure function is not sensitive to the low-frequency behavior. Under the assumption of ergodicity (ensemble averages equal statistical averages), which is generally assumed reasonable, not only autocorrelations but also cross-correlations of differences can be obtained, based on the structure function.

Because the tropospheric delay is an integral function of the refractivity, the auto- or cross-correlation functions of tropospheric delay differences follow from double integration. And since the tropospheric delay differences are zero-mean, the obtained correlation matrices also serve as covariance matrices of constraints that can be added to a GPS observation model to strengthen it. The double integrals cannot be solved analytically but can be computed numerically. Three types of single-differenced tropospheric delays are considered: inter-station and inter-epoch single differences, and residual STDs.

Inter-station single-differenced tropospheric delays do not have a strict constant power-law behavior, but a power that declines from 5/3th to 2/3th over increasing distances with a turn-over point at about the same distance as the (effective wet) tropospheric height of a few kilometers. Within certain distance regimes smaller or larger than the turn-over point, we can consider the structure function to have a (nearly) strict power-law behavior.

To obtain a structure function for the inter-epoch single-differenced tropospheric delays (relative ZTD observable) we need to assume a certain constant wind velocity. The structure function is a function of time with a turn-over point that depends on the wind velocity. Only in case of a random-walk process, a power-law process with power 1, efficient processing of GPS data is possible in the presence of relative constraints on zenith delays.

---

[1]Stochastic process of three variables with an autocorrelation function that depends only on distance.

Based on the stochastic properties of the wet refractive index, one can also derive a stochastic double-integral model for residual STDs. A numeric, polynomial, model was derived in [*Emardson and Jarlemark*, 1999], but this model turned out to be (probably intrinsically) indefinite for all tested satellite configurations. Approximating the double integrals by (crude) summations gave a satisfying positive-definite covariance matrix for the same tested configurations. If the correlations are discarded, one is left with a positive-definite covariance matrix that implies zenith angle dependent weighting.

The covariance matrices of the constraints depend on the scaling parameter C_N^2 and effective height H. Although we assumed these parameters to be constant, under different atmospheric conditions they may vary and are therefore still weak points in the described models when estimated parameters are sensitive to it.

An overview of the most important models as derived in this part is given in Table 19.1.

relative ZTD constraints	Eqs. (18.24), (18.20), (18.12) and (18.17); Table 18.1
residual STDs (EJ)	Eqs. (18.26) and (18.28); Table 18.2
residual STDs (summation model)	Eqs. (18.32) and (18.4)

Table 19.1   Equations/tables with described stochastic models. EJ: model of [*Emardson and Jarlemark*, 1999].

state of the art and contribution

The power-law behavior of the tropospheric delay is known and the use of temporal constraints in GPS data processing is common practice in scientific GPS software. The spatiotemporal constraints based on the new TL2 model are by the author. Most of the derivations are based on what is present in literature, but the framework of derivations is by the author as well as App. G. The derivation of a covariance matrix for residual STDs is after Emardson and Jarlemark (1999), but the positive-definite Model (18.32) is by the author.

Now that we have obtained stochastic models for the residual STDs, and the spatiotemporal constraints of the troposphere-weighted model, all model components are available to be built in a filter. The next part describes several of these filters for the models at hand.

Fourier transform of a power-law function

This appendix gives a derivation of $\mathcal{F}\{|r|^p\}$. For a more complete description of these kinds of Fourier transforms and their existence, see [*Lighthill*, 1959].

The Fourier transform of $r^p \exp(-\epsilon r)H(r)$, with

$$H(r) \doteq \begin{cases} 0, \ r < 0; \\ 1, \ r \geq 0, \end{cases} \tag{F.1}$$

the Heaviside function and ϵ some constant, reads

$$\begin{aligned} \mathcal{F}\{r^p \exp(-\epsilon r)H(r)\} &= \int_0^\infty r^p \exp(-r(\epsilon + j\kappa))dr \\ &= (\epsilon + j\kappa)^{-p-1}\Gamma(p+1), \end{aligned} \tag{F.2}$$

since for some constant $c = \epsilon + j\kappa$:

$$\int_0^\infty r^p \exp(-cr)dr = c^{-p-1} \int_0^\infty r^p \exp(-r)dr = c^{-p-1}\Gamma(p+1). \tag{F.3}$$

If we let $\epsilon \to 0$, we have:

$$\begin{aligned} \mathcal{F}\{r^p H(r)\} &= (j\kappa)^{-p-1}\Gamma(p+1); \\ \mathcal{F}\{(-r)^p H(-r)\} &= (-j\kappa)^{-p-1}\Gamma(p+1). \end{aligned} \tag{F.4}$$

And finally:

$$\begin{aligned} \mathcal{F}\{|r|^p\} &= \mathcal{F}\{r^p H(r) + (-r)^p H(-r)\} \\ &= [(j\kappa)^{-p-1} + (-j\kappa)^{-p-1}] \cdot \Gamma(p+1) \\ &= [j^{p+1} + j^{-p-1}] \cdot |\kappa|^{-p-1} \cdot \Gamma(p+1) \\ &= [\exp(\tfrac{1}{2}\pi j(p+1)) + \exp(-\tfrac{1}{2}\pi j(p+1))] \cdot |\kappa|^{-(p+1)} \cdot \Gamma(p+1) \\ &= 2\cos(\tfrac{1}{2}\pi(p+1)) \cdot |\kappa|^{-(p+1)} \cdot \Gamma(p+1) \\ &= -2\sin(\tfrac{p\pi}{2})\Gamma(p+1) \cdot |\kappa|^{-(p+1)}. \end{aligned} \tag{F.5}$$

Autocorrelation function in case of a power-law spectrum

Assume we have the following power spectrum of a real and wide-sense stationary stochastic process $\underline{X}(r)$ [m]:

$$S_X(\kappa) = \begin{cases} S_0 \cdot |\kappa|^{-(p+1)} \text{ for } |\kappa| \geq \kappa_0; \\ S_0 \cdot \kappa_0^{-(p+1)} \quad \text{ for } |\kappa| \leq \kappa_0, \end{cases} \tag{G.1}$$

for which $0 < p < 2$, S_0 a positive constant [m^{2-p}], and $\kappa_0 r \ll 1$ for $0 < r < r_{\max}$. The autocorrelation function of $\underline{X}(r)$ then reads:

$$\begin{aligned} R_X(r) &\doteq \frac{1}{\pi} \int_0^\infty \cos(\kappa r) S_X(\kappa) d\kappa \\ &= \frac{1}{\pi} \int_0^{\kappa_0} \cos(\kappa r) S_0 \kappa_0^{-(p+1)} d\kappa + \frac{1}{\pi} \int_{\kappa_0}^\infty \cos(\kappa r) S_0 \kappa^{-(p+1)} d\kappa \\ &\doteq I_1 + I_2. \end{aligned} \tag{G.2}$$

The first integral is easily obtained as:

$$I_1 = \frac{S_0 \kappa_0^{-p}}{\pi} \cdot \frac{\sin \kappa_0 r}{\kappa_0 r}. \tag{G.3}$$

For the second integral we use the change of variables $u \doteq \kappa r$ and $\cos u = 1 - 2\sin^2(u/2)$ to obtain:

$$\begin{aligned} I_2 &= \frac{S_0}{\pi} r^p \int_{\kappa_0 r}^\infty u^{-(p+1)}(1 - 2\sin^2(\frac{u}{2})) du \\ &= \frac{S_0}{\pi} r^p \left[\int_{\kappa_0 r}^\infty u^{-(p+1)} du - 2 \int_0^\infty u^{-(p+1)} \sin^2(\frac{u}{2}) du + 2 \int_0^{\kappa_0 r} u^{-(p+1)} \sin^2(\frac{u}{2}) du \right] \\ &\doteq \frac{S_0}{\pi} r^p \left[I_{21} + I_{22} + I_{23} \right]. \end{aligned} \tag{G.4}$$

Because $p > 0$ we have:

$$I_{21} = \left[\frac{u^{-p}}{-p} \right]_{\kappa_0 r}^\infty = \frac{(\kappa_0 r)^{-p}}{p}. \tag{G.5}$$

The second integral reads [*Gradsteyn et al.*, 1994], formula 3.823:

$$\begin{aligned} I_{22} &= -2 \cdot -\frac{\Gamma(-p) \cos(-\frac{p\pi}{2})}{2^{1-p} \cdot (\frac{1}{2})^{-p}} \\ &= \Gamma(-p) \cos(\frac{p\pi}{2}) \\ &= \frac{-\pi}{2\Gamma(p+1) \sin(\frac{p\pi}{2})}. \end{aligned} \tag{G.6}$$

The last integral is found by the change of variables $u \leftarrow u/2$:

$$
\begin{aligned}
I_{23} &= 2 \int_0^{\frac{1}{2}\kappa_0 r} 2^{1-p} u^{1-p} \sin^2(u) 2 du \\
&= 2^{1-p} \int_0^{\frac{1}{2}\kappa_0 r} u^{1-p} u^{-2} \sin^2(u) du.
\end{aligned}
\tag{G.7}
$$

Since we integrate from 0 to $\frac{1}{2}\kappa_0 r \ll 1$, for all u we have $u^{-2}\sin^2(u) \approx 1$. The integral may therefore be approximated by:

$$
\begin{aligned}
I_{23} &\approx 2^{1-p} \int_0^{\frac{1}{2}\kappa_0 r} u^{1-p} du \\
&= 2^{1-p} \left[\frac{u^{2-p}}{2-p} \right]_0^{\frac{1}{2}\kappa_0 r} \\
&= \frac{(\kappa_0 r)^{2-p}}{2(2-p)},
\end{aligned}
\tag{G.8}
$$

since $2 - p > 0$. Combining the results of Eqs. (G.2–G.6) and (G.8) and approximating by using $\kappa_0 r \ll 1$ gives:

$$
\begin{aligned}
R_X(r) &= \frac{S_0 \kappa_0^{-p}}{\pi} \cdot \frac{\sin(\kappa_0 r)}{\kappa_0 r} + \frac{S_0}{\pi} r^p \left(\frac{(\kappa_0 r)^{-p}}{p} - \frac{\pi}{2\Gamma(p+1)\sin(\frac{p\pi}{2})} + \frac{(\kappa_0 r)^{2-p}}{2(2-p)} \right) \\
&\approx \frac{S_0 \kappa_0^{-p}}{\pi} + \frac{S_0 \kappa_0^{-p}}{p\pi} - \frac{S_0}{2\Gamma(p+1)\sin(\frac{p\pi}{2})} r^p \\
&= \frac{(p+1)S_0}{p\pi \kappa_0^p} - \frac{S_0}{2\Gamma(p+1)\sin(\frac{p\pi}{2})} r^p.
\end{aligned}
\tag{G.9}
$$

Bibliography

Abramowitz, M., and I.A. Stegun (1970), *Handbook of mathematical functions*, 9th edn., Dover Publications, New York.

Agnew, D.C. (1992), The time-domain behavior of power-law noises, *Geophysical Research Letters*, *19*, 333–336.

Baarda, W. (1973), S-transformations and criterion matrices, *Publications on Geodesy, Vol. 5, No. 1*, Netherlands Geodetic Commission.

Emardson, T.R., and P.O.J. Jarlemark (1999), Atmospheric modelling in GPS analysis and its effect on the estimated geodetic parameters, *Journal of Geodesy*, *73*, 322–331.

Gradinarsky, L.P. (2002), *Sensing Atmospheric Water Vapor Using Radio Waves*, Ph.D. dissertation, Chalmers University of Technology.

Gradsteyn, I.S., I.M. Ryzhik, and A. Jeffrey (1994), *Table of Integrals, Series, and Products*, 5th edn., Boston.

Hanssen, R.F. (2001), *Radar Interferometry: Data interpretation and error analysis*, Kluwer, Dordrecht.

Jarlemark, P.O.J. (2002), Personal communications.

Kolmogorov, A.N. (1941), The local structure of turbulence in incompressible viscous fluid for very large Reynolds' numbers, *Doklady Akad. Nauk SSSR*, *30*.

Lanyi, G. (1984), Tropospheric delay effects in radio interferometry, *TDA prog. rep., 42-78, pp. 152-159*, JPL Pasadena.

Lighthill, M.J. (1959), *Introduction to Fourier analysis and generalised functions*, Cambridge University Press, Cambridge, 2nd edn.

Moler, C., and P.J. Costa (1997), *Symbolic Math Toolbox, For Use with Matlab*, The Math Works, Inc.

Obukhov, A.M. (1941), On the distribution of energy in the spectrum of turbulent flow, *Doklady Akad. Nauk SSSR*, *32*.

Peebles, P.Z. (1987), *Probability, random variables, and random signal principles*, McGraw-Hill, New York, 2nd edn.

Saastamoinen, J. (1972), Atmospheric correction for the troposphere and stratosphere in radio ranging of satellites, *The use of artificial satellites for geodesy, Geophys. Monogr. Ser.*, *15*, 247–251.

Tatarski, V.I. (1961), *Wave propagation in a turbulent medium*, McGraw-Hill, New York.

Taylor, G.I. (1938), The spectrum of turbulence, in *Proceedings of the Royal Society London, Series A, CLXIV*, pp. 476–490.

Treuhaft, R.N., and G.E. Lanyi (1987), The effect of the dynamic wet troposphere on radio interferometric measurements, *Radio Science*, *22*, 251–265.

Yaglom, A.M. (1962), *An introduction to the theory of stationary random functions*, Prentice-Hall, London.

RECURSIVE GPS DATA PROCESSING

Symbols in Part IV

symbol	meaning
A_k	design matrix of measurement update
A_k	$m_k \times n$ partial design matrix of global parameters
$\bar{A}_k$	Cholesky-factor transformed A_k
$\forall$	for all
α_q	level of significance (for q-dimensional test)
B_k	$m_k \times n_\beta$ partial design matrix of batch parameters
BA_k	combined partial design matrix $[B_k, A_k]$
β_k	constrained (batch) parameters
$\hat{\beta}_{k|k-1}$	predicted value of β_k
$\hat{\beta}_{k|k}$	filtered value of β_k
βx_k	concatenated vector $[\beta_k', \ x']'$
C_v	model-error specifying matrix
γ	power
$\underline{d}_k$	pseudo-observable of $\underline{\beta}_k - \underline{\beta}_{k-1}$
δk	latency $\delta k = k - k_0$ of model error
$\partial_x A_k(x^0)$	Jacobian (gradient matrix) of global parameters evaluated in x^0
Δx	increments/observed-minus-computed $(x - x^0)$
ϵ	convergence criterion
ε	residuals
$\Phi_{k,k-1}$	state transition matrix, epoch $k-1$ to k
$\Psi_{k,k-1}$	state transition matrix $\Psi_{k,k-1} = \mathrm{diag}[\Phi_{k,k-1}, I_n]$
H_0	null hypothesis
H_A	alternative hypothesis
I_m	$m \times m$ identity matrix
k	current epoch number $(k = 1, ..., p)$
k_0	first epoch a model error occured
κ	running epoch counter $(\kappa = k_0, ..., k)$
K_k	Kalman gain matrix
λ	noncentrality parameter
m_k	number of observations at epoch k
$\overline{m}_k$	$m_k - n_{t_k}$, number of transformed observations at epoch k

symbol	meaning
n	number of global parameters
n_β	number of constrained parameters
n_{t_k}	number of temporal parameters (of epoch k)
N	phase ambiguities
p	number of epochs
P_{T_k}	projector on T_k
$P_{T_k}^\perp$	projector on orthogonal subspace of T_k
$\mathcal{P}$	permutation matrix
q	degrees of freedom of test statistic
Q_y	covariance matrix of y
$Q, Q_I, Q_1, ...$	orthonormal matrices (Q factors)
r	station coordinates
$R_{\hat{x}}$	upper-triangular matrix
$R_{\hat{\beta}\hat{x}}$	rectangular matrix as part of upper-triangular matrix
$\mathbb{R}^m$	m-dimensional space of reals
t_k	temporal parameters
T_k	$m_k \times n_{t_k}$ partial design matrix of temporal parameters
$T_k^\perp$	null space of T_k'
$\underline{\Theta}(q)$	q-dimensional test statistic
$\underline{\Theta}_k^0(q)$	q-dimensional local test statistic of epoch k
$\underline{\Theta}_k^{\delta k}(q)$	q-dimensional global test statistic of epochs $k - \delta k$ to k
$\mathcal{T}$	ambiguity transformation matrix (to change pivot satellite)
v_k	vector of predicted residuals
x	global parameters
x^0	approximate value of x
$\hat{x}$	estimates of x
x_k	vector with all parameters at epoch k
$X_{k|k-1}^{\delta k}, X_{k|k}^{\delta k}$	response matrices
y_k	observations at epoch k
y_k	stacked vector; in SRIF: $\mathrm{y}_k = [y_k', z'_{\hat{\beta}x_{k|k-1}}]'$
$\overline{y}_k$	Cholesky-factor transformed y_k
$\underline{y}_k$	observables at epoch k
$z_{\hat{x}}$	transformed state vector
∇	model error

Chapter 20

Introduction to Part IV

models of
interest

As mentioned in the 'General introduction', high quality GPS-derived heights
are our main interest. We aim at a 5-mm accuracy. To obtain this accuracy,
we need static networks and use long observation time spans. The tropospheric
delay is our main concern because of its strong correlation with the height.
In our observation models of interest, heights, horizontal coordinates, and
ambiguities are unchanging/global parameters, and the slowly changing Zenith
Tropospheric Delays (ZTDs) form a second group of parameters, which are
(soft-)constrained in the troposphere-weighted models; see Part II. An extreme
case of the troposphere-weighted model is the troposphere-float model, which does
not include constraints. In the troposphere-fixed model the ZTDs are assumed
a-priori known.

recursive
estimation

Although strictly speaking GPS (pseudo-)observables are time correlated,
they are usually assumed uncorrelated because implementation of correlation
would make a batch processing very time consuming.[1] For recursive estimation,
uncorrelated observables are a condition. Normally, the very argument for using
a recursive processing method is the need for parameter estimates in real-time.
Although for static network processing a batch procedure may be more efficient,
recursive estimation has the advantage that the impact of the observation time
span is easier to analyze in the stage of model design. Intermediate estimates of
global parameters can be obtained as well as the final estimates. Especially for
this reason, estimation (and testing) procedures are developed. An advantage of
recursive processing is that the software is more flexible in the sense that new
parameters can be introduced when needed and old parameters can be discarded
when not needed anymore. Another advantage is that the normal matrices are
smaller than in batch processing. But when temporal parameters, forming the
third group of parameters, are pre-eliminated, this size difference only concerns
batch parameters and ambiguities. (Although there may be many of these in large
networks, especially for long observation times and frequent ZTD estimation.)
For those applications where one is only interested in real-time estimates of ZTDs,
a recursive procedure is even indispensable.

pre-
elimination

Because of the presence of the quickly changing temporal parameters in
untransformed GPS observation models, a straightforward implementation of
standard recursive methods is not possible. We consider two recursive methods:
the Kalman Filter (KF) in Chap. 21, and the recursive Square Root Information
Filter (SRIF) in Chap. 22. Since temporal parameters are (usually) not the
parameters of interest (nuisance parameters) they are pre-eliminated in the

---

[1]Exponential correlation is an exception.

extended procedures of Chaps. 21 and 22.

outline   The Kalman Filter method of Chap. 21 uses an orthogonal projector on
              the column space of the temporal parameters, whereas the recursive SRIF of
              Chap. 22 only solves the lower-right part of an upper-triangular system, which
              concerns the nontemporal parameters. Chapter 23 deals with some technical
              implementation problems like nonlinearity, computation speed, and changing
              satellite configurations. Chapter 24 gives a pre-elimination transformation using
              the actual null space of the column space of the temporal parameters.   This
              null space is in fact the same space as spanned by the orthogonal projector of
              Chap. 21. This method is rather commonly used, but rarely put in a more general
              perspective. In an observation model where receiver and satellite clock errors and
              ionospheric delays are temporal parameters, this boils down to double differencing
              and obtaining linear inter-frequency combinations. Chapter 24 also describes
              testing and reliability aspects in Kalman filtering. Both the estimation and
              testing benefit from Cholesky-factor transformations to avoid matrix inversions.
              Finally, Chap. 25 ends with the conclusions.

assumption   All displayed procedures are in terms of partial design matrices of the three types
              of parameters and covariance matrices of the observables and constraints. These
              matrices are assumed known when implementing a procedure in software. In
              Part II the partial design matrices are given for several models, and in Part III
              the covariance matrices of troposphere constraints are given.

Kalman filtering with pre-elimination

21.1   Introduction

In this chapter we describe recursive parameter estimation in linear models using a Kalman Filter procedure with observation equations. We start with a review on least-squares (LSQ) estimation in Sect. 21.2. In the sections that follow we consider three different partitioned models for which we formulate a recursive estimation method. In the first model (Sect. 21.3) there are only global parameters. These are the parameters we are after, so they are all to be determined. The second model (Sect. 21.4) also incorporates temporal parameters. These parameters are considered nuisance parameters and are not explicitly solved for by the described procedure. Our troposphere-fixed and float models are of this type. The third and last model (Sect. 21.5) incorporates constrained parameters as well as global and temporal parameters. Our troposphere-weighted model is of this type. The constrained parameters are also considered nuisance parameters but they are treated differently. An estimation procedure for the latter two models, in which temporal parameters are pre-eliminated, is what we aim at. The first model serves as a steppingstone. By partitioning our observation model, the vector of LSQ estimates of the global parameters is derived. The concepts in this chapter date back to [*Kalman*, 1960] and we will therefore speak of a Kalman Filter, although this is not strictly what was described by Kalman. This term is however used to distinguish it from the SRIF method as described in the next chapter.

Kalman

21.2   Least-squares estimation

BLUE

The Best Linear Unbiased Estimator (BLUE) of parameters x in the full-rank Gauss–Markov model

$$\mathbf{E}\{\underline{y}\} = Ax \;\; ; \;\; \mathbf{D}\{\underline{y}\} = Q_y, \tag{21.1}$$

with A an $m \times n$ design matrix $(m \geq n)$ is known to be [*Koch*, 1987]:

$$\hat{\underline{x}} = Q_{\hat{x}} A' Q_y^{-1} \underline{y} \;\; ; \;\; Q_{\hat{x}} = (A' Q_y^{-1} A)^{-1}. \tag{21.2}$$

This estimator is also the least-squares estimator if we use Q_y^{-1} as weight matrix in the LSQ method. Because we will further assume this is the case, we speak of LSQ estimation.

LSQ

If $\varepsilon \doteq y - Ax$ is the vector of residuals, then $x = \hat{x}$ is the vector that minimizes the sum of squared residuals $\| \varepsilon \|^2 \doteq \varepsilon' Q_y^{-1} \varepsilon$. Let

P_A

$$P_A \doteq A(A'Q_y^{-1}A)^{-1}A'Q_y^{-1} \tag{21.3}$$

$P_A^\perp$ be a projector that projects onto the range space of A, $P_A A = A$, and let $P_A^\perp \doteq I - P_A$ be its orthogonal complement for which $P_A^\perp A = 0$. The sum of squared residuals is then:

$$\parallel \varepsilon \parallel^2 = \parallel P_A \varepsilon \parallel^2 + \parallel P_A^\perp \varepsilon \parallel^2 = \parallel P_A y - Ax \parallel^2 + \parallel P_A^\perp y \parallel^2 . \tag{21.4}$$

Since $\parallel P_A y - A\hat{x} \parallel^2= 0$, $\hat{x}$ minimizes $\parallel \varepsilon \parallel^2$, and the sum of least-squares residuals reads $\parallel \hat{\varepsilon} \parallel^2=\parallel P_A^\perp y \parallel^2$. The product $P_A y$ has a meaning in itself. This projection of the observations on the column space of A is the LSQ estimate of the observations:

$$\begin{aligned} \hat{y} &= A\hat{x} &= P_A y; \\ Q_{\hat{y}} &= P_A Q_y P_A' &= P_A Q_y. \end{aligned} \tag{21.5}$$

21.3   The model with global parameters only

Let the full-rank Gauss–Markov model of Eq. (21.1) be partitioned as $A = [A_1', ..., A_p']'$, $y = [y_1', ..., y_p']'$, and $Q_y = \text{diag}[Q_{y_1}, ..., Q_{y_p}]$. In this model, y_k is an $m_k \times 1$ vector of observations at epoch $k = 1, ..., p$, with corresponding covariance matrix Q_{y_k}; A_k is an $m_k \times n$ partial design matrix, and x is the $n \times 1$ vector of global parameters.

We consider the following partitioned system to find a LSQ estimate of x recursively [*Teunissen*, 2001]:

$$\boxed{\mathbf{E}\{\begin{bmatrix} \underline{y}_k \\ \underline{\hat{x}}_{k-1} \end{bmatrix}\} = \begin{bmatrix} A_k \\ I_n \end{bmatrix} x \;\; ; \;\; \mathbf{D}\{\begin{bmatrix} \underline{y}_k \\ \underline{\hat{x}}_{k-1} \end{bmatrix}\} = \begin{bmatrix} Q_{y_k} & \\ & Q_{\hat{x}_{k-1}} \end{bmatrix},} \tag{21.6}$$

where $\underline{\hat{x}}_{k-1}$ is the LSQ estimator of x, based on the first $k-1$ epochs of data, that is combined with the new observables of epoch k. Straightforward application of Eq. (21.2) then gives:

$$\hat{x}_k = Q_{\hat{x}_k} Q_{\hat{x}_{k-1}}^{-1} \hat{x}_{k-1} + Q_{\hat{x}_k} A_k' Q_{y_k}^{-1} y_k, \tag{21.7}$$

where the covariance matrix of $\hat{x}_k$ reads:

$$\boxed{Q_{\hat{x}_k} = \left(Q_{\hat{x}_{k-1}}^{-1} + A_k' Q_{y_k}^{-1} A_k \right)^{-1}.} \tag{21.8}$$

Equation (21.7) may also be written as:

$$\boxed{\begin{aligned} \hat{x}_k &= Q_{\hat{x}_k} \left(Q_{\hat{x}_{k-1}}^{-1} + A_k' Q_{y_k}^{-1} A_k \right) \hat{x}_{k-1} \; + \; Q_{\hat{x}_k} A_k' Q_{y_k}^{-1}(y_k - A_k \hat{x}_{k-1}) \\ &= \hat{x}_{k-1} + Q_{\hat{x}_k} A_k' Q_{y_k}^{-1}(y_k - A_k \hat{x}_{k-1}) \\ &= \hat{x}_{k-1} + K_k v_k, \end{aligned}} \tag{21.9}$$

K_k, v_k where K_k is the Kalman gain matrix and v_k is the vector of predicted residuals, also called innovations:

$$\boxed{\begin{aligned} K_k &= Q_{\hat{x}_k} A_k' Q_{y_k}^{-1}; \\ v_k &= y_k - A_k \hat{x}_{k-1}. \end{aligned}} \tag{21.10}$$

The left panel in Fig. 21.1 shows the complete estimation procedure.
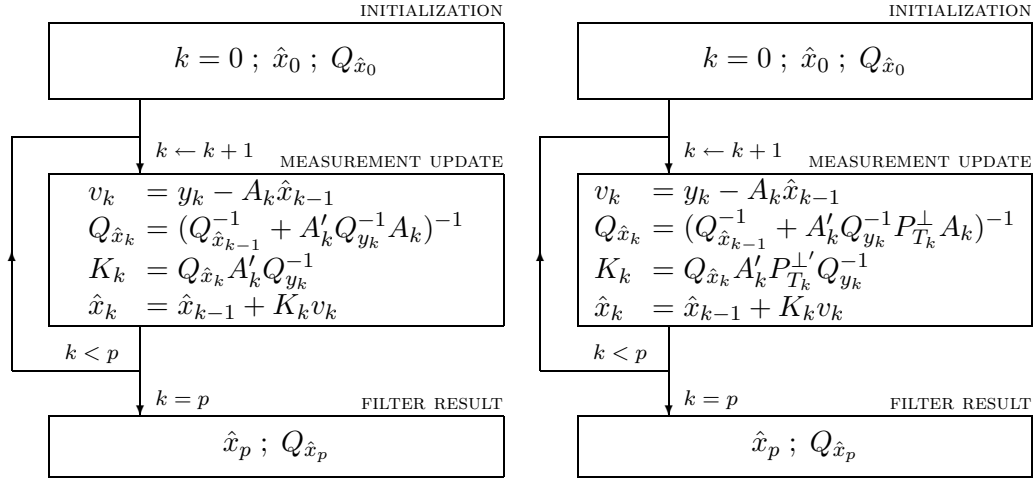
Figure 21.1   Recursive Kalman Filter procedures corresponding with the linear model. Left panel: With global parameters only. Right panel: With global and temporal parameters.

Q_{v_k} The covariance matrix of the predicted residuals follows from the propagation law of (co-)variances:

$$Q_{v_k} = Q_{y_k} + A_k Q_{\hat{x}_{k-1}} A'_k. \tag{21.11}$$

Application of the matrix lemma Eq. (H.2) of App. H, using the expressions for $Q_{\hat{x}_k}$ and K_k of Eqs. (21.8) and (21.10), gives a useful analytic expression for the inverse of this matrix:

$$Q_{v_k}^{-1} = Q_{y_k}^{-1}(I_{m_k} - A_k K_k), \tag{21.12}$$

which has a special significance in the testing for model errors; see Chap. 24.

21.4   Including temporal parameters

In this section the formulas for recursive parameter estimation are derived for the model with both global and temporal parameters. Both the troposphere-fixed and the troposphere-float model are of this form. The model reads:

$$\mathbf{E}\{\underline{y}\} = [T, \ A] \begin{bmatrix} t \\ x \end{bmatrix} \quad ; \quad \mathbf{D}\{\underline{y}\} = \mathrm{diag}[Q_{y_1}, ..., Q_{y_p}], \tag{21.13}$$

with A as in the previous section, $T \doteq \mathrm{diag}([T_1, ..., T_p])$, and $t \doteq [t'_1, ..., t'_p]'$. T_k is an $m_k \times n_{t_k}$ partial design matrix, and t_k is an $n_{t_k} \times 1$ vector of temporal parameters. After solving the parameters of the first $k-1$ epochs in the above given model, the solutions for the global parameters can be updated with the following observation model:

$$\mathbf{E}\{\begin{bmatrix} \underline{y}_k \\ \underline{\hat{x}}_{k-1} \end{bmatrix}\} = \begin{bmatrix} T_k & A_k \\ 0 & I_n \end{bmatrix} \begin{bmatrix} t_k \\ x \end{bmatrix} \quad ; \quad \mathbf{D}\{\begin{bmatrix} \underline{y}_k \\ \underline{\hat{x}}_{k-1} \end{bmatrix}\} = \begin{bmatrix} Q_{y_k} & \\ & Q_{\hat{x}_{k-1}} \end{bmatrix}. \tag{21.14}$$

Model (21.6) is thus extended with temporal nuisance parameters we do not wish to solve for. Two partial design matrices can be distinguished in the partitioned Model (21.14):

$$A_I = \begin{bmatrix} T_k \\ 0 \end{bmatrix} \quad ; \quad A_{II} = \begin{bmatrix} A_k \\ I_n \end{bmatrix}. \tag{21.15}$$

Solving for the global parameters only, gives the following LSQ estimate [*Teunissen*, 1999]:

$$\hat{x}_k = (\bar{A}'_k Q^{-1}_{\bar{y}_k} \bar{A}_k)^{-1} \bar{A}'_k Q^{-1}_{\bar{y}_k} \bar{y}_k, \tag{21.16}$$

where $\bar{y}_k \doteq [y_k', \hat{x}'_{k-1}]'$ is the stacked vector of all observations with covariance matrix $Q_{\bar{y}_k} \doteq \text{diag}[Q_{y_k}, Q_{\hat{x}_{k-1}}]$ and the design matrix $\bar{A}_k$ is the orthogonal projection of the second partial design matrix onto the first:

$$\bar{A}_k \doteq P^{\perp}_{A_I} A_{II} \quad \text{with} \quad P^{\perp}_{A_I} \doteq I_{m_k+n} - A_I (A'_I Q^{-1}_{\bar{y}_k} A_I)^{-1} A'_I Q^{-1}_{\bar{y}_k}. \tag{21.17}$$

Straightforward computation now gives the design matrix $\bar{A}_k$ and projector $P^{\perp}_{A_I}$:

$$\bar{A}_k = \begin{bmatrix} P^{\perp}_{T_k} A_k \\ I_n \end{bmatrix}; \ P^{\perp}_{A_I} = \begin{bmatrix} P^{\perp}_{T_k} \\ & I_n \end{bmatrix}, \tag{21.18}$$

$P_{T_k}, \ P^{\perp}_{T_k}$ where we use the orthogonal projector on T_k:

$$P^{\perp}_{T_k} \doteq I_{m_k} - P_{T_k} \quad \text{with} \quad P_{T_k} \doteq T_k (T'_k Q^{-1}_{y_k} T_k)^{-1} T'_k Q^{-1}_{y_k}. \tag{21.19}$$

With Eq. (21.18), the LSQ estimate of the vector of global parameters from Eq. (21.16) is worked out as:

$$\hat{x}_k = Q_{\hat{x}_k} \left(Q^{-1}_{\hat{x}_{k-1}} \hat{x}_{k-1} + A'_k P^{\perp'}_{T_k} Q^{-1}_{y_k} y_k \right). \tag{21.20}$$

And its covariance matrix reads:

$$\boxed{Q_{\hat{x}_k} = \left(Q^{-1}_{\hat{x}_{k-1}} + A'_k Q^{-1}_{y_k} P^{\perp}_{T_k} A_k \right)^{-1}.} \tag{21.21}$$

Compare these results with Eqs. (21.7–21.8). The only difference is the presence of the orthogonal projector on T_k. In a similar way as in the previous section, Eq. (21.20) may also be written as:

$$\boxed{\hat{x}_k = \hat{x}_{k-1} + K_k v_k,} \tag{21.22}$$

where K_k and v_k are again the Kalman gain matrix and the vector of predicted residuals respectively:

$$\boxed{\begin{aligned} K_k &= Q_{\hat{x}_k} A'_k P^{\perp'}_{T_k} Q^{-1}_{y_k}; \\ v_k &= y_k - A_k \hat{x}_{k-1}. \end{aligned}} \tag{21.23}$$

The right-hand panel in Fig. 21.1 shows the complete procedure for this model. Contrary to the model with global parameters only, Eq. (H.2) cannot be applied to obtain a simple expression for $Q^{-1}_{v_k}$ as in Eq. (21.12).

21.5   The model with constrained parameters

When time constraints are included in the model, we have the following Gauss–Markov model:

$$\mathbf{E}\left\{ \begin{bmatrix} \underline{d} \\ \underline{y} \end{bmatrix} \right\} = \begin{bmatrix} 0 & E & 0 \\ T & B & A \end{bmatrix} \begin{bmatrix} t \\ \beta \\ x \end{bmatrix} \quad ; \quad \mathbf{D}\left\{ \begin{bmatrix} \underline{d} \\ \underline{y} \end{bmatrix} \right\} = \begin{bmatrix} Q_d & \\ & Q_y \end{bmatrix}, \tag{21.24}$$

with $B \doteq \mathrm{diag}([B_1,...,B_p])$, $d \doteq [d_1',...,d_p']'$, and $Q_d \doteq \mathrm{diag}[Q_{d_1},...,Q_{d_p}]$. The first row of the functional model looks like (cf. Part II, where $\Phi_{k,k-1} = I_{n_\beta}$):

$$\mathbf{E}\{\underbrace{\begin{bmatrix} \underline{d}_1 \\ \vdots \\ \underline{d}_p \end{bmatrix}}_{\underline{d}}\} = \underbrace{\begin{bmatrix} -\Phi_{1,0} & I_{n_\beta} & & \\ & -\Phi_{2,1} & I_{n_\beta} & \\ & & \ddots & \ddots \\ & & & -\Phi_{p,p-1} & I_{n_\beta} \end{bmatrix}}_{E} \underbrace{\begin{bmatrix} \beta_0 \\ \vdots \\ \beta_p \end{bmatrix}}_{\beta}, \tag{21.25}$$

transition matrix

with β_k an $n_\beta \times 1$ vector of constrained ZTD parameters at the kth epoch, B_k the corresponding $m_k \times n_\beta$ partial design matrix, $\Phi_{k,k-1}$ $(k = 1,...,p)$ the $n_\beta \times n_\beta$ state transition matrices, and $d_k = 0$ an $n_\beta \times 1$ vector of constraints. In this model we use constraints on the β parameters every epoch.

The recursive processing can now be split into two steps per epoch: a time update and a measurement update. At the kth epoch we have the following observation model:

$$\mathbf{E}\{\begin{bmatrix} \hat{\underline{\beta}}_{k-1|k-1} \\ \hat{\underline{x}}_{k-1} \\ \underline{d}_k \\ \underline{y}_k \end{bmatrix}\} = \begin{bmatrix} I_{n_\beta} & & & \\ & & & I_n \\ -\Phi_{k,k-1} & I_{n_\beta} & & \\ T_k & & B_k & A_k \end{bmatrix} \begin{bmatrix} t_k \\ \beta_{k-1} \\ \beta_k \\ x \end{bmatrix} \; ;$$

$$\mathbf{D}\{\begin{bmatrix} \hat{\underline{\beta}}_{k-1|k-1} \\ \hat{\underline{x}}_{k-1} \\ \underline{d}_k \\ \underline{y}_k \end{bmatrix}\} = \begin{bmatrix} Q_{\hat{\beta}_{k-1|k-1}} & Q_{\hat{\beta}_{k-1|k-1}\hat{x}_{k-1}} & & \\ Q_{\hat{x}_{k-1}\hat{\beta}_{k-1|k-1}} & Q_{\hat{x}_{k-1}} & & \\ & & Q_{d_k} & \\ & & & Q_{y_k} \end{bmatrix} . \tag{21.26}$$

$\hat{\underline{\beta}}_{k-1|k-1}$

The first index of $\hat{\underline{\beta}}_{k-1|k-1}$ indicates the epoch for which the estimator holds and the second index refers to the last epoch that is used to determine the estimator. The model therefore implies that we have an estimator of β_{k-1} based on all observations up to and including epoch $k-1$.

time update

In the first step, the time update, we solve for the partial system of the first three rows:

$$\mathbf{E}\{\begin{bmatrix} \hat{\underline{\beta}}_{k-1|k-1} \\ \hat{\underline{x}}_{k-1} \\ \underline{d}_k \end{bmatrix}\} = \begin{bmatrix} I_{n_\beta} & 0 & 0 \\ 0 & 0 & I_n \\ -\Phi_{k,k-1} & I_{n_\beta} & 0 \end{bmatrix} \begin{bmatrix} \beta_{k-1} \\ \beta_k \\ x \end{bmatrix} . \tag{21.27}$$

Because we have a square and invertible matrix here, the estimators simply follow by inversion:

$$\begin{bmatrix} \hat{\underline{\beta}}_{k-1|k-1} \\ \hat{\underline{\beta}}_{k|k-1} \\ \hat{\underline{x}}_k \end{bmatrix} = \begin{bmatrix} I_{n_\beta} & 0 & 0 \\ \Phi_{k,k-1} & 0 & I_{n_\beta} \\ 0 & I_n & 0 \end{bmatrix} \begin{bmatrix} \hat{\underline{\beta}}_{k-1|k-1} \\ \hat{\underline{x}}_{k-1} \\ \underline{d}_k \end{bmatrix} , \tag{21.28}$$

$\hat{\underline{\beta}}_{k|k-1}$

where $\hat{\underline{\beta}}_{k|k-1}$ is the predicted estimator of β_k; we only used the observations up to epoch $k-1$ and the additional constraints to determine this estimator. Since $\hat{\underline{\beta}}_{k-1|k-1}$ is not used in the measurement update that follows, we can restrict the time update to the last two rows of Eq. (21.28):

$$\begin{bmatrix} \hat{\underline{\beta}}_{k|k-1} \\ \hat{\underline{x}}_k \end{bmatrix} = \begin{bmatrix} \Phi_{k,k-1} & 0 \\ 0 & I_n \end{bmatrix} \begin{bmatrix} \hat{\underline{\beta}}_{k-1|k-1} \\ \hat{\underline{x}}_{k-1} \end{bmatrix} + \begin{bmatrix} \underline{d}_k \\ 0 \end{bmatrix} . \tag{21.29}$$

Although the estimator $\hat{\underline{\beta}}_{k|k-1}$ differs from $\hat{\underline{\beta}}_{k-1|k-1}$ by $\underline{d}_k$, the estimates are the same if $\Phi_{k,k-1} = I_{n_\beta}$, since $d_k = 0$.

$\hat{\underline{\beta x}}_{k|k-1}$
BA_k
$\hat{\underline{\beta x}}_{k|k}$
$\Psi_{k,k-1}$

We now introduce the following short-hand notation:

$$\hat{\underline{\beta x}}_{k|k-1} \doteq \begin{bmatrix} \hat{\underline{\beta}}_{k|k-1} \\ \hat{\underline{x}}_k \end{bmatrix} \; ; BA_k \quad \doteq \begin{bmatrix} B_k, & A_k \end{bmatrix} ;$$

$$\hat{\underline{\beta x}}_{k|k} \quad \doteq \begin{bmatrix} \hat{\underline{\beta}}_{k|k} \\ \hat{\underline{x}}_k \end{bmatrix} \quad ; \Psi_{k,k-1} \doteq \begin{bmatrix} \Phi_{k,k-1} & 0 \\ 0 & I_n \end{bmatrix} .$$

(21.30)

With this notation the stochastic model for Eq. (21.29) becomes:

$$\boxed{\mathbf{D}\{\hat{\underline{\beta x}}_{k|k-1}\} \doteq Q_{\hat{\beta x}_{k|k-1}} = \Psi_{k,k-1} Q_{\hat{\beta x}_{k-1|k-1}} \Psi'_{k,k-1} + \begin{bmatrix} Q_{d_k} & 0 \\ 0 & 0 \end{bmatrix},}$$

(21.31)

which simply follows from application of the propagation law of (co)variances. A specification of the matrix Q_{d_k} is given in Part III. Using the predicted states and the new measurements of the kth epoch we find the model for the measurement update measurement update as:

$$\boxed{\begin{aligned} &\mathbf{E}\{\begin{bmatrix} \underline{y}_k \\ \hat{\underline{\beta x}}_{k|k-1} \end{bmatrix}\} = \begin{bmatrix} T_k \; BA_k \\ 0 \; I_{n_\beta+n} \end{bmatrix} \begin{bmatrix} t_k \\ \begin{pmatrix} \beta_k \\ x \end{pmatrix} \end{bmatrix} ; \\ &\mathbf{D}\{\begin{bmatrix} \underline{y}_k \\ \hat{\underline{\beta x}}_{k|k-1} \end{bmatrix}\} = \begin{bmatrix} Q_{y_k} & \\ & Q_{\hat{\beta x}_{k|k-1}} \end{bmatrix} . \end{aligned}}$$

(21.32)

The approach for deriving the formulas for the gain matrix and the vector of predicted residuals is similar to the one in Sect. 21.4. Model (21.32) is in fact the same as Model (21.14) if

$$BA_k \leftarrow A_k \quad ; \quad \hat{\beta x}_{k|k-1} \leftarrow \hat{x}_{k-1} \quad ; \quad Q_{\hat{\beta x}_{k|k-1}} \leftarrow Q_{\hat{x}_{k-1}}.$$

(21.33)

The measurement update therefore reads:

$$\boxed{\hat{\beta x}_{k|k} = \hat{\beta x}_{k|k-1} + K_k v_k,}$$

(21.34)

with gain matrix and vector of predicted residuals:

$$\boxed{\begin{aligned} K_k &= Q_{\hat{\beta x}_{k|k}} BA'_k P_{T_k}^{\perp'} Q_{y_k}^{-1}; \\ v_k &= y_k - BA_k \; \hat{\beta x}_{k|k-1}. \end{aligned}}$$

(21.35)

The complete recursive estimation procedure for the model with constrained parameters is given in Fig. I.1 in App. I.

Pre-elimination in a recursive SRIF

22.1 Introduction

SRIF

Because Kalman filtering involves matrix inversions of squared design-matrix terms[1], it is sensitive to roundoff errors and therefore the algorithm may not be stable when the design matrices are ill-conditioned. An alternative is given by the recursive Square Root Information Filter (SRIF) [*Bierman*, 1977]. This method does not propagate a covariance matrix like $Q_{\hat{x}_k}$, but a square root $R_{\hat{x}_k}$ of its inverse: $Q_{\hat{x}_k}^{-1} = R_{\hat{x}_k}{}'R_{\hat{x}_k}$. An extensive description, which includes testing aspects, is given in [*Tiberius*, 1998]. In this chapter, we review the estimation part for linear models, and deal with the pre-elimination of temporal parameters.

QR

Section 22.2 gives a short review of the use of QR decomposition in the SRIF method. As in the previous chapter, we deal with three different models: the model with global parameters only (Sect. 22.3), the model with global and temporal parameters (Sect. 22.4), and the model with global, temporal, and constrained parameters (Sect. 22.5).

22.2 Least squares and QR decomposition

Consider the general full-rank Gauss–Markov model:

$$\mathbf{E}\{\underline{y}\} = Ax \quad ; \quad \mathbf{D}\{\underline{y}\} = I_m, \tag{22.1}$$

where A is an $m \times n$ matrix, y is an $m \times 1$, and x an $n \times 1$ vector.

The more general model where $\mathbf{D}\{\underline{y}\} = Q_y$ can be transformed to obtain Model (22.1) as follows:

$$\begin{aligned} y &\leftarrow (R_y^{-1})'y = R_y{}'\backslash y; \\ A &\leftarrow (R_y^{-1})'A = R_y{}'\backslash A, \end{aligned} \tag{22.2}$$

substitution, \

with $R_y = Chol(Q_y)$ the upper-triangular Cholesky factor of $Q_y = R_y{}'R_y$. The backslash stands for substitution; in other words, no actual matrix inversion is needed to apply this transformation.

The LSQ estimate of x can be obtained by a QR decomposition [*Golub and Van Loan*, 1989] of the information array $[A, y]$:

$$Q \begin{bmatrix} R \\ 0 \end{bmatrix} = [A, y]. \tag{22.3}$$

---

[1] $A_k{}'Q_{y_k}^{-1}A_k$ in the model of observation equations as described in the previous chapter or $A_k Q_{\hat{x}_{k-1}} A_k{}'$ in the model of condition equations as described in Chap. 24.

orthonormal
Q

R is an upper-triangular matrix and Q is an orthonormal and square $m \times m$ matrix: $QQ' = Q'Q = I$.

QR decomposition is a computationally simple method that utilizes either repeated Householder or Givens transformations [*Golub and Van Loan*, 1989]. Figure 22.1 shows MATLAB code [*Matlab*, 1997] to apply this decomposition without explicitly determining the Q matrix. In fact, a column number may be transported into this subroutine up to where the decomposition should be applied (incomplete decomposition).

If we split Q into $[Q_I, Q_{II}]$, with an $m \times (n+1)$ matrix Q_I and an $m \times (m-n-1)$ matrix Q_{II}, and premultiply Eq. (22.3) by Q', we find:

$$\begin{bmatrix} R \\ 0 \end{bmatrix} = \begin{bmatrix} Q_I'A & Q_I'y \\ Q_{II}'A & Q_{II}'y \end{bmatrix} = \begin{bmatrix} R_{\hat{x}} & z_{\hat{x}} \\ 0 & \hat{\rho} \\ 0 & 0 \end{bmatrix}, \tag{22.4}$$

where $\hat{\rho}$ is a scalar. The upper row just follows by definition: $Q_I'A \doteq R_{\hat{x}}$, and $Q_I'y \doteq z_{\hat{x}}$. The middle row of the right-hand side is a consequence of the upper-triangular structure of R, which also means that $R_{\hat{x}}$ must be upper triangular. Note that $R_{\hat{x}}$ may have negative diagonal elements unlike a Cholesky factor.

If $\varepsilon \doteq y - Ax$ is the vector of residuals, $\hat{x}$ is the vector x that minimizes $\| \varepsilon \|^2$ and $\hat{\varepsilon} \doteq y - A\hat{x}$ is the vector of LSQ residuals. Using the property $QQ' = I$ and the QR decomposition of Eq. (22.4), we find [*Bierman*, 1977]:

$$\| \varepsilon \|^2 = \varepsilon'\varepsilon = \varepsilon'QQ'\varepsilon = \| Q'\varepsilon \|^2 = \| Q' \, [A, \, y] \begin{bmatrix} -x \\ 1 \end{bmatrix} \|^2$$

$$= \| \begin{bmatrix} R_{\hat{x}} & z_{\hat{x}} \\ 0 & \hat{\rho} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -x \\ 1 \end{bmatrix} \|^2 \quad = \| z_{\hat{x}} - R_{\hat{x}}x \|^2 + \hat{\rho}^2. \tag{22.5}$$

Since $R_{\hat{x}}$ is a square and invertible matrix, $\hat{x} = R_{\hat{x}}^{-1}z_{\hat{x}} \equiv R_{\hat{x}} \backslash z_{\hat{x}}$ sets $\| z_{\hat{x}} - R_{\hat{x}}\hat{x} \|^2 = 0$, and therefore minimizes $\| \varepsilon \|^2$. Consequently, $\hat{\rho}^2 = \| \hat{\varepsilon} \|^2$ is the sum of squared LSQ residuals. The hat notation in $R_{\hat{x}}$, $z_{\hat{x}}$, and $\hat{\rho}$ is now apparent.

QR decomposition of the information array $[A, y]$ is very practical to obtain the observation equation $z_{\hat{x}} = R_{\hat{x}} \, \hat{x}$. From a theoretical point of view, a decomposition of the design matrix into $A = [Q_I, \; Q_{II}] \begin{bmatrix} R_{\hat{x}} \\ 0 \end{bmatrix}$ is easier to understand (with different Q_{II} as above). Since Q_I spans the column space of A, and Q_I and Q_{II} are orthonormal, we have the projectors $P_A = Q_I Q_I'$ and $P_A^{\perp} = Q_{II}Q_{II}'$. The vector of LSQ residuals then reads $\hat{\varepsilon} = Q_{II}Q_{II}'y$, and thus $Q_I'\hat{\varepsilon} = 0$. Therefore:

$$\begin{aligned} y &= & A \, \hat{x} + \hat{\varepsilon} \\ y &= Q_I \, R_{\hat{x}} \, \hat{x} + \hat{\varepsilon} \\ z_{\hat{x}} \doteq Q_I'y &= & R_{\hat{x}} \, \hat{x} \, . \end{aligned} \tag{22.6}$$

In fact we can obtain this equation by an incomplete QR decomposition of the first n columns of $[A, \, y]$, that is, without the last column.

Since $R_{\hat{x}}$ and $z_{\hat{x}}$ are the result of a QR decomposition of $[A, \; y]$, the symbolic notation:

$$[R_{\hat{x}}, \; z_{\hat{x}}] \leftarrow QR([A, \; y]) \tag{22.7}$$

is used henceforth.

```
function [A] = qrpart(A,n1);

[m,n] = size(A);
if n>m, A = [A;zeros(n-m,n)]; m=n; end
if nargin==1, n1=n; end
v = zeros(m,1);
for j=1:n1
  v(j:m) = house(A(j:m,j));
  A(j:m,j:n) = rowhouse(A(j:m,j:n),v(j:m));
  A(j+1:m,j) = 0;
end
A = A(1:n1,1:n1);

function v = house(x);

n = length(x);
normx = norm(x);
v = x;
s = sign(x(1));
if s==0, s=1; end
if normx
  beta = x(1) + s*normx;
  v(2:n) = v(2:n)/beta;
end
v(1) = 1;

function A = rowhouse(A,v);

beta = -2/(v'*v);
w = beta*A'*v;
A = A + v*w';
```

Figure 22.1   Partial QR decomposition in MATLAB code using Householder transformations (after Golub and van Loan). An in-place memory technique is used; the output A is in fact the upper-triangular matrix R.

22.3   The model with global parameters only

The partitioned model with global parameters only reads:

$$\mathbf{E}\{\underline{y}\} = Ax \quad ; \quad \mathbf{D}\{\underline{y}\} = I_{mp}, \tag{22.8}$$

where $y \doteq [y_1', ..., y_p']'$ and $A \doteq [A_1', ..., A_p']'$. In this model, y_k is an $m_k \times 1$ vector of observations at epoch k, A_k is an $m_k \times n$ partial design matrix, and x is the $n \times 1$ vector of global parameters.

This model may be solved recursively by using repeated QR decompositions. We assume that $m_1 > n$ and that A_1 is of full rank. Although this is not strictly necessary, since we can also start the recursion after a few epochs, this assures that the first epoch of data is sufficient to obtain the following observation equations:

$$\boxed{\mathbf{E}\{\underline{z}_{\hat{x}_1}\} = R_{\hat{x}_1}x \quad ; \quad \mathbf{D}\{\underline{z}_{\hat{x}_1}\} = I_n,} \tag{22.9}$$

where $z_{\hat{x}_1} \doteq Q_{I,1}{}' \; y_1$, and $A_1 = Q_1 R_1$ is the QR decomposition of A_1. Q_1 is an orthonormal $m_1 \times m_1$ matrix, $Q_{I,1}$ is an $m_1 \times n$ submatrix of Q_1 (compare with Sect. 22.2), $R_1 = [R_{\hat{x}_1}{}', \; 0]'$, and $R_{\hat{x}_1}$ is an upper-triangular $n \times n$ matrix.

After processing $k - 1$ epochs of data $(k \geq 2)$, the following measurement-update model incorporates all information until epoch k:

$$\mathbf{E}\{\begin{bmatrix} \underline{y}_k \\ \underline{z}_{\hat{x}_{k-1}} \end{bmatrix}\} = \begin{bmatrix} A_k \\ R_{\hat{x}_{k-1}} \end{bmatrix} x \quad ; \quad \mathbf{D}\{\begin{bmatrix} \underline{y}_k \\ \underline{z}_{\hat{x}_{k-1}} \end{bmatrix}\} = \begin{bmatrix} I_{m_k} & 0 \\ 0 & I_n \end{bmatrix}. \tag{22.10}$$

The QR decomposition of the information array now becomes:

$$\begin{bmatrix} A_k & y_k \\ R_{\hat{x}_{k-1}} & z_{\hat{x}_{k-1}} \end{bmatrix} = Q_k R_k \quad ; \quad R_k \doteq \begin{bmatrix} R_{\hat{x}_k} & z_{\hat{x}_k} \\ 0 & \hat{\rho}_k \\ 0 & 0 \end{bmatrix}. \tag{22.11}$$

Or symbolically:

$$[R_{\hat{x}_k}, \; z_{\hat{x}_k}] \leftarrow QR(\begin{bmatrix} A_k & y_k \\ R_{\hat{x}_{k-1}} & z_{\hat{x}_{k-1}} \end{bmatrix}). \tag{22.12}$$

So both $R_{\hat{x}_k}$ and $z_{\hat{x}_k}$ now result as part of the larger matrix R_k. The computation of Q_k may again be omitted and the global parameters are regained by $\hat{x}_k = R_{\hat{x}_k} \backslash z_{\hat{x}_k}$.

22.4   Incorporation of temporal parameters

Incorporation of temporal parameters leads to an extension of Model (22.8):

$$\mathbf{E}\{\underline{y}\} = [T, \; A] \begin{bmatrix} t \\ x \end{bmatrix} \quad ; \quad \mathbf{D}\{\underline{y}\} = I_{mp}, \tag{22.13}$$

where $T \doteq \mathrm{diag}([T_1, ..., T_p])$ and $t \doteq [t_1', ..., t_p']'$. The dimensions are the same as in the previous section. T_k is an $m_k \times n_{t_k}$ partial design matrix, and t_k is an $n_{t_k} \times 1$ vector of temporal parameters. Like in the previous model, we assume that after only one epoch of data processing we arrive at the general model of the measurement update for epoch k:

$$\mathbf{E}\{\begin{bmatrix} \underline{y}_k \\ \underline{z}_{\hat{x}_{k-1}} \end{bmatrix}\} = \begin{bmatrix} T_k & A_k \\ 0 & R_{\hat{x}_{k-1}} \end{bmatrix} \begin{bmatrix} t_k \\ x \end{bmatrix} \quad ; \quad \mathbf{D}\{\begin{bmatrix} \underline{y}_k \\ \underline{z}_{\hat{x}_{k-1}} \end{bmatrix}\} = \begin{bmatrix} I_{m_k} & 0 \\ 0 & I_n \end{bmatrix}. \tag{22.14}$$

Again, the design matrix can be augmented to obtain an information array for the QR decomposition:

$$\begin{bmatrix} T_k & A_k & y_k \\ 0 & R_{\hat{x}_{k-1}} & z_{\hat{x}_{k-1}} \end{bmatrix} = Q_k R_k \quad ; \quad R_k \doteq \begin{bmatrix} R_{\hat{t}_k} & R_{\hat{t}_k \hat{x}_k} & z_{\hat{t}_k} \\ 0 & R_{\hat{x}_k} & z_{\hat{x}_k} \\ 0 & 0 & \hat{\rho}_k \\ 0 & 0 & 0 \end{bmatrix}. \tag{22.15}$$

The LSQ estimates of both the global and temporal parameters can be regained from the (transformed) observation equations:

$$\mathbf{E}\{\begin{bmatrix} \underline{z}_{\hat{t}_k} \\ \underline{z}_{\hat{x}_k} \end{bmatrix}\} = \begin{bmatrix} R_{\hat{t}_k} & R_{\hat{t}_k \hat{x}_k} \\ 0 & R_{\hat{x}_k} \end{bmatrix} \begin{bmatrix} t_k \\ x \end{bmatrix} \quad ; \quad \mathbf{D}\{\begin{bmatrix} \underline{z}_{\hat{t}_k} \\ \underline{z}_{\hat{x}_k} \end{bmatrix}\} = \begin{bmatrix} I_{n_{t_k}} & 0 \\ 0 & I_n \end{bmatrix}. \tag{22.16}$$

But since we are only interested in the global parameters, we may include only the last row in the next measurement-update step. This would not have been possible directly if the vector of parameters would have been in the order $[x', t_k']'$ instead! Mind that we cannot recover the covariance matrix of the temporal parameters by $R_{\hat{t}_k}$ alone: $Q_{\hat{t}_k} \neq R_{\hat{t}_k}^{-1}(R_{\hat{t}_k}^{-1})'$, whereas for the global parameters: $Q_{\hat{x}_k} = R_{\hat{x}_k}^{-1}(R_{\hat{x}_k}^{-1})'$.

22.5   Incorporation of constrained parameters

Extending Model (22.13) with constrained parameters gives:

$$\mathbf{E}\left\{\begin{bmatrix} \underline{d} \\ \underline{y} \end{bmatrix}\right\} = \begin{bmatrix} 0 & E & 0 \\ T & B & A \end{bmatrix}\begin{bmatrix} t \\ \beta \\ x \end{bmatrix} \quad ; \quad \mathbf{D}\left\{\begin{bmatrix} \underline{d} \\ \underline{y} \end{bmatrix}\right\} = \begin{bmatrix} I_{n_\beta p} & \\ & I_{m_k p} \end{bmatrix}, \tag{22.17}$$

with B, β, d, and E as in the previous chapter; the first row of the functional model is shown in Eq. (21.25). Without the first row of equations in Eq. (22.17), the model would basically be the same as in the previous section, but because we make a distinction between β and x, the model for the measurement update of epoch k reads (not using the first row of equations):

$$\mathbf{E}\left\{\begin{bmatrix} \underline{y}_k \\ \underline{z}_{\hat{\beta}_{k|k-1}} \\ \underline{z}_{\hat{x}_{k-1}} \end{bmatrix}\right\} = \begin{bmatrix} T_k & B_k & A_k \\ 0 & R_{\hat{\beta}_{k|k-1}} & R_{\hat{\beta}_{k|k-1}\hat{x}_{k-1}} \\ 0 & 0 & R_{\hat{x}_{k-1}} \end{bmatrix}\begin{bmatrix} t_k \\ \beta_k \\ x \end{bmatrix} ;$$

$$\mathbf{D}\left\{\begin{bmatrix} \underline{y}_k \\ \underline{z}_{\hat{\beta}_{k|k-1}} \\ \underline{z}_{\hat{x}_{k-1}} \end{bmatrix}\right\} = \begin{bmatrix} I_{m_k} & & \\ & I_{n_\beta} & \\ & & I_n \end{bmatrix}. \tag{22.18}$$

Note that we use the predicted z vectors that result from a time-update step. The QR decomposition leaves us with a smaller model of filtered z vectors:

$$\mathbf{E}\left\{\begin{bmatrix} \underline{z}_{\hat{t}_k} \\ \underline{z}_{\hat{\beta}_{k|k}} \\ \underline{z}_{\hat{x}_k} \end{bmatrix}\right\} = \begin{bmatrix} R_{\hat{t}_k} & R_{\hat{t}_k\hat{\beta}_{k|k}} & R_{\hat{t}_k\hat{x}_k} \\ 0 & R_{\hat{\beta}_{k|k}} & R_{\hat{\beta}_{k|k}\hat{x}_k} \\ 0 & 0 & R_{\hat{x}_k} \end{bmatrix}\begin{bmatrix} t_k \\ \beta_k \\ x \end{bmatrix} ;$$

$$\mathbf{D}\left\{\begin{bmatrix} \underline{z}_{\hat{t}_k} \\ \underline{z}_{\hat{\beta}_{k|k}} \\ \underline{z}_{\hat{x}_k} \end{bmatrix}\right\} = \begin{bmatrix} I_{n_{t_k}} & & \\ & I_{n_\beta} & \\ & & I_n \end{bmatrix}. \tag{22.19}$$

Because there is no need to solve for the temporal parameters, only the last two rows are used in the time update for epoch $k + 1$, which also incorporates a row from Eq. (21.25) [*Tiberius, 1997*]:

$$\mathbf{E}\left\{\begin{bmatrix} \underline{d}_{k+1} \\ \underline{z}_{\hat{\beta}_{k|k}} \\ \underline{z}_{\hat{x}_k} \end{bmatrix}\right\} = \begin{bmatrix} -\Phi_{k,k-1} & I_{n_\beta} & 0 \\ R_{\hat{\beta}_{k|k}} & 0 & R_{\hat{\beta}_{k|k}\hat{x}_k} \\ 0 & 0 & R_{\hat{x}_k} \end{bmatrix}\begin{bmatrix} \beta_k \\ \beta_{k+1} \\ x \end{bmatrix} ;$$

$$\mathbf{D}\left\{\begin{bmatrix} \underline{d}_{k+1} \\ \underline{z}_{\hat{\beta}_{k|k}} \\ \underline{z}_{\hat{x}_k} \end{bmatrix}\right\} = \begin{bmatrix} I_{n_\beta} & & \\ & I_{n_\beta} & \\ & & I_n \end{bmatrix}. \tag{22.20}$$

The corresponding QR decomposition of this square system reads:

$$
\mathbf{E}\left\{\begin{bmatrix} \underline{z}_{\hat{\beta}_{k|k}} \\ \underline{z}_{\hat{\beta}_{k+1|k}} \\ \underline{z}_{\hat{x}_{k+1}} \end{bmatrix}\right\} = \begin{bmatrix} R_{\hat{\beta}_{k|k}} & R_{\hat{\beta}_{k|k}\hat{\beta}_{k+1|k}} & R_{\hat{\beta}_{k|k}\hat{x}_{k+1}} \\ 0 & R_{\hat{\beta}_{k+1|k}} & R_{\hat{\beta}_{k+1|k}\hat{x}_{k+1}} \\ 0 & 0 & R_{\hat{x}_{k+1}} \end{bmatrix} \begin{bmatrix} \beta_k \\ \beta_{k+1} \\ x \end{bmatrix};
$$
$$
\mathbf{D}\left\{\begin{bmatrix} \underline{z}_{\hat{\beta}_{k|k}} \\ \underline{z}_{\hat{\beta}_{k+1|k}} \\ \underline{z}_{\hat{x}_{k+1}} \end{bmatrix}\right\} = \begin{bmatrix} I_{n_\beta} & & \\ & I_{n_\beta} & \\ & & I_n \end{bmatrix}.
$$

(22.21)

Only the last two rows of these equations are used in the measurement update of epoch $k + 1$, cf. Eq. (22.18). Note that the information array of Eq. (22.20) is square and $R_{\hat{x}_k}$ is already upper triangular. A complete QR decomposition is therefore not necessary. Transformation of the first $2n_\beta$ columns is sufficient and $R_{\hat{x}_{k+1}} = R_{\hat{x}_k}$.

Implementation aspects

23.1 Introduction

In the previous chapters, procedures are described for recursive estimation of parameters in linear observation models. In this chapter we review some technical problems when implementing these procedures. First of all, the GPS observation model is not linear because of the nonlinear coordinates. Section 23.2 illustrates how to deal with this. Section 23.3 shows how to avoid matrix inversions and how to speed up the computations. Finally, Sect. 23.4 shows how to handle ambiguities under changing satellite configurations. This implies that: (1) ambiguities are transformed when pivot satellites change; (2) old ambiguities are expired when satellites set; and (3) new ambiguities are introduced when satellites rise.

23.2 Nonlinearity

Because the coordinates appear as nonlinear parameters in the observation models, we need iterative estimation procedures for both the KF and the SRIF approach. Since we have three different parameter combinations, two estimation methods, and both linear and nonlinear versions, there are in total twelve different procedures. In Table 23.1, the figure numbers are shown of the flow charts. Most of them can be found in App. I.

parameters	Kalman linear	Kalman nonlinear	SRIF nonlinear
x	21.1	I.2	I.3
x, t_k	21.1	I.4	I.5
x, t_k, β_k	I.1	I.6	I.7

Table 23.1   Figure numbers of estimation procedures. No figures are shown of the linear SRIF procedures (the QR decomposition is the only step in the measurement update).

Fortunately, the global parameters, more specifically the coordinates, are the only nonlinear parameters. In the model with global parameters only, we may form 'observed-minus-computed' observation increments $\Delta y_k = y_k - A_k(x^0)$ and $\Delta \hat{x}_{k-1} = x_{k-1} - x^0$ for the 'parameter observations' of the previous epoch, where $x^0 \equiv x_k^0$. Using a Taylor expansion we find that:

Taylor expansion

$$\mathbf{E}\{\underline{\Delta y}_k\} = \partial_x A_k(x^0)\Delta x_k + \mathcal{O}(\| \Delta x_k \|^2_{Q_{x_k}}), \tag{23.1}$$

with $\partial_x A_k(x^0)$ the gradient matrix or Jacobian, $\Delta x_k = x_k - x^0$ the parameter increments, and $\mathcal{O}(\| \Delta x_k \|^2_{Q_{x_k}})$ an error term. The derivation of the estimation

procedure is rather straightforward using the increments and the gradient matrix as above. Figure I.2 shows the estimation procedure for the KF approach. Each iteration step, the LSQ estimate is updated as $\hat{x}_k = x^0 + \Delta\hat{x}_k$. Because $\hat{x}_k$ serves as approximate values for the next iteration as well as for the next recursion step, the update is written as $x^0 \leftarrow x^0 + \Delta\hat{x}_k$. Only after the last recursion we find the final estimate in the procedure: $\hat{x}_p \leftarrow x^0$.

convergence check   Each iteration step of Fig. I.2 ends with the convergence check:

$$\Delta\hat{x}_k' Q_{\hat{x}_k}^{-1} \Delta\hat{x}_k < \epsilon. \tag{23.2}$$

The right-hand side of Eq. (23.2) is a dimensionless scalar that serves as a fixed criterion. If the weighted sum of parameter increments on the left-hand side is small enough, the iteration is ended, and the procedure continues with the next epoch or ends when the last epoch is reached.

Figure I.3 shows the nonlinear SRIF procedure for the model with global parameters only. In the linear SRIF procedures we propagate the vector $z_{\hat{x}_{k-1}}$ instead of the state vector. This 'z vector' serves as 'parameter observation' vector, and like $\hat{x}_{k-1}$ in the KF approach, we need a linearized version in the recursion. As long as there is no need to know the actual state vector, in the linear SRIF there is also no need to compute it. In the nonlinear model however, we must recover the state vector because we need it in the gradient matrix and for the observation increments.

Since $\Delta\hat{x}_k = R_{\hat{x}_k}^{-1} \Delta z_{\hat{x}_k}$ and $Q_{\hat{x}_k}^{-1} = R_{\hat{x}_k}' R_{\hat{x}_k}$, the convergence check of Eq. (23.2) is equivalent with:

$$\Delta z_{\hat{x}_k}' \Delta z_{\hat{x}_k} < \epsilon. \tag{23.3}$$

Figures I.4 and I.5 show the nonlinear KF and SRIF estimation procedures of the model with both global and temporal parameters. These procedures do not actually solve for the temporal parameters. Because they also occur as linear parameters we do not need to have approximate values and we do not need to update the observation increments Δy_k with $T_k\, t_k^0$.

$\partial_{\beta x} BA_k(\beta x^0)$ Figures I.6 and I.7 show the procedures of the nonlinear models with all parameter types. In these procedures we use the notation $\partial_{\beta x} BA_k(\beta x^0) \doteq [B_k,\ \partial_x A_k(x^0)]$.
$z_{\hat{\beta} x_{k|k}}$ In the latter figure we use the notation $z_{\hat{\beta} x_{k|k}} \doteq [z'_{\hat{\beta}_{k|k}},\ z'_{\hat{x}_k}]'$. The nonlinearity only plays a role in the measurement update. A slightly more complicated convergence check than in Eq. (23.3) is used that prevents propagation of errors in the constrained parameters: $\|\Delta z_{\hat{\beta} x_{k|k}}\|^2 < \epsilon$.

23.3   Computation speed and sparsity considerations

23.3.1   Cholesky-factor transformation

SRIF   As we have seen in Chap. 22, in the SRIF approach we need to compute the Cholesky factor of the covariance matrix of the observables, and use it to transform the observables and partial design matrices.[1] Because the covariance matrix is block diagonal, the Cholesky factor is also block diagonal and each of the blocks

---

[1]Alternatively we could also use generalized QR decompositions [*Björck*, 1996], but this is beyond our scope.

may be computed individually. The individual Cholesky factorizations and transformations can be implemented in the recursion. The observation equations in the time and measurement update,

$$
\begin{aligned}
\mathbf{E}\{\underline{y}_k\} &= T_k t_k + B_k \beta_k + A_k x \quad ; \quad \mathbf{D}\{\underline{y}_k\} = Q_{y_k} = R_{y_k}{}' R_{y_k}; \\
\mathbf{E}\{\underline{d}_k\} &= I_{n_\beta} \beta_k - \Phi_{k,k-1} \beta_{k-1} \quad ; \quad \mathbf{D}\{\underline{d}_k\} = Q_{d_k} = R_{d_k}{}' R_{d_k},
\end{aligned}
\tag{23.4}
$$

with Cholesky factors R_{y_k} and R_{d_k}, are then transformed as

$$
\begin{aligned}
\mathbf{E}\{(R_{y_k}^{-1})'\underline{y}_k\} &= (R_{y_k}^{-1})'T_k t_k + (R_{y_k}^{-1})'B_k \beta_k + (R_{y_k}^{-1})'A_k x; \\
\mathbf{D}\{(R_{y_k}^{-1})'\underline{y}_k\} &= I_{m_k}; \\
\mathbf{E}\{(R_{d_k}^{-1})'\underline{d}_k\} &= (R_{d_k}^{-1})'\beta_k - (R_{d_k}^{-1})'\Phi_{k,k-1}\beta_{k-1}; \\
\mathbf{D}\{(R_{d_k}^{-1})'\underline{d}_k\} &= I_{n_\beta}.
\end{aligned}
\tag{23.5}
$$

In other words, new observations and partial design matrices result. Because we use $d_k = 0$, the transformation of d_k is only formal. Recall from the previous chapter that we do not actually compute the inverse of Cholesky factors but use substitutions instead.

KF   When using the KF approach, we do not necessarily have to transform the observations and partial design matrices. In the measurement update however, the inverse of the covariance matrix of the observations pops up several times. To avoid this actual inversion and some matrix products, we can compute a Cholesky factor here as well, and use it to transform the measurement-update equations. In the KF procedures, Q_{y_k} is then replaced by an identity matrix. More on this topic can be found in Chap. 24.

example of sizes   Figure 23.1 shows typical sizes of the partial design matrices for the troposphere-weighted model without troposphere gradients or slant delay parameters, when $l = 4$ observation types are used, $m = 7$ satellites ($m_k = m$, for all k) are observed in a network of $n = 5$ stations. In this case we have $mnl = 140$ observations, $3n = 15$ coordinates $+ 2(m-1)(n-1) = 48$ ambiguities (63 global parameters), $n = 5$ ZTD parameters, and $mn = 35$ ionosphere $+ 3(m + n - 1) = 33$ clock parameters (68 temporal parameters).

operation	flops	kflops	operation	flops	kflops
$Chol(Q_{y_k})$	$\frac{1}{6}m_k^3$	457	$Q_{y_k}^{-1}$	m_k^3	2744
$R_{y_k}' \backslash [T_k\ BA_k\ y_k]$	$\frac{1}{2}m_k^2 \cdot$	1392	$Q_{y_k}^{-1} \cdot BA_k$	$m_k^2(n_\beta + n)$	1431
	$(n_{t_k} + n_\beta + n + 1)$				
$T_k' \cdot T_k$	$\frac{1}{2}m_k n_{t_k}^2$	324	$T_k' \cdot Q_{y_k}^{-1}$	$m_k^2 n_{t_k}$	1333
$Chol(T_k'T_k)$	$\frac{1}{6}n_{t_k}^3$	52	$(T_k'Q_{y_k}^{-1}) \cdot T_k$	$\frac{1}{2}m_k n_{t_k}^2$	324
$R_{T_k}' \backslash T_k'$	$\frac{1}{2}m_k n_{t_k}^2$	324	$(T_k'Q_{y_k}^{-1}T_k)^{-1}$	$n_{t_k}^3$	314
$\widetilde{R}_{T_k} \cdot BA_k$	$m_k n_{t_k}(n_\beta + n)$	695	P_{T_k}	$\frac{1}{2}m_k n_{t_k}(m_k + n_{t_k})$	990
$\widetilde{R}_{T_k}' \cdot (\widetilde{R}_{T_k}BA_k)$	$m_k n_{t_k}^2$	647	$P_{T_k} \cdot BA_k$	$m_k^2(n_\beta + n)$	1431
total:		3891			8567

Table 23.2   Operation count in floating point operations (1 flops = 1 multiplication + 1 addition [*Björck*, 1996]) for two ways of computing. Left: With Cholesky factorization and substitution. Right: By inverting and multiplying. The numbers in kiloflops are for the example of page 173. m_k: number of observations. n_{t_k}: number of temporal parameters. n_β: number of constrained parameters. n: number of global parameters.
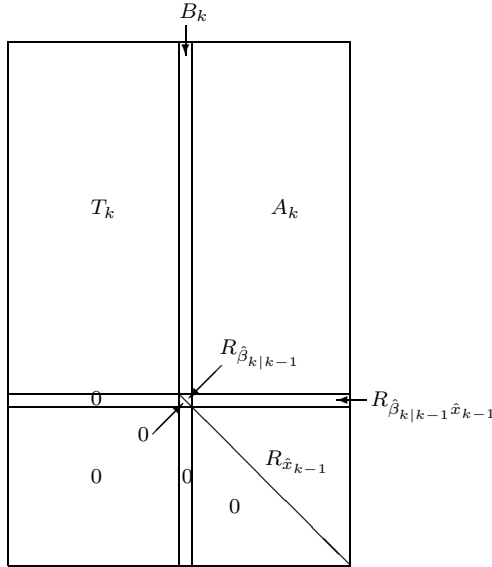
Figure 23.1   Mirror image of the sizes of the design matrix in the SRIF measurement update. Troposphere-weighted model; $l = 4$ observation types, $n = 5$ stations, $m = 7$ satellites.

23.3.2   Computation of the product $P_{T_k}^{\perp} BA_k$

In the KF approach, the temporal parameters are pre-eliminated by using the projector $P_{T_k}^{\perp}$. Because T_k is a large matrix, the computation of the projector may be time consuming. Since we only use the product $P_{T_k}^{\perp} BA_k$ (or $P_{T_k}^{\perp} A_k$), this section shows how this product can be computed efficiently.

If we use the Cholesky transformations of the previous subsection, we have the projector $P_{T_k} = T_k(T_k{}'T_k)^{-1}T_k{}'$. Instead of the time consuming inversion of the matrix $T_k{}'T_k$, we compute its Cholesky factor R_{T_k}:

$$R_{T_k}{}' R_{T_k} = T_k{}'T_k, \tag{23.6}$$

with which the projector P_{T_k} becomes:

$$P_{T_k} = T_k \, R_{T_k}^{-1} \, (R_{T_k}^{-1})' \, T_k{}' = \widetilde{R}_{T_k}' \, \widetilde{R}_{T_k}, \tag{23.7}$$

where

$$\widetilde{R}_{T_k} \doteq (R_{T_k}^{-1})' \, T_k{}' \equiv R_{T_k}{}'\backslash T_k{}' \tag{23.8}$$

follows from substitution. Note that $\widetilde{R}_{T_k}$ is not triangular.

Table 23.2 shows the smaller operation count for the approach with Cholesky factorizations and transformations when compared with inversions and multiplications.

If Q_{y_k} is diagonal, which is often assumed, the Cholesky-factor transformed T_k and BA_k are sparse. In this case the products $T_k'T_k$ and $\widetilde{R}_{T_k} BA_k$ can be computed faster. The computation time can than be further reduced by avoiding fill-in elements in the Cholesky factorization of $T_k'T_k$ with an ordering algorithm like the column minimum degree ordering, see for example [*George and Liu*, 1989]. With this algorithm, we obtain a column permutation matrix [2] $\mathcal{P}_c$. The projector

---

[2]Permutation matrices have the property $\mathcal{P}^{-1} = \mathcal{P}'$.

P_{T_k} may be computed by replacing T_k by $T_k \mathcal{P}_c$, since

$$
\begin{aligned}
P_{T_k} &= T_k \; (T_k{}' \, T_k)^{-1} \; T_k{}' \\
&= T_k \; \mathcal{P}_c \; \mathcal{P}_c^{-1} \; (T_k{}' \, T_k)^{-1} \; (\mathcal{P}_c^{-1})' \; \mathcal{P}_c' \; T_k{}' \\
&= T_k \; \mathcal{P}_c \; (\mathcal{P}_c{}' \, T_k{}' \, T_k \, \mathcal{P}_c)^{-1} \; \mathcal{P}_c{}' \; T_k{}'.
\end{aligned}
\tag{23.9}
$$

In other words, P_{T_k} is invariant for the column ordering of T_k. By using the permutated matrix, the Cholesky factor is sparse and therefore also the substitution step is faster. The most time consuming step is the full matrix multiplication of Eq. (23.7), which generally results in a full matrix. Therefore we first compute the still sparse product $\widetilde{R}_{T_k} BA_k$. Finally, the product is simply computed as:

$$
P_{T_k}^{\perp} BA_k = BA_k - \widetilde{R}_{T_k}'(\widetilde{R}_{T_k} BA_k).
\tag{23.10}
$$

Summarizing the steps above gives:

$$
\begin{aligned}
&1. \; T_k && \leftarrow T_k \mathcal{P}_c; \\
&2. \; R_{T_k} && = Chol(T_k' \, T_k); \\
&3. \; \widetilde{R}_{T_k} && = R_{T_k}{}' \backslash T_k{}'; \\
&4. \; P_{T_k}^{\perp} BA_k && = BA_k - \widetilde{R}_{T_k}'(\widetilde{R}_{T_k} BA_k).
\end{aligned}
\tag{23.11}
$$

Some timing results are shown in Table 23.3.

operation	time [ms] (with column ordering)	time [ms] (without column ordering)
$R_{T_k}{}' \backslash T_k{}'$	5.5	9.6
$\widetilde{R}_{T_k} \cdot BA_k$	1.8	2.5
$\widetilde{R}_{T_k}' \cdot (\widetilde{R}_{T_k} BA_k)$	14.3	19.1

Table 23.3   Computation time of operations in milliseconds for a diagonal covariance matrix of the observations. The matrix sizes correspond to the example of page 173.

23.3.3   Sparse QR decomposition

In the SRIF approach, the QR decompositions are the most time consuming. Especially the QR decomposition of the measurement update is slow because of the large information array. When T_k, B_k, and A_k are sparse, it may pay off to use a sparse QR decomposition. In [*Matstoms*, 1994] a multifrontal QR decomposition is described. This method uses the sparsity of the system and permutations implies both a column ($\mathcal{P}_c$) and row permutation ($\mathcal{P}_r$).

The system $\Delta y_k = A_k \; \Delta \hat{x}_k + \hat{\varepsilon}_k$, with $\Delta x_k \doteq [\Delta t_k', \; \Delta \beta_k', \; \Delta x']'$ the vector of all parameters at epoch k, $\Delta y_k \doteq [\Delta y_k', \; \Delta z_{\hat{\beta} x_{k|k-1}}']'$ the vector with all observations, and A_k the complete design matrix for epoch k (as in Fig. 23.1), is solved as:

$$
\begin{aligned}
\mathcal{P}_r \, \Delta y_k &= \mathcal{P}_r \, A_k \, \mathcal{P}_c \, \mathcal{P}_c' \, \Delta \hat{x}_k + \mathcal{P}_r \hat{\varepsilon}_k \\
\mathcal{P}_r \, \Delta y_k &= \overline{Q}_I \, \overline{R}_k \, \mathcal{P}_c' \, \Delta \hat{x}_k + \mathcal{P}_r \hat{\varepsilon}_k \\
\Delta \overline{z}_k \doteq \overline{Q}_I{}' \, \mathcal{P}_r \, \Delta y_k &= \overline{R}_k \, \mathcal{P}_c' \, \Delta \hat{x}_k \; ,
\end{aligned}
\tag{23.12}
$$

or symbolically: $[\overline{R}_k \mathcal{P}_c', \; \Delta \overline{z}_k] \leftarrow QR(\mathcal{P}_r A_k \mathcal{P}_c, \; \mathcal{P}_r \Delta y_k)$.
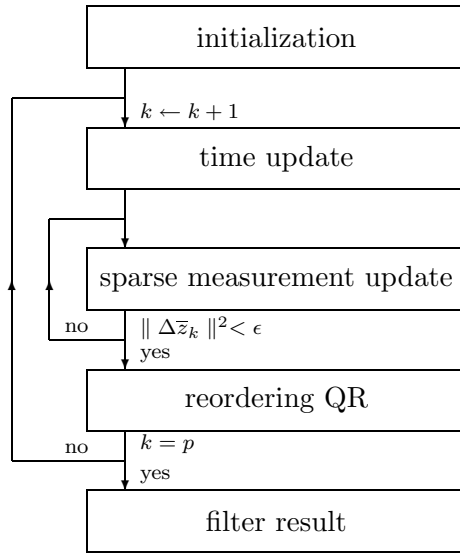
Figure 23.2   Flow chart for sparse QR decomposition.

Although this method is useful for solving $\Delta \hat{x}_k = \mathcal{P}_c\ \overline{R}_k^{-1}\ \Delta \overline{z}_k$, it has some disadvantages related to the permutation.

First, we cannot use the convergence check (see Fig. I.7)

$$\| \Delta z_{\hat{\beta} x_{k|k}} \|^2 = \Delta z'_{\hat{\beta} x_{k|k}} \Delta z_{\hat{\beta} x_{k|k}} < \epsilon, \tag{23.13}$$

because the $\Delta \overline{z}_k$ vector of Eq. (23.12) has mixed elements that are not directly related to a lower-right part of an upper-triangular matrix. Therefore, we have to check the norm of the complete $\Delta \overline{z}_k$ vector, and this implies that we have to solve for the temporal parameters as well to meet the convergence check. Because usually we do not have good approximate values for the temporal parameters, since they are hard to obtain from the previous epoch, at least a second iteration step is needed to meet the convergence criterion.

Second, in the next time update we need an unpermutated system to relate the upper-triangular matrix to the parameters. We can retain this reordering QR unpermutated system by a new reordering QR decomposition of the smaller matrix $\overline{R}_k\ \mathcal{P}'_c = Q_I R_k$:

$$\begin{aligned} \Delta \overline{z}_k &= \overline{R}_k\ \mathcal{P}'_c\ \Delta \hat{x}_k \\ \Delta \overline{z}_k &= Q_I\ R_k\ \Delta \hat{x}_k \\ \Delta z_k \doteq Q_I'\ \Delta \overline{z}_k &= \quad R_k\ \Delta \hat{x}_k \\ R_k^{-1}\ \Delta z_k &= \qquad \Delta \hat{x}_k\ . \end{aligned} \tag{23.14}$$

In other words, by using the (full) QR decomposition of the information array $[\overline{R}_k\ \mathcal{P}'_c,\ \Delta \overline{z}_k]$, we obtain the unpermutated array $[R_k,\ \Delta z_k] = QR([\overline{R}_k \mathcal{P}'_c,\ \Delta \overline{z}_k])$. This reordering QR decomposition is only needed once the convergence criterion is met; see Fig. 23.2.

Despite the disadvantages of the sparse QR decomposition, it is still preferred in terms of computation time for large systems. Table 23.4 shows some typical computation times per epoch. In this case the sparse QR decomposition is preferred in a network of five stations, whereas both methods are comparable for the baseline case. If a looser convergence criterion were used of, for example, $\epsilon = 10^{-4}$,
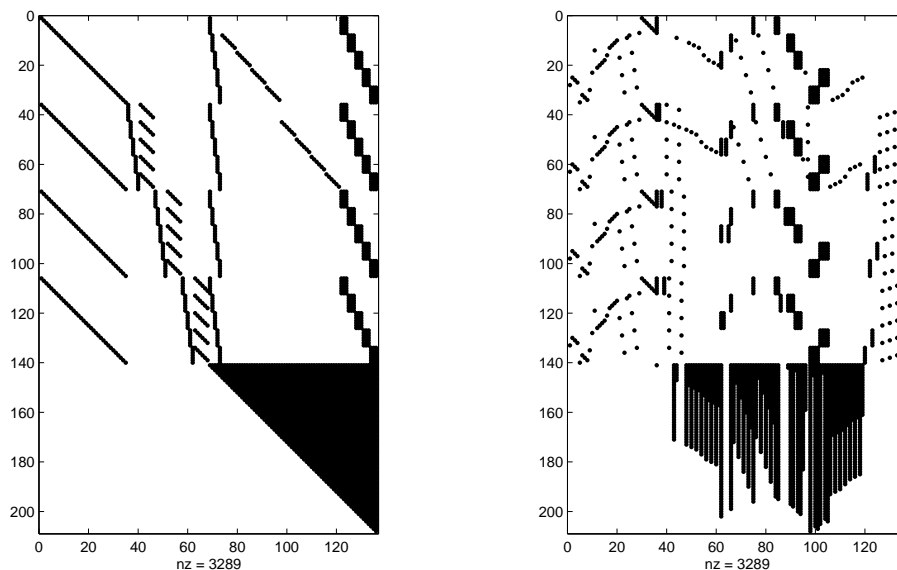
Figure 23.3   Sparsity of design matrix of Fig. 23.1. Left panel: Unpermutated design matrix for full QR decomposition. Right panel: Permutated design matrix for sparse QR decomposition.

the full decomposition would be preferable, because then one iteration suffices. In general, sparse decomposition is more attractive when the design matrix becomes larger. Clearly, in terms of computation time, the Kalman filtering has the best performance. Note that the measurement update may be one of the most time-consuming steps in the processing, but all methods have (nearly) the same overhead.

m_k	measurement update	calls	time/ call [s]	$n = 2$ time [s]	time/ call [s]	$n = 5$ time [s]
10	sparse QR	2	0.11	0.22	0.41	0.82
10	reordering QR	1	0.15	<u>0.15</u>	1.21	<u>1.21</u>
10	sparse + reordering QR			**0.37**		**2.03**
10	full QR	2	0.19	**0.38**	2.42	**4.84**
10	Kalman	2	0.01	**0.02**	0.13	**0.26**
7	sparse QR	2	0.08	0.16	0.27	0.54
7	reordering QR	1	0.09	<u>0.09</u>	0.51	<u>1.02</u>
7	sparse + reordering QR			**0.25**		**1.56**
7	full QR	2	0.11	**0.22**	0.89	**1.78**
7	Kalman	2	0.01	**0.02**	0.06	**0.13**

Table 23.4   Computation times of the measurement update of one epoch in MATLAB (version 6.1) on a Pentium III, 450 MHz PC for the troposphere-weighted model with $n = 2$ and $n = 5$ stations, and $m_k = 7$ and $m_k = 10$ satellites (for all k). Convergence criterion: $\epsilon = 10^{-6}$.

23.4   Changing satellite configurations

23.4.1   Changing the pivot

Because of changing satellite configurations, new satellites rise and old ones set, we need to change the pivot satellite of the double-differenced ambiguities once in a while. When a new satellite serves as pivot for the ambiguities, it will also serve

as pivot for the satellite clock errors. We assume that the network does not change during the observation period and that there are no missing data. Therefore the same reference station may be used as pivot for the whole time span. In principle, it does not matter which (visible) satellite is used as pivot. It may however be practical to use the satellite with the smallest zenith angle, because this satellite is likely to have the least problems observing: high signal-to-noise ratio, no blocking objects. Using the zenith angle as criterion for the choice of pivot means that pivots are usually changed when we are still observing the old pivot and when we have been observing the new pivot for some while.

Changing the pivot involves a transformation of the double-differenced ambiguities. This transformation should precede the measurement update and can either precede or follow the time update. How to draft the transformation matrix is best understood with the help of an example. Suppose we process a baseline (stations 1 and 2) and observe to five satellites[3] (1,...,5) with four L1 ambiguities using the second satellite as pivot. We now wish to transform the ambiguities so that the fourth satellite becomes pivot. The transformation then reads:

$$
\underbrace{\begin{bmatrix} \hat{N}_{12,1}^{41} \\ \hat{N}_{12,1}^{43} \\ \hat{N}_{12,1}^{42} \\ \hat{N}_{12,1}^{45} \end{bmatrix}}_{\text{new amb.}} = \underbrace{\begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}}_{\text{transf. matrix}} \underbrace{\begin{bmatrix} \hat{N}_{12,1}^{21} \\ \hat{N}_{12,1}^{23} \\ \hat{N}_{12,1}^{24} \\ \hat{N}_{12,1}^{25} \end{bmatrix}}_{\text{old amb.}}.
\tag{23.15}
$$

Clearly, we can always draft the transformation matrix in such a way that it is an identity matrix with one column replaced by a vector of -1s, since in general:

$$
\begin{aligned}
\hat{N}_{rs,l}^{oj} - \hat{N}_{rs,l}^{op} &= \hat{N}_{rs,l}^{pj}; \\
-\hat{N}_{rs,l}^{op} &= \hat{N}_{rs,l}^{po},
\end{aligned}
\tag{23.16}
$$

if p is the new pivot and o is the old pivot. The column with -1s always occurs where the ambiguity is formed between the old and new pivot. When the ambiguities are ordered, we can easily find the column for which this is the case.

Because in Eq. (23.15) the new vector of ambiguities is not ordered any more, we need to premultiply this vector by a row permutation matrix. The transformation matrix on the right-hand side of Eq. (23.15) therefore also has to be premultiplied by this permutation matrix to obtain the final transformation matrix, say $\mathcal{T}_{12,1}$. In a network with n stations where we observe on two frequencies, the transformation matrix for all ambiguities becomes: $\mathcal{T} = I_{n-1} \otimes I_2 \otimes \mathcal{T}_{12,1}$. If we denote the old ambiguities by $\hat{N}$ and the new ones by $\hat{N}^{new}$, the transformation of the vector with all global and constrained parameters looks like:

$$
\begin{bmatrix} \hat{\beta} \\ \hat{N}^{new} \\ \hat{r} \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ 0 & \mathcal{T} & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{N} \\ \hat{r} \end{bmatrix} \quad ; \quad \mathbf{D}\left\{ \begin{bmatrix} \underline{\hat{\beta}} \\ \underline{\hat{N}}^{new} \\ \underline{\hat{r}} \end{bmatrix} \right\} = \begin{bmatrix} Q_{\hat{\beta}} & \mathcal{T}Q_{\hat{\beta}\hat{N}} & Q_{\hat{\beta}\hat{r}} \\ Q_{\hat{N}\hat{\beta}}\mathcal{T}' & \mathcal{T}Q_{\hat{N}}\mathcal{T}' & Q_{\hat{N}\hat{r}}\mathcal{T}' \\ Q_{\hat{r}\hat{\beta}} & \mathcal{T}Q_{\hat{r}\hat{N}} & Q_{\hat{r}} \end{bmatrix},
\tag{23.17}
$$

where $\hat{\beta}$ is either the vector of predicted or filtered ZTDs, depending on when the transformation is applied, and $\hat{r}$ is the vector with coordinates.

---

[3]As in Sect. 10.3 of Part II we assume that the same satellites are visible for all stations.

In the SRIF approach, we do not use the (transformed) covariance matrix explicitly. We then transform the root of its inverse:

$$
\begin{bmatrix} z_{\hat{\beta}} \\ z_{\hat{N}} \\ z_{\hat{r}} \end{bmatrix} = \begin{bmatrix} R_{\hat{\beta}} & R_{\hat{\beta}\hat{N}}\mathcal{T}^{-1} & R_{\hat{\beta}\hat{r}} \\ 0 & R_{\hat{N}}\mathcal{T}^{-1} & R_{\hat{N}\hat{r}} \\ 0 & 0 & R_{\hat{r}} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \mathcal{T}\hat{N} \\ \hat{r} \end{bmatrix}.
\tag{23.18}
$$

The z vector is thus left unchanged.

23.4.2   Expiring parameters

Ambiguities that are no longer active become dormant, when they refer to satellites that are no longer visible. If one is not interested in these parameters, they may be excluded from the processing [*Wu and Muellerschoen*, 1992].

KF approach
Expiring these parameters is trivial in the KF approach. Since they occur explicitly in the observation equations, the expired parameters and their corresponding elements in the covariance matrix can be excluded in either the time or measurement update.

SRIF approach
In the SRIF approach the dormant parameters are not explicitly present in the observation equations and expiring them would mess up the upper-triangular system in which they are present. To exclude them from the processing, they need to be (among) the first in the vector of parameters, which implies a column permutation of the upper-triangular matrix, thus leaving the system intact. Although the system is now not triangular anymore, this is not a problem because after a reordering QR decomposition like in Eq. (23.14), the triangular system is regained and the top rows may now be excluded from further processing. In this reordering QR decomposition we thus combine two permutations: one to regain the parameter order and to eliminate the temporal parameters, and one to eliminate the dormant ambiguities.

If we name the second, expiring, permutation matrix $\mathcal{P}_{exp}$, the reordering QR decomposition is replaced by $Q_I\ R_k\ =\ \overline{R}_k\ \mathcal{P}_c'\ \mathcal{P}_{exp}'$, and instead of Eq. (23.14) we have:

$$
\begin{aligned}
\Delta\overline{z}_k &= \overline{R}_k\ \mathcal{P}_c'\ \mathcal{P}_{exp}{}'\ \mathcal{P}_{exp}\ \Delta\hat{\mathrm{x}}_k \\
\Delta\overline{z}_k &= \qquad Q_I\ R_k\ \mathcal{P}_{exp}\ \Delta\hat{\mathrm{x}}_k \\
\Delta z_k \doteq Q_I{}'\ \Delta\overline{z}_k &= \qquad\qquad R_k\ \mathcal{P}_{exp}\ \Delta\hat{\mathrm{x}}_k \\
R_k^{-1}\ \Delta z_k &= \qquad\qquad\quad \mathcal{P}_{exp}\ \Delta\hat{\mathrm{x}}_k\ ,
\end{aligned}
\tag{23.19}
$$

or symbolically: $[R_k,\ \Delta z_k] \leftarrow QR([\overline{R}_k\mathcal{P}_c'\mathcal{P}_{exp}',\ \Delta\overline{z}_k])$. The vector $\mathcal{P}_{exp}\ \Delta\hat{\mathrm{x}}_k$ has the temporal parameters and dormant ambiguities to be expired on top.

23.4.3   Introducing new parameters

Because every once in a while new satellites rise, new ambiguities have to be introduced. Since they occur as linear parameters in the observation equations, we do not need good approximate values. New parameters are introduced in the measurement update.

SRIF approach
The introduction is easiest in the SRIF approach. In the measurement update, we use the observation equation $z_{\hat{x}_{k-1}^{old}} = R_{\hat{x}_{k-1}^{old}}\, x^{old}$, where the old global parameters are denoted by x^{old}.

These observation equations may be extended for the new parameters as:

$$z_{\hat{x}_{k-1}^{old}} = \begin{bmatrix} R_{\hat{x}_{k-1}^{old}} & 0 \end{bmatrix} \begin{bmatrix} x^{old} \\ x^{new} \end{bmatrix}. \tag{23.20}$$

Because the complete information array also includes the matrix A_k with columns referring to the new parameters, this does not introduce a rank deficiency.

KF approach   In the KF approach, we have to consider the observation model with both old and new parameters. For example, the model with temporal and global parameters reads:

$$\mathbf{E}\left\{\begin{bmatrix} \underline{y}_k \\ \underline{\hat{x}}_{k-1}^{old} \end{bmatrix}\right\} = \begin{bmatrix} T_k & A_k^{old} & A_k^{new} \\ 0 & I & 0 \end{bmatrix} \begin{bmatrix} t_k \\ x^{old} \\ x^{new} \end{bmatrix} \;;\; \mathbf{D}\left\{\begin{bmatrix} \underline{y}_k \\ \underline{\hat{x}}_{k-1}^{old} \end{bmatrix}\right\} = \begin{bmatrix} Q_{y_k} & \\ & Q_{\hat{x}_{k-1}^{old}} \end{bmatrix}. \tag{23.21}$$

Like in Sect. 21.4, this model can be partitioned. We use the partition:

$$A_I = \begin{bmatrix} T_k \\ 0 \end{bmatrix} \;;\; A_{II} = \begin{bmatrix} A_k^{old} & A_k^{new} \\ I & 0 \end{bmatrix}. \tag{23.22}$$

After a derivation along the lines of Sect. 21.4, we find for the covariance matrix of all (old and new) global parameters:

$$Q_{\hat{x}_k} = \left(\begin{bmatrix} Q_{\hat{x}_{k-1}^{old}}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + A_k' Q_{y_k}^{-1} P_{T_k}^{\perp} A_k\right)^{-1}, \tag{23.23}$$

where $A_k \doteq [A_k^{old},\ A_k^{new}]$. The measurement update can be derived as:

$$\begin{bmatrix} \hat{x}_k^{old} \\ \hat{x}_k^{new} \end{bmatrix} = \begin{bmatrix} \hat{x}_{k-1}^{old} \\ 0 \end{bmatrix} + K_k v_k, \tag{23.24}$$

with gain matrix K_k and predicted residuals v_k:

$$\begin{aligned} K_k &= Q_{\hat{x}_k} A_k' P_{T_k}^{\perp}{}' Q_{y_k}^{-1}; \\ v_k &= y_k - A_k^{old} \hat{x}_{k-1}^{old}. \end{aligned} \tag{23.25}$$

In the actual software implementation, we would mix the old and new parameters and use an index referring to them instead of appending the new parameters. Maintaining a strict order is important for the ambiguity transformation of Sect. 23.4.1.

Condition equations, testing, and reliability

24.1 Introduction

The Kalman-filtering procedures as described in Chaps. 21 and 22 were in terms of observation equations. Temporal parameters were pre-eliminated using a projector $P_{T_k}^\perp$.

The first part of this chapter, Sect. 24.2, deals with the linear Kalman condition and observation equations. Section 24.2.1 shows the measurement-update equations of the models with global and constrained parameters only. Section 24.2.2 extends the equations with Cholesky-factor transformations to avoid matrix inversions. When including temporal parameters, a matrix $T_k^{\perp'}$ that spans the null space of T_k is needed in the condition equations. Section 24.2.3 then shows that this matrix can be used to pre-eliminate the temporal parameters instead of using the projector $P_{T_k}^{\perp'}$. This section also shows the pre-elimination in the SRIF approach. The nonlinear versions of all the estimation procedures are given in App. I.

Because the conditions in an observation model allow for testing, the second part of this chapter, Sect. 24.3, deals with testing and reliability aspects in Kalman filtering. For the SRIF implementation is referred to [*Tiberius*, 1998]. Strict procedures are worked out for the Kalman–Cholesky filtering with pre-elimination. Flow charts of these procedures can also be found in App. I.

24.2 Condition equations

24.2.1 Kalman filtering

Section 21.2 showed a model of observation equations, Eq. (21.1), for which the LSQ estimate of the observations in Eq. (21.5) was in terms of the $m \times n$ design matrix A. There is also an equivalent model of condition equations:

$$A^{\perp'} \mathbf{E}\{\underline{y}\} = 0 \quad ; \quad \mathbf{D}\{\underline{y}\} = Q_y, \tag{24.1}$$

$A^\perp$

where $A^\perp$ is the $m \times (m-n)$ orthogonal complement of A: $A^{\perp'}A = A'A^\perp = 0$ and $A \oplus A^\perp$ spans $\mathbb{R}^m$. As shown in [*Teunissen*, 1999], the projector P_A can also be written in terms of $A^\perp$: $P_A = P_{Q_y A^\perp}^\perp = I_m - P_{Q_y A^\perp}$. With Eq. (21.5) we therefore

find the following estimates for the observations in terms of $A^\perp$:

$$\begin{aligned}\hat{y} &= (I_m - Q_y A^\perp (A^{\perp'} Q_y A^\perp)^{-1} A^{\perp'}) \, y; \\ Q_{\hat{y}} &= Q_y - Q_y A^\perp (A^{\perp'} Q_y A^\perp)^{-1} A^{\perp'} Q_y.\end{aligned} \quad (24.2)$$

The model of condition equations that corresponds to the kth epoch of the recursive model with global parameters only, Eq. (21.6), reads:

global parameters

$$[I_{m_k}, \; -A_k] \, \mathbf{E}\{ \begin{bmatrix} \underline{y}_k \\ \hat{\underline{x}}_{k-1} \end{bmatrix} \} = 0 \quad ; \quad \mathbf{D}\{ \begin{bmatrix} \underline{y}_k \\ \hat{\underline{x}}_{k-1} \end{bmatrix} \} = \begin{bmatrix} Q_{y_k} & \\ & Q_{\hat{x}_{k-1}} \end{bmatrix}, \quad (24.3)$$

where $[I_{m_k}, \; -A_k]$ plays the role of $A^{\perp'}$ in Eq. (24.1), and A_k is an $m_k \times n$ matrix. Applying Eq. (24.2) then gives after some derivation the expressions on the right-hand side of Fig. 24.1. Note that we use $Q_{v_k}^{-1}$ to compute K_k. We therefore need to explicitly invert Q_{v_k} and cannot use the expression of Eq. (21.12) that is used in the observation equations.

OBSERVATION EQUATIONS　　$v_k = y_k - A_k \hat{x}_{k-1}$　　CONDITION EQUATIONS

$$\begin{aligned} Q_{\hat{x}_k} &= (Q_{\hat{x}_{k-1}}^{-1} + A_k' Q_{y_k}^{-1} A_k)^{-1} \\ K_k &= Q_{\hat{x}_k} A_k' Q_{y_k}^{-1} \\ Q_{v_k}^{-1} &= Q_{y_k}^{-1}(I_{m_k} - A_k K_k) \end{aligned} \qquad \begin{aligned} Q_{v_k}^{-1} &= (Q_{y_k} + A_k Q_{\hat{x}_{k-1}} A_k')^{-1} \\ K_k &= Q_{\hat{x}_{k-1}} A_k' Q_{v_k}^{-1} \\ Q_{\hat{x}_k} &= (I_n - K_k A_k) Q_{\hat{x}_{k-1}} \end{aligned}$$

$$\hat{x}_k = \hat{x}_{k-1} + K_k v_k$$

Figure 24.1　Kalman measurement-update equations of the model with linear global parameters only.

constrained parameters

In the model where also constrained parameters are included, the equations for the measurement update are basically the same as in Fig. 24.1; we use the predicted state vector $\hat{\beta x}_{k|k-1}$ instead of the filtered state vector $\hat{x}_{k-1}$, and design matrix BA_k instead of A_k as in Sect. 21.5; see Fig. 24.2.

OBSERVATION EQUATIONS　　$v_k = y_k - BA_k \hat{\beta x}_{k|k-1}$　　CONDITION EQUATIONS

$$\begin{aligned} Q_{\hat{\beta x}_{k|k}} &= (Q_{\hat{\beta x}_{k|k-1}}^{-1} + BA_k' Q_{y_k}^{-1} BA_k)^{-1} \\ K_k &= Q_{\hat{\beta x}_{k|k}} BA_k' Q_{y_k}^{-1} \\ Q_{v_k}^{-1} &= Q_{y_k}^{-1}(I_{m_k} - BA_k K_k) \end{aligned} \qquad \begin{aligned} Q_{v_k}^{-1} &= (Q_{y_k} + BA_k Q_{\hat{\beta x}_{k|k-1}} BA_k')^{-1} \\ K_k &= Q_{\hat{\beta x}_{k|k-1}} BA_k' Q_{v_k}^{-1} \\ Q_{\hat{\beta x}_{k|k}} &= (I_{n_\beta + n} - K_k BA_k) Q_{\hat{\beta x}_{k|k-1}} \end{aligned}$$

$$\hat{\beta x}_{k|k} = \hat{\beta x}_{k|k-1} + K_k v_k$$

Figure 24.2　Kalman measurement-update equations of the model with linear global and constrained parameters.

24.2.2　Kalman–Cholesky filtering

Kalman–Cholesky

As mentioned in Sect. 23.3.1, we may avoid inversion of the covariance matrix of the observations by computing its Cholesky factor and use it to transform the observations and design matrix. This Kalman–Cholesky approach is suitable for the model of observation equations. In the model with condition equations this Cholesky-factor transformation is of no use, because there is no Q_{y_k} to be inverted, but instead we can avoid inversion of Q_{v_k}.[1] With the Cholesky factor

---

[1]In the model of observation equations, $Q_{v_k}^{-1}$ is explicitly computed without actually inverting, whereas this is not possible in the model of condition equations.

of $Q_{v_k} = R'_{v_k} R_{v_k}$, the observations and the design matrix are transformed. Not only the estimation speed benefits from these transformations but also the speed of hypothesis testing. Figure 24.3 shows how the transformations are integrated in the estimation procedure. Note that the predicted residuals are not explicitly transformed, but new predicted residuals $\overline{v}_k \equiv (R_{y_k}^{-1})' v_k$ and $\overline{v}_k \equiv (R_{v_k}^{-1})' v_k$ are directly obtained from the transformed observations and design matrix.

<div align="center">CHOLESKY-FACTOR TRANSFORMATIONS</div>

$$R_{y_k} = Chol(Q_{y_k})$$
$$\overline{y}_k \leftarrow R'_{y_k} \backslash y_k$$
$$\overline{BA}_k \leftarrow R'_{y_k} \backslash BA_k$$

$$Q_{v_k} = Q_{y_k} + BA_k Q_{\hat{\beta} x_{k|k-1}} BA'_k$$
$$R_{v_k} = Chol(Q_{v_k})$$
$$\overline{y}_k \leftarrow R'_{v_k} \backslash y_k$$
$$\overline{BA}_k \leftarrow R'_{v_k} \backslash BA_k$$

$$\overline{v}_k = \overline{y}_k - \overline{BA}_k \hat{\beta} x_{k|k-1}$$

OBSERVATION EQUATIONS   CONDITION EQUATIONS

$$Q_{\hat{\beta} x_{k|k}} = (Q_{\hat{\beta} x_{k|k-1}}^{-1} + \overline{BA}'_k \overline{BA}_k)^{-1}$$
$$\overline{K}_k = Q_{\hat{\beta} x_{k|k}} \overline{BA}'_k$$
$$Q_{v_k}^{-1} = I_{m_k} - \overline{BA}_k \overline{K}_k$$

$$\overline{K}_k = Q_{\hat{\beta} x_{k|k-1}} \overline{BA}'_k$$
$$Q_{\hat{\beta} x_{k|k}} = (I_{n_\beta+n} - \overline{K}_k \overline{BA}_k) Q_{\hat{\beta} x_{k|k-1}}$$

$$\hat{\beta} x_{k|k} = \hat{\beta} x_{k|k-1} + \overline{K}_k \overline{v}_k$$

Figure 24.3   Kalman–Cholesky measurement update of the model with linear global and constrained parameters. Overbars are shown for clarity; they indicate a difference with the original quantities of Fig. 24.2 caused by the Cholesky-factor transformation.

temporal parameters   When temporal parameters are included, we need the orthogonal complement $T_k^\perp$ to obtain the model of condition equations. $T_k^\perp$ spans the null space of T'_k, or stated otherwise: $T_k^{\perp'} T_k = 0$, and $T_k \oplus T_k^\perp$ spans $\mathbb{R}^{m_k}$. The model of condition equations that now corresponds to the model of observation equations, Eq. (21.14), reads:

$$T_k^{\perp'}[I_{m_k}, -A_k] \mathbf{E}\left\{ \begin{bmatrix} y_k \\ \hat{\underline{x}}_{k-1} \end{bmatrix} \right\} = 0 \quad ; \quad \mathbf{D}\left\{ \begin{bmatrix} y_k \\ \hat{\underline{x}}_{k-1} \end{bmatrix} \right\} = \begin{bmatrix} Q_{y_k} & \\ & Q_{\hat{x}_{k-1}} \end{bmatrix}. \tag{24.4}$$

24.2.3   Pre-elimination transformation

Usually a null-space computation is numerically expensive. For GPS observation models however, $T_k^\perp$ can be determined analytically; see Part II. Instead of this strict model of condition equations, Eq. (24.4), we can also use $T_k^\perp$ to obtain an alternative model of observation equations:

$$\mathbf{E}\left\{ \begin{bmatrix} T_k^{\perp'} \underline{y}_k \\ \hat{\underline{x}}_{k-1} \end{bmatrix} \right\} = \begin{bmatrix} T_k^{\perp'} A_k \\ I_n \end{bmatrix} x \quad ; \quad \mathbf{D}\left\{ \begin{bmatrix} T_k^{\perp'} \underline{y}_k \\ \hat{\underline{x}}_{k-1} \end{bmatrix} \right\} = \begin{bmatrix} T_k^{\perp'} Q_{y_k} T_k^\perp & \\ & Q_{\hat{x}_{k-1}} \end{bmatrix}. \tag{24.5}$$

pre-elimination transformations   We can now recognize the same model as Eq. (21.6) if we apply the pre-elimination transformations: $y_k \leftarrow T_k^{\perp'} y_k$, $Q_{y_k} \leftarrow T_k^{\perp'} Q_{y_k} T_k^\perp$ and $A_k \leftarrow T_k^{\perp'} A_k$ (in the model with constrained parameters the last transformation becomes $BA_k \leftarrow T_k^{\perp'} BA_k$). Both approaches have the advantage over the Kalman filtering of Chap. 21 of a smaller Q_{v_k} to be inverted, since the number of observations, and therefore **$\overline{m}_k$**   also the number of predicted residuals, at epoch k reduces to $\overline{m}_k \doteq m_k - n_{t_k}$, with n_{t_k} the number of temporal parameters. On the other hand, when Q_{y_k} is

sparse, its Cholesky factor is easily computed whereas the Cholesky factor of the full covariance matrix of the transformed observations, $T_k^{\perp'}Q_{y_k}T_k^{\perp}$, is more time consuming. Combining the Cholesky and pre-elimination transformation results in the flow charts of Figs. I.8 and I.10 (for all parameter types and nonlinear global parameters).

SRIF   In the SRIF approach, the double transformation can be applied in a similar way as in the Kalman–Cholesky approach. We use the following observation model for the measurement update (all parameter types, linear global parameters):

$$
\boxed{
\begin{aligned}
\mathbf{E}\Big\{ \begin{bmatrix} (R_{y_k}^{-1})'T_k^{\perp'}\underline{y}_k \\ \underline{z}_{\hat{\beta}x_{k|k-1}} \end{bmatrix} \Big\} &= \begin{bmatrix} (R_{y_k}^{-1})'T_k^{\perp'}BA_k \\ R_{\hat{\beta}x_{k|k-1}} \end{bmatrix} \beta x_k; \\[2ex]
\mathbf{D}\Big\{ \begin{bmatrix} (R_{y_k}^{-1})'T_k^{\perp'}\underline{y}_k \\ \underline{z}_{\hat{\beta}x_{k|k-1}} \end{bmatrix} \Big\} &= \begin{bmatrix} I_{m_k-n_{t_k}} & \\ & I_{n_\beta+n} \end{bmatrix},
\end{aligned}
}
\tag{24.6}
$$

where R_{y_k} is the upper-triangular Cholesky factor of $T_k^{\perp'}Q_{y_k}T_k^{\perp} = R_{y_k}'R_{y_k}$. The advantage of the transformation in Eq. (24.6) is that a much smaller matrix is used in the QR decomposition. For the complete estimation procedure with nonlinear global parameters, see Fig. I.12.

24.2.4   Observation equations or condition equations?

Often the choice for using observation equations or condition equations is based on dimensional considerations. If the number of parameters is small compared with the number of observations, the observation equations are preferred because the covariance matrix of the parameters is easy to compute. Otherwise, especially when testing is involved, the condition equations seem more appealing. In GPS data processing there is another consideration that is more important. Whenever a new satellite is tracked, a new (global) ambiguity parameter is introduced. The observation model for epoch k is then of the form:

$$
\mathbf{E}\Big\{ \begin{bmatrix} \underline{y}_k \\ \underline{\hat{x}}_{k-1} \end{bmatrix} \Big\} = \begin{bmatrix} A_k & a_k \\ I & 0 \end{bmatrix} \begin{bmatrix} x_k^{old} \\ x_k^{new} \end{bmatrix}.
\tag{24.7}
$$

In this model we do not have an orthogonal complement of the design matrix that is in terms of A_k, like in Eq. (24.3). This orthogonal complement therefore first needs to be derived, which may be time consuming and complicates software implementation. Therefore the observation equations are preferred over the condition equations.

24.3   Testing and reliability

The concepts of testing and reliability [*Baarda*, 1968] are known for geodetic networks for some time. A recursive approach for models with time-varying parameters is given in [*Teunissen and Salzmann*, 1989], [*Teunissen*, 1990c], [*Teunissen*, 1990a], and [*Teunissen*, 1990b]. Here a recursive description is given for a GPS model which also includes global and temporal parameters.

To detect model errors we can do several hypothesis tests. Under the null hypothesis we assume the observation model to be correct. Under an alternative hypothesis we assume that something specific is wrong with the model or leave

H_0

the model error unspecified. Either way, from the model of condition equations, we see that under the null hypothesis H_0, we assume:

$$\mathrm{H}_0: \quad \mathbf{E}\{\underline{v}\} = 0 \quad ; \quad \mathbf{D}\{\underline{v}\} = Q_v, \tag{24.8}$$

H_A

where $v = [v_1', ..., v_p']'$ is the concatenated vector of predicted residuals of all epochs. In [*Teunissen and Salzmann*, 1989], a proof is given that $\mathbf{E}\{\underline{v}_k \underline{v}_l'\} = 0$, for all $k \neq l$; in other words: $Q_v = \mathrm{diag}[Q_{v_1}, ..., Q_{v_p}]$. Under an alternative hypothesis H_A, we assume the $(\sum_{\kappa=1}^p \overline{m}_\kappa) \times 1$ vector v to be contaminated with a bias caused by a model error:

$$\mathrm{H}_A: \quad \mathbf{E}\{\underline{v}\} = C_v \nabla \quad ; \quad \mathbf{D}\{\underline{v}\} = Q_v, \tag{24.9}$$

with $C_v = [C_{v_1}', ..., C_{v_p}']'$ a $(\sum_{\kappa=1}^p \overline{m}_\kappa) \times q$ full-rank matrix that specifies the location of the model error and ∇ the $q \times 1$ model error itself.

Under the alternative hypothesis we have, cf. Eq. (21.2), the estimator of the model error:

$$\hat{\underline{\nabla}} = Q_{\hat{\nabla}} C_v' Q_v^{-1} \underline{v} \quad ; \quad Q_{\hat{\nabla}}^{-1} = C_v' Q_v^{-1} C_v. \tag{24.10}$$

Because of the block-diagonal structure of Q_v, this can also be written as:

$$\hat{\underline{\nabla}} = Q_{\hat{\nabla}} \cdot \sum_{\kappa=1}^p C_{v_\kappa}' Q_{v_\kappa}^{-1} \underline{v}_\kappa \quad ; \quad Q_{\hat{\nabla}}^{-1} = \sum_{\kappa=1}^p C_{v_\kappa}' Q_{v_\kappa}^{-1} C_{v_\kappa}. \tag{24.11}$$

As shown in [*Teunissen*, 1990c], a Uniformly Most Powerful Invariant (UMPI) test for testing H_0 against H_A can be given as:

$$\text{Reject } \mathrm{H}_0 \text{ in favor of } \mathrm{H}_A \text{ if} \qquad \hat{\nabla}' Q_{\hat{\nabla}}^{-1} \hat{\nabla} > df_{1-\alpha}, \tag{24.12}$$

$\underline{\Theta}(q)$

where $df_{1-\alpha}$ is the critical value corresponding to a level of significance α and (cumulative) distribution function df. If we assume that $\underline{v}$ and, therefore also $\hat{\underline{\nabla}}$, are normally distributed, the q-dimensional test statistic

$$\underline{\Theta}(q) \doteq \hat{\underline{\nabla}}' Q_{\hat{\nabla}}^{-1} \hat{\underline{\nabla}} \tag{24.13}$$

has a χ^2 distribution:

$$\mathrm{H}_0: \quad \underline{\Theta}(q) \sim \chi^2(q, 0) \quad ; \quad \mathrm{H}_A: \quad \underline{\Theta}(q) \sim \chi^2(q, \lambda), \tag{24.14}$$

noncentrality parameter

with $\lambda = \nabla' Q_{\hat{\nabla}}^{-1} \nabla$ the noncentrality parameter. In that case the critical value $df_{1-\alpha} \equiv \chi_{1-\alpha}^2(q, 0)$ is the q-dimensional upper α probability point of the central χ^2 distribution.

We distinguish global and local tests [*Teunissen and Salzmann*, 1989]. In global tests, $C_{v_\kappa} = 0$ for $\kappa < k_0$ and $\kappa > k$, where k is the current epoch of recursive processing, and $k_0 < k$ is the first epoch some model error occurs. Then $\delta k \doteq k - k_0$ is the number of epochs before epoch k that a model error first occurred. A local test statistic is a global test statistic for which $\delta k = 0$, which means that $C_{v_\kappa} = 0$, for all $\kappa \neq k$.

We also distinguish overall model and slippage tests. In overall model tests, we assume there is a model error in one or more epochs without specifying it, whereas in slippage tests the model error is specified.

GOM, LOM   In the Global Overall Model (GOM) and Local Overall Model (LOM) tests we have: $q = \sum_{\kappa=k_0}^{k} \overline{m}_\kappa$ and $q = \overline{m}_k$ respectively. In other words, $[C'_{v_{k_0}}, ..., C'_{v_k}]'$ and C_{v_k} become square and invertible, and the GOM and LOM test statistic reduce to:

$$\boxed{\text{GOM} : \ \underline{\Theta}_k^{\delta k}(\sum_{\kappa=k_0}^{k} \overline{m}_\kappa) = \sum_{\kappa=k_0}^{k} \underline{v}'_\kappa Q_{v_\kappa}^{-1} \underline{v}_\kappa \ ; \ \text{LOM} : \ \underline{\Theta}_k^0(\overline{m}_k) = \underline{v}'_k Q_{v_k}^{-1} \underline{v}_k.} \quad (24.15)$$

recursive   The GOM test value is suitable for a recursive computation (omitting the sizes of
GOM   the tests for readability):

$$\Theta_k^{\delta k}(.) = \Theta_{k-1}^{\delta k-1}(.) + \Theta_k^0(.). \quad (24.16)$$

OMT   If $k = p$ and $\delta k = p - 1$ the most general Overall Model Test (OMT) is obtained. If we divide the χ^2-distributed quantity of Eq. (24.15) by the degrees of freedom, we obtain an F-distributed OMT. This OMT is in fact an unbiased estimate of the variance factor of unit weight of the covariance matrix of the observations if the observation model is correct; see for example [*Teunissen*, 2000]. Note that the degrees of freedom are given as the sum of the transformed observations. Strictly speaking this is only the case when no new ambiguities are introduced.[2] Whenever new ambiguities are introduced the redundancy reduces by the same amount.

GS, LS   In the Global Slippage (GS) tests and Local Slippage (LS) tests, the size of q is respectively: $q < \sum_{\kappa=k_0}^{k} \overline{m}_\kappa$ and $q < \overline{m}_k$, but usually $q = 1$. The test statistics reduce to:

$$\boxed{\text{GS} : \ \underline{\Theta}_k^{\delta k}(q) = \hat{\underline{\nabla}}_k^{\delta k'} Q_{\hat{\nabla}_k^{\delta k}}^{-1} \hat{\underline{\nabla}}_k^{\delta k} \ ; \ \text{LS} : \ \underline{\Theta}_k^0(q) = \hat{\underline{\nabla}}_k^{0'} Q_{\hat{\nabla}_k^0}^{-1} \hat{\underline{\nabla}}_k^0,} \quad (24.17)$$

where we use the following short-hand notation:

$$
\begin{aligned}
\hat{\underline{\nabla}}_k^{\delta k} &\doteq Q_{\hat{\nabla}_k^{\delta k}} \cdot \begin{bmatrix} C_{v_{k_0}} \\ \vdots \\ C_{v_k} \end{bmatrix}' \begin{bmatrix} Q_{v_{k_0}}^{-1} & & \\ & \ddots & \\ & & Q_{v_k}^{-1} \end{bmatrix} \begin{bmatrix} \underline{v}_{k_0} \\ \vdots \\ \underline{v}_k \end{bmatrix} = Q_{\hat{\nabla}_k^{\delta k}} \cdot \sum_{\kappa=k_0}^{k} C'_{v_\kappa} Q_{v_\kappa}^{-1} \underline{v}_\kappa; \\
Q_{\hat{\nabla}_k^{\delta k}}^{-1} &\doteq \begin{bmatrix} C_{v_{k_0}} \\ \vdots \\ C_{v_k} \end{bmatrix}' \begin{bmatrix} Q_{v_{k_0}}^{-1} & & \\ & \ddots & \\ & & Q_{v_k}^{-1} \end{bmatrix} \begin{bmatrix} C_{v_{k_0}} \\ \vdots \\ C_{v_k} \end{bmatrix} = \sum_{\kappa=k_0}^{k} C'_{v_\kappa} Q_{v_\kappa}^{-1} C_{v_\kappa}.
\end{aligned}
$$

$$(24.18)$$

recursive GS Like the GOM test values, the GS test values can also be obtained recursively, but the procedure is less straightforward. The overall model test statistics of Eq. (24.15) are in terms of the predicted residuals and their covariance matrix that result from the recursive estimation. The slippage test statistics, on the other hand, are still in terms of the estimator of the model error. Actual computation of the test quantities therefore seems to involve filtering with the model under the alternative hypothesis. However, a parallel processing can be implemented to estimate the model errors without much computational effort; see for example

---

[2]We assume (good) a-priori values for the coordinates, which serve as observations in the first epoch; in other words, the coordinates are both parameters and observables and therefore do not affect the redundancy.

[*Salzmann*, 1993]. In Eq. (24.9) we can recognize the model of global parameters only, cf. Eqs. (21.1) and (21.6). The matrix C_{v_k} plays the role of A_k, v_k plays the role of y_k, and $\hat{\nabla}_k^{\delta k}$ plays the role of $\hat{x}_k$. We know that in this model a recursive estimation is possible. If we use observation equations, the update equation of the as parameters occurring model errors reads:

$$\hat{\nabla}_\kappa^{\delta\kappa} = \hat{\nabla}_{\kappa-1}^{\delta\kappa-1} + K_\kappa^{\delta\kappa}(v_\kappa - C_{v_\kappa}\hat{\nabla}_{\kappa-1}^{\delta\kappa-1}), \tag{24.19}$$

with

$$K_\kappa^{\delta k} = Q_{\hat{\nabla}_\kappa^{\delta\kappa}}C'_{v_\kappa}Q_{v_\kappa}^{-1} \tag{24.20}$$

the gain vector, and

$$Q_{\hat{\nabla}_\kappa^{\delta\kappa}}^{-1} = Q_{\hat{\nabla}_{\kappa-1}^{\delta\kappa-1}}^{-1} + C'_{v_\kappa}Q_{v_\kappa}^{-1}C_{v_\kappa}. \tag{24.21}$$

The way an error in the predicted residual is specified by C_{v_κ} depends on the type of model error we assume. Three types of model errors that seem to be reasonable to test for are:
(a) an outlier in the observations;
(b) a slip in the state vector or constraints;
(c) a sensor slip.

response matrices

Figure 24.4 shows how these model errors can be propagated with C_{v_κ}, ($\kappa = k_0, ..., k$) using response matrices $X_{\kappa|\kappa-1}$ and $X_{\kappa|\kappa}$. Because these matrices depend on the epoch the first model error occurred, we use the superscripts $\delta\kappa$ ($C_{v_\kappa}^{\delta\kappa}$, $X_{\kappa|\kappa-1}^{\delta\kappa}$, and $X_{\kappa|\kappa}^{\delta\kappa}$). The matrices are obtained from the recursive formulas:

$$\begin{aligned}
\hat{\beta}x_{\kappa|\kappa-1} &= \Psi_{\kappa,\kappa-1}\hat{\beta}x_{\kappa-1|\kappa-1} + \begin{bmatrix} d_k \\ 0 \end{bmatrix} & \text{(time update)}; \\
v_\kappa &= y_\kappa - BA_\kappa\hat{\beta}x_{\kappa|\kappa-1} & \text{(predicted residuals)}; \\
\hat{\beta}x_{\kappa|\kappa} &= \hat{\beta}x_{\kappa|\kappa-1} + K_\kappa v_\kappa & \text{(measurement update)}.
\end{aligned} \tag{24.22}$$
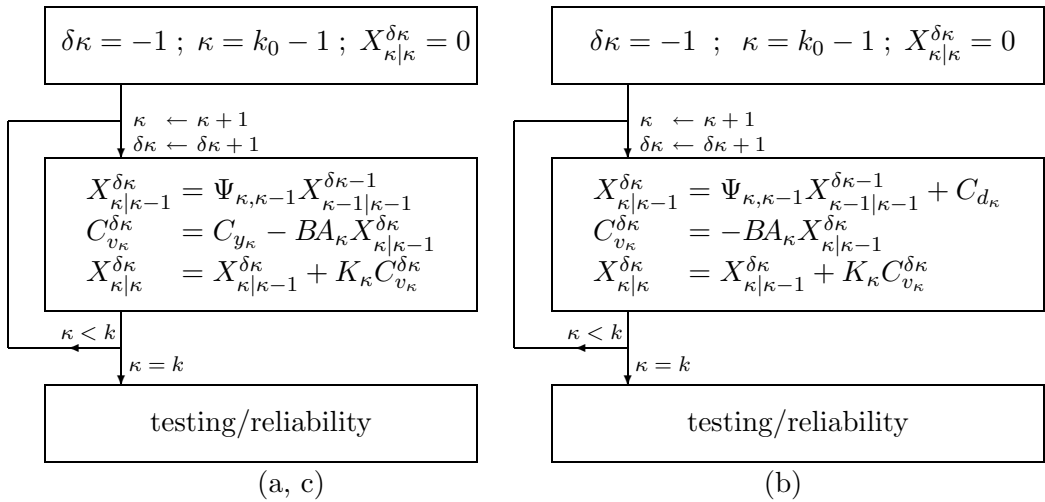
Figure 24.4   Recursive update of $C_{v_\kappa}^{\delta\kappa}$ and response matrices $X_{\kappa|\kappa-1}^{\delta\kappa}$ and $X_{\kappa|\kappa}^{\delta\kappa}$ for three different types of tests. (a) Outlier in observation; $C_{y_\kappa} = 0$ for $\kappa \neq k_0$, and $C_{y_\kappa} \neq 0$ for $\kappa = k_0$. (b) Slip in state vector; $C_{d_\kappa} = 0$ for $\kappa < k_0$, and $C_{d_\kappa} \neq 0$ for $\kappa \geq k_0$. (c) A sensor slip: $C_{y_\kappa} = 0$ for $\kappa < k_0$, and $C_{y_\kappa} \neq 0$ for $\kappa \geq k_0$.

one dimensional

Note that when all of the model errors are one dimensional, $C_{v_\kappa}^{\delta\kappa}$ and the response

matrices are vectors, and $\hat{\nabla}_\kappa^{\delta\kappa}$ and $Q_{\hat{\nabla}_\kappa^{\delta\kappa}}$ are scalars. The one-dimensional tests are relatively easy to compute. Moreover, we may assume there is an outlier in just one of the (pseudo-)observations or a slip in just one of the state-vector elements. So for each of the (pseudo-)observations and state-vector elements we formulate an alternative hypothesis. In other words, we have m_{k_0} tests for outliers (a), and $n_\beta + n$ tests for slips (b). Although in the latter case the most reasonable tests seem to concern the n global parameters, namely a slip in one of the ambiguities or coordinates, the n_β tests for slips in the troposphere constraints may detect errors in the stochastic model, more specifically whether a correct Power Spectral Density is used. We have characteristic matrices C_{y_κ} and C_{d_κ} of the form $[0, ..., 0, 1, 0, ..., 0]'$, where the position of the '1' corresponds with the index of a (pseudo-)observation or state-vector element. This way of testing **data snooping** is known as data snooping, although this usually only concerns the outlier tests (a).

B-method A method that has become more and more common in geodetic practice is the B-method of testing [*Baarda*, 1968]. This procedure can also be implemented in a recursive data processing approach [*Teunissen and Salzmann*, 1989]. In the B-method we choose a level of significance α_1 and power γ for the one-dimensional tests, and determine the corresponding noncentrality parameter λ_0. If the $\underline{v}_k$ are normally distributed (as assumed), then:

$$\sqrt{\lambda_0} = N_{1-\frac{1}{2}\alpha_1} + N_\gamma, \tag{24.23}$$

where $N_{1-\frac{1}{2}\alpha_1}$ and N_γ are (positive) upper probability points of the standard normal distribution. Often, $\alpha_1 = 0.001$ and $\gamma = 0.8$ are used, which corresponds to $\lambda_0 = 17.0746$. In the B-method all (q-dimensional) tests use the same noncentrality parameter and power. The critical value for a q-dimensional test follows from the noncentral χ^2 distribution:

$$\chi^2_{1-\gamma}(q, \lambda_0) = \chi^2_{1-\alpha_q}(q, 0). \tag{24.24}$$

As a consequence, the level of significance α_q, which follows from the central χ^2 distribution, differs for different q.

A nice property of the one-dimensional tests is that the size of a one-dimensional model error ∇ that can just be detected by this test, can be computed uniquely from $\lambda_0 = \nabla' C_v' Q_v^{-1} C_v \nabla$, using Eq. (24.10):

$$\boxed{|\nabla| = \left(\frac{\lambda_0}{C_v' Q_v^{-1} C_v} \right)^{\frac{1}{2}} = \sqrt{\lambda_0} \cdot \sigma_{\hat{\nabla}}.} \tag{24.25}$$

MDB This $|\nabla|$ is also known as Minimal Detectable Bias (MDB) and is a measure **reliability** of internal reliability. External reliability describes the impact of a model error (MDB) on (functions of) the state-vector elements. Figure 24.5 shows how the MDBs and external reliability are computed recursively along with the GS test quantity, using both the Cholesky and pre-elimination transformations from the previous section.

At any epoch k we can test for model errors that started occurring at any epoch $k_0 < k$. By actually doing these tests, the total number of (global) tests would grow quadratically with the number of epochs. Instead, it was proposed in **windowing** [*Teunissen and Salzmann*, 1989] to use a windowing technique, in which we only test up to a fixed latency $\delta k = k - k_0$. The idea is that after this latency the

INITIALIZATION

$$\delta\kappa \quad = -1 \ ; \ \kappa \quad = k_0 - 1$$
$$X^{\delta\kappa}_{\kappa|\kappa} = 0$$
$$Q^{-1}_{\hat{\nabla}^{\delta\kappa}_{\kappa}} = 0 \quad ; \ \hat{\nabla}^{\delta\kappa}_{\kappa} = 0$$

$$\kappa \ \leftarrow \kappa + 1$$
$$\delta\kappa \leftarrow \delta\kappa + 1$$

RECURSION OBS. EQ.

$$\Theta^0_\kappa(.) \quad = \overline{v}'_\kappa Q^{-1}_{\overline{v}_k} \overline{v}_\kappa$$
$$\Theta^{\delta\kappa}_\kappa(.) \quad = \Theta^{\delta\kappa-1}_{\kappa-1}(.) + \Theta^0_\kappa(.)$$

$$X^{\delta\kappa}_{\kappa|\kappa-1} = \Psi_{\kappa,\kappa-1} X^{\delta\kappa-1}_{\kappa-1|\kappa-1}$$
$$C_{\overline{y}_\kappa} \quad \leftarrow R'_{y_\kappa} \backslash C_{y_\kappa}$$
$$C^{\delta\kappa}_{\overline{v}_\kappa} \quad = C_{\overline{y}_\kappa} - \overline{BA}_\kappa X^{\delta\kappa}_{\kappa|\kappa-1}$$
$$X^{\delta\kappa}_{\kappa|\kappa} \quad = X^{\delta\kappa}_{\kappa|\kappa-1} + \overline{K}_\kappa C^{\delta\kappa}_{\overline{v}_\kappa}$$
$$Q^{-1}_{\hat{\nabla}^{\delta\kappa}_\kappa} = Q^{-1}_{\hat{\nabla}^{\delta\kappa-1}_{\kappa-1}} + C^{\delta\kappa'}_{\overline{v}_\kappa} Q^{-1}_{\overline{v}_\kappa} C^{\delta\kappa}_{\overline{v}_\kappa}$$
$$\overline{K}^{\delta\kappa}_\kappa \quad = Q_{\hat{\nabla}^{\delta\kappa}_\kappa} C^{\delta\kappa'}_{\overline{v}_\kappa} Q^{-1}_{\overline{v}_\kappa}$$
$$\hat{\nabla}^{\delta\kappa}_\kappa \quad = \hat{\nabla}^{\delta\kappa-1}_{\kappa-1} +$$
$$\overline{K}^{\delta\kappa}_\kappa (\overline{v}_\kappa - C^{\delta\kappa}_{\overline{v}_\kappa} \hat{\nabla}^{\delta\kappa-1}_{\kappa-1})$$

$$\kappa \ \leftarrow \kappa + 1$$
$$\delta\kappa \leftarrow \delta\kappa + 1$$

RECURSION COND. EQ.

$$\Theta^0_\kappa(.) \quad = \overline{v}'_\kappa \overline{v}_\kappa$$
$$\Theta^{\delta\kappa}_\kappa(.) \quad = \Theta^{\delta\kappa-1}_{\kappa-1}(.) + \Theta^0_\kappa(.)$$

$$X^{\delta\kappa}_{\kappa|\kappa-1} = \Psi_{\kappa,\kappa-1} X^{\delta\kappa-1}_{\kappa-1|\kappa-1}$$
$$C_{\overline{y}_\kappa} \quad \leftarrow R'_{v_\kappa} \backslash C_{y_\kappa}$$
$$C^{\delta\kappa}_{\overline{v}_\kappa} \quad = C_{\overline{y}_\kappa} - \overline{BA}_\kappa X^{\delta\kappa}_{\kappa|\kappa-1}$$
$$X^{\delta\kappa}_{\kappa|\kappa} \quad = X^{\delta\kappa}_{\kappa|\kappa-1} + \overline{K}_\kappa C^{\delta\kappa}_{\overline{v}_\kappa}$$
$$Q^{-1}_{\hat{\nabla}^{\delta\kappa}_\kappa} = Q^{-1}_{\hat{\nabla}^{\delta\kappa-1}_{\kappa-1}} + C^{\delta\kappa'}_{\overline{v}_\kappa} C^{\delta\kappa}_{\overline{v}_\kappa}$$
$$\overline{K}^{\delta\kappa}_\kappa \quad = Q_{\hat{\nabla}^{\delta\kappa}_\kappa} C^{\delta\kappa'}_{\overline{v}_\kappa}$$
$$\hat{\nabla}^{\delta\kappa}_\kappa \quad = \hat{\nabla}^{\delta\kappa-1}_{\kappa-1} +$$
$$\overline{K}^{\delta\kappa}_\kappa (\overline{v}_\kappa - C^{\delta\kappa}_{\overline{v}_\kappa} \hat{\nabla}^{\delta\kappa-1}_{\kappa-1})$$

$$\kappa < k \qquad \qquad \kappa < k$$
$$\kappa = k \quad \text{TESTING/RELIABILITY} \quad \kappa = k$$

$$\Theta^{\delta k}_k(.) \quad \equiv \Theta^{\delta\kappa}_\kappa(.)$$
$$\Theta^{\delta k}_k(1) \quad = (\hat{\nabla}^{\delta k}_k / \sigma_{\hat{\nabla}^{\delta k}_k})^2$$
$$|\nabla^{\delta k}_k| \quad = \sqrt{\lambda_0} \cdot \sigma_{\hat{\nabla}^{\delta k}_k}$$
$$\nabla\hat{\beta} x^{\delta k}_{k|k} = X^{\delta k}_{k|k} \cdot |\nabla^{\delta k}_k|$$
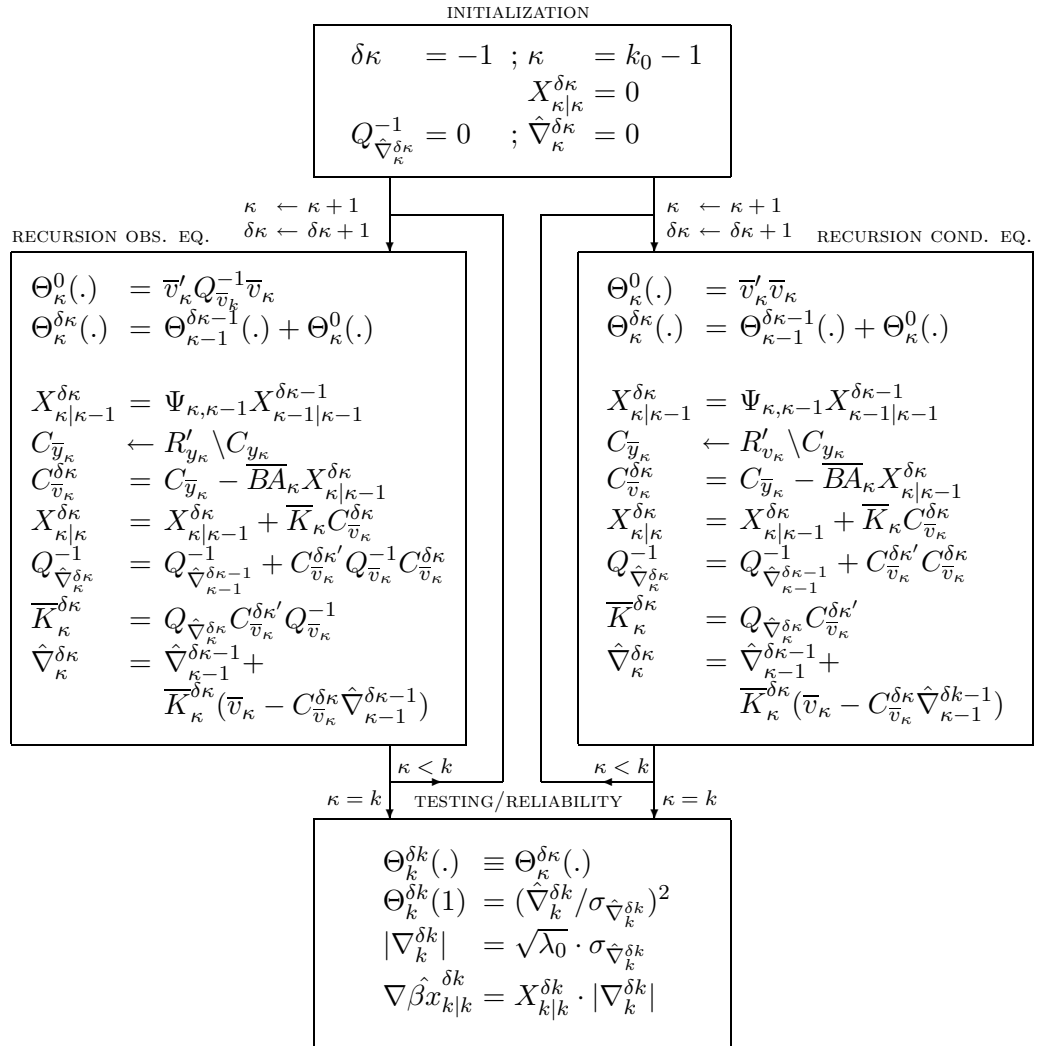
Figure 24.5   Recursive computation of a GOM test quantity $\Theta^{\delta k}_k(.)$, a one-dimensional GS test quantity $\Theta^{\delta k}_k(1)$, MDB $|\nabla^{\delta k}_k|$, and external reliability $\nabla\hat{\beta} x^{\delta k}_{k|k}$ in the model with linear global and constrained parameters. The left panel shows the procedure for the model of observation equations where the observations and design matrix are transformed with the Cholesky factor R_{y_κ} of the covariance matrix of the observations. The new predicted residuals are $\overline{v}_\kappa \equiv (R^{-1}_{y_\kappa})' v_\kappa$. $Q^{-1}_{\overline{v}_\kappa}$ is computed and used explicitly. The panel on the right shows the implementation for the condition equations where we use Cholesky-factor transformed predicted residuals $\overline{v}_\kappa \equiv (R^{-1}_{v_\kappa})' v_\kappa$; $Q_{\overline{v}_\kappa} = I_{m_k}$. A specific example of outlier tests (a) is given.

increase in detection power for model errors slows down rapidly. The number of tests then only grows linear with the number of epochs.

The flow chart of Fig. 24.5 shows the recursion from a theoretic point of view. We start at some epoch k_0, let the epoch counter κ run to epoch k, and compute the GS test quantity and reliability measures at this epoch. In a practical implementation however, we need to do all computations at one specific epoch k; we compute the GS test quantity and reliability measures with latency δk, but also start the recursion for testing at δk epochs later, and continue the recursion for the epochs between epochs k and $k + \delta k$. The flow charts of Figs. I.9 and I.11 show possible implementations where pre-elimination transformations

are applied in the model with nonlinear global parameters. The running latency counter $\delta\kappa$ runs from the fixed δk down to and including 0. Although this reverse order is not strictly necessary, it assures that an in-place computation technique is possible. For example: $\hat{\nabla}_k^{\delta\kappa}$ replaces $\hat{\nabla}_{k-1}^{\delta\kappa-1}$ in memory.

Chapter 25

Conclusions of Part IV

This part deals with recursive estimation and testing for GPS observation models as given in Parts II and III. Recursive techniques are considered useful for analyzing the performance of observation models in the design stage, because estimates are obtained after each new epoch of data. Two filters are described: the recursive Square Root Information Filter (SRIF) and the Kalman Filter (KF). The latter one can be formulated in observation equations and in condition equations. In both the SRIF and KF, we can apply Cholesky-factor and pre-elimination transformations.

A transformation of observations and design matrices with the Cholesky factor of the covariance matrix of the observations is a natural part of the estimation procedure in the SRIF. In the KF with observation equations this is a computation-time reducing technique. The KF with condition equations benefits from another Cholesky-factor transformation: with the Cholesky factor of the covariance matrix of the predicted residuals.

In the most general observation models, like the troposphere-weighted models, three types of parameters are present: global, constrained, and temporal parameters. Because temporal parameters are considered nuisance parameters, different pre-elimination methods are considered. In the SRIF we can solve for part of a triangular system of equations. In the KF with observation equations we can use a projector on the orthogonal subspace of the temporal parameters. In both the SRIF and the KF, pre-elimination is however more straightforward using the orthogonal subspace itself.[1] For the latter method, testing procedures for the KF are given as well as reliability measures that follow from these procedures. An overview of methods and transformations as dealt with in this part are given in Table 25.1.

state of the art and contribution   The Kalman Filter and SRIF are known filters which have been used for GPS data processing for some time. Quality control for recursive models with time-changing (batch) parameters has also been described before. The specific design of the filters in this part are by the author. They combine techniques to deal with (global and) temporal parameters, Cholesky-factor transformations and implementation aspects. The reliability measures that can be obtained with these filters are also by the author.

Now that filter procedures are available, we can build them into software and analyze the results for different observation scenarios. This is done in the next part, where we come back to the questions of Chap. 15.

---

[1] In the ionosphere-float model with both satellite and receiver clocks as temporal parameters, this boils down to forming linear inter-frequency combinations and double differencing; see Part II.

method	pre-elimination transformation	testing	Cholesky-factor transformation
Kalman, obs. eqs.	$P_{T_k}^{\perp}$ (I.6)	no	no
SRIF	solve part of triangular system (I.7)	no	no (identity matrix)
Kalman, obs. eqs.	$T_k^{\perp}$ (I.8)	yes (I.9)	yes
Kalman, cond. eqs.	$T_k^{\perp}$ (I.10)	yes (I.11)	yes
SRIF	$T_k^{\perp}$ (I.12)	no	yes

Table 25.1   Main procedures as described in this part. All procedures involve three types of parameters: global, constrained, and temporal parameters. T_k: partial design matrix of temporal parameters; $T_k^{\perp}$: space orthogonal to T_k; $P_{T_k}^{\perp}$: projector on space orthogonal to T_k. Between parentheses are shown the figure numbers of flow charts.

Appendix H

Matrix lemmas

The inverse of a regular and square partitioned matrix reads:

$$
\begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1} = \begin{bmatrix} (N_{11} - N_{12}N_{22}^{-1}N_{21})^{-1}; & -N_{11}^{-1}N_{12}(N_{22} - N_{21}N_{11}^{-1}N_{12})^{-1} \\ -N_{22}^{-1}N_{21}(N_{11} - N_{12}N_{22}^{-1}N_{21})^{-1}; & (N_{22} - N_{21}N_{11}^{-1}N_{12})^{-1} \end{bmatrix},
$$
(H.1)

where

$$
(N_{11} - N_{12}N_{22}^{-1}N_{21})^{-1} = N_{11}^{-1} + N_{11}^{-1}N_{12}(N_{22} - N_{21}N_{11}^{-1}N_{12})^{-1}N_{21}N_{11}^{-1}. \quad \text{(H.2)}
$$

Proof: see, for example: [*Koch*, 1987].

Flow charts

$$k = 0 \; ; \; \hat{\beta} x_{0|0} \; ; \; Q_{\hat{\beta} x_{0|0}}$$

$k \leftarrow k + 1$

TIME UPDATE

$$\hat{\beta} x_{k|k-1} = \Psi_{k,k-1} \, \hat{\beta} x_{k-1|k-1}$$

$$Q_{\hat{\beta} x_{k|k-1}} = \Psi_{k,k-1} Q_{\hat{\beta}_{k-1|k-1}} \Psi'_{k,k-1} + \begin{bmatrix} Q_{d_k} & 0 \\ 0 & 0 \end{bmatrix}$$

MEASUREMENT UPDATE

$$v_k = y_k - BA_k \, \hat{\beta} x_{k|k-1}$$

$$Q_{\hat{\beta} x_{k|k}} = (Q_{\hat{\beta} x_{k|k-1}}^{-1} + BA'_k Q_{y_k}^{-1} P_{T_k}^{\perp} BA_k)^{-1}$$

$$K_k = Q_{\hat{\beta} x_{k|k}} BA'_k P_{T_k}^{\perp'} Q_{y_k}^{-1}$$

$$\hat{\beta} x_{k|k} = \hat{\beta} x_{k|k-1} + K_k v_k$$

$k < p$

$k = p$

FILTER RESULT

$$\hat{x}_p \; ; \; Q_{\hat{x}_p}$$
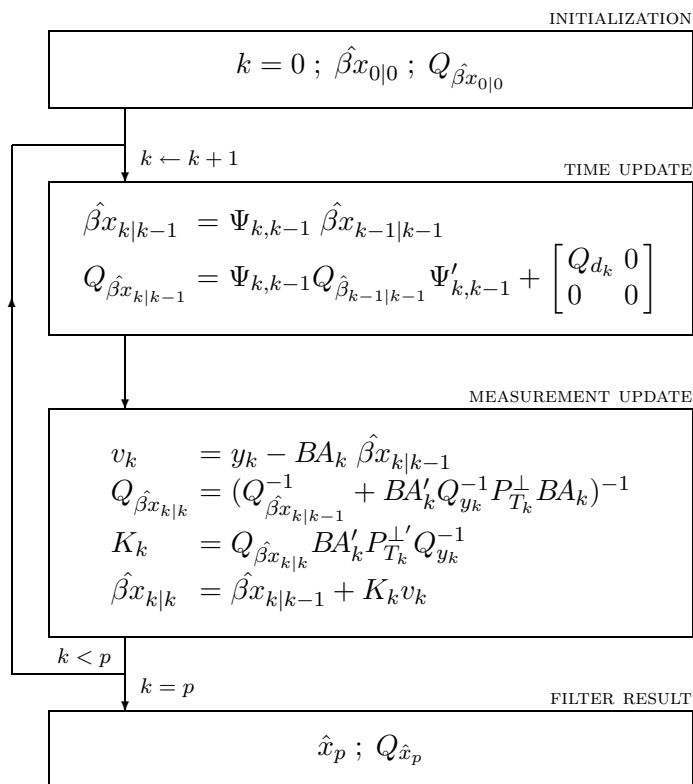
Figure I.1   Recursive KF procedure corresponding with the linear model with constrained and temporal parameters.

$$k = 0 \; ; \; x^0 \leftarrow \hat{x}_0 \; ; \; Q_{\hat{x}_0}$$

$$k \leftarrow k + 1$$

MEASUREMENT UPDATE

$$
\begin{aligned}
\Delta y_k &= y_k - A_k(x^0) \\
\Delta \hat{x}_{k-1} &= \hat{x}_{k-1} - x^0 \\
v_k &= \Delta y_k - \partial_x A_k(x^0) \Delta \hat{x}_{k-1} \\
Q_{\hat{x}_k} &= (Q_{\hat{x}_{k-1}}^{-1} + \partial_x A_k(x^0)' Q_{y_k}^{-1} \partial_x A_k(x^0))^{-1} \\
K_k &= Q_{\hat{x}_k} \partial_x A_k(x^0)' Q_{y_k}^{-1} \\
\Delta \hat{x}_k &= \Delta \hat{x}_{k-1} + K_k v_k \\
x^0 &\leftarrow x^0 + \Delta \hat{x}_k
\end{aligned}
$$

no $\qquad \| \Delta \hat{x}_k \|_{Q_{\hat{x}_k}}^2 < \epsilon$

$k < p \qquad$ yes

$k = p \qquad$ FILTER RESULT

$$\hat{x}_p \leftarrow x^0 \; ; \; Q_{\hat{x}_p}$$
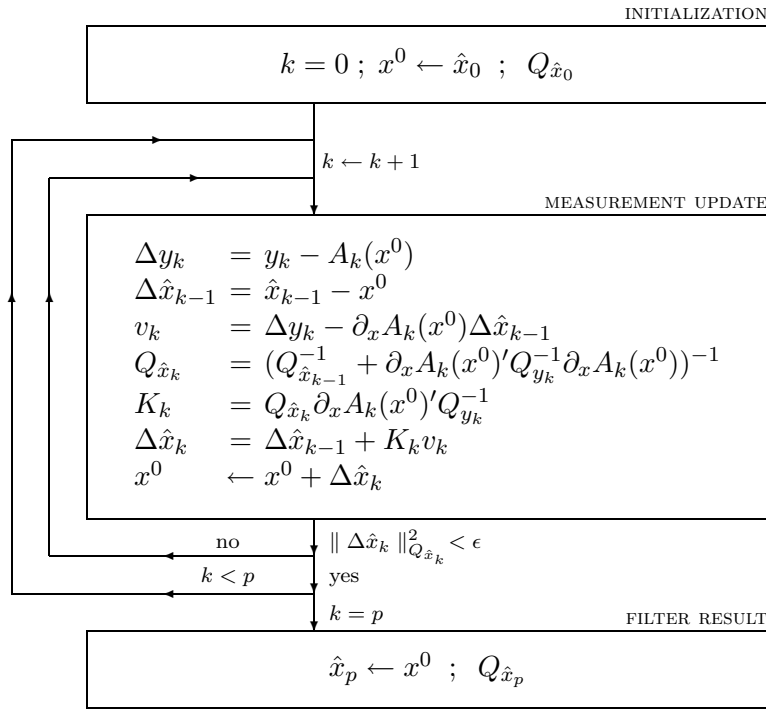
Figure I.2   Recursive KF algorithm for the model with nonlinear global parameters.

INITIALIZATION

$$k = 0 \; ; \; x^0 \leftarrow \hat{x}_0 \; ; \; [R_{\hat{x}_0}, \; z_{\hat{x}_0}]$$

$$k \leftarrow k + 1$$

MEASUREMENT UPDATE

$$
\begin{aligned}
\Delta y_k &= y_k - A_k(x^0) \\
\Delta z_{\hat{x}_{k-1}} &= z_{\hat{x}_{k-1}} - R_{\hat{x}_{k-1}} x^0
\end{aligned}
$$

$$
\begin{bmatrix} R_{\hat{x}_k} & \Delta z_{\hat{x}_k} \\ 0 & \Delta \hat{\rho}_k \end{bmatrix} = QR\left(\begin{bmatrix} \partial_x A_k(x^0) & \Delta y_k \\ R_{\hat{x}_{k-1}} & \Delta z_{\hat{x}_{k-1}} \end{bmatrix} \right)
$$

$$
\begin{aligned}
\Delta \hat{x}_k &= R_{\hat{x}_k} \backslash \Delta z_{\hat{x}_k} \\
x^0 &\leftarrow x^0 + \Delta \hat{x}_k \\
z_{\hat{x}_k} &= R_{\hat{x}_k} x^0
\end{aligned}
$$

no $\qquad \| \Delta z_{\hat{x}_k} \|^2 < \epsilon$

$k < p \qquad$ yes

$k = p \qquad$ FILTER RESULT

$$\hat{x}_p \leftarrow x^0 \; ; \; Q_{\hat{x}_p} = R_{\hat{x}_p}^{-1} (R_{\hat{x}_p}^{-1})'$$
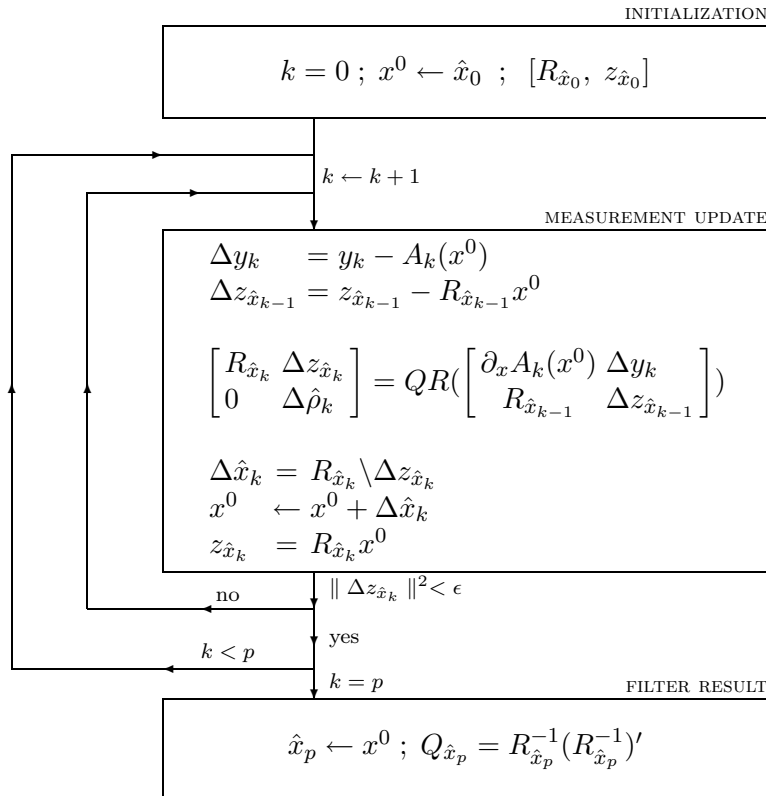
Figure I.3   Recursive SRIF algorithm for the model with nonlinear global parameters. $Q_{y_k} = I_{m_k}$ assumed.
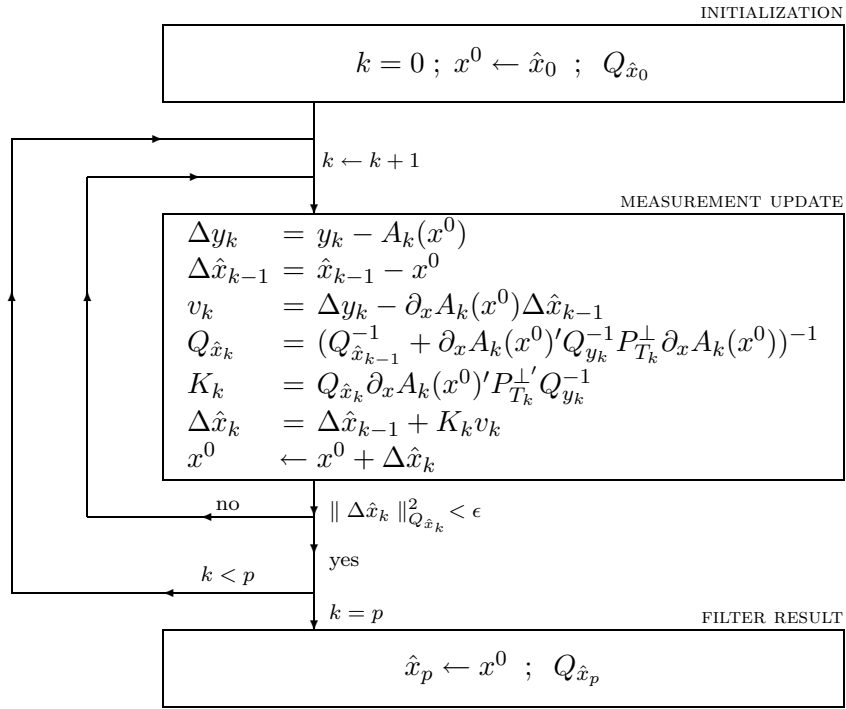
Figure I.4   Recursive KF algorithm for the model with nonlinear global parameters and linear temporal parameters.
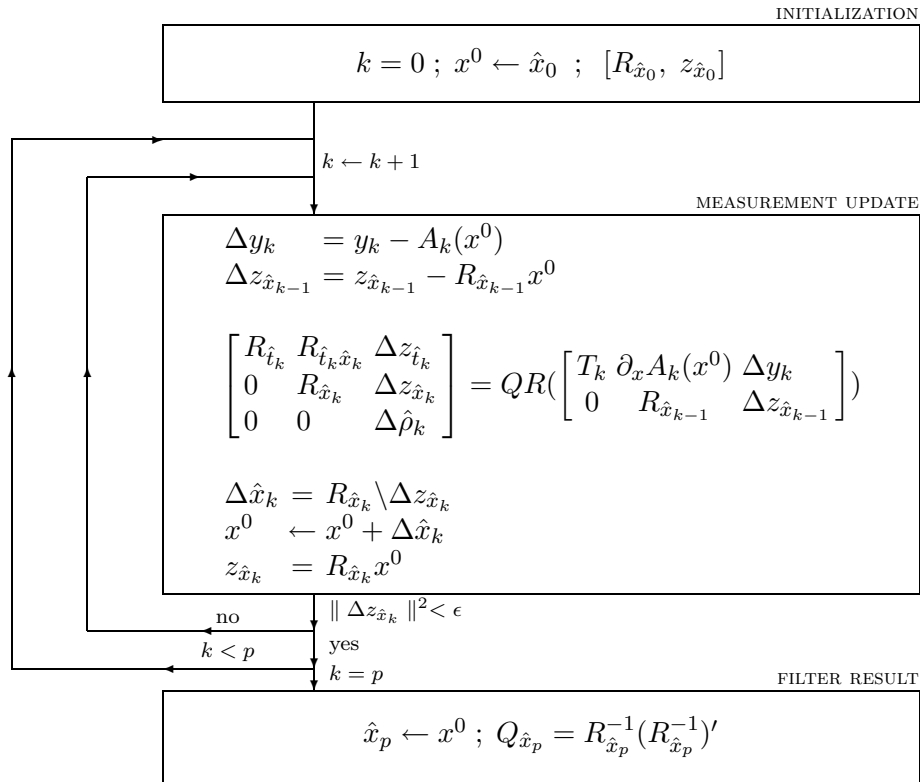
Figure I.5   Recursive SRIF algorithm for the model with nonlinear global parameters and linear temporal parameters. $Q_{y_k} = I_{m_k}$ assumed.

$$k = 0 \; ; \; \beta x^0 \leftarrow \hat{\beta x}_{0|0} \; ; \; Q_{\hat{\beta x}_{0|0}}$$

$k \leftarrow k + 1$

$$\hat{\beta x}_{k|k-1} = \Psi_{k,k-1}\hat{\beta x}_{k-1|k-1} = \beta x^0$$

$$Q_{\hat{\beta x}_{k|k-1}} = \Psi_{k,k-1}Q_{\hat{\beta x}_{k-1|k-1}}\Psi'_{k,k-1} + \begin{bmatrix} Q_{d_k} & 0 \\ 0 & 0 \end{bmatrix}$$

$$\begin{aligned}
\Delta y_k &= y_k - BA_k(\beta x^0) \\
\Delta\hat{\beta x}_{k|k-1} &= \hat{\beta x}_{k|k-1} - \beta x^0 \\
v_k &= \Delta y_k - \partial_{\beta x}BA_k(\beta x^0)\,\Delta\hat{\beta x}_{k|k-1} \\
Q_{\hat{\beta x}_{k|k}} &= (Q_{\hat{\beta x}_{k|k-1}}^{-1} + \\
&\quad \partial_{\beta x}BA_k(\beta x^0)'\,Q_{y_k}^{-1}\,P_{T_k}^{\perp}\,\partial_{\beta x}BA_k(\beta x^0))^{-1} \\
K_k &= Q_{\hat{\beta x}_{k|k}}\partial_{\beta x}BA'_k(\beta x^0)'P_{T_k}^{\perp'}Q_{y_k}^{-1} \\
\Delta\hat{\beta x}_{k|k} &= \Delta\hat{\beta x}_{k|k-1} + K_k\,v_k \\
\beta x^0 &\leftarrow \beta x^0 + \Delta\hat{\beta x}_{k|k}
\end{aligned}$$

no $\| \Delta\hat{\beta x}_{k|k} \|^2_{Q_{\hat{\beta x}_{k|k}}} < \epsilon$

yes

$k < p$

$k = p$

$$\hat{x}_p \leftarrow x^0 \; ; \; Q_{\hat{x}_p}$$

Figure I.6   Recursive KF procedure corresponding with the model with nonlinear global parameters and linear constrained and temporal parameters.

$$k = 0 \; ; \; \beta x^0 \leftarrow \hat{\beta} x_{0|0} \; ; \; \left[R_{\hat{\beta} x_{0|0}} \; z_{\hat{\beta} x_{0|0}} \right]$$

$k \leftarrow k + 1$

$$\begin{bmatrix} R_{\hat{\beta}_{k-1|k-1}} & R_{\hat{\beta}_{k-1|k-1}\hat{\beta}_{k|k-1}} & R_{\hat{\beta}_{k-1|k-1}\hat{x}_{k|k-1}} & z_{\hat{\beta}_{k-1|k-1}} \\ 0 & R_{\hat{\beta}_{k|k-1}} & R_{\hat{\beta}_{k|k-1}\hat{x}_{k|k-1}} & z_{\hat{\beta}_{k|k-1}} \\ 0 & 0 & R_{\hat{x}_{k|k-1}} & z_{\hat{x}_{k|k-1}} \end{bmatrix}$$

$$= QR(\begin{bmatrix} -\Phi_{k,k-1} & I_{n_\beta} & 0 & d_k = 0 \\ R_{\hat{\beta}_{k-1|k-1}} & 0 & R_{\hat{\beta}_{k-1|k-1}\hat{x}_{k-1|k-1}} & z_{\hat{\beta}_{k-1|k-1}} \\ 0 & 0 & R_{\hat{x}_{k-1|k-1}} & z_{\hat{x}_{k-1|k-1}} \end{bmatrix})$$

$$\Delta y_k \qquad = y_k - BA_k(\beta x^0)$$
$$\Delta z_{\hat{\beta} x_{k|k-1}} = z_{\hat{\beta} x_{k|k-1}} - R_{\hat{\beta} x_{k|k-1}} \beta x^0$$

$$\begin{bmatrix} R_{\hat{t}_{k|k}} & R_{\hat{t}_{k|k}\hat{\beta} x_{k|k}} & \Delta z_{\hat{t}_{k|k}} \\ 0 & R_{\hat{\beta} x_{k|k}} & \Delta z_{\hat{\beta} x_{k|k}} \\ 0 & 0 & \Delta \hat{\rho}_{k|k} \end{bmatrix} = QR(\begin{bmatrix} T_k & \partial_{\beta x} BA_k(\beta x^0) & \Delta y_k \\ 0 & R_{\hat{\beta} x_{k|k-1}} & \Delta z_{\hat{\beta} x_{k|k-1}} \end{bmatrix})$$

$$\Delta \hat{\beta} x_{k|k} = R_{\hat{\beta} x_{k|k}} \backslash \Delta z_{\hat{\beta} x_{k|k}}$$
$$\beta x^0 \qquad \leftarrow \beta x^0 + \Delta \hat{\beta} x_{k|k}$$
$$z_{\hat{\beta} x_{k|k}} \quad = R_{\hat{\beta} x_{k|k}} \beta x^0$$

$\| \Delta z_{\hat{\beta} x_{k|k}} \|^2 < \epsilon$

no

yes

$k < p$

$k = p$

$$\hat{x}_p \leftarrow x^0 \; ; \; Q_{\hat{x}_p} = R_{\hat{x}_p}^{-1} (R_{\hat{x}_p}^{-1})'$$

Figure I.7   Recursive SRIF algorithm for the model with nonlinear global parameters and linear constrained and temporal parameters. $Q_{y_k} = I_{m_k}$ and $Q_{d_k} = I_{n_\beta}$ assumed.
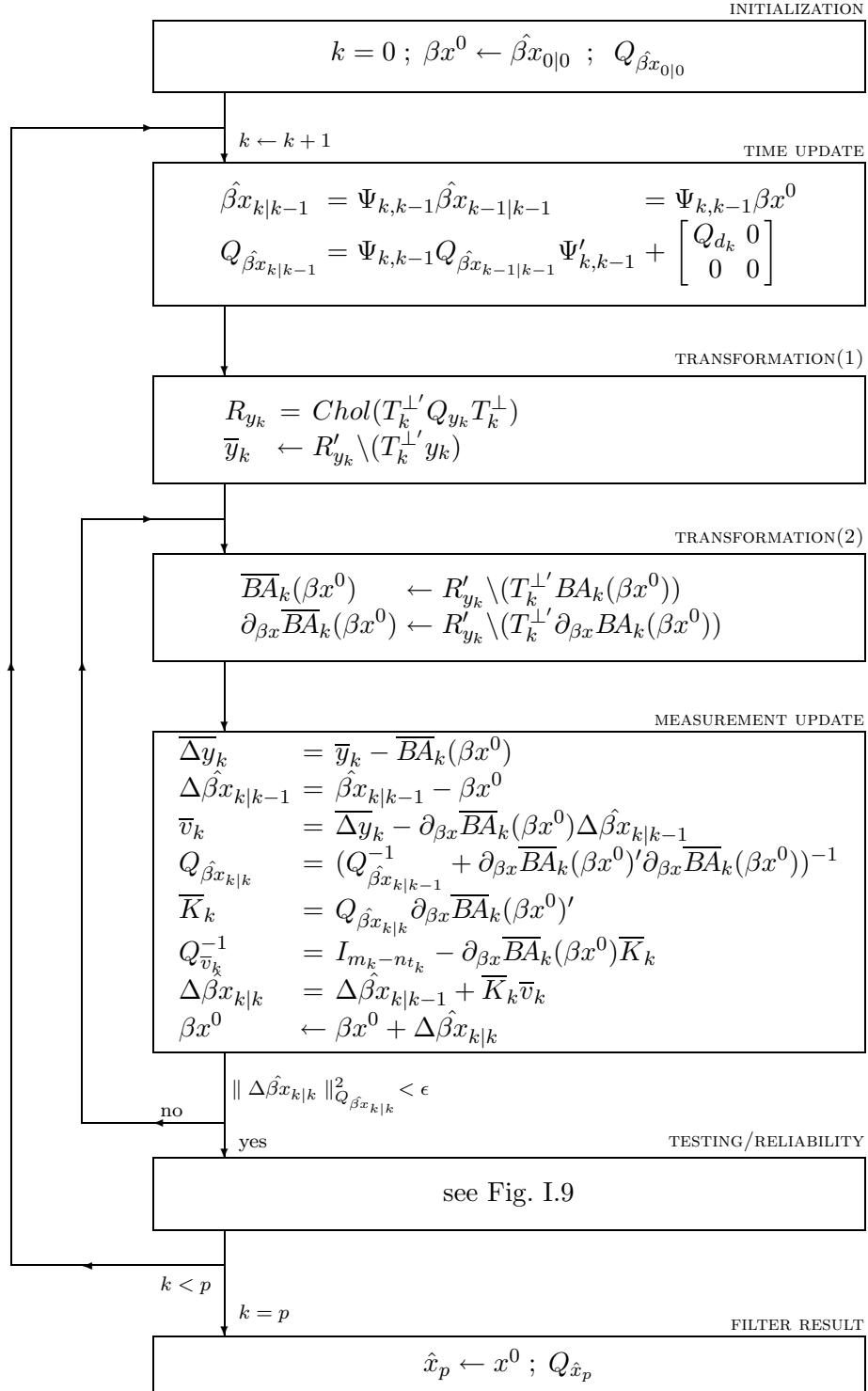
Figure I.8   Recursive Kalman–Cholesky Filter procedure corresponding to the model with nonlinear global parameters and linear constrained and temporal parameters using the model of observation equations.

FILTER INITIALIZATION

$$k = 0 \; ; \; \text{see Fig. I.8}$$

$k \leftarrow k + 1$ FILTERING

see Fig. I.8

$\delta\kappa \leftarrow \min(\delta k, k - 1)$

$\delta\kappa = 0$

$\delta\kappa > 0$ GLOBAL RECURSION

$$
\begin{aligned}
X_{k|k-1}^{\delta\kappa} &= \Psi_{k,k-1} X_{k-1|k-1}^{\delta\kappa-1} & (n_\beta + n \times 1)\\
C_{\overline{v}_k}^{\delta\kappa} &= -\partial_{\beta x}\overline{BA}_k(\beta x^0) X_{k|k-1}^{\delta\kappa} & (m_{k_0} - n_{t_{k_0}} \times 1)\\
X_{k|k}^{\delta\kappa} &= X_{k|k-1}^{\delta\kappa} + \overline{K}_k C_{\overline{v}_k}^{\delta\kappa} & (n_\beta + n \times 1)\\
Q_{\hat{\nabla}_k^{\delta\kappa}}^{-1} &= Q_{\hat{\nabla}_{k-1}^{\delta\kappa-1}}^{-1} + C_{\overline{v}_k}^{\delta\kappa'} Q_{\overline{v}_k}^{-1} C_{\overline{v}_k}^{\delta\kappa} & (1 \times 1)\\
\overline{K}_k^{\delta\kappa} &= Q_{\hat{\nabla}_k^{\delta\kappa}} C_{\overline{v}_k}^{\delta\kappa'} Q_{\overline{v}_k}^{-1} & (1 \times m_{k_0} - n_{t_{k_0}})\\
\hat{\nabla}_k^{\delta\kappa} &= \hat{\nabla}_{k-1}^{\delta\kappa-1} + \overline{K}_k^{\delta\kappa}(\overline{v}_k - C_{\overline{v}_k}^{\delta\kappa} \hat{\nabla}_{k-1}^{\delta\kappa-1}) & (1 \times 1)
\end{aligned}
$$

$\delta\kappa \leftarrow \delta\kappa - 1$

INITIALIZATION

$$
\begin{aligned}
(a)\ C_{y_k} &= [0,...,0,1,0,...,0]' & (m_k \times 1)\\
(a)\ C_{\overline{y}_k} &\leftarrow R'_{y_k}\backslash(T_k^{\perp'} C_{y_k}) & (m_k - n_{t_k} \times 1)\\
(a)\ C_{\overline{v}_k}^0 &= C_{\overline{y}_k} & (m_k - n_{t_k} \times 1)\\
(a)\ X_{k|k}^0 &= \overline{K}_k C_{\overline{v}_k}^0 & (n_\beta + n \times 1)\\
(b)\ C_{d_k} &= [0,...,0,1,0,...,0]' & (n_\beta + n \times 1)\\
(b)\ X_{k|k-1}^0 &= C_{d_k} & (n_\beta + n \times 1)\\
(b)\ C_{\overline{v}_k}^0 &= -\partial_{\beta x}\overline{BA}_k(\beta x^0) X_{k|k-1}^0 & (m_k \times 1)\\
(b)\ X_{k|k}^0 &= X_{k|k-1}^0 + \overline{K}_k C_{\overline{v}_k}^0 & (n_\beta + n \times 1)\\
Q_{\hat{\nabla}_k^0}^{-1} &= C_{\overline{v}_k}^{0'} Q_{\overline{v}_k}^{-1} C_{\overline{v}_k}^0 & (1 \times 1)\\
\overline{K}_k^0 &= Q_{\hat{\nabla}_k^0} C_{\overline{v}_k}^{0'} Q_{\overline{v}_k}^{-1} & (1 \times m_k - n_{t_k})\\
\hat{\nabla}_k^0 &= \overline{K}_k^0 \overline{v}_k & (1 \times 1)
\end{aligned}
$$

LS OR GS TESTING/RELIABILITY

$$
\begin{aligned}
\Theta_k^{\delta k}(1) &= (\hat{\nabla}_k^{\delta k}/\sigma_{\hat{\nabla}_k^{\delta k}})^2 & (1 \times 1 \text{ test quantity})\\
|\nabla_k^{\delta k}| &= \sqrt{\lambda_0}\cdot\sigma_{\hat{\nabla}_k^{\delta k}} & (1 \times 1 \text{ MDB})\\
\nabla\hat{\beta x}_{k|k}^{\delta k} &= X_{k|k}^{\delta k}\cdot|\nabla_k^{\delta k}| & (n_\beta + n \times 1 \text{ external reliability})
\end{aligned}
$$

$k < p$

$k = p$

end

Figure I.9   Recursive computation of GS test quantities and reliability measures in the Kalman–Cholesky Filter procedure corresponding to the model with nonlinear global parameters and linear constrained and temporal parameters using the model of observation equations. Specific examples of outlier testing (a) and slippage testing (b).

INITIALIZATION

$$k = 0 \;\; ; \;\; \beta x^0 \leftarrow \hat{\beta x}_{0|0} \;\; ; \;\; Q_{\hat{\beta x}_{0|0}}$$

$k \leftarrow k+1$ TIME UPDATE

$$\hat{\beta x}_{k|k-1} = \Psi_{k,k-1}\hat{\beta x}_{k-1|k-1} \qquad = \Psi_{k,k-1}\beta x^0$$
$$Q_{\hat{\beta x}_{k|k-1}} = \Psi_{k,k-1}Q_{\hat{\beta x}_{k-1|k-1}}\Psi'_{k,k-1} + \begin{bmatrix} Q_{d_k} & 0 \\ 0 & 0 \end{bmatrix}$$

TRANSFORMATION

$$Q_{v_k} \qquad\qquad = T_k^{\perp'} Q_{y_k} T_k^{\perp}$$
$$\qquad\qquad\qquad + \partial_{\beta x} BA_k(\beta x^0) Q_{\hat{\beta x}_{k|k-1}} \partial_{\beta x} BA_k(\beta x^0)'$$
$$R_{v_k} \qquad\qquad = Chol(Q_{v_k})$$
$$\overline{y}_k \qquad\qquad \leftarrow R'_{v_k} \backslash (T_k^{\perp'} y_k)$$
$$\overline{BA}_k(\beta x^0) \qquad \leftarrow R'_{v_k} \backslash (T_k^{\perp'} BA_k(\beta x^0))$$
$$\partial_{\beta x}\overline{BA}_k(\beta x^0) \leftarrow R'_{v_k} \backslash (T_k^{\perp'} \partial_{\beta x} BA_k(\beta x^0))$$

MEASUREMENT UPDATE

$$\overline{\Delta y}_k \qquad = \overline{y}_k - \overline{BA}_k(\beta x^0)$$
$$\Delta\hat{\beta x}_{k|k-1} = \hat{\beta x}_{k|k-1} - \beta x^0$$
$$\overline{v}_k \qquad\qquad = \overline{\Delta y}_k - \partial_{\beta x}\overline{BA}_k(\beta x^0)\Delta\hat{\beta x}_{k|k-1}$$
$$\overline{K}_k \qquad\qquad = Q_{\hat{\beta x}_{k|k-1}} \partial_{\beta x}\overline{BA}_k(\beta x^0)'$$
$$Q_{\hat{\beta x}_{k|k}} \qquad = (I_{n_\beta+n} - \overline{K}_k \partial_{\beta x}\overline{BA}_k(\beta x^0))Q_{\hat{\beta x}_{k|k-1}}$$
$$\Delta\hat{\beta x}_{k|k} \qquad = \Delta\hat{\beta x}_{k|k-1} + \overline{K}_k\overline{v}_k$$
$$\beta x^0 \qquad\qquad \leftarrow \beta x^0 + \Delta\hat{\beta x}_{k|k}$$

$\| \Delta\hat{\beta x}_{k|k} \|^2_{Q_{\hat{\beta x}_{k|k}}} < \epsilon$

no

yes   TESTING/RELIABILITY

see Fig. I.11

$k < p$

$k = p$ FILTER RESULT

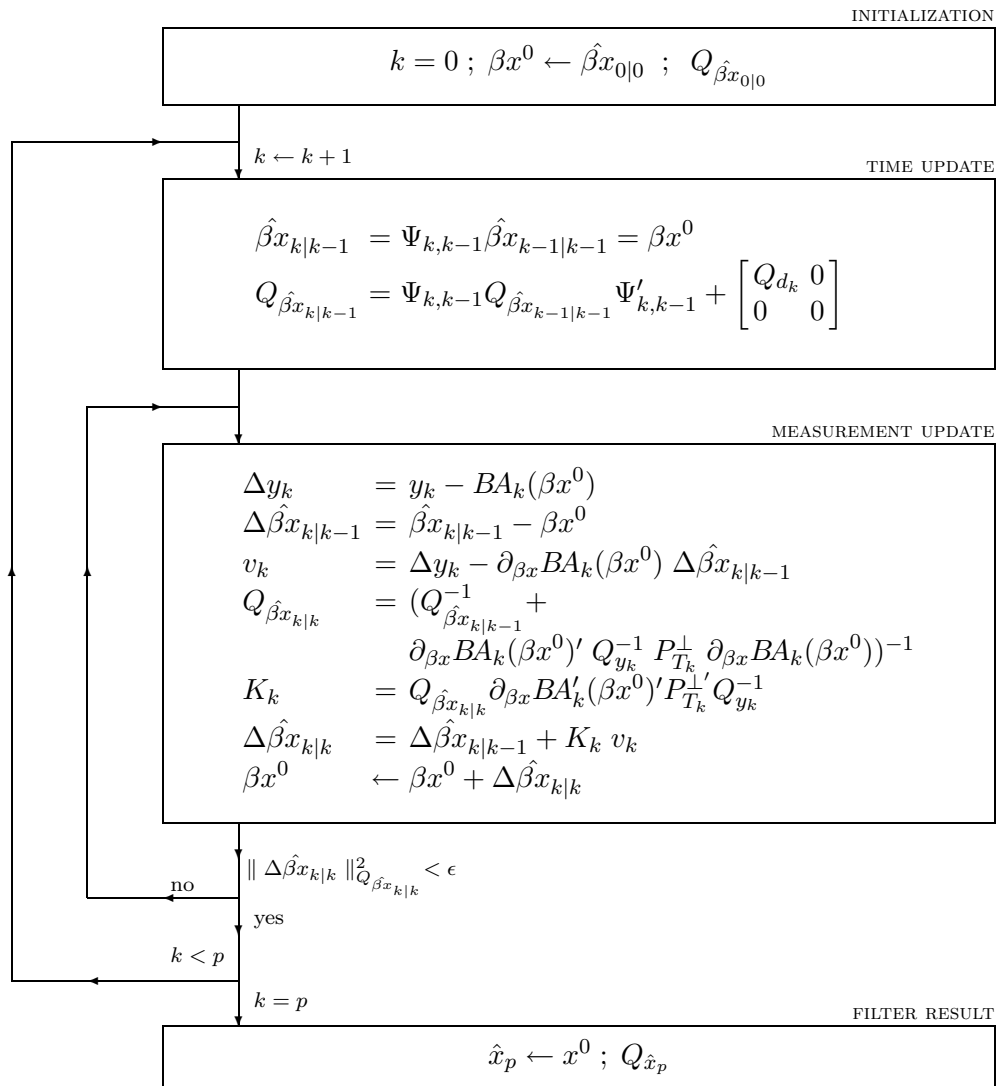$$\hat{x}_p \leftarrow x^0 \; ; \; Q_{\hat{x}_p}$$

Figure I.10   Recursive Kalman–Cholesky Filter procedure corresponding to the model with nonlinear global parameters and linear constrained and temporal parameters using the model of condition equations.
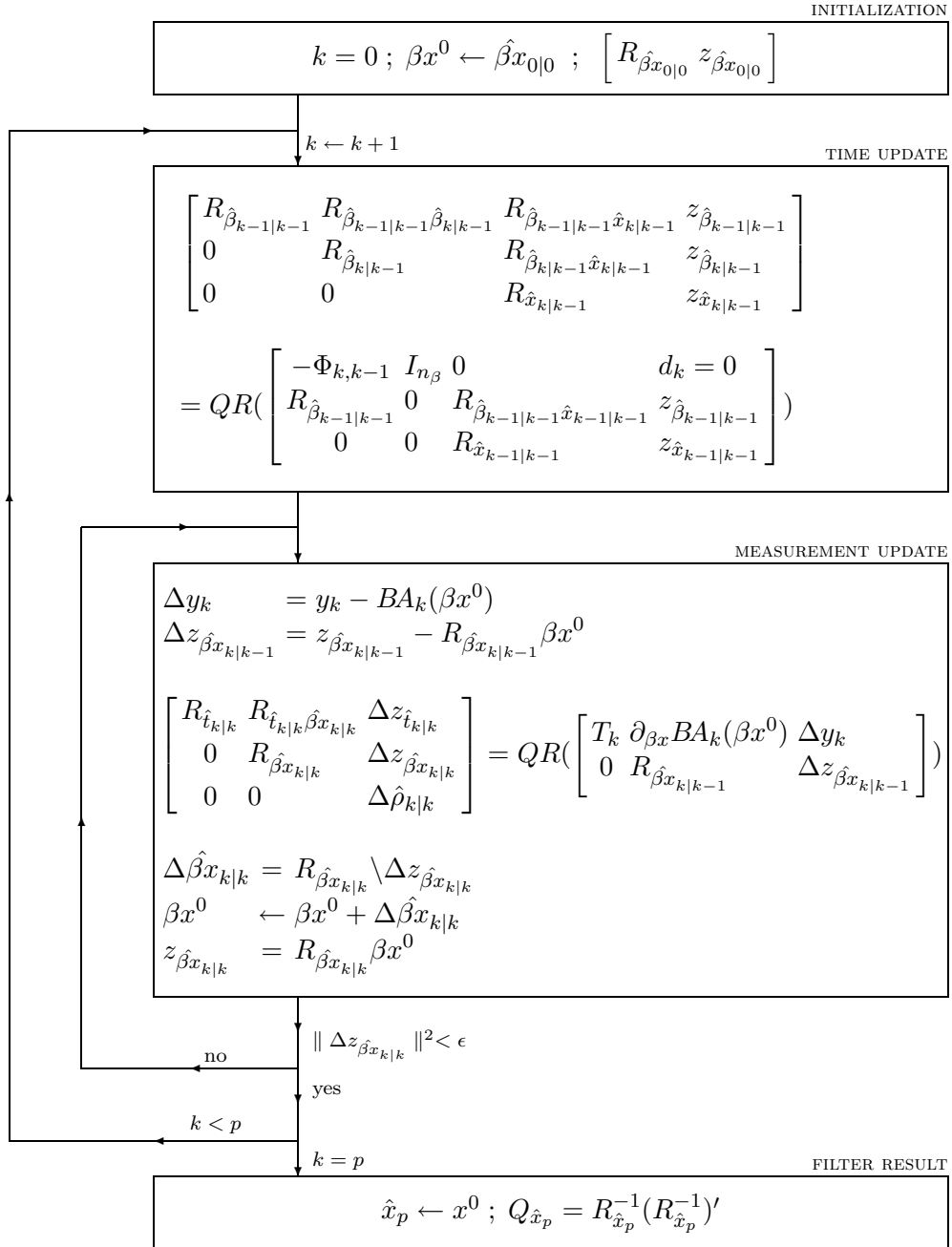
FILTER INITIALIZATION

$$k = 0 \; ; \; \text{see Fig. I.10}$$

$k \leftarrow k + 1$ FILTERING

see Fig. I.10

$\delta\kappa \leftarrow \min(\delta k, k-1)$

$\delta\kappa = 0$

$\delta\kappa > 0$ GLOBAL RECURSION

$$
\begin{aligned}
X_{k|k-1}^{\delta\kappa} &= \Psi_{k,k-1} X_{k-1|k-1}^{\delta\kappa-1} & (n_\beta + n \times 1) \\
C_{\overline{v}_k}^{\delta\kappa} &= -\partial_{\beta x}\overline{BA}_k(\beta x^0) X_{k|k-1}^{\delta\kappa} & (m_{k_0} - n_{t_{k_0}} \times 1) \\
X_{k|k}^{\delta\kappa} &= X_{k|k-1}^{\delta\kappa} + \overline{K}_k C_{\overline{v}_k}^{\delta\kappa} & (n_\beta + n \times 1) \\
Q_{\hat{\nabla}_k^{\delta\kappa}}^{-1} &= Q_{\hat{\nabla}_{k-1}^{\delta\kappa-1}}^{-1} + C_{\overline{v}_k}^{\delta\kappa'} C_{\overline{v}_k}^{\delta\kappa} & (1 \times 1) \\
\overline{K}_k^{\delta\kappa} &= Q_{\hat{\nabla}_k^{\delta\kappa}} C_{\overline{v}_k}^{\delta\kappa'} & (1 \times m_{k_0} - n_{t_{k_0}}) \\
\hat{\nabla}_k^{\delta\kappa} &= \hat{\nabla}_{k-1}^{\delta\kappa-1} + \overline{K}_k^{\delta\kappa}(\overline{v}_k - C_{\overline{v}_k}^{\delta\kappa}\hat{\nabla}_{k-1}^{\delta\kappa-1}) & (1 \times 1)
\end{aligned}
$$

$\delta\kappa \leftarrow \delta\kappa - 1$

INITIALIZATION

$$
\begin{aligned}
(a)\; C_{y_k} &= [0,...,0,1,0,...,0]' & (m_k \times 1) \\
(a)\; C_{\overline{y}_k} &\leftarrow R_{v_k}' \backslash (T_k^{\perp'} C_{y_k}) & (m_k - n_{t_k} \times 1) \\
(a)\; C_{\overline{v}_k}^0 &= C_{\overline{y}_k} & (m_k - n_{t_k} \times 1) \\
(a)\; X_{k|k}^0 &= \overline{K}_k C_{\overline{v}_k}^0 & (n_\beta + n \times 1) \\
(b)\; C_{d_k} &= [0,...,0,1,0,...,0]' & (n_\beta + n \times 1) \\
(b)\; X_{k|k-1}^0 &= C_{d_k} & (n_\beta + n \times 1) \\
(b)\; C_{\overline{v}_k}^0 &= -\partial_{\beta x}\overline{BA}_k(\beta x^0) X_{k|k-1}^0 & (m_k \times 1) \\
(b)\; X_{k|k}^0 &= X_{k|k-1}^0 + \overline{K}_k C_{\overline{v}_k}^0 & (n_\beta + n \times 1) \\
Q_{\hat{\nabla}_k^0}^{-1} &= C_{\overline{v}_k}^{0'} C_{\overline{v}_k}^0 & (1 \times 1) \\
\overline{K}_k^0 &= Q_{\hat{\nabla}_k^0} C_{\overline{v}_k}^{0'} & (1 \times m_k - n_{t_k}) \\
\hat{\nabla}_k^0 &= \overline{K}_k^0 \overline{v}_k & (1 \times 1)
\end{aligned}
$$

LS OR GS TESTING/RELIABILITY

$$
\begin{aligned}
\Theta_k^{\delta k}(1) &= (\hat{\nabla}_k^{\delta k}/\sigma_{\hat{\nabla}_k^{\delta k}})^2 & (1 \times 1 \text{ test quantity}) \\
|\nabla_k^{\delta k}| &= \sqrt{\lambda_0} \cdot \sigma_{\hat{\nabla}_k^{\delta k}} & (1 \times 1 \text{ MDB}) \\
\nabla\hat{\beta x}_{k|k}^{\delta k} &= X_{k|k}^{\delta k} \cdot |\nabla_k^{\delta k}| & (n_\beta + n \times 1 \text{ external reliability})
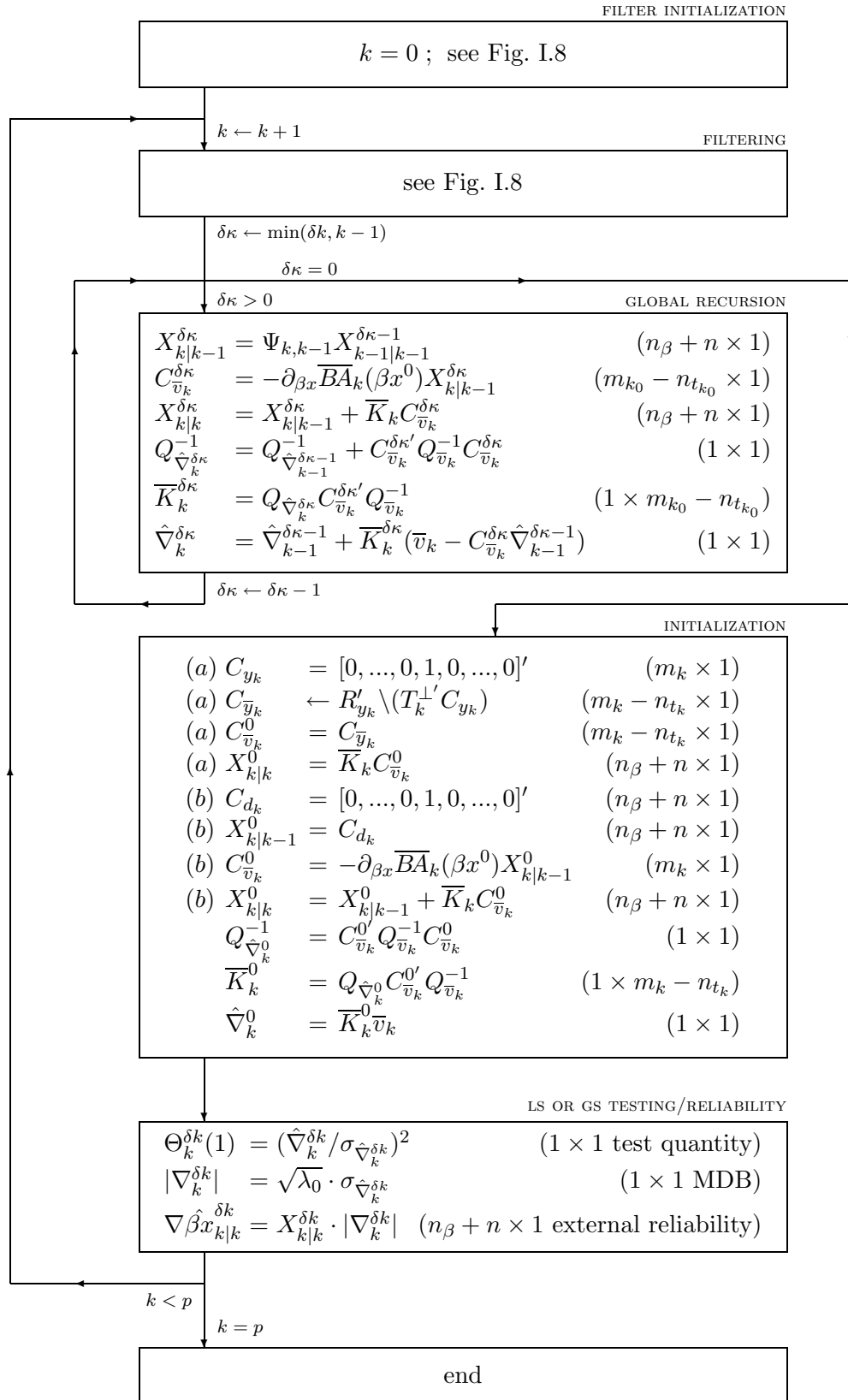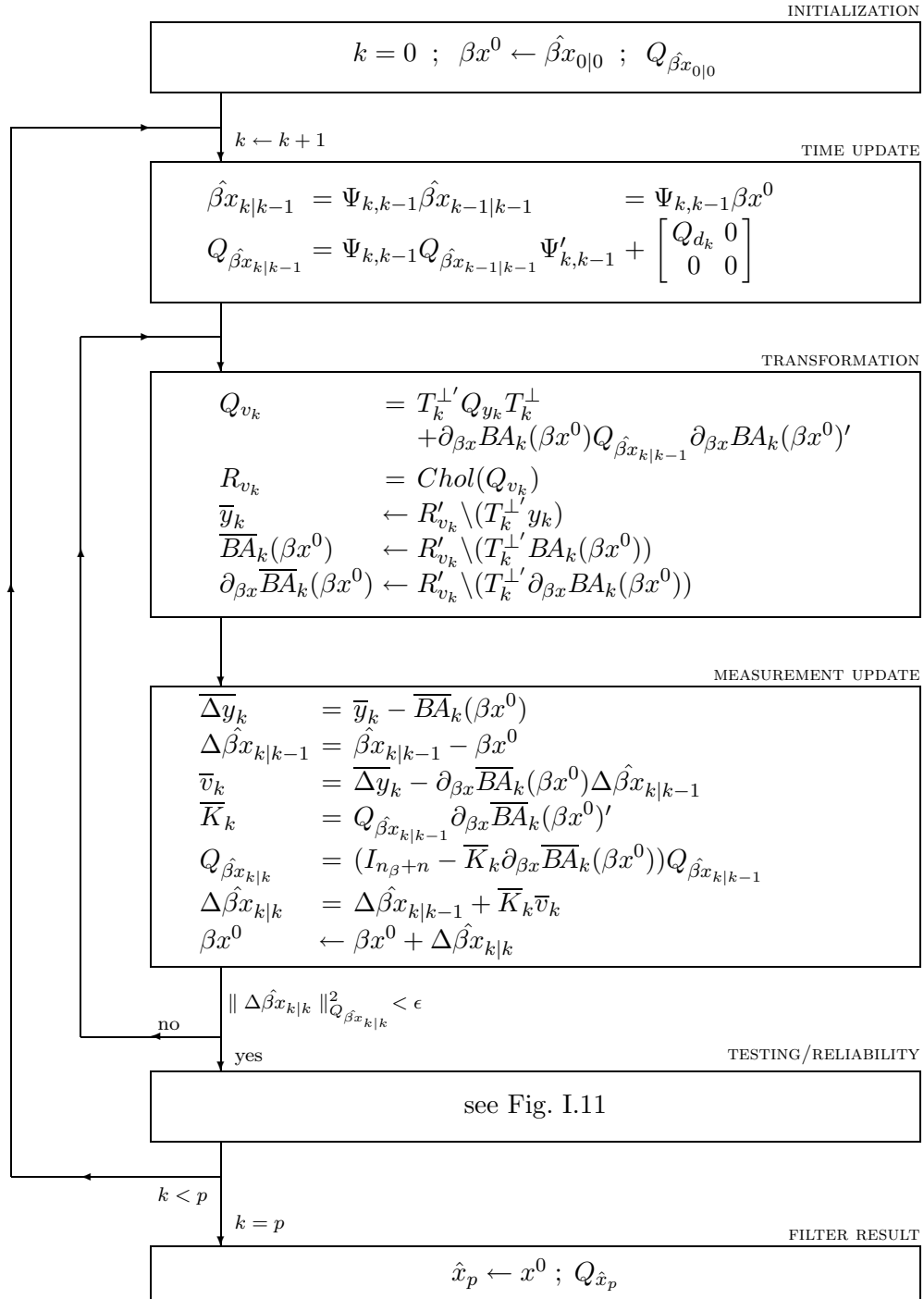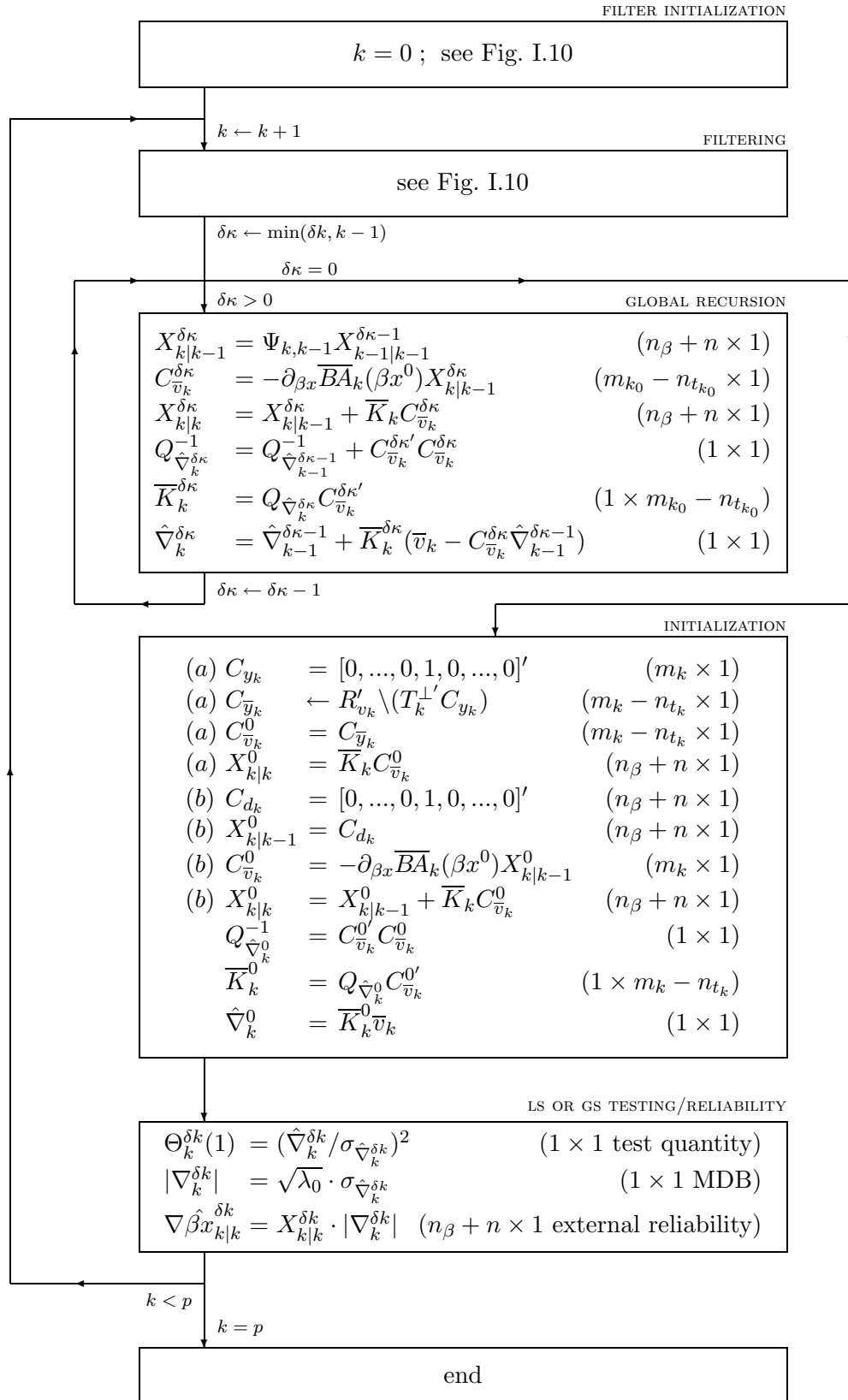\end{aligned}
$$

$k < p$

$k = p$

end

Figure I.11   Recursive computation of GS test quantities and reliability measures in the Kalman–Cholesky Filter procedure corresponding to the model with nonlinear global parameters and linear constrained and temporal parameters using the model of condition equations. Specific examples of outlier testing (a) and slippage testing (b).
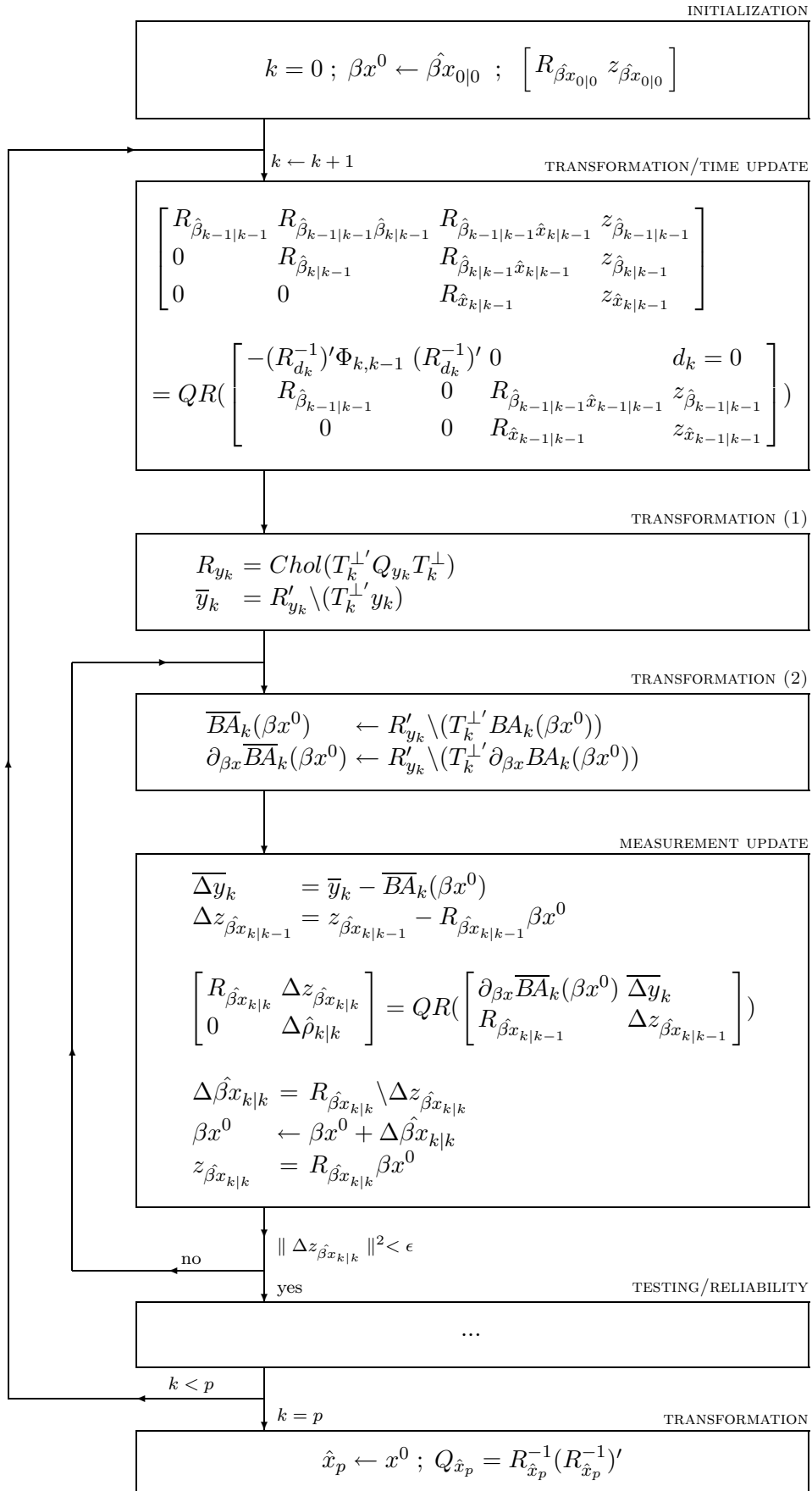
INITIALIZATION

$$k = 0 \;;\; \beta x^0 \leftarrow \hat{\beta} x_{0|0} \;\;;\;\; \left[\, R_{\hat{\beta} x_{0|0}} \;\; z_{\hat{\beta} x_{0|0}} \,\right]$$

$k \leftarrow k+1$ TRANSFORMATION/TIME UPDATE

$$\begin{bmatrix} R_{\hat{\beta}_{k-1|k-1}} & R_{\hat{\beta}_{k-1|k-1}\hat{\beta}_{k|k-1}} & R_{\hat{\beta}_{k-1|k-1}\hat{x}_{k|k-1}} & z_{\hat{\beta}_{k-1|k-1}} \\ 0 & R_{\hat{\beta}_{k|k-1}} & R_{\hat{\beta}_{k|k-1}\hat{x}_{k|k-1}} & z_{\hat{\beta}_{k|k-1}} \\ 0 & 0 & R_{\hat{x}_{k|k-1}} & z_{\hat{x}_{k|k-1}} \end{bmatrix}$$

$$= QR(\begin{bmatrix} -(R_{d_k}^{-1})'\Phi_{k,k-1} & (R_{d_k}^{-1})' & 0 & d_k = 0 \\ R_{\hat{\beta}_{k-1|k-1}} & 0 & R_{\hat{\beta}_{k-1|k-1}\hat{x}_{k-1|k-1}} & z_{\hat{\beta}_{k-1|k-1}} \\ 0 & 0 & R_{\hat{x}_{k-1|k-1}} & z_{\hat{x}_{k-1|k-1}} \end{bmatrix})$$

TRANSFORMATION (1)

$$R_{y_k} = Chol(T_k^{\perp'} Q_{y_k} T_k^{\perp})$$
$$\overline{y}_k = R'_{y_k} \backslash (T_k^{\perp'} y_k)$$

TRANSFORMATION (2)

$$\overline{BA}_k(\beta x^0) \leftarrow R'_{y_k} \backslash (T_k^{\perp'} BA_k(\beta x^0))$$
$$\partial_{\beta x} \overline{BA}_k(\beta x^0) \leftarrow R'_{y_k} \backslash (T_k^{\perp'} \partial_{\beta x} BA_k(\beta x^0))$$

MEASUREMENT UPDATE

$$\overline{\Delta y}_k = \overline{y}_k - \overline{BA}_k(\beta x^0)$$
$$\Delta z_{\hat{\beta} x_{k|k-1}} = z_{\hat{\beta} x_{k|k-1}} - R_{\hat{\beta} x_{k|k-1}} \beta x^0$$

$$\begin{bmatrix} R_{\hat{\beta} x_{k|k}} & \Delta z_{\hat{\beta} x_{k|k}} \\ 0 & \Delta \hat{\rho}_{k|k} \end{bmatrix} = QR(\begin{bmatrix} \partial_{\beta x} \overline{BA}_k(\beta x^0) & \overline{\Delta y}_k \\ R_{\hat{\beta} x_{k|k-1}} & \Delta z_{\hat{\beta} x_{k|k-1}} \end{bmatrix})$$

$$\Delta \hat{\beta} x_{k|k} = R_{\hat{\beta} x_{k|k}} \backslash \Delta z_{\hat{\beta} x_{k|k}}$$
$$\beta x^0 \leftarrow \beta x^0 + \Delta \hat{\beta} x_{k|k}$$
$$z_{\hat{\beta} x_{k|k}} = R_{\hat{\beta} x_{k|k}} \beta x^0$$

no $\| \Delta z_{\hat{\beta} x_{k|k}} \|^2 < \epsilon$

yes   TESTING/RELIABILITY

...

$k < p$ $k = p$ TRANSFORMATION

$$\hat{x}_p \leftarrow x^0 \;;\; Q_{\hat{x}_p} = R_{\hat{x}_p}^{-1}(R_{\hat{x}_p}^{-1})'$$

Figure I.12   Recursive SRIF algorithm for the model with nonlinear global parameters and linear constrained and temporal parameters.

Bibliography

Baarda, W. (1968), A testing procedure for use in geodetic networks, *Publ. on geodesy, new series, vol. 5(1)*, Netherlands Geodetic Commission.

Bierman, G.J. (1977), *Factorization Methods for Discrete Sequential Estimation*, Academic Press, Orlando.

Björck, Å. (1996), *Numerical Methods for Least Squares Problems*, Siam, Philadelphia.

George, A., and J.W.H. Liu (1989), The evolution of the minimum degree ordering algorithm, *SIAM Review, 31*, 1–19.

Golub, G.H., and C.F. van Loan (1989), *Matrix computations*, John Hopkins University Press, Baltimore.

Kalman, R.E. (1960), A new approach to linear filtering and prediction problems, *Trans. ASME Journal of Basic Engineering, 82*, 35–45.

Koch, K.-R. (1987), *Parameter Estimation and Hypothesis Testing in Linear Models*, Springer, Berlin.

Matlab (1997), *The language of technical computing*, The MathWorks, Inc.

Matstoms, P. (1994), Sparse QR factorization in matlab, *ACM Transactions on Mathematical Software, 20*, 136–159.

Salzmann, M. (1993), *Least Squares Filtering and Testing for Geodetic Navigation Applications*, Ph.D. dissertation, Delft University of Technology.

Teunissen, P.J.G. (1990a), An integrity and quality control procedure for use in multisensor integration, in *Proc. of ION-90, Colorado Springs*, pp. 513–522.

Teunissen, P.J.G. (1990b), On the local covergence of the iterated extended Kalman filter: Contributions to geodetic theory and methodology, in *Proc. of the XXth general assembly of the IUGG/IAG section IV, Vienna*, pp. 177–184.

Teunissen, P.J.G. (1990c), Quality control in integrated navigation systems, in *Proc. of the IEEE PLANS '90, Las Vegas*, pp. 158–165.

Teunissen, P.J.G. (1999), *Adjustment theory: An introduction*, Series on Mathematical Geodesy and Positioning, Delft University Press.

Teunissen, P.J.G. (2000), *Testing theory: An introduction*, Series on Mathematical Geodesy and Positioning, Delft University Press.

Teunissen, P.J.G. (2001), *Dynamic data processing: Recursive least-squares*, Series on Mathematical Geodesy and Positioning, Delft University Press.

Teunissen, P.J.G., and M.A. Salzmann (1989), A recursive slippage test for use in state-space filtering, *Manuscripta Geodaetica, 14*, 383–390.

Tiberius, C.C.J.M. (1997), Implementation of the square root information filter for recursive data processing, *internal publication*, DUT, Delft Geodetic Computing Centre, Delft.

Tiberius, C.C.J.M. (1998), *Recursive data processing for kinematic GPS surveying*, Ph.D. dissertation, Delft University of Technology.

Wu, S.C., and R.J. Muellerschoen (1992), Efficient filtering schemes for GPS data with a large number of carrier phase biases, in *6th Int. Geod. Symp. on Sat. Pos., Ohio*, pp. 265–272.

SIMULATIONS

Symbols in Part V

symbol	meaning
A_k	partial design matrix of global parameters
β_k	constrained (batch) parameters
$\hat{\beta}_{k\|k-1}$	predicted value of β_k
$\hat{\beta}_{k\|k}$	filtered value of β_k
βx_k	concatenated vector $[\beta'_k,\ x']'$
B_k	partial design matrix of batch (constrained) parameters
BA_k	combined partial design matrix $[B_k,\ A_k]$
C_n	$n \times (n-1)$ matrix $[0,\ I_{n-1}]'$
C_N	constant
$d,\ d(k),\ d_k$	vector of spatiotemporal constraints
$D^z(k)$	ZTDs at epoch k
δs	satellite clock errors
$\Delta D(k)$	residual STDs at epoch k
Δt	sampling rate
e_n	$n \times 1$ vector with ones
H	effective height
I_n	$n \times n$ identity matrix
k	current epoch number ($k = 1, ..., p$)
K_k	gain matrix
m_k	number of satellites at epoch k
M_r	mapping values for receiver r
$M_w(.)$	wet mapping function
n	number of stations
p	number of epochs
$\widetilde{p}$	batch size
PSD_sim	PSD as used in simulation
PSD_est	PSD as used in estimation

symbol	meaning
Q_d, $Q_d(k)$, Q_{d_k}	covariance matrix of spatiotemporal constraints
$\overline{Q}_d$	co-factor matrix of spatiotemporal constraints
$Q_{\Delta D}(k)$	covariance matrix of residual STDs
$Q_y(k)$, Q_{y_k}	covariance matrix of observations
$Q_{y,}(k)$	covariance matrix of observations in absence of residual STDs
$\tilde{Q}$	'real' covariance matrix (instead of formal Q in filter)
R_d, $R_d(k)$	Cholesky factor of Q_d, $Q_d(k)$
$R_y(k)$	Cholesky factor of $Q_y(k)$
$\rho(k)$	vector of receiver–satellite ranges
σ_d^2	variance of spatiotemporal constraints
σ_ϕ^2	variance of phase observations
σ_p^2	variance of code observations
T_k	partial design matrix of temporal parameters
$T_k^\perp$	null space of T_k'
w_d, w_y	generated standard normally distributed white noise
x	global parameters
$X_{k|k}^{\delta k}$	response matrix
$y(k)$, y_k	vector of observations at epoch k
$y,(k)$	observations in absence of residual STDs
z	zenith angle

Chapter 26

Introduction to Part V

To analyze the impact of model choices on the height estimates, GPS analysis software was written which simulates and filters GPS observations in the presence of (simulated) tropospheric delays, based on the theory, models, and filtering techniques as described in the previous parts.

Implementation of a filter for simulated GPS observations and delays is in several ways easier than for actual observations. We do not need to read RINEX data or compute accurate orbits; some error sources, like phase center variations etc., do not need to be modeled; the observations can be assumed unbiased; missing observations, outliers, and cycle slips do not need to be treated; the number of stations can be considered constant (no missing data), but the changes in the satellite configuration do need to be dealt with.

Chapter 27 describes the models that are implemented, why they are implemented, and how this is done. We distinguish simulation and estimation aspects. In Chap. 28 we give the results of a simulation with default settings. The simulation results include precision, coordinate and ZTD errors, test quantities, and reliabilities. The results are generally in good agreement with the theory as presented in the previous parts. Chapter 29 covers a selection of simulations where we vary some of the modeling aspects and study the sensitivity of some of the simulation results, mainly the precision. Chapter 30 studies the baseline length for which the troposphere-fixed model is applicable. We end this part with some conclusions and recommendations in Chap. 31.

In contrast to the previous parts, this part cannot be read independently of the other parts.

Software implementation

27.1   Introduction

This chapter describes the software as used for the simulations in Chaps. 28 and 29. The software is written in MATLAB [*Matlab*, 1997] because the models and procedures of Parts II–IV are relatively easy to implement in MATLAB and because it is sufficiently fast for research purposes.

Section 27.2 describes the modeling as implemented in the software, Sect. 27.3 describes the simulation part of the software, and Sect. 27.4 deals with the estimation/filter part.

27.2   Modeling

No a-priori model was implemented because it does not play a role in the filter. As mapping function we use NMFw [*Niell*, 1996], because we assume the hydrostatic delay to be modeled correctly based on surface pressure values and because NMFw is independent of surface meteorological parameters.

A recursive filtering approach was chosen to be implemented because the impact of the observation time is easier analyzed than with a batch method. Kalman–Cholesky Filtering was implemented instead of Square Root Information Filtering[1] because it is faster and because the testing and reliability aspects were already worked out in Part IV. Temporal parameters are pre-eliminated using the null space of their partial design matrix (double-differencing, ionosphere-free linear combination) rather than with the orthogonal projector (undifferenced observations) for the same reasons. The model with observation equations was implemented rather than the model with condition equations because the latter model has complications with respect to changes in the satellite configuration.

Only the troposphere-weighted and fixed models are implemented. The float and constant models are simply special cases of the weighted model. Zenith delays can be estimated every epoch or every predefined number of epochs (batches). Spatiotemporal constraints on zenith delays can be applied with a corresponding covariance matrix and a scaling Power Spectral Density (PSD). The constraints may be spatially uncorrelated (classic approach) or correlated according to the polynomial model of Eq. (18.17) and Table 18.1 of Part III. This is called TL2 correlation after Treuhaft and Lanyi and the scale height of H = 2 km. Residual STDs can be left unmodeled or be pre-eliminated. In the

TL2

---
[1]SRIF was only implemented in an early version.

```
read input;
prepare settings;
read satellite coordinates for each epoch;
for k=1:p
  compute ranges, zenith angles, etc.;
  model design:  Ry(k) = Chol(Tk⊥'Qy(k)Tk⊥), Qd and
  Bk ← Ry(k)'\(Tk⊥'Bk);
  observation simulation yk;
  observation transformation yk ← Ry(k)'\(Tk⊥'yk);
  if k==1, initialization of Qβx0|0; end
  time update;
  if new pivot satellite, ambiguity transformation; end
  while true
    draft and transformation of design matrix:
    Āk(x0) ← Ry(k)'\(Tk⊥'∂xĀk(x0)),
    ∂βxBAk(βx0) ← [Bk, ∂xĀk(x0))];
    approximate values yk0 = BAk(βx0);
    observation increment Δyk = yk − yk0;
    measurement update;
    update global parameters;
    if stop criterion < ε
      save standard deviations of state vector;
      testing and reliability, save results;
      break;
    else
      update ranges;
    end
  end
  save filtered state vector;
end
make plots;
```

Figure 27.1   Software structure.

latter case the covariance matrix of the observables has an additional term. The covariance matrix of the residual STDs is based on the summation model of Eq. (18.32) in Part III, with $C_N = 0.24$ m$^{-\frac{1}{3}}$ and H = 2 km. Gradient parameters are not implemented because modeling STDs was expected to be more promising.

ionosphere-float   The ionosphere-float model (ionosphere-free linear combination) was implemented and epoch-wise changing satellite clocks were assumed because these models are most often used; other ionosphere or clock models were not implemented to restrict the number of simulation options. Both phase and code observations are simulated although the impact of the code observations is very small.

ambiguities   Although the software does not resolve integer ambiguities it can compute both an ambiguities-fixed or float solution.

An overview of the chosen modeling aspects and their implementation is given in Table 27.1. Figure 27.1 shows the structure of the software as implemented, which is basically the flow chart of Fig. I.8 in Part IV.

modeling aspect	implementation
a-priori model	none
mapping function	NMFw
filtering technique	Kalman–Cholesky, observation equations
pre-elimination temp. par.	with null space of T_k
ZTD modeling	troposphere-weighted or troposphere-fixed
batch size	free choice
PSD	free choice
temp. corr. troposphere	random walk
spatial corr. troposphere	TL2 correlation or uncorrelated
residual STDs	unmodeled or pre-eliminated (summation model)
troposphere gradients	no
ionosphere	float
satellite clocks	epoch-wise changing
observations	dual-frequency phase and code
ambiguities	float or fixed; no resolution
coordinates first station	fixed

Table 27.1   Model implementation aspects.

27.3   Simulation

Two data types are simulated for each epoch: n absolute ZTDs and $4m_k n$ observations, where n is the number of stations and m_k the number of satellites at epoch $k = 1, ..., p$. Both data types are described in this section.

ZTDs   ZTDs can be simulated following any functional relation in time. However, a more realistic behavior of the ZTDs is assumed to be a random-walk process driven by a Power Spectral Density 'PSD_sim' $[\mathrm{m}^2/\mathrm{s}]$, with for each epoch either spatially uncorrelated or spatially correlated values based on the TL2 model. New simulated ZTD values, $D^z(k)$, are generated every epoch k as:

$$D^z(k) \quad = D^z(k-1) + R'_d w_d(k);$$

$$Q_d = R'_d R_d = \begin{cases} \sigma_d^2 \, \overline{Q}_d \ (\text{TL2 correlation}) \text{ or} \\ \sigma_d^2 \, I_n \ (\text{uncorrelated}); \end{cases} \tag{27.1}$$

$$\sigma_d^2 \quad = \text{PSD\_sim} \cdot \Delta t,$$

R_d, $w_d(k)$ where R_d is the Cholesky factor of Q_d, $w_d(k)$ is a generated Gaussian white-noise
Δt process with zero expectation and unit variance, and Δt [s] is the sampling rate. Note that spatiotemporal constraints $d(k) \equiv d$ and their covariance matrix $Q_d(k) \equiv Q_d$ do not change from epoch to epoch since d only depends on the number of stations, $\overline{Q}_d$ only depends on the station coordinates, and Δt is a fixed value. If either the number of stations or Δt changes, we would have to write $d(k)$ and $Q_d(k)$. The co-factor matrix $\overline{Q}_d$ may be an identity matrix as in the 'classic' approach of, for example, the Bernese software [*Beutler et al.*, 2000], or a full matrix based on Eqs. (18.17) and (18.20), and Table 18.1 in Part III.

observations   Two-frequency phase and two-frequency code observations, $y(k) = [y_{,1}(k), ..., y_{,4}(k)]'$, are simulated every epoch as the sum of actual ranges,

tropospheric slant delays, and additional noise:

$$y(k) = e_4 \otimes \rho(k) + e_4 \otimes M_w(z(k)) \cdot D^z(k) + R_y(k)'w_y(k), \qquad (27.2)$$

where $e_4 = [1, 1, 1, 1]'$. The ranges $\rho(k)$ are computed from the known receiver and satellite coordinates. The slant delays are computed from the simulated zenith delays $D^z(k)$ and the wet mapping function 'NMFw' of Niell, $M_w(z(k))$ [*Niell*, 1996]. Ambiguities are assumed zero and have therefore no active role in the simulated observations. Other parameters, like clock errors and ionospheric delays, do not need to be simulated because they are pre-eliminated. Note that we may add an a-priori ZTD to the simulated observations, but this will have no effect when the same delay is subtracted from the observations in the processing. In the simulation we therefore do not need to account for an a-priori delay unless we wish to analyze the impact of an incorrect one [*Kleijer*, 2001]. The effect of using a sec z mapping function instead of NMFw is about 3 mm height difference per 10 cm zenith delay difference by a cut-off angle of 80 degrees.[2] The same amount can be expected when using a hydrostatic mapping function like NMFh instead of NMFw (difference between NMFh and NMFw is about the same as for NMFw and sec z).

The additional observation noise is based on $w_y(k)$, a generated Gaussian white-noise process with zero expectation and unit variance, and $R_y(k)$, the Cholesky factor of the covariance matrix of the observations, for which there are four options:

$$
\begin{array}{lll}
1. & Q_y(k) = 0 & \text{(zero)}; \\
2. & Q_y(k) = Q_{y,}(k) = \text{diag}[\sigma_\phi^2 I_{2m_k n}, \sigma_p^2 I_{2m_k n}] & \text{(unit)}; \\
3. & Q_y(k) = \text{diagonal of } (Q_{y,}(k) + e_4 e_4' \otimes Q_{\Delta D}(k)) & \text{(diagonal)}; \\
4. & Q_y(k) = Q_{y,}(k) + e_4 e_4' \otimes Q_{\Delta D}(k), & \text{(full)}
\end{array}
\qquad (27.3)
$$

where $Q_{y,}(k)$ is the covariance matrix of the GPS observables in absence of residual STDs, and $Q_{\Delta D}(k)$ the covariance matrix of the residual STD pseudo-observables, which loosens the isotropy assumption of the tropospheric delays introduced by the mapping functions. Clearly, $Q_y(k)$ is epoch dependent because it depends on the number of satellites in $Q_{y,}(k)$ and the satellite configuration in $Q_{\Delta D}(k)$.

The last option is assumed to be the most realistic one. The third option discards any correlation between observations and reduces to a form of zenith angle dependent weighting. The second option does not assume any STD-related noise; the matrix $Q_{y,}(k)$ consists of two scaled unit matrices with variances for phase observations (σ_ϕ^2) and for code observations (σ_p^2). All the zenith angle dependent noise is thus assumed to be related to the troposphere. The first option does not assume any noise at all and is only a benchmark option.

Table 27.2 shows all the simulation options as implemented in the software.

27.4   Estimation

Basically the flow charts of Fig. I.8 for estimation and Fig. I.9 for testing/reliability measures of Part IV are implemented.

---

[2]This can be seen from a simulation with errorless observations and using the troposphere-float model with absolute delays (no biases introduced).

symbol	range	default	unit	description
$M_w(z)$	NMFw	NMFw	-	mapping function
Δt	$\mathbb{R}^+$	30	[s]	sampling rate/time interval of epochs
PSD_sim	$(0-\infty)$	1.5e-7	[m^2/s]	Power Spectral Density
$\overline{Q}_d$	TL2/uncorr.	TL2	-	cov. matrix spatiotemp. constraints
$Q_y(k)$	1/2/3/4	4	-	covariance matrix observations
σ_ϕ	$(0-\infty)$	0.002	[m]	standard deviation phase observations
σ_p	$(0-\infty)$	0.2	[m]	standard deviation code observations

Table 27.2   Simulation options. Options for $Q_y(k)$ correspond to Eq. (27.3). Units relate to the option ranges, $Q_y(k)$ for example has unit [m^2], but option 4 has no unit. The value for the PSD is taken from Part III, but is considered too high (too 'variable') according to [*Jarlemark et. al., 1998*], where a value of 1e-8 m^2/s was found to be a moderate value, based on WVR data.

Table 27.3 summarizes the estimation options as implemented in the software. Note that some of the defaults differ from what is considered 'standard' nowadays: The batch size is often larger, the spatiotemporal constraints are usually uncorrelated, and the covariance matrix of the observations is often a diagonal matrix with some zenith angle dependency.

In the measurement update, both observations $y_,(k)$ and residual STD pseudo-observations $\Delta D(k)$ are processed with respective covariance matrices $Q_{y,}(k)$ and $Q_{\Delta D}(k)$. Since pre-elimination of the residual STD pseudo-observations is applied, the transformed observations have covariance matrix

$$T_k^{\perp '}Q_y(k)T_k^\perp = T_k^{\perp '}\{Q_{y,}(k) + e_4 e_4' \otimes Q_{\Delta D}(k)\}T_k^\perp \tag{27.4}$$

according to the transformation of Eq. (14.12) in Part II, with $T_k^{\perp '}$ the transformation matrix that pre-eliminates the temporal parameters. The analytic expression for this covariance matrix is used in the software.

symbol	range	default	unit	description
z_{cut}	[0–90]	80	[deg]	zenith cut-off angle
-	0/1	0	-	ambiguities fixed (1) or float (0)
$M_w(z)$	NMFw/sec	NMFw	-	mapping function
-	0/1	0	-	absolute (1) or relative (0) ZTD estim.
Δt	30*[1–p]	30	[s]	sampling rate
$\widetilde{p}$	[1–p]	1	[-]	batch size
PSD_est	$\mathbb{R}^+$	1.5e-7	[m^2/s]	Power Spectral Density
$\overline{Q}_d$	TL2/uncorr.	TL2	-	cov. matrix spatiotemp. constraints
$Q_y(k)$	2/3/4	4	-	covariance matrix observations
σ_ϕ	$\mathbb{R}^+$	0.002	[m]	standard deviation phase observations
σ_p	$\mathbb{R}^+$	0.2	[m]	standard deviation code observations
ϵ	$\mathbb{R}^+$	1e-6	[-]	convergence criterion
-	$\mathbb{N}^+$	10	[-]	maximum number of iterations

Table 27.3   Estimation/filter options. Options for $Q_y(k)$ correspond with Eq. (27.3). The mapping function may be Niell's wet mapping function (NMFw) or a simple $\sec z$ function.

time update   In the time update, the spatiotemporal constraints d with covariance matrix $Q_d =$

$\sigma_d^2 \overline{Q}_d$ are processed. The scaling variance factor is obtained from:

$$\sigma_d^2 = \text{PSD\_est} \cdot \Delta t \cdot \widetilde{p}, \tag{27.5}$$

where $\widetilde{p}$ is the batch size, and PSD_est the assumed PSD for the parameter esti-
mation. The variance factor σ_d^2 is independent of the baseline length. Like in the
simulation, $\overline{Q}_d$ may be an identity matrix or imply full TL2 correlation. In the lat-
ter case, the covariance matrix depends on the baseline length. In the terminology
of Part II, the following troposphere models are distinguished:

$$\underset{\text{constant}}{0} \quad < \underset{\text{weighted}}{Q_d = \sigma_d^2 \, \overline{Q}_d} < \underset{\text{float}}{\infty}.$$

A small value of PSD_est can therefore be considered a troposphere-constant
model, and a large value of PSD_est or $\widetilde{p}$, a troposphere-float model.

Q_d is an $n \times n$ matrix of n constraints on absolute ZTDs. As shown in
Part II, estimating absolute ZTDs introduces one near rank deficiency per batch.
When estimating relative ZTDs, the time update should therefore be based on
the observation Eq. (13.16) of Part II. The variance-update part of the time
update in this case is not implemented with Q_d but with $(C_n' Q_d^{-1} C_n)^{-1}$ instead,
as explained below.

The Gauss–Markov model of the time update when estimating relative ZTDs and
not considering global parameters reads:

$$\mathbf{E}\Big\{ \begin{bmatrix} \underline{\hat{\beta}}_{k-1|k-1} \\ \underline{d} \end{bmatrix} \Big\} = \begin{bmatrix} I_{n-1} & 0 \\ -C_n & C_n \end{bmatrix} \begin{bmatrix} \beta_{k-1} \\ \beta_k \end{bmatrix};$$

$$\mathbf{D}\Big\{ \begin{bmatrix} \underline{\hat{\beta}}_{k-1|k-1} \\ \underline{d} \end{bmatrix} \Big\} = \begin{bmatrix} Q_{\hat{\beta}_{k-1|k-1}} & 0 \\ 0 & Q_d \end{bmatrix}, \tag{27.6}$$

with the $n \times (n-1)$ matrix $C_n \doteq [0, \, I_{n-1}]'$. The LSQ estimators of β_k, the vector
with relative ZTDs that act as batch parameters at epoch k, and corresponding
covariance matrix then read:

$$\begin{aligned} \underline{\hat{\beta}}_{k|k-1} &= \underline{\hat{\beta}}_{k-1|k-1} + (C_n' Q_d^{-1} C_n)^{-1} C_n' Q_d^{-1} \underline{d}; \\ Q_{\hat{\beta}_{k|k-1}} &= Q_{\hat{\beta}_{k-1|k-1}} + (C_n' Q_d^{-1} C_n)^{-1}, \end{aligned} \tag{27.7}$$

which means that if $d = 0$, the LSQ prediction of β_k equals the filtered value at
epoch $k - 1$: $\hat{\beta}_{k|k-1} = \hat{\beta}_{k-1|k-1}$.

Including the global parameters in the Gauss–Markov model of the time
update affects neither the estimates of the constrained nor the global parameters.
This can be shown by *free variates* as shown in App. J. The complete time update
then reads:

$$\boxed{\begin{aligned} \hat{\beta x}_{k|k-1} &= \hat{\beta x}_{k-1|k-1}; \\ Q_{\hat{\beta x}_{k|k-1}} &= Q_{\hat{\beta x}_{k-1|k-1}} + \begin{bmatrix} (C_n' Q_d^{-1} C_n)^{-1} & 0 \\ 0 & 0 \end{bmatrix}, \end{aligned}} \tag{27.8}$$

where βx_k is a concatenated state vector with batch and global parameters.
Estimating relative instead of absolute ZTDs thus only involves replacing Q_d by
$(C_n' Q_d^{-1} C_n)^{-1}$, cf. Eq. (21.31) of Part IV.

In the simple baseline case $(n = 2)$, Q_d reads

$$Q_d = \sigma_d^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},$$

with $|\rho| < 1$ the correlation coefficient. As shown in Fig. 18.5 of Part III, for TL2 correlation, $\rho > 0$ decreases with baseline length. When estimating the relative ZTDs, $(C_2' Q_d^{-1} C_2)^{-1} = \sigma_d^2 (1 - \rho^2)$ is used in the time update instead of Q_d. Clearly, this scalar is smaller for larger ρ. In other words, tighter constraints are applied for shorter baselines.

Means of quality assessment

28.1   Introduction

The advantage of simulations over real data processing is that it is easier to study the filter behavior itself. The sensitivity of the estimates, or other simulation results, to any of the filter options can be analyzed. We can however not conclude which of the models is best, based on simulations alone.

In this chapter, we study various ways to assess and evaluate a model on the basis of a simulation with a benchmark model. One day of GPS data is simulated for a baseline[1] with the default simulation options of Tables 27.2 and 28.1. The default options of Tables 27.3 and 28.1 were used in the parameter estimation. The filter defaults match the simulation defaults and are considered reasonable options from a theoretical point of view, like the TL2 correlation and residual STD model, or based on previous experiences with GPS data, like the standard deviations of the observations.

We have an ideal situation where the filter options most likely result in the best estimates since the chosen filter parameters match those used in the simulation. By changing any of the filter options, the quality of the estimates is likely to be degraded. In actual data processing, the results may be even worse because of orbit errors and unmodeled or incorrectly modeled error sources like multipath and loading effects.

Station 1:	Delft (X=3924687.708, Y=301132.769, Z=5001910.770)
Station 2:	Delft50 (50 km east of station Delft)
Date:	11 October 2000
Orbits:	YUMA almanac (week 59; http://www.navcen.uscg.gov/gps/default.htm)

Table 28.1   Baseline, date, and orbits of the simulation with the benchmark model.

Figure 28.1 shows a sky plot for the 24-hour simulated data set and the number of satellites for each of the 2880 epochs (30-second sampling). Figure 28.2 shows the simulated ZTDs of the two stations and their differences. Note that the ZTDs are residual delays (an a-priori delay has to be added) and may therefore be negative. The rather negative value of about -10 cm at the end of the day is caused by the random-walk assumption, which implies unbounded processes. In actual tropospheric circumstances this is possible, but will not happen often.

---

[1]Simulation and filtering in a network of stations is also possible, but the results are harder to interpret.
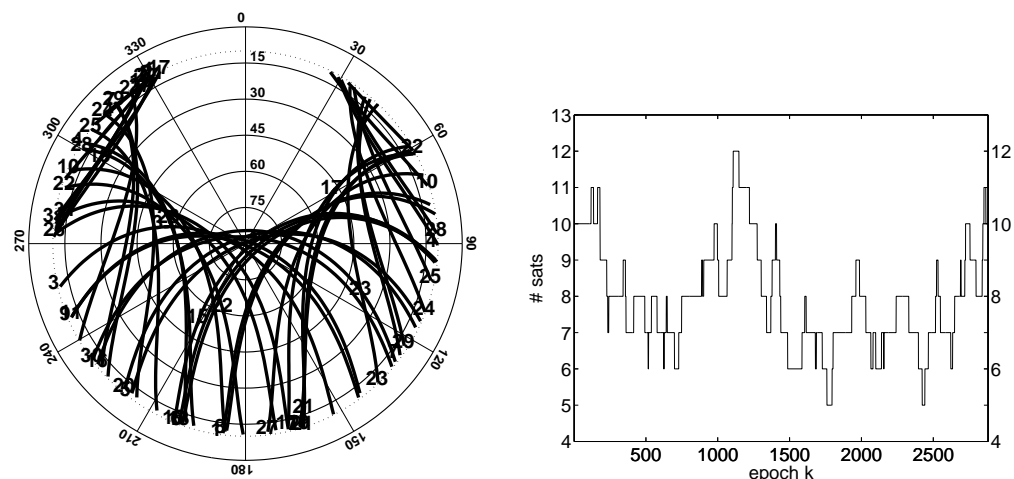
Figure 28.1   Left panel: Sky plot of station Delft, 11 Oct. 2000. Right panel: Number of visible satellites.
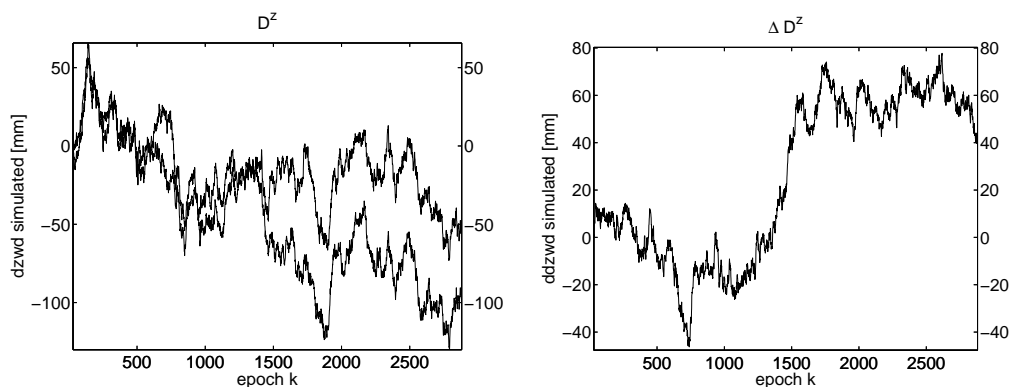

Figure 28.2   Left panel: Simulated ZTDs of stations Delft and Delft50. Right panel: Difference of simulated ZTDs (Delft50 minus Delft).

As long as only relative constraints are used in the filter this does not affect the simulation results.

The following means of quality assessment are discussed in this chapter:
1. errors of the estimated parameters;
2. precision of the estimated parameters;
3. test values;
4. MDBs (internal reliability);
5. external reliability measures.

These simulation results all depend on the chosen baseline and time of observation. The error and test values also depend on the actual noise generated.

Section 28.2 deals with the error and formal precision (standard deviation) of the estimated parameters. The first one can be computed because we know the simulated 'real' value and may be caused by the observation noise or biases. Section 28.3 deals with testing and reliability aspects.

28.2   Precision and error

The primary simulation output is the precision and error (difference between simulated and estimated values) of the coordinate estimates, especially the height component. Since we fixed the coordinates of the first station, the only coordinates we are looking at are the coordinates of the second station; the errors in these coordinates are in fact the baseline errors.
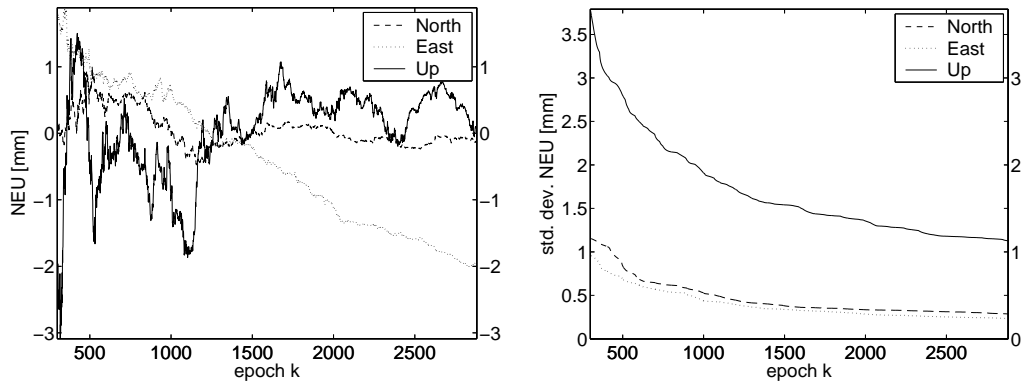
Figure 28.3   Left panel: Error in filtered North, East, and Up component. Right panel: Standard deviation of filtered North, East, and Up component. Plots start at epoch 300 ($\equiv$ 2.5 h) to avoid 'empty plots'.

Figure 28.3 shows the errors and standard deviations of the filtered coordinates. The errors in the North and Up component agree well with the 2σ boundaries. The East component however, shows a significant (negative) deviation at the end of the day. Since the baseline is east–west orientated this is a clear indication of a scale error, which is caused by the error in the (also negative) absolute troposphere delay, see Fig. 28.2, where a zero a-priori delay was assumed. A similar error in the North component was found in a simulation for a north–south orientated baseline using the same options and noise. The scaling effect of the absolute delay error was earlier described in [*Beutler et al.*, 1988]. In our case the error is introduced by the fixing of the zenith delays of one station.

As shown in Part II, there is a near rank deficiency in small networks related to satellite clocks and ZTDs. A simplified model with only these parameters and 1 observation type, 2 stations, m satellites, and 1 epoch reads:

$$\mathbf{E}\left\{\begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \end{bmatrix}\right\} = \begin{bmatrix} I_m & M_1 & 0 \\ I_m & 0 & M_2 \end{bmatrix} \begin{bmatrix} \delta s \\ D_1^z \\ D_2^z \end{bmatrix} = \begin{bmatrix} I_m & 0 & 0 \\ I_m & M_{12} & M_2 \end{bmatrix} \begin{bmatrix} \overline{\delta s} \\ D_1^z \\ D_{12}^z \end{bmatrix}, \tag{28.1}$$

where δs is the vector with satellite clock errors, y_1 and y_2 are the observations at stations 1 and 2 respectively, M_1 and M_2 are vectors with mapping values, $\overline{\delta s} \doteq \delta s + M_1 D_1^z$ are lumped satellite clock errors, $D_{12}^z \doteq D_2^z - D_1^z$ and $M_{12} \doteq M_2 - M_1$. Eliminating the column with small mapping values M_{12} then introduces a bias. The size of this bias changes with satellite configuration and is large for either large D_1^z or M_{12}. The worst case for one specific satellite is when the second station is in the same plane as the first station, the satellite, and the earth's center, as shown in App. K. Figure 28.4 shows the zenith angles as function of distance for this case and the corresponding mapping difference M_{12} for a simple secant mapping function. Because in general the stations are not in
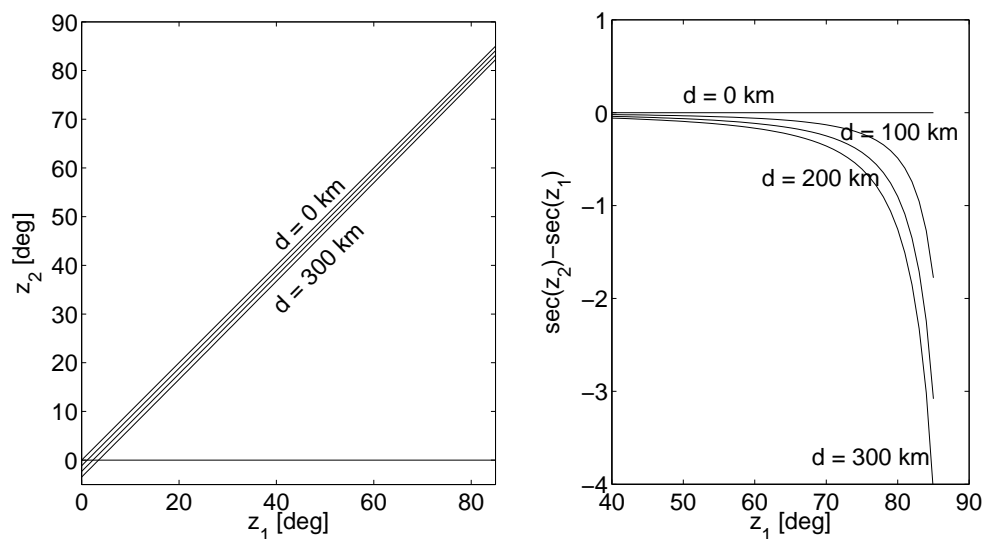
the same plane, the bias is also azimuth dependent.

Figure 28.4   Left panel: Relation between zenith angles of a GPS satellite for two stations separated by distance d = 0, 100, 200, and 300 km corresponding with Fig. K.1. Right panel: Function $M_{12} = \sec z_2 - \sec z_1$ for the same geometry.
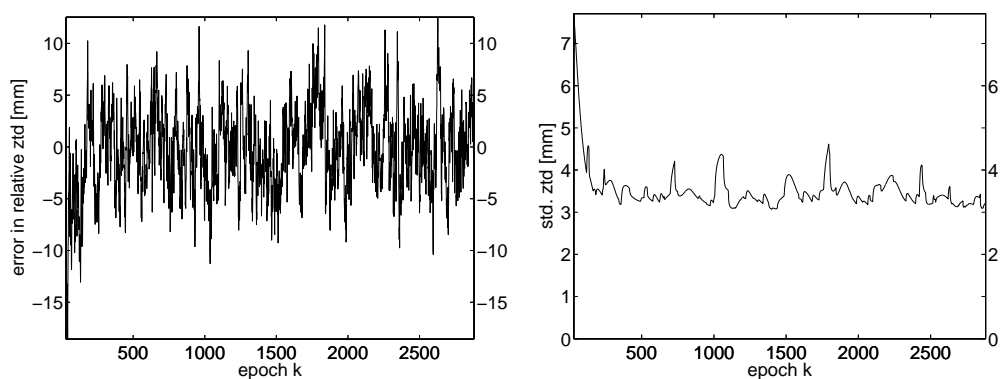

Figure 28.5   Left panel: Filtered minus simulated relative ZTD. Right panel: Standard deviation of filtered relative ZTD.

Figure 28.5 shows the error and standard deviation of the filtered relative ZTDs. After an initialization phase where the standard deviations drop from 7 to 3–4 mm, they remain rather constant at this level owing to the variance-update part of the time updates. Most of the errors in the ZTDs are within the 2σ boundaries of 6–8 mm.

28.3   Testing and reliability

Because the filtered parameters may be deteriorated by model errors, hypothesis tests are desired for detection of these errors. We distinguish six types of tests:
1. Global Overall Model (GOM) tests (1);
2. tests for outliers in the carrier phase observations ($2m_{k_0}n$);
3. tests for outliers in the code observations ($2m_{k_0}n$);
4. tests for slips in the ZTDs (n);
5. tests for (cycle) slips in the ambiguities ($2(m_{k_0} - 1)(n - 1)$);

6. tests for slips in the coordinates ($3n$),

with the number of tests in epoch k shown between parentheses; n is the number of stations[2]; m_{k_0} is the number of satellites in epoch k_0, the first epoch a model error starts occurring. Each of these tests are computed $\delta k + 1$ times, with $\delta k = k - k_0$ the maximum latency.

Except for the GOM tests, all tests are one dimensional. For each of these tests we may therefore compute the MDB as a measure of internal reliability. The external reliability we are most interested in, is described by the impact of an MDB on the height of the second station (coordinates of the first station are kept fixed). Below, we show the MDBs and external reliability for tests 2–5, but not for the last test because a change in antenna position is less likely to occur in (only) 24-hour data sets. The fourth test can detect sudden changes in the atmosphere, but is probably more theoretic than practical.

Although we may very well do these hypothesis tests in actual data processing, the actual testing (comparing test values with critical values) is not executed for our simulated data set since we do not introduce any outliers or slips to be detected. The test *values* however may be of concern as an indication of sensitivity to changes in the filter settings. For example: the values of the tests for slips in the ZTDs give an indication of an error in the stochastic model of the constraints, especially the PSD, when using zero latencies.

Because the simulation model matches the filter model, all tests have a central χ^2 distribution, and the 'square roots' of the test quantities ($\hat{\underline{\nabla}}_k / \sigma_{\hat{\nabla}_k}$) have standard normal distributions. Plots of the test values confirmed this and are not shown.

More interesting to study are the plots of the MDBs. Because there are many tests per epoch, visualization is restricted to showing only the maximum values of the MDBs and external reliabilities for outliers and ambiguity (cycle) slips. For the MDBs and external reliabilities of ZTD slips, one station suffices because those of the other station are nearly the same (in this case). The testing/reliability options are shown in Table 28.2.

symbol	range	default	unit	description
δk	[0–p]	3	[-]	maximum latency testing/reliability
α_1	[0–1]	0.001	[-]	level of significance
γ	[0–1]	0.8	[-]	power of all tests

Table 28.2   Testing/reliability options.

Figure 28.6 shows the maximum MDBs of the phase and code outliers. In both cases the MDBs are smaller for larger latencies, which means that the outliers can be better detected after a couple of extra epochs when more observations are available. This makes sense and can be seen from the testing formulas of Part IV since extra data causes $Q_{\hat{\nabla}}^{-1}$ to increase and therefore $Q_{\hat{\nabla}}$ and thus the MDBs decrease. The MDBs of the phase outliers are sensitive to changes in the satellite configuration; the large peaks occur where new satellites are being tracked. A latency of just one epoch largely reduces these MDBs.

---

[2]In the notation of Part II, not Part IV.

Contrary to the phase outliers, the MDBs of code outliers are insensitive to the satellite configuration because of their much lower weighting in the filter.
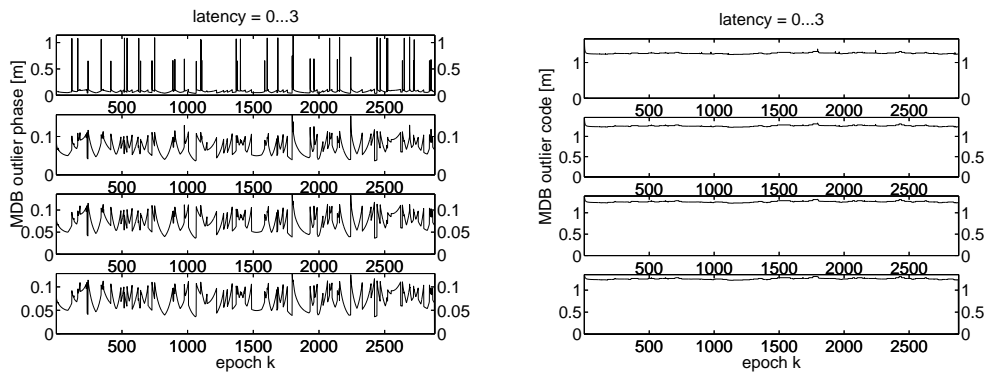

Figure 28.6   Internal reliability. Left panel: Maximum MDBs of phase outliers. Right panel: Maximum MDBs of code outliers. The latency increases from 0 in the top graphs to 3 in the bottom graphs.
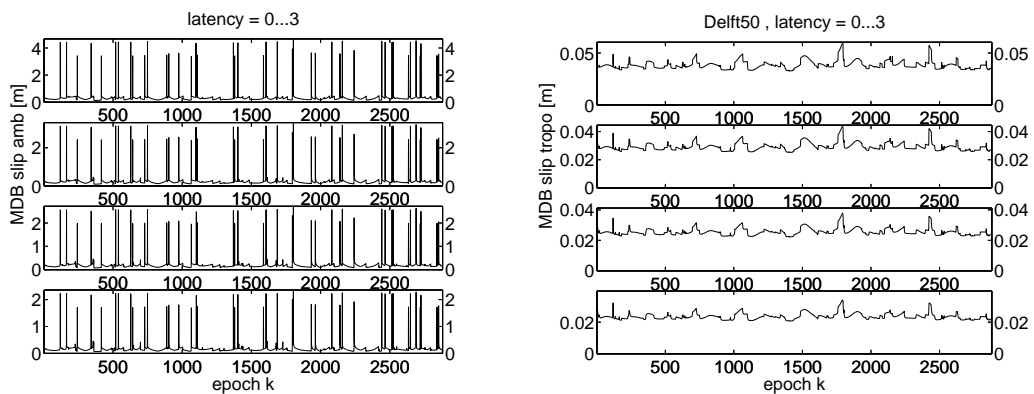

Figure 28.7   Internal reliability. Left panel: Maximum MDBs of ambiguity slips. Right panel: MDBs of ZTD slips of station Delft50.

Figure 28.7 shows the maximum MDBs of ambiguity slips and the MDBs of ZTD slips of one of the stations. Like for the MDBs of phase outliers, those of ambiguity slips show peaks when new satellites are being tracked. The ambiguity slips can only marginally better be detected after a couple of extra epochs since little new information is gathered that can be used to determine (an expired) ambiguity that has slipped immediately, that is, one epoch after tracking.
The MDBs of the ZTD slips have only a small dependence on the satellite geometry.

Figure 28.8 shows the maximum external reliabilities (impact on height component only) of phase and code outliers for epoch 300 and higher. The external reliabilities for the earlier epochs are much higher. Large observation time spans have a clear positive influence on the reliability. Increasing the latency has not much effect. Errors with the size of an MDB in the phase observations have a significant impact; those in code observations are negligible.

Figure 28.9 shows the external reliabilities of ZTD and ambiguity slips. Both decrease with larger observation time spans, but increase with the latency. This can be explained as follows: Although the MDBs decrease, their impact on the height increases, since the error is maintained for more epochs.
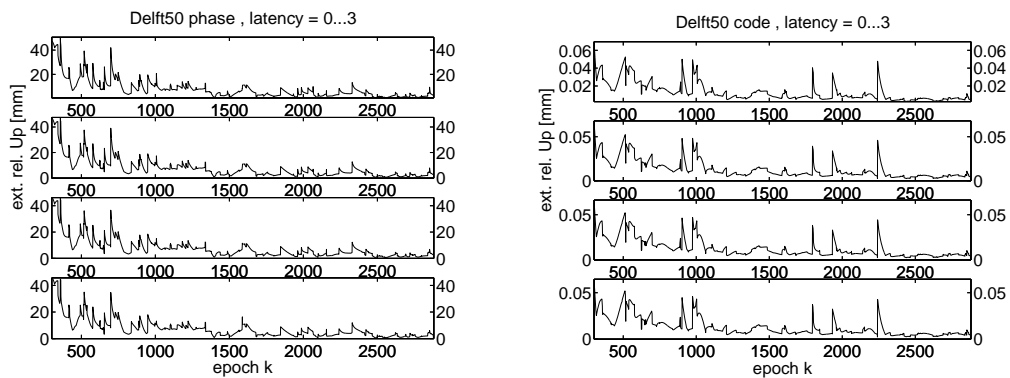
Figure 28.8   External reliability. Left panel: Maximum error of phase outlier on the filtered (baseline) height. Right panel: Maximum error of code outlier on the filtered (baseline) height. Plots start at epoch 300.
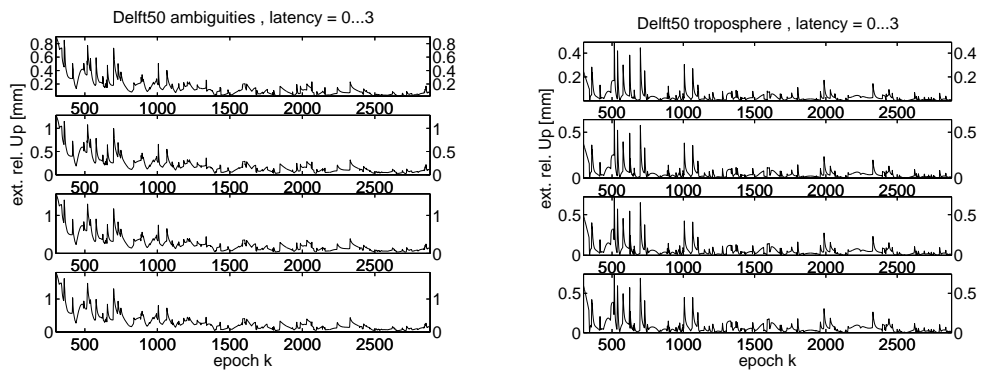

Figure 28.9   External reliability. Left panel: Maximum error of ambiguity slip on the filtered (baseline) height. Right panel: Maximum error of ZTD slip on the filtered (baseline) height. Plots start at epoch 300.

Simulation scenarios

29.1 Introduction

This chapter describes a selection of simulation results as found with the software described in Chap. 27. Section 29.2 first compares the formal precision of coordinates and ZTDs for several models. Section 29.3 then compares the impact of a wrong covariance matrix on the actual precision. Section 29.4 gives some of the results of Sects. 29.2 and 29.3 for an increased standard deviation of the phase observables.

29.2 Formal precision

In this section we only compare the formal standard deviations of the relative ZTDs and coordinates in several scenarios. The scenarios are subdivided into seven groups and are shown in Table 29.1. This table also shows Overall Model Test (OMT) values of specific simulations per scenario; see Sect. 24.3 in Part IV. Especially interesting to compare are the scenarios within these groups.

coordinates   As can be seen from Fig. 28.3, the standard deviations of the coordinates are slowly decreasing with the number of epochs. Except for the scenarios where we mix fixed and float ambiguities (see Fig. 29.3), the quotients of coordinate standard deviations of two scenarios do not depend on the observation time strongly. Therefore only the values after 24 hours are presented in Table 29.1.

ZTDs   After an initial decrease, the standard deviations of the ZTDs fluctuate around a stable value; see Fig. 28.5. The standard deviations in Table 29.1 are therefore mean values after the initial decrease (the period 2.5–24 hours was used).

OMT   The OMT values are shown rather as estimates of the variance of unit weight than as test values. If we use them as test values, we need to know the degrees of freedom. For the scenario indicated by '∗2', for example, we have a redundancy r=59708 and critical value for a 5% level of significance of 1.0095. The precision of this estimate is [*Koch*, 1987]: $\hat{\sigma}_{\mathrm{OMT}} = \mathrm{OMT} \cdot \sqrt{2/r} = $ 6e-3.

The OMT/variance of unit weight is sometimes used to scale the covariance matrix of the observations and consequently also the covariance matrix of the estimated parameters; the standard deviations are thus scaled with the square root of the OMT. This gives a better precision description when the a-priori variance factor is unknown. The formal standard deviations are unscaled, meaning that the a-priori variance factor is assumed known.

		ZTD [mm]	North [mm]	East [mm]	Up [mm]	OMT [-]
1.	1 epoch batch, tropo-float (∗5)	8.75	0.29	0.23	1.14	0.95
	1 epoch batch, tropo-weighted (∗2)	*4.81*	*0.29*	*0.23*	*1.13*	*1.00*
	12 epochs batch, tropo-float	4.23	0.29	0.23	1.13	(1.01)
	12 epochs batch, tropo-weighted	3.85	0.29	0.23	1.13	(1.01)
	240 epochs batch, tropo-float	1.29	0.28	0.23	1.09	(1.19)
	240 epochs batch, tropo-weighted	1.29	0.28	0.23	1.09	(1.19)
	2880 ep. batch, tropo-constant	0.50	0.28	0.23	1.08	(4.39)
	tropo-fixed (∗4)	0.00	0.28	0.23	0.57	(5.12)
2.	250 km, absolute troposphere	4.95	0.29	0.33	1.21	1.00
	250 km, relative troposphere	4.88	0.29	0.24	1.15	1.05
	50 km, absolute troposphere	4.82	0.29	0.30	1.15	0.99
	50 km, relative troposphere (∗2)	*4.81*	*0.29*	*0.23*	*1.13*	*1.00*
	10 km, absolute troposphere	4.80	0.28	0.29	1.12	0.99
	10 km, relative troposphere	4.80	0.28	0.23	1.12	0.99
3.	250 km, absolute troposphere	4.26	0.29	0.33	1.22	1.00
	250 km, relative troposphere	3.95	0.29	0.24	1.15	1.05
	50 km, absolute troposphere	3.54	0.29	0.30	1.15	0.99
	50 km, relative troposphere	*3.43*	*0.29*	*0.23*	*1.13*	*1.00*
	10 km, absolute troposphere	2.83	0.28	0.29	1.11	0.99
	10 km, relative troposphere	2.80	0.28	0.23	1.11	0.99
4.	tropo-float, amb.-float, abs.	9.46	0.30	0.34	1.28	0.90
	tropo-float, amb.-float, rel. (∗5)	8.75	0.29	0.23	1.14	0.95
	tropo-float, amb.-fixed, rel.	8.65	0.23	0.18	0.94	0.95
	tropo-fixed, amb.-float (∗4)	0.00	0.28	0.23	0.57	(1.72)
	tropo-fixed, amb.-fixed	0.00	0.23	0.18	0.47	(1.87)
5.	Q_y diag/H2, amb.-float (∗1)	6.97	0.56	0.43	2.43	0.75
	Q_y diag/H2, amb.-fixed	6.61	0.45	0.31	1.95	0.75
	Q_y full/H2, amb.-float (∗2)	*4.81*	*0.29*	*0.23*	*1.13*	*1.00*
	Q_y full/H2, amb.-fixed	4.68	0.23	0.18	0.94	1.00
	Q_y unit, amb.-float (∗3)	2.53	0.17	0.14	0.75	1.84
	Q_y unit, amb.-fixed	2.44	0.13	0.10	0.63	1.86
6.	Q_y diag/H2 (∗1)	6.97	0.56	0.43	2.43	0.75
	Q_y diag/H1	4.75	0.32	0.25	1.41	0.82
	Q_y full/H2 (∗2)	*4.81*	*0.29*	*0.23*	*1.13*	*1.00*
	Q_y full/H1	3.34	0.21	0.17	0.86	1.00
	Q_y unit (∗3)	2.53	0.17	0.14	0.75	1.84
7.	Q_y diag/H2, 70 deg	10.03	0.62	0.46	4.47	0.76
	Q_y diag/H2, 80 deg (∗1)	6.97	0.56	0.43	2.43	0.75
	Q_y full/H2, 70 deg	6.54	0.33	0.26	2.21	1.00
	Q_y full/H2, 80 deg (∗2)	*4.81*	*0.29*	*0.23*	*1.13*	*1.00*
	Q_y unit, 70 deg	5.14	0.25	0.19	1.75	1.25
	Q_y unit, 80 deg (∗3)	2.53	0.17	0.14	0.75	1.84

Table 29.1   Formal standard deviations of parameters, and Overall Model Test values. Standard deviations after processing 24 hours for coordinates and mean values for ZTDs over 2.5–24 hours. Asterisks indicate identical scenarios (∗1–3) or near identical scenarios (∗4–5). The scenarios are according to the defaults of Table 27.3 except for spatial correlation. A-priori standard deviation of phase observables: 2 mm. Scenarios under **3** and **4**: TL2 correlation, otherwise uncorrelated (in simulation and estimation). Italicization: (near) default scenarios.
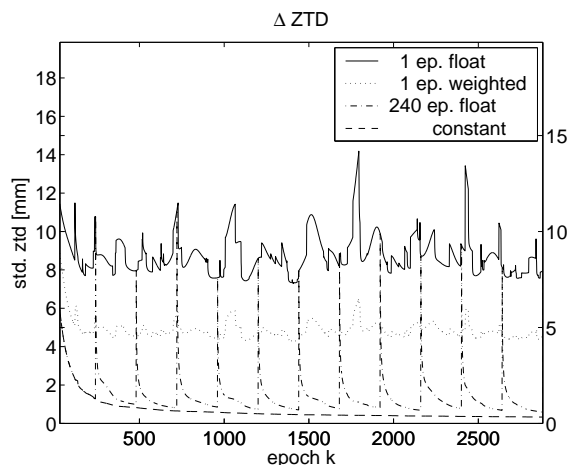
Figure 29.1 Formal standard deviations of relative ZTDs for four different scenarios: the troposphere-float and a troposphere-weighted model for a batch size of 1, the troposphere-float model for a batch size of 240, and the troposphere-constant model. Defaults of Table 27.3; no spatial correlation assumed.

When the filter model differs from the model used in the simulation, the expected OMT value may deviate from 1. If this model difference is other than the a-priori variance factor, scaling the covariance matrix by the OMT gives an estimated but incorrect description.

1st group: weighting and batch size

In the first group, the troposphere-float, weighted, constant, and fixed models are compared. The troposphere-float and weighted models have batch sizes of 1, 12, and 240 epochs. The batch size has a strong influence on the formal precision of the ZTDs: The more epochs/observations are used to estimate a ZTD, the more precise this ZTD will be. Large batches show smaller differences between troposphere-weighted and float models because the constraints are relatively weaker.

The troposphere-constant model is a limiting case for both the weighted and float models, cf. Fig. 29.1. The standard deviations of the horizontal coordinates are hardly affected by the batch size or weighting. The Up component is just slightly more affected. App. L shows why the troposphere weighting has so little effect on the precision of the height. Only fixing the troposphere parameters has a large effect on the formal precision of the Up component.

In the specific simulation of Table 29.1, we see that the troposphere-float model performs best when the standard deviations are scaled with the square root of the OMT (e.g.:$\sqrt{0.95} \cdot 1.14 = 1.11 < \sqrt{1.19} \cdot 1.09 = 1.19$). Because the OMT values depend on the simulated troposphere, this is not necessarily so in general. Especially the OMT value for the troposphere-fixed model depends heavily on the (unbounded) random-walk simulated ZTDs. Because of this, in Chap. 30 we reconsider the applicability of this model.

Figure 29.2 illustrates the effect of weighting and batch size on the ZTD precision. The numbers differ from Table 29.1 because this figure is based on a processing of only four hours and two-hour means to limit computation time.

2nd/3rd group: baseline lengths

With the scenarios of the second and third group we can compare the influence of the baseline length. In the third group TL2 correlation is assumed
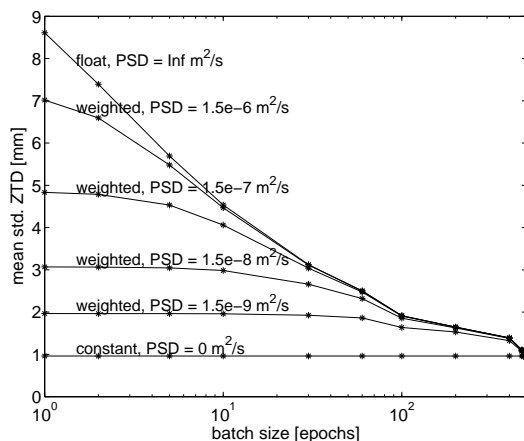
Figure 29.2   Mean formal standard deviations of ZTDs as function of batch size and Power Spectral Density, based on simulations from 0.00–4.00 h and means over 2.00–4.00 h. Defaults of Table 27.3; no spatial correlation assumed. Computed values only at asterisks; lines between asterisks are shown for clarity.

(both in simulation and estimation), whereas in the second group no correlation is assumed. For three baseline lengths, both absolute troposphere (one parameter per station) and relative troposphere (parameters of one station fixed to zero) are estimated, using different runs.

In the model without spatial correlation the precision of the ZTDs is only slowly decreasing with increasing baseline length, probably due to the decreasing number of visible low-elevation satellites common to both stations. The same effect can be seen in the height component. In the model with spatial correlation, the higher correlation between troposphere parameters for short baselines causes the relative ZTDs to be more precise, but the precision of the estimated heights is hardly affected. With or without correlation, in both cases estimating relative delays instead of absolute delays results in only slightly more precise ZTDs. The precision of the East component gains more from estimating relative delays, which is due to the east–west orientation of the baseline. By estimating relative delays, the formal precision of the baseline length improves, but, as we saw in Sect. 28.2, this may cause a bias in the baseline length.

For the Up component, the difference in precision when estimating either absolute or relative delays is very small. To avoid biases in the estimates, estimating absolute delays is therefore preferred. This finding agrees with [*Brunner and McCluskey*, 1991].

4th group: fixing parameters

In the fourth group we compare the fixing and estimation of troposphere and ambiguity parameters. (Note that the difference in OMT values for scenarios *4 in groups 1 and 4 are due to a different ZTD simulation.) Whereas in the relative model only the ZTDs of one station are fixed, in the troposphere-fixed model the ZTDs of all (both) stations are fixed. In other words, the standard deviations of the (relative) ZTDs are zero by definition. Clearly, fixing both ZTDs and ambiguities gives the best precision obtainable, especially for the height. Fixing ZTDs has a larger impact than fixing ambiguities, but this is however only true for long time spans as illustrated in Fig. 29.3. The observation model (*2) is strong enough for successful fixing of the ambiguities. After an initialization time of 20 epochs a maximum probability of unsuccessful fixing (fail rate) of the

ambiguities based on integer bootstrapping [*Teunissen*, 1998] was found to be 3e-5. When using the diagonal covariance matrix of the observations (model $*1$) the fail rates are much higher. A maximum was found of 0.21 directly after the rise of a new satellite; see Fig. 29.3. After a few epochs the fail rate gradually drops.
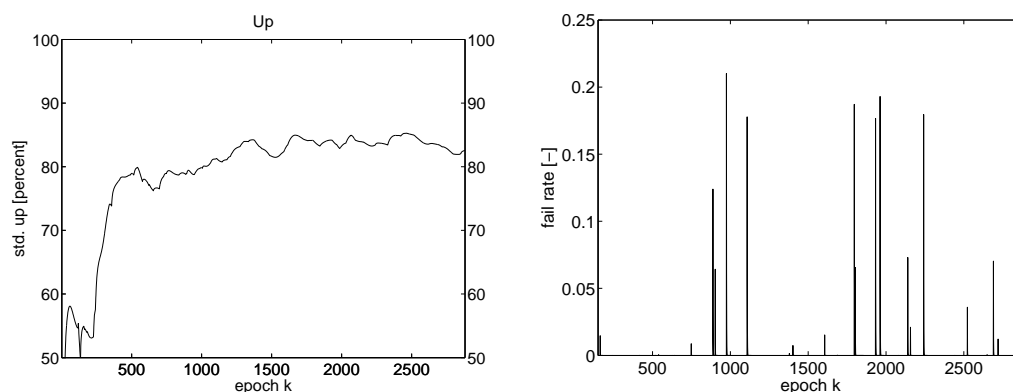

Figure 29.3   Left panel: Formal standard deviation of filtered Up component in the ambiguity-fixed model in percent of the standard deviations in the ambiguities-float model (troposphere-float model, relative ZTDs estimated). Fixing ambiguities has a large effect on the standard deviation for short time spans but the effect levels out to a fixed improvement for long time spans. Right panel: Fail rates for model $*1$ (diagonal covariance matrix) for epochs 20–2880.

5th–7th group: cov. matrix of observations
In the fifth to seventh group we compare the impact of different covariance matrices of the observations. The three covariance matrices correspond to Eq. (27.3): a scaled unit covariance matrix for the phase and code observations, a full covariance matrix in which residual STDs are stochastically modeled and a diagonal covariance matrix with zenith angle dependent weighting. The latter one is the diagonal part of the full covariance matrix. In Table 29.1, H2 and H1 indicate effective heights of 2 and 1 km.

5th group
The results of the fifth group show that, like in the fourth group, (correct) ambiguity fixing has a more beneficial effect on the coordinates than on the ZTDs. Furthermore, they show that the unit covariance matrix gives the best formal precision for both ZTDs and coordinates. This is however a too optimistic covariance matrix as reflected in the OMT. That the unit covariance matrix, $Q_{y,}(k)$, gives better formal parameter precision than the full covariance matrix, $Q_{y,}(k) + e_4 e_4' \otimes Q_{\Delta D}(k)$, is understandable since the latter matrix is the sum of two positive-definite matrices and therefore implies lower formal precision of the observations. The diagonal covariance matrix turns out to give less precise formal parameter estimates than the full covariance matrix, which is not so obvious (scaling with the OMT does not change this). Apparently, the combination of design matrix and the (positive) correlation between observations to different slant directions and between observation types has a beneficial effect. We can understand this from the pre-elimination step where differences are formed between observations to different satellites and between frequencies (linear combination). When the undifferenced observations are positively correlated, the differenced observations are more precise than when the undifferenced observations are uncorrelated.

6th group
From the scenarios of the sixth group we can see the strong dependence of the effective height (2 km, 1 km) used in the stochastic modeling of the residual slant delays. It acts as a scaling factor (power 8/3th) of the residual STD part of

the covariance matrix. Note that the unit covariance matrix is actually either a full or diagonal covariance matrix with an effective height of zero.

7th group

The seventh group shows the importance of the geometry-improving low-elevation satellites for both the precision of the ZTDs and the height. The horizontal coordinates also improve in precision owing to the increase of the number of observations. As can be seen from Table 29.2, which shows the improvements in percentages, the improvement is largest for the unit covariance matrix since in the other two cases, observations to low satellites are down weighted. The effect is smaller when scaling with the OMT, because noisy observations are then too highly weighted.

Q_y	ZTD [%]	Up [%]
diagonal/H2	31	46
full/H2	26	49
unit	51 (40)	57 (48)

Table 29.2   Percentages improvement in formal standard deviations, $(1 - \sigma_{80}/\sigma_{70}) \times 100\%$, when increasing the zenith cut-off angle from 70 to 80 degrees for different covariance matrices of the observations. Scenarios of seventh group of Table 29.1. Between parentheses: percentages when taking the OMT into account.

percentage improvement

In Table 29.3, different modeling options are compared that may improve the formal precision. The table shows the improvement in terms of percentage of the parameter standard deviations. Since these numbers depend on combinations of options, they are only rough numbers that are derived from Table 29.1.

impact of options

Table 29.3 shows that three options could be important for improving the precision of the height. Fixing the troposphere may be very beneficial for the formal standard deviation, but this may deteriorate the actual precision when (relative) troposphere values deviate from the fixed values, which is especially likely for long baselines. Increasing the zenith cut-off angle is also very beneficial for the precision of the height. Increasing it further than 80° is even more beneficial, see Fig. 29.4. On the other hand this may introduce (larger) errors caused by, for example, multipath. The last and most revealing improvement option is the full covariance matrix that incorporates residual STDs. This option not only improves the precision of the height, but also of the horizontal coordinates.

fixing ambiguities

Fixing ambiguities also gives a significant improvement of the precision of heights and horizontal coordinates, but for long observation times this is less pronounced than the first three options. Ambiguity resolution is also important in the presence of many cycle slips (for example at low elevations).

batch size

The other three options are of smaller significance. The batch size has hardly any impact on the formal coordinate precision. This suggests that we might as well use small batches to avoid mapping of troposphere errors in the height component at small cost of formal precision reduction. In general, the best batch size is hard to determine a priori, because it depends on the actual troposphere behavior. Small batches avoid the errors introduced by the extreme weather conditions, whereas on quiet days larger batches can be used. However, to be on the save side, it is better to use small batches, as illustrated by the OMT

Improvement	ZTD [%]	N/E [%]	Up [%]
Troposphere fixing vs. rel. troposphere est.	100	0–3	50
Cut-off angle ($z = 80°$ vs. $z = 70°$)	30	10	50
Residual STD model vs. diagonal (H = 2 km)	30	45	50
Ambiguities fixed vs. float	0–5	25	20
Batch size (1 day vs. 1 epoch)	94	0–3	5
Relative instead of absolute troposphere*	0–10	0–30	0–5
Spatial correlation* vs. no correlation	15–40	0–1	0–1

Table 29.3   Indicative percentage of improvement in formal standard deviation for different modeling techniques. This table is based on Table 29.1. * Baseline length: 10–250 km.

values in Table 29.1.
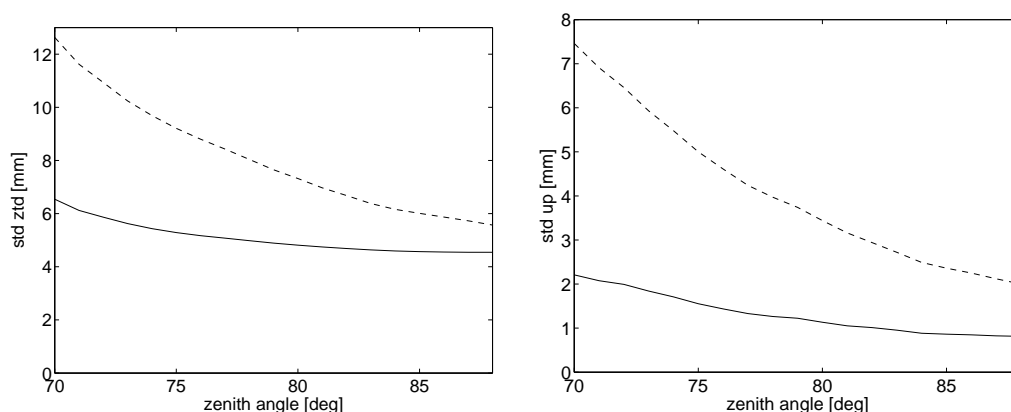
Figure 29.4   Formal standard deviations of filtered parameters as function of zenith cut-off angles. Left panel: Mean values for relative ZTDs over 2.5–24 h. Right panel: For height differences after 24 h. Solid lines: A-priori standard deviation of phase observables of 2 mm. Dashed lined: A-priori standard deviation of phase observables of 10 mm. Defaults of Table 27.3; no spatial correlation assumed.

abs./rel. troposphere

Fixing the troposphere of one of the stations (estimating relative delays) introduces a bias that depends on the size of the (error in the fixed) tropospheric delay, whereas estimating absolute delays deteriorates the formal coordinate precision. Either way the impact of estimating absolute or relative delays is largest on the horizontal coordinates. The precision of the height is only weakly affected.

spatial correlation

The impact of introducing spatial correlation on the coordinate precision is very small. It has larger impact on the estimated ZTDs. Because of this small effect most scenarios stick to the more conventional modeling without spatial correlation assumed.

29.3   Impact of the stochastic model

In the previous section we showed that the formal parameter precision based on the full covariance matrix of the observations is better than based on the diagonal covariance matrix. Since a diagonal covariance matrix is 'standard' in GPS data processing (although with a different zenith angle dependent weighting) it is

interesting to see what happens with the actual precision when we assume the full covariance matrix to describe the actual precision and use the diagonal matrix in the filter. But since the full covariance matrix is still a theoretical model, it still needs to be validated. If the diagonal matrix would actually describe a realistic precision and the full matrix were to be used in the filter, what would then be the impact on the coordinate precision? In both cases we would use another covariance matrix than the 'real' one. So in both cases the parameter precision would be lower than optimal, because only when the inverse of the real covariance matrix is used as weight matrix, the least-squares filter gives minimum variance [*Teunissen*, 1999].
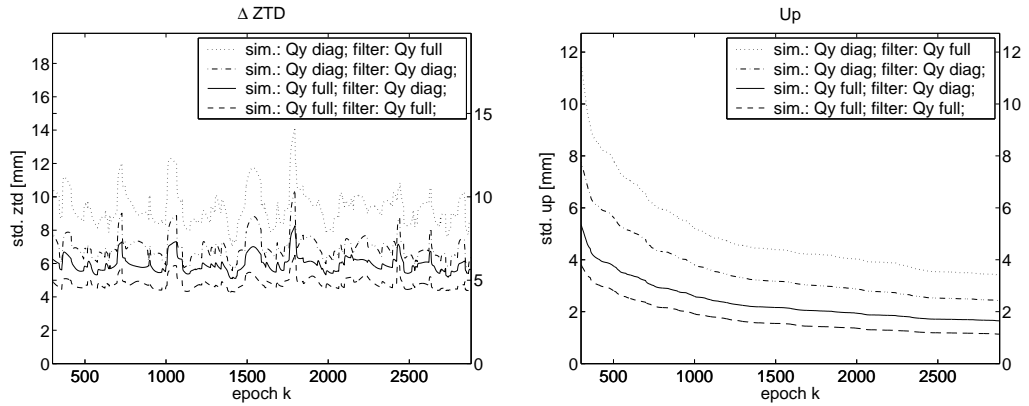
Figure 29.5 Standard deviations of filtered parameters for four scenarios: filtering with assumed real covariance matrix of observations (full/diagonal) and with wrong covariance matrix (ditto). Left panel: Relative ZTDs. Right panel: Baseline heights. Plots start at epoch 300 ($\equiv 2.5$ h).

In the two above given situations, we want to compare the 'actual' covariance matrices of the parameters using two different filters. A simultaneous computation with different covariance propagation can be implemented as shown below.

propagation of variances

The measurement-update equation reads in the notation of Part IV:

$$\hat{\beta}x_{k|k} = K_k\, y_k + (I - K_k BA_k)\, \hat{\beta}x_{k|k-1}, \tag{29.1}$$

with K_k the gain matrix, BA_k the design matrix, y_k the observations, and $\hat{\beta}x_{k|k-1}$ and $\hat{\beta}x_{k|k}$ the predicted and filtered state vector respectively. The covariance matrix of the filtered state vector, $Q_{\hat{\beta}x_{k|k}}$, can be found by using the propagation law of variances as:

$$Q_{\hat{\beta}x_{k|k}} = K_k Q_{y_k} K_k' + (I - K_k BA_k)\, Q_{\hat{\beta}x_{k|k-1}}\, (I - K_k BA_k)'. \tag{29.2}$$

Since the gain matrix is computed with $Q_{\hat{\beta}x_{k|k}}$ itself, this equation is not used in the processing. But if the actual precision of the observations can be described by $\widetilde{Q}_{y_k}$, while we are filtering with an assumed covariance matrix Q_{y_k}, then the actual parameter precision can be described by:

$$\boxed{\widetilde{Q}_{\hat{\beta}x_{k|k}} = K_k \widetilde{Q}_{y_k} K_k' + (I - K_k BA_k)\, \widetilde{Q}_{\hat{\beta}x_{k|k-1}}\, (I - K_k BA_k)',} \tag{29.3}$$

which can be computed simultaneously with the formal but assumed unrealistic $Q_{\hat{\beta}x_{k|k}}$.

diagonal or full covariance matrix

The standard deviations of the filtered ZTDs and heights, as computed by Eq. (29.3), are shown in Fig. 29.5 for all four cases. Table 29.4 gives the standard deviations of the coordinates after processing 24 hours and mean values for the ZTDs (2.5–24 hours) as well as OMT values. Figure 29.5 and Table 29.4 confirm that processing with the wrong covariance matrix deteriorates the parameter precision. They also show that when we process with the diagonal covariance matrix when the full matrix is realistic it would still give us a better precision than when the diagonal matrix would be realistic. In other words, the correlation between observations has a favorable effect on the precision whether we model it or not. The OMT value of 0.75 confirms that the (transformed) observations appear to be more precise when they are correlated. The actual precision improvement when the full matrix would indeed be the true one reduces therefore considerably $(100 \times (1 - 1.13/1.65) = 32\%$ instead of $100 \times (1 - 1.13/2.43) = 53\%$ in the height; see Table 29.4). But when we use the full covariance matrix when the diagonal matrix is more realistic we would deteriorate the precision of the parameters $(100 \times (1 - 3.42/2.43) = -41\%$ for the height). These numbers (see sixth column of Table 29.4) would be quite different when using the OMT for scaling the covariance matrix, indicating the roughness this method.

$\widetilde{Q}_{y_k}$ (sim.)	Q_{y_k} (filter)	ZTD [mm]	North [mm]	East [mm]	Up [mm]	OMT [-]
diag	full	9.39	0.70	0.53	3.42 (2.04)	3.25
diag	diag ($*1$)	6.97	0.56	0.43	2.43	1.00
full	diag	5.97	0.37	0.29	1.65 (2.10)	0.75
full	full ($*2$)	4.81	0.29	0.23	1.13	1.00

Table 29.4   Actual standard deviations (computed with Eq. (29.3)) and OMT values for four different scenarios where covariance matrix Q_{y_k} is used in the filter while $\widetilde{Q}_{y_k}$ describes the precision of the simulated observations. Standard deviations of coordinates after 24 hours and mean values for ZTDs over 2.5–24 hours. Defaults of Table 27.3; no spatial correlation assumed. Asterisks indicate scenarios in Table 29.1. Between parentheses in sixth column: standard deviations we would get by scaling with the OMT ($\sqrt{3.25} \cdot 1.13 = 2.04$, $\sqrt{0.75} \cdot 2.43 = 2.10$).

$\widetilde{Q}_{y_k}$ (simulated)	Q_{y_k} (filter)	ZTD [mm]	North [mm]	East [mm]	Up [mm]	OMT [-]
full/H2	full/H1	5.47	0.32	0.25	1.22	1.34
full/H1.9	full/H2	4.82	0.29	0.23	1.13	1.02
full/H2	full/H2	4.81	0.29	0.23	1.13	1.00
full/H2	full/H1.9	4.67	0.29	0.23	1.10	0.98
full/H1	full/H2	3.78	0.22	0.18	0.91	0.88
full/H1	full/H1	3.34	0.21	0.17	0.86	1.00

Table 29.5   Standard deviations and OMT values for six different scenarios where covariance matrix Q_{y_k} is used in the filter while $\widetilde{Q}_{y_k}$ describes the precision of the simulated observations. Standard deviations of coordinates after 24 hours and mean values for ZTDs over 2.5–24 hours. Defaults of Table 27.3; no spatial correlation assumed.

impact of effective heights

Table 29.5 shows the effect of using an effective height of 1 km and 1.9 km respectively, instead of 2 km. The precision of the estimates only shows to be sensitive to a large difference in this parameter.

cov. matrix of constraints

Above we saw the impact on the parameter precision when we filter with a

wrong covariance matrix of the observations. In a very analogue way we may study the impact of a wrong covariance matrix of pseudo-observations like the spatiotemporal constraints on ZTDs. If the 'real' covariance matrix is $\widetilde{Q}_d$ whereas Q_d (or $(C'_n \widetilde{Q}_d^{-1} C_n)^{-1}$ and $(C'_n Q_d^{-1} C_n)^{-1}$ respectively; see App. J) is used in the filtering, the covariance matrix of the predicted state vector is computed by

$$\widetilde{Q}_{\hat{\beta} x_{k|k-1}} = \widetilde{Q}_{\hat{\beta} x_{k-1|k-1}} + \begin{bmatrix} \widetilde{Q}_d & 0 \\ 0 & 0 \end{bmatrix}, \tag{29.4}$$

with which the covariance matrix of the filtered state vector is computed using Eq. (29.3).

PSD   Table 29.6 shows the impact of a wrong scaling of the covariance matrix of these constraints caused by a wrongly assumed PSD.

PSD in sim. [m²/s]	PSD in filter [m²/s]	std. ZTD [%]	std. N/E [%]	std. Up [%]	test value [-]	OMT [-]
1.5e-7	1.5e-7	100	100	100	0.99	1.00
1.5e-7	1.5e-5 (u)	169	100	100	0.79	0.96
1.5e-7	1.5e-6 (u)	129	100	100	0.80	0.98
1.5e-7	1.5e-8 (o)	137	101	104	1.20	1.02
1.5e-7	1.5e-9 (o)	249	128	149	1.56	1.08
1.5e-5	1.5e-7 (o)	632	107	121	4.10	2.13
1.5e-6	1.5e-7 (o)	213	101	102	1.56	1.10
1.5e-8	1.5e-7 (u)	80	100	100	0.91	0.99
1.5e-9	1.5e-7 (u)	78	100	100	0.91	0.99

Table 29.6   Actual standard deviations (using Eqs. (29.4) and (29.3), with $\widetilde{Q}_{y_k} = Q_{y_k}$) in percentages of reference case (top row) for different scenarios where some PSD is used in the filter while another PSD describes the stochastic behavior of the simulated ZTDs. (u): underconstraining; (o): overconstraining. Test value: Mean value of square root of troposphere slippage tests, station Delft, latency 0 (test quantity has a standard normal distribution). Standard deviations of coordinates after 24 hours and mean values for ZTDs over 2.5–24 hours. Defaults of Table 27.3; no spatial correlation assumed.

In (most of) the previous simulations we assumed a filtering with a PSD of 1.5e-7 m²/s whereas the ZTDs were driven by a PSD of the same size. The second to fifth row of Table 29.6 show the impact on the precision of the ZTDs and coordinates when we filter with a ten or hundred times smaller or larger PSD. The sixth to last row show the impact when the actual driving PSD is ten or hundred times smaller or larger than in the filter.

The first five rows of Table 29.6 show that filtering with the wrong PSD results in a lower parameter precision. The ZTD is the most sensitive parameter to the choice of the PSD, and the Up component is more sensitive than the horizontal coordinates. The precision of the Up component is hardly affected by underconstraining. Overconstraining has more impact, but a factor ten still has not much impact.

The last four rows again show the precision-deteriorating impact of over-constraining, but they also show that the ZTDs can be estimated more precisely when the troposphere has a less fluctuating behavior than assumed (the pseudo-

observations/constraints are more precise than assumed). The impact on the Up component is however negligible.

The results of Table 29.6 confirm the simulation results of [*Jarlemark et al.*, 1998] with the GIPSY software [*Webb and Zumberge*, 1993]. They used actual RMS values of the parameter estimates after hundred troposphere simulations.

Table 29.6 also shows troposphere slippage tests for one specific simulation. The stronger the overconstraining, the larger these test values. The slippage test is a tool that can be used to detect and avoid overconstraining. A better tool would however be an Overall Model Test for constraints only.

29.4   Increased noise level

The simulations of the previous section show that, in absence of biases, a (nearly) baseline-length independent (at least up to 250 km) intrinsic accuracy of GPS baseline heights is obtainable at the 1-mm level for 24-hour data sets. The precision that was derived from a time series analysis of daily estimates in the AGRS.NL network (see 'General introduction') was in order of 5 mm standard deviation. The difference is probably due to several error sources that were not considered in the simulations. Some important error sources are multipath and unmodeled or incorrect phase center variations. Other error sources may be: errors in the mapping function; wrongly used a-priori hydrostatic delays; wrong fixing of coordinates of the reference station; orbit errors; atmosphere and ocean loading effects; delays caused by rain drops, ice, etc.; higher order ionospheric delays; undetected observation errors; and possibly antenna instability. Although not all of these error sources are large and reasonable candidates to account for the found difference in precision, they may accumulate in unexpected ways. The sum of all these error sources add up to a lower precision of the observations. Often this is modeled by a scaling of the covariance matrix of the GPS (phase) observables. Although this is strictly not correct because these error sources tend to give time-correlated observations, in this section we assume this to be a reasonable choice to see what happens under these extreme conditions. In Tables 29.7 and 29.9–29.11 the standard deviation of the phase observables is assumed to be 1 cm. The other options are the same as in the previous sections.

Improvement	ZTD [%]	N/E [%]	Up [%]
Troposphere fixing vs. rel. troposphere est.	100	0–6	60
Cut-off angle ($z = 80°$ vs. $z = 70°$)	40	25	50
Residual STD model vs. diagonal (H = 2 km)	10	10	15
Ambiguities fixed vs. float	0–5	20	15
Batch size (1 day vs. 1 epoch)	85	5	5
Relative instead of absolute troposphere*	0–10	0–35	0–5
Spatial correlation* vs. no correlation	15–40	0–1	0–3

Table 29.7   Indicative percentage of improvement in formal standard deviation for different modeling techniques. This table is based on Table 29.9. Standard deviation phase observables: <u>1 cm</u>.
* Baseline length: 10–250 km.

In Table 29.9 we see how this increased standard deviation translates into a

lower formal precision of the parameter estimates. Table 29.7 shows the impact of modeling options on the formal precision. We see an increased impact of troposphere fixing and a decreased effect of ambiguity fixing. Unfortunately the impact of the residual STDs is also decreased largely. This is caused by the relatively low weighting of these constraints with respect to the phase observables. The covariance matrix is dominated by the diagonal part containing the variances of the phase observables. This also has effect on the impact of the zenith cut-off angle as can be seen in Fig. 29.4. Since the observations at lower elevations are relatively less down-weighted, increasing this angle above 80 degrees has more impact.

Tables 29.10 and 29.11 show that the decreased impact of the residual STDs also implies less sensitivity to a wrong choice of covariance matrix (full versus diagonal; wrongly chosen effective height). The OMT values are closer to one, indicating 'less stress' in the model. The Up component improves less by using a full covariance matrix than in the scenarios with 2 mm standard deviation for the phase observables: $\pm 15\%$ formal precision improvement and $\pm 10\%$ actual improvement.

In Table 29.12 we see the impact of under- and overconstraining. Again, overconstraining has a larger impact than underconstraining. The effect of constraining is however not unambiguously larger or smaller than in Sect. 29.3.

29.5   MDBs

The better formal accuracy of the observations when using a full covariance matrix instead of a diagonal one has consequences for the testing and reliability. As shown by Figs. 29.6 and 29.7 in the model with a full covariance matrix, the MDBs for both outlier and cycle slip detection are smaller than with the diagonal covariance matrix. In other words, formally, smaller errors can be detected when modeling residual STDs. Both the mean values of the maximum MDBs per epoch of phase outliers and of cycle slips reduce (depending on the latency 1–4); see Table. 29.8.

std. phase obs. [mm]	cov. matrix	MDBs phase outlier [cm]	MDBs ambiguity slip [cm]
2	diag	26–38	101–150
2	full	7–20	64– 80
10	diag	29–41	108–163
10	full	14–27	80–110

Table 29.8   Means of maximum MDBs per epoch for outliers in phase observables and ambiguity slips for four different covariance matrices (ranges for latencies 1–4). Defaults used of Table 27.3; no correlation assumed.
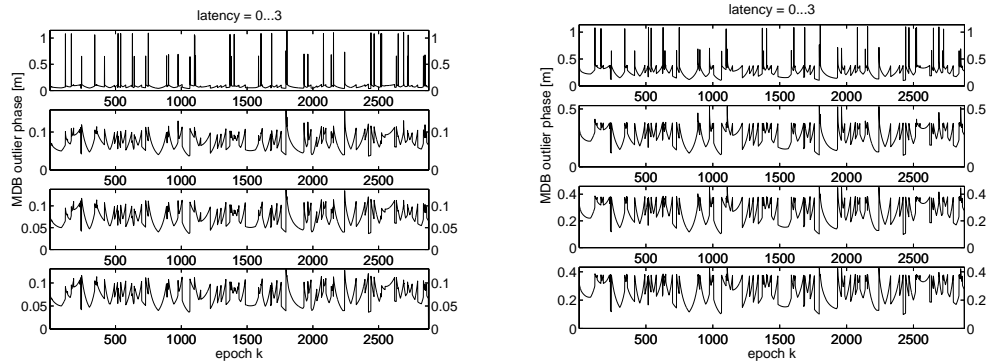
Figure 29.6   Maximum MDBs of outliers in phase observations for latencies 0–3. Left panel: With full covariance matrix of observations. Right panel: With diagonal covariance matrix. Defaults according to Table 27.3; no spatial correlation assumed.
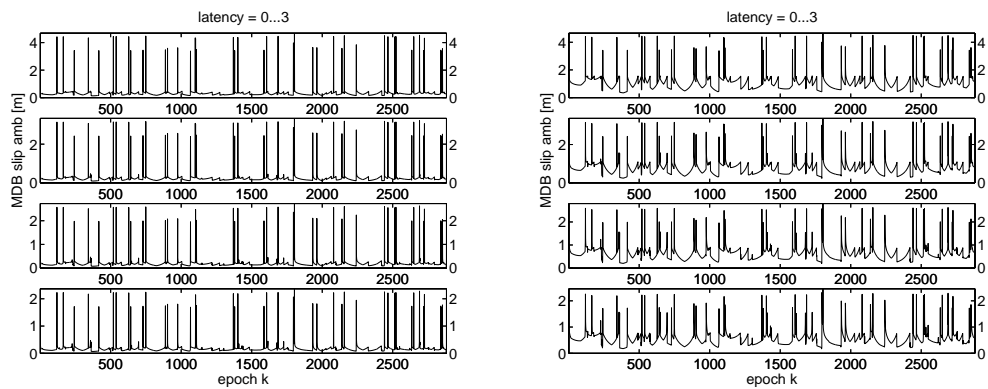

Figure 29.7   Maximum MDBs of ambiguity slips for latencies 0–3. Left panel: With full covariance matrix of observations. Right panel: With diagonal covariance matrix. Defaults of Table 27.3; no spatial correlation assumed.

		ZTD [mm]	North [mm]	East [mm]	Up [mm]	OMT [-]
1.	1 epoch batch, tropo-float (∗5)	18.05	0.80	0.66	3.55	0.95
	1 epoch batch, tropo-weighted (∗2)	*7.32*	*0.79*	*0.65*	*3.44*	*1.00*
	12 epochs batch, tropo-float	8.79	0.79	0.65	3.47	(1.00)
	12 epochs batch, tropo-weighted	6.71	0.79	0.65	3.43	(1.00)
	240 epochs batch, tropo-float	2.64	0.76	0.63	3.25	(1.05)
	240 epochs batch, tropo-weighted	2.63	0.76	0.63	3.25	(1.05)
	2880 ep. batch, tropo-constant	1.14	0.75	0.62	3.23	(1.98)
	tropo-fixed (∗4)	0.00	0.74	0.62	1.43	(2.12)
2.	250 km, absolute troposphere	7.70	0.82	1.01	3.64	1.00
	250 km, relative troposphere	7.56	0.81	0.67	3.52	1.01
	50 km, absolute troposphere	7.34	0.79	0.93	3.47	1.00
	50 km, relative troposphere (∗2)	*7.32*	*0.79*	*0.65*	*3.44*	*1.00*
	10 km, absolute troposphere	7.27	0.78	0.91	3.42	1.00
	10 km, relative troposphere	7.27	0.78	0.64	3.41	1.00
3.	250 km, absolute troposphere	6.60	0.82	1.01	3.63	1.00
	250 km, relative troposphere	6.04	0.80	0.66	3.47	1.01
	50 km, absolute troposphere	5.32	0.79	0.92	3.42	1.00
	50 km, relative troposphere	*5.15*	*0.78*	*0.65*	*3.37*	*1.00*
	10 km, absolute troposphere	4.22	0.77	0.89	3.31	0.99
	10 km, relative troposphere	4.17	0.77	0.63	3.30	0.99
4.	tropo-float, amb.-float, abs.	23.75	0.85	1.12	4.18	0.90
	tropo-float, amb.-float, rel. (∗5)	18.05	0.80	0.66	3.55	0.95
	tropo-float, amb.-fixed, rel.	17.83	0.66	0.51	3.08	0.95
	tropo-fixed, amb.-float (∗4)	0.00	0.74	0.62	1.43	(1.19)
	tropo-fixed, amb.-fixed	0.00	0.62	0.50	1.22	(1.24)
5.	Q_y diag/H2, amb.-float (∗1)	8.26	0.92	0.72	4.12	0.89
	Q_y diag/H2, amb.-fixed	7.93	0.79	0.59	3.58	0.89
	Q_y full/H2, amb.-float (∗2)	*7.32*	*0.79*	*0.65*	*3.44*	*1.00*
	Q_y full/H2, amb.-fixed	6.97	0.65	0.51	3.00	1.00
	Q_y unit, amb.-float (∗3)	6.66	0.73	0.61	3.25	1.05
	Q_y unit, amb.-fixed	6.29	0.59	0.47	2.83	1.05
6.	Q_y diag/H2 (∗1)	8.26	0.92	0.72	4.12	0.89
	Q_y diag/H1	7.19	0.79	0.65	3.53	0.95
	Q_y full/H2 (∗2)	*7.32*	*0.79*	*0.65*	*3.44*	*1.00*
	Q_y full/H1	6.78	0.74	0.61	3.29	1.00
	Q_y unit (∗3)	6.66	0.73	0.61	3.25	1.05
7.	Q_y diag/H2, 70 deg	13.50	1.15	0.86	8.29	0.93
	Q_y diag/H2, 80 deg (∗1)	8.26	0.92	0.72	4.12	0.89
	Q_y full/H2, 70 deg	12.63	1.05	0.80	7.46	1.00
	Q_y full/H2, 80 deg (∗2)	*7.32*	*0.79*	*0.65*	*3.44*	*1.00*
	Q_y unit, 70 deg	12.40	1.02	0.78	7.31	1.01
	Q_y unit, 80 deg (∗3)	6.66	0.73	0.61	3.25	1.05

Table 29.9 Formal standard deviations of parameters, and Overall Model Test values. Standard deviations after processing 24 hours for coordinates and mean values for ZTDs over 2.5–24 hours. Asterisks indicate identical scenarios (∗1–3) or near identical scenarios (∗4–5). The scenarios are according to the defaults of Table 27.3 except for spatial correlation and a-priori standard deviation of phase observables: 1 cm. Scenarios under **3** and **4**: TL2 correlation, otherwise uncorrelated (in simulation and estimation). Italicization: (near) default scenarios.

$\widetilde{Q}_{y_k}$ (simulated)	Q_{y_k} (filter)	ZTD [mm]	North [mm]	East [mm]	Up [mm]	OMT [-]
diagonal	full	9.45	1.05	0.83	4.83	1.33
diagonal	diagonal (∗1)	8.26	0.92	0.74	4.12	1.00
full	diagonal	7.86	0.85	0.69	3.77	0.89
full	full (∗2)	7.32	0.79	0.65	3.44	1.00

Table 29.10   Standard deviations and OMT values for four different scenarios where covariance matrix Q_{y_k} is used in the filter while $\widetilde{Q}_{y_k}$ describes the precision of the simulated observations. Standard deviations of coordinates after 24 hours and mean values for ZTDs over 2.5–24 hours. Defaults of Table 27.3; no spatial correlation assumed; a-priori standard deviation of phase observables: <u>1 cm</u>. Asterisks indicate scenarios in Table 29.1.

$\widetilde{Q}_{y_k}$ (simulated)	Q_{y_k} (filter)	ZTD [mm]	North [mm]	East [mm]	Up [mm]	OMT [-]
full/H2	full/H1	7.39	0.80	0.66	3.47	1.03
full/H2	full/H2	7.32	0.79	0.65	3.44	1.00
full/H1	full/H2	6.85	0.75	0.62	3.31	0.97
full/H1	full/H1	6.78	0.74	0.61	3.29	1.00

Table 29.11   Standard deviations and OMT values for four different scenarios where covariance matrix Q_{y_k} is used in the filter while $\widetilde{Q}_{y_k}$ describes the precision of the simulated observations. Standard deviations of coordinates after 24 hours and mean values for ZTDs over 2.5–24 hours. Defaults of Table 27.3; no spatial correlation assumed; a-priori standard deviation of phase observables: <u>1 cm</u>.

PSD in sim. [m^2/s]	PSD in filter [m^2/s]	std. ZTD [%]	std. N/E [%]	std. Up [%]	test value [-]	OMT [-]
1.5e-7	1.5e-7	100	100	100	0.99	1.00
1.5e-7	1.5e-5 (u)	195	101	102	0.64	0.96
1.5e-7	1.5e-6 (u)	130	100	101	0.86	0.98
1.5e-7	1.5e-8 (o)	135	105	108	1.11	1.01
1.5e-7	1.5e-9 (o)	241	154	142	1.30	1.04
1.5e-5	1.5e-7 (o)	653	130	179	3.17	1.51
1.5e-6	1.5e-7 (o)	219	103	110	1.34	1.04
1.5e-8	1.5e-7 (u)	79	100	99	0.95	0.99
1.5e-9	1.5e-7 (u)	76	100	99	0.95	0.99

Table 29.12   Standard deviations in percentages of reference case (top row) for different scenarios where some PSD is used in the filter while another PSD describes the stochastic behavior of the simulated ZTDs. (u): underconstraining; (o): overconstraining. Test value: Mean value of square root of troposphere slippage tests, station Delft, latency 0 (test quantity has a standard normal distribution). Standard deviations of coordinates after 24 hours and mean values for ZTDs over 2.5–24 hours. Defaults of Table 27.3; no spatial correlation assumed; a-priori standard deviation of phase observables: <u>1 cm</u>.

Chapter 30

The troposphere-fixed model

By fixing the ZTDs at both stations of a baseline, we introduce a bias when the fixed values are incorrect. In Sect. 28.2 we saw that, as long as the ZTDs at both stations are equally biased, this mainly has a scaling effect. A wrongly fixed relative ZTD however, has impact on the estimated height differences. When the fixed relative ZTDs are correct on the other hand, the precision of the estimated parameters, such as the height component, is better.

If we denote the variances of the Up component for the troposphere-fixed and weighted model by $\sigma^2_{Up,fixed}$ and $\sigma^2_{Up,weighted}$ respectively, and the height bias introduced by the troposphere-fixed model by ∇Up, the mean square error (MSE)[1] for both models is equal when:

$$\sigma^2_{Up,fixed} + \nabla Up^2 = \sigma^2_{Up,weighted}. \tag{30.1}$$

The distance for which the MSEs are equal, is the distance where it becomes advantageous to switch from the troposphere-fixed model to the troposphere-weighted model. We look for this distance.

In [*Beutler et al.*, 1988] it was shown that the bias in the Up component is about $c = 3$ times the bias in the relative ZTD, ∇D^z_{12}:

$$\nabla Up = c \cdot \nabla D^z_{12}. \tag{30.2}$$

This constant c was derived for a 70 degrees zenith cut-off angle. The factor three is a wide-spread rule-of-thumb value and is still quite reasonable for lower elevations if a proper weighting is used.[2] We found empirical values for this factor for a zenith cut-off angle of 80 degrees and modeling residual STDs by simulation of the 24-hour data set of the previous chapters, where we fixed the ZTDs of both stations by different values and filtered with the troposphere-fixed model. These c values are shown in Table 30.1.

If we assume that the ZTDs at both ends of the baseline are fixed to the same value, while the actual difference is equal to a realistic mean value based on the structure function of the TL2 model in Part III, Eqs. (18.12) and (18.17), then:

$$(\nabla D^z_{12})^2 \approx \sigma^2_{D^z_{12}} \equiv \mathbf{D}_{D^z}(\rho), \tag{30.3}$$

with ρ the baseline length. For $\rho < 4$ km we can approximate this function by the linear relation:

$$\sigma^2_{D^z_{12}} \approx \sigma^2_0 \cdot \rho, \tag{30.4}$$

---

[1]The MSE is defined as $\mathbf{E}\{\| \hat{x} - x \|^2_2\}$ for $\mathbf{E}\{\hat{x}\} = x + \nabla x$ [*Bain and Engelhardt*, 1989]. In our case x is the baseline Up component.

[2]In [*Beutler et al.*, 1988] a $\sec^2 z$ weighting of the observations was used, but their derivation was intended for a more realistic $\cos^2 z$ weighting, as later used in [*Sjöberg*, 1992].

where $\sigma_0^2 \approx 10 \text{ mm}^2/\text{km}$. An empirical approximation of the variance of the Up component as function of the number of epochs $p = 300, ..., 2880$ can be found as (compare with the right panel of Fig. 28.3):

$$\sigma_{Up}^2(p) \approx \frac{s_0^2}{p}, \tag{30.5}$$

for either the troposphere-fixed or weighted model. The empirical constants s_0 are determined by a LSQ fit; they are given in Table 30.1 for four scenarios.

σ_ϕ [mm]	model	c [-]	s_0 [mm]
2	troposphere-fixed (∗4)	2.8	32
2	troposphere-weighted (∗2)	-	61
10	troposphere-fixed (∗4)	3.9	80
10	troposphere-weighted (∗2)	-	188

Table 30.1   Empirical constants of Eqs. (30.2) and (30.5), and formal standard deviations of baseline heights for four scenarios. Scenarios correspond to numbers ∗2 and ∗4 in Tables 29.1 and 29.9 (no correlation assumed).

Combining Eqs. (30.1–30.5) finally gives the distance for which Eq. (30.1) holds:

$$\rho = \frac{(s_0^2)_{weighted} - (s_0^2)_{fixed}}{p \, \sigma_0^2 \, c^2}. \tag{30.6}$$

Figure 30.1 shows this relation between epochs p and distance ρ for a-priori standard deviations of the phase observations of 2 mm and 10 mm. Apparently, under the assumptions in this chapter, for long observation time spans the troposphere-weighted model is already advantageous for baseline lengths smaller than 2.5 km. If the ZTD differences are not constant, but vary around a constant value, the troposphere-fixed model may however be suitable for longer baseline lengths.
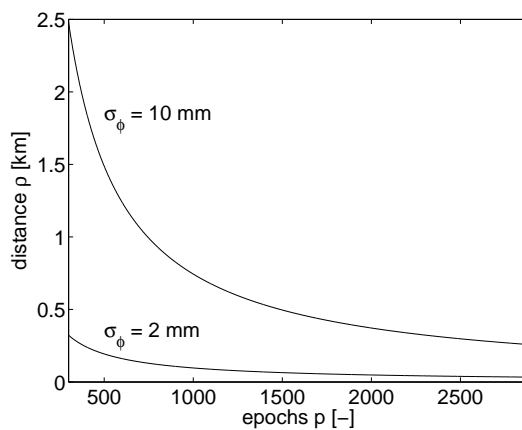
Figure 30.1   Baseline length ρ as function of the number of epochs p (30-s sampling) for which the troposphere-fixed model has the same MSE as the troposphere-weighted model for a constant bias in the relative ZTD equal to its empirical standard deviation as function of distance (TL2 model). Larger biases would give smaller distances and vice versa.

Conclusions of Part V

spatiotemporal
constraints

The formal precision of the filtered baseline heights is not sensitive to the spatiotemporal constraints. Spatial correlation has hardly any contribution and the troposphere-float, weighted, and constant models give comparable formal precision. The batch size is also of minor influence on the formal height precision. The only effect of the constraints that was shown is the severe precision-deteriorating effect of overconstraining, although only when the constraints are more than ten times too tight. Underconstraining has little effect, which suggests that we might as well use the troposphere-float model to avoid overconstraining (or the troposphere-weighted model with very soft constraints). The use of spatiotemporal constraints showed a much larger impact on the filtered ZTDs, which means that weather-prediction applications may benefit much more by an advanced application of these constraints.

troposphere
fixing

A much larger step in precision can be found between troposphere fixing and estimating. Troposphere fixing gives a formal standard deviation of the baseline height that is about twice as small as in the troposphere-float models. The actual precision of the height, however, depends on the actual troposphere delay differences, which generally increase with distance. Because we do not know these delay differences beforehand, by using a troposphere-float model one avoids large surprises; in general the precision may be lower, but a bias caused by the troposphere can be avoided. Based on the assumption of a constant delay difference under mean atmospheric conditions, maximum distances were found for application of the troposphere-fixed model. As a rule of thumb, the troposphere-fixed model should not be used for distances larger than ± 1 km. Instead the troposphere-float model is preferred where absolute ZTDs are parameterized.

residual
STDs

The simulation results show that the residual STD model is promising. A formal coordinate precision improvement with respect to using a diagonal covariance matrix was found of about 50% for default settings, where the standard deviation of the phase observables was set to 2 mm. The actual improvement was about 30%, because the assumed correlations have a positive effect whether we model them or not. The exact improvement depends on the noise level of the observations. For a standard deviation of the phase observables of 10 mm, these percentages are only about 15% and 10% respectively. The improvement also depends on the effective height. Real data analysis is necessary to validate the model and a fine tuning of the model is needed. This validation may prove to be a difficult task in the presence of unmodeled error sources. Since they were not assumed in the simulations, the actual precision improvement depends on them. Further research on the size, modeling, and filter propagation of other

error sources, especially the zenith angle dependent errors, remains necessary.

cut-off angle

ambiguity
fixing

Other model options that may improve the height precision are: increasing the zenith cut-off angle and fixing the ambiguities. Because the increase of zenith cut-off angles above about 80 degrees may also cause biases, caused by, for example, multipath, this should be done with care. Besides, when using the residual STD model, the observations to low-elevation satellites are strongly down-weighted. Correct ambiguity fixing should, if possible, always be striven for. Even for long observation times it gives 15–20% improvement in the standard deviations of the filtered baseline heights.

Time update for relative ZTD estimation

The full Gauss–Markov model of the time update, including both constrained and global parameters when estimating relative ZTDs, reads:

$$\mathbf{E}\left\{\begin{bmatrix}\underline{d}\\ \underline{\hat{\beta}}_{k-1|k-1}\\ \underline{\hat{x}}_{k-1}\end{bmatrix}\right\} = \begin{bmatrix}-C_n & C_n & 0\\ I_{n-1} & 0 & 0\\ 0 & 0 & I_{n_x}\end{bmatrix}\begin{bmatrix}\beta_{k-1}\\ \beta_k\\ x\end{bmatrix};$$

$$\mathbf{D}\left\{\begin{bmatrix}\underline{d}\\ \underline{\hat{\beta}}_{k-1|k-1}\\ \underline{\hat{x}}_{k-1}\end{bmatrix}\right\} = \begin{bmatrix}Q_d & 0 & 0\\ 0 & Q_{\hat{\beta}_{k-1|k-1}} & Q_{\hat{\beta}_{k-1|k-1}\hat{x}_{k-1}}\\ 0 & Q_{\hat{x}_{k-1}\hat{\beta}_{k-1|k-1}} & Q_{\hat{x}_{k-1}}\end{bmatrix},$$

(J.1)

with $C_n' \doteq [0,\ I_{n-1}]$. The observables $\underline{\hat{x}}_{k-1}$ of the global parameters are stochastically Related to $\underline{\hat{\beta}}_{k-1|k-1}$ but not functionally. These observables are called *free* *variates*, a special type of y^R *variates* as defined in [*Teunissen*, 1999].

free variates

y^R variates

The general model in which the free variates occur reads (in terms of the weight matrix W):

$$\mathbf{E}\left\{\begin{bmatrix}\underline{y}\\ \underline{y}^R\end{bmatrix}\right\} = \begin{bmatrix}A & 0\\ 0 & I\end{bmatrix}\begin{bmatrix}x\\ \mathbf{E}\{\underline{y}^R\}\end{bmatrix} \quad;\quad \mathbf{D}\left\{\begin{bmatrix}\underline{y}\\ \underline{y}^R\end{bmatrix}\right\} = \begin{bmatrix}W_y & W_{yR}\\ W_{Ry} & W_R\end{bmatrix}^{-1}.$$

(J.2)

The normal equations of this model read [*Teunissen*, 1999]:

$$\begin{bmatrix}A'W_yA & A'W_{yR}\\ W_{Ry}A & W_R\end{bmatrix}\begin{bmatrix}\hat{x}\\ \hat{y}^R\end{bmatrix} = \begin{bmatrix}A'W_y & A'W_{yR}\\ W_{Ry} & W_R\end{bmatrix}\begin{bmatrix}y\\ y^R\end{bmatrix}.$$

(J.3)

From the second row, we obtain:

$$W_R\,\hat{y}^R = W_R\,y^R + W_{Ry}\,(y - A\hat{x}),$$

(J.4)

or with $\hat{\varepsilon} \doteq y - A\hat{x}$ and $\hat{\varepsilon}^R \doteq y^R - \hat{y}^R$:

$$\hat{\varepsilon}^R = -W_R^{-1}\,W_{Ry}\,\hat{\varepsilon}.$$

(J.5)

If the LSQ corrections of the model without the y^R variates are zero, $\hat{\varepsilon} = 0$, then also the LSQ corrections of the y^R variates are zero: $\hat{\varepsilon}^R = 0$. Since the LSQ corrections of the observations in Eq. (27.6) are zero (with $d = 0$, verify), this also holds for Model (J.1). In other words, including global parameters does not affect the estimates in the time update:

$$\boxed{\hat{\beta x}_{k|k-1} = \hat{\beta x}_{k-1|k-1}.}$$

(J.6)

The covariance matrix of the predicted estimates is also not affected by the extension of Model (27.6) with global parameters. First define:

$$\begin{bmatrix} Q_{\hat{\beta}_{k-1|k-1}} & Q_{\hat{\beta}_{k-1|k-1}\hat{x}_{k-1}} \\ Q_{\hat{x}_{k-1}\hat{\beta}_{k-1|k-1}} & Q_{\hat{x}_{k-1}} \end{bmatrix}^{-1} = \begin{bmatrix} W_{\hat{\beta}_{k-1|k-1}} & W_{\hat{\beta}_{k-1|k-1}\hat{x}_{k-1}} \\ W_{\hat{x}_{k-1}\hat{\beta}_{k-1|k-1}} & W_{\hat{x}_{k-1}} \end{bmatrix}. \tag{J.7}$$

In terms of this weight matrix, the normal matrix of Model (J.1) reads:

$$\begin{bmatrix} W_{\hat{\beta}_{k-1|k-1}} & +C_n'Q_d^{-1}C_n \ ; & -C_n'Q_d^{-1}C_n \ ; & W_{\hat{\beta}_{k-1|k-1}\hat{x}_{k-1}} \\ & -C_n'Q_d^{-1}C_n \ ; & +C_n'Q_d^{-1}C_n \ ; & 0 \\ W_{\hat{x}_{k-1}\hat{\beta}_{k-1|k-1}} & ; & 0 & ; & W_{\hat{x}_{k-1}} \end{bmatrix}. \tag{J.8}$$

The covariance matrix of the estimated parameters of Model (J.1) now follows as the inverse of the normal matrix:

$$\begin{bmatrix} Q_{\hat{\beta}_{k-1|k-1}} & Q_{\hat{\beta}_{k-1|k-1}} & Q_{\hat{\beta}_{k-1|k-1}\hat{x}_{k-1}} \\ Q_{\hat{\beta}_{k-1|k-1}} & Q_{\hat{\beta}_{k-1|k-1}} + (C_n'Q_d^{-1}C_n)^{-1} & Q_{\hat{\beta}_{k-1|k-1}\hat{x}_{k-1}} \\ Q_{\hat{x}_{k-1}\hat{\beta}_{k-1|k-1}} & Q_{\hat{x}_{k-1}\hat{\beta}_{k-1|k-1}} & Q_{\hat{x}_{k-1}} \end{bmatrix}, \tag{J.9}$$

which can be verified by using the definition of Eq. (J.7). The covariance matrix of the total vector of predicted estimates therefore follows from the lower-right corner of Eq. (J.9) as

$$\mathbf{D}\left\{ \begin{bmatrix} \underline{\hat{\beta}}_{k|k-1} \\ \underline{\hat{x}}_k \end{bmatrix} \right\} = \begin{bmatrix} Q_{\hat{\beta}_{k-1|k-1}} + (C_n'Q_d^{-1}C_n)^{-1} & Q_{\hat{\beta}_{k-1|k-1}\hat{x}_{k-1}} \\ Q_{\hat{x}_{k-1}\hat{\beta}_{k-1|k-1}} & Q_{\hat{x}_{k-1}} \end{bmatrix}. \tag{J.10}$$

From Eqs. (J.6) and (J.10) therefore follows that the time update may be summarized as:

$$\boxed{\begin{aligned} \hat{\beta}x_{k|k-1} &= \hat{\beta}x_{k-1|k-1}; \\ Q_{\hat{\beta}x_{k|k-1}} &= Q_{\hat{\beta}x_{k-1|k-1}} + \begin{bmatrix} (C_n'Q_d^{-1}C_n)^{-1} & 0 \\ 0 & 0 \end{bmatrix}. \end{aligned}} \tag{J.11}$$

To compute $(C_n'Q_d^{-1}C_n)^{-1}$, we partition Q_d and its inverse as:

$$Q_d = \begin{bmatrix} Q_{d_{11}} & Q_{d_{12}} \\ Q_{d_{21}} & Q_{d_{22}} \end{bmatrix} \quad ; \quad Q_d^{-1} = \begin{bmatrix} (Q_d^{-1})_{11} & (Q_d^{-1})_{12} \\ (Q_d^{-1})_{21} & (Q_d^{-1})_{22} \end{bmatrix}, \tag{J.12}$$

where $Q_{d_{11}}$ is a scalar and $Q_{d_{22}}$ is $(n-1) \times (n-1)$. Then we have $C_n'Q_d^{-1}C_n = (Q_d^{-1})_{22}$, and using the lemma of App. H of Part IV, we find:

$$\boxed{(C_n'Q_d^{-1}C_n)^{-1} = Q_{d_{22}} - Q_{d_{21}}Q_{d_{11}}^{-1}Q_{d_{12}},} \tag{J.13}$$

which is easily computed and needs to be computed only once.

Zenith angles

Consider two points P_1 and P_2 on the earth's surface, the center of the earth M, and satellite S. P_2 is in the same plane as M, S, and P_1 (see Fig. K.1). Let z_1 and z_2 be the zenith angles of P_1S and P_2S. From the distance d between P_1 and P_2 and the earth's radius $R \approx 6370$ km, we derive the angle $\phi = d/R$. The zenith angle z_2 can now be computed for known zenith angle z_1 as follows.
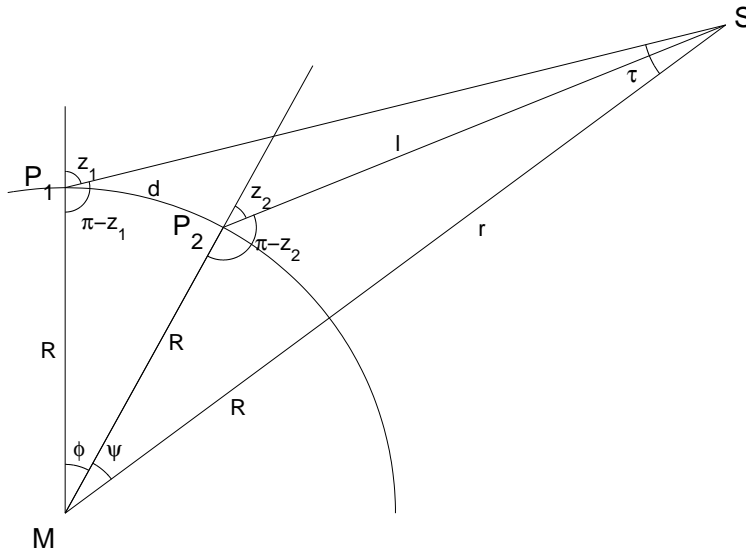

Figure K.1   Zenith angles z_1 and z_2 to satellite S for two stations P_1 and P_2 separated by distance d.

With the sine rule for triangle MP_1S we obtain the top angle τ:

$$\sin \tau = R \cdot \sin z_1 / (R + r), \tag{K.1}$$

where $r \approx 20200$ km. From triangle MP_1S we then obtain:

$$\psi = z_1 - \phi - \tau. \tag{K.2}$$

Applying the cosine rule on triangle MP_2S then gives:

$$l^2 = R^2 + (R + r)^2 - 2R(R + r) \cos \psi. \tag{K.3}$$

And finally, with the sine rule for triangle MP_2S:

$$\sin z_2 = (R + r) \cdot \sin \psi / l. \tag{K.4}$$

Temporal ZTD constraints

Unlike what was expected beforehand, temporal ZTD constraints have little effect on the baseline height precision. This appendix clarifies this by looking into the normal matrix. We consider the following simplified Gauss–Markov model of p epochs with mp single-differenced GPS observations y to m satellites, and $p-1$ constraints d:

$$\mathbf{E}\{\begin{bmatrix} \underline{d} \\ \underline{y} \end{bmatrix}\} = \begin{bmatrix} 0 & E \\ C & S \end{bmatrix} \begin{bmatrix} h \\ D^z \end{bmatrix} \quad ; \quad \mathbf{D}\{\begin{bmatrix} \underline{d} \\ \underline{y} \end{bmatrix}\} = \begin{bmatrix} \sigma_d^2 I_{p-1} & 0 \\ 0 & \sigma_y^2 I_{mp} \end{bmatrix}, \tag{L.1}$$

containing only the baseline height h and the $p \times 1$ vector with zenith delays D^z.

When using a $\sec z$ mapping function, the partial design matrices read:

$$C' \doteq [\cos z(1)', ..., \cos z(p)']; \qquad E \doteq \begin{bmatrix} -1 & 1 & & 0 \\ & -1 & 1 & \\ & & \ddots & \ddots \\ 0 & & & -1 & 1 \end{bmatrix}, \tag{L.2}$$
$$S \doteq \mathrm{diag}[\sec z(1), ..., \sec z(p)];$$

where $z(k)$ is a vector with m zenith angles at epoch k and $\cos z(k)$ and $\sec z(k)$ are vectors with cosine and secant values. The normal matrix then contains the elements

$$\begin{aligned} C'C &= \textstyle\sum_{k=1}^{p} \sum_{s=1}^{m} \cos^2 z_s(k) &&\doteq p\, m\, \beta; \\ S'S &= \mathrm{diag}[\textstyle\sum_{s=1}^{m} \sec^2 z_s(1), ..., \sum_{s=1}^{m} \sec^2 z_s(p)] \approx m\, \gamma\, I_p; \\ C'S &= m\, e_p', \end{aligned} \tag{L.3}$$

where e_p is a $p \times 1$ vector with all ones. We write the mean value of $\sum_{s=1}^{m} \cos^2 z_s(k)$ over all epochs as $m\,\beta$ and approximate $\sum_{s=1}^{m} \sec^2 z_s(k)$ by a constant $m\,\gamma$ for each epoch. The normal matrix then reads:

$$N = \begin{bmatrix} \sigma_y^{-2} m\, \beta\, p & \sigma_y^{-2}\, m\, e_p' \\ \sigma_y^{-2}\, m\, e_p & \sigma_y^{-2}\, m\, \gamma\, I_p + \sigma_d^{-2}\, E'E \end{bmatrix}. \tag{L.4}$$

The variance of the height can be computed using Eq. (H.1) as:

$$\sigma_{\hat{h}}^2 = (\sigma_y^{-2}\, m\, \beta\, p - \sigma_y^{-4}\, m^2\, e_p'\, (\sigma_y^{-2}\, m\, \gamma I_p + \sigma_d^{-2} E'E)^{-1} e_p)^{-1}. \tag{L.5}$$

We use Eq. (H.2) to find:

$$(\sigma_y^{-2}\, m\, \gamma\, I_p + \sigma_d^{-2} E'E)^{-1} = \sigma_y^2\, \gamma^{-1}\, m^{-1}(I_p - E'(EE' - \sigma_d^2 \sigma_y^{-2}\, m\, \gamma I_{p-1})^{-1} E). \tag{L.6}$$

Since $E\, e_p = 0$ and $e_p'\, I_p\, e_p = p$, substituting Eq. (L.6) into Eq. (L.5) finally gives:

$$\boxed{\sigma_{\hat{h}}^2 = \sigma_y^2\, m^{-1}\, p^{-1}(\beta - \gamma^{-1})^{-1}.} \tag{L.7}$$

In other words: the precision of the height component is insensitive to the value of the constraint variance σ_d^2. In the simulation there is a small effect of the

constraint; see first group of Fig. 29.1, where the troposphere-float, weighted and constant model are compared. This is because the approximation of $S'S$ by $\gamma\, I_p$ is not exact. Also in the actual model, receiver clock parameters are introduced. Although this does not change the structure of the normal matrix (after pre-elimination), the correlation term becomes an approximation of a scaled vector with ones.

We also see that the precision increases with the precision and the number of observations. The constants β and γ^{-1} depend on the zenith cut-off angle. Both decrease with increasing zenith cut-off angle (z_{cut}), but the difference $\beta - \gamma^{-1}$ increases. This means that the precision of the height increases with increasing zenith cut-off angles. We can see this by approximating:

$$\beta = m^{-1} \sum_{s=1}^{m} \cos^2 z_s(k) \approx \frac{\int_0^{z_{cut}} \cos^2 z\, \sin z\, dz}{\int_0^{z_{cut}} \sin z\, dz} = \frac{1}{3} \frac{1 - \cos^3 z_{cut}}{1 - \cos z_{cut}};$$

$$\gamma \approx m^{-1} \sum_{s=1}^{m} \sec^2 z_s(k) \approx \frac{\int_0^{z_{cut}} \sec^2 z\, \sin z\, dz}{\int_0^{z_{cut}} \sin z\, dz} = \frac{\cos^{-1} z_{cut} - 1}{1 - \cos z_{cut}},$$

$$(L.8)$$

using the Jacobian[1] $|\partial(x, y, h)/\partial(r, z, \alpha)| = \sin z$ that accounts for an assumed equal distribution of satellites over the sky. For more on this technique, see [*Beutler et al.*, 1988] and [*Santerre*, 1991]. Figure L.1 shows β and γ^{-1} as function of the zenith cut-off angle.
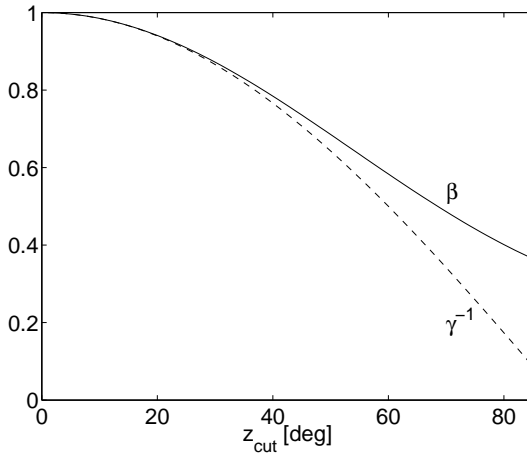
Figure L.1   Constants β and γ^{-1} as function of the zenith cut-off angle.

With Eq. (H.2), the covariance matrix of the ZTDs can be derived as:

$$Q_{\hat{D}^z} = (\sigma_y^{-2}\, m\, \gamma\, I_p + \sigma_d^{-2} E'E)^{-1} + \gamma^{-2}\sigma_{\hat{h}}^2 e_p\, e_p'. \tag{L.9}$$

We can now distinguish two extreme cases: (1) the troposphere-float case, where $\sigma_d^{-2} \to 0$; and (2) the troposphere-constant case, where $\sigma_d^2 \to 0$. From Eq. (L.9) we immediately see that in the first case:

$$Q_{\hat{D}^z} = \sigma_y^2\, m^{-1}\, \gamma^{-1} I_p + \gamma^{-2}\sigma_{\hat{h}}^2 e_p\, e_p'. \tag{L.10}$$

In the second case, we apply Eq. (L.6) and recognize that $I_p - E(EE')^{-1}E = p^{-1}e_p\, e_p'$, so that:

$$Q_{\hat{D}^z} = (\sigma_y^2\, m^{-1}\, \gamma^{-1}\, p^{-1} + \gamma^{-2}\sigma_{\hat{h}}^2)e_p\, e_p'. \tag{L.11}$$

---

[1]With partial derivatives of Cartesian coordinates x, y, and h with respect to spherical coordinates: radius r, zenith angle z, and azimuth α.

Bibliography

Bain, L.J., and M. Engelhardt (1989), *Introduction to Probability and Mathematical Statistics*, PWS-KENT Publishing Company, Boston.

Beutler, G., I. Bauersima, W. Gurtner, M. Rothacher, T. Schildknecht, and A. Geiger (1988), Atmospheric refraction and other important biases in GPS carrier phase observations, in *Atmospheric effects on geodetic space measurements*, edited by F.K. Brunner, pp. 15–43, Monograph 12, The University of NSW.

Beutler, G., E. Brockmann, R. Dach, P. Fridez, W. Gurtner, U. Hugentobler, J. Johnson, L. Mervart, M. Rothacher, S. Schaer, T. Springer, and R. Weber (2000), Bernese GPS Software, Version 4.2, *software documentation*, AUIB.

Brunner, F.K., and S. McCluskey (1991), Tropospheric zenith delay parameters: How many should be estimated in GPS processing?, *Aust. J. Geod. Photogram. Surv.*, *55*, 67–75.

Jarlemark, P.O.J., T.R. Emardson, and J.M. Johansson (1998), Wet delay variability calculated from radiometric measurements and its role in space geodetic parameter estimation, *Radio Science*, *33*, 719–730.

Kleijer, F. (2001), Mapping function induced bias in tropospheric delay estimation using GPS, *Phys. Chem. of the Earth (A)*, *26*, 467–470.

Koch, K.-R. (1987), *Parameter Estimation and Hypothesis Testing in Linear Models*, Springer, Berlin.

Matlab (1997), *The language of technical computing*, The MathWorks, Inc.

Niell, A.E. (1996), Global mapping functions for the atmosphere delay at radio wavelengths, *Journal of Geophysical Research (B)*, *101*, 3227–3246.

Santerre, R. (1991), Impact of GPS satellite sky distribution, *Manuscripta Geodaetica*, *16*, 28–53.

Sjöberg, L.E. (1992), Systematic tropospheric errors in geodetic positioning with the global positioning system, *Manuscripta Geodaetica*, *17*, 201–209.

Teunissen, P.J.G. (1998), Succes probability of integer GPS ambiguity rounding and bootstrapping, *Journal of Geodesy*, *72*, 606–612.

Teunissen, P.J.G. (1999), *Adjustment theory: An introduction*, Series on Mathematical Geodesy and Positioning, Delft University Press.

Webb, F.H., and J.F. Zumberge (1993), An introduction to the GIPSY/OASIS-II, *JPL publ., D-11088*, JPL.

Chapter 32

Conclusions and recommendations

GPS leveling

This thesis focused on troposphere modeling and filtering for precise GPS leveling. GPS leveling is expected to be a cost-effective technique that can replace some of the traditional spirit and hydrostatic levelings. It is especially suited for leveling over water and long distances and can be used for, for example, subsidence monitoring and maintenance of national height datums. For the latter application, the precision of GPS-derived orthometric (or normal) heights also depends on the precision of the local (quasi-)geoid. The precision of GPS heights differences is considered optimal when it is comparable to geoid height differences; in the Netherlands this is in the order of 5 mm.

troposphere modeling

To obtain this precision in short observation time spans, a judicious modeling of the tropospheric delay is required because of the strong impact of tropospheric errors on the height component. Although other error sources also contribute to the precision, we restricted ourselves to the question: How should the tropospheric delay be modeled to obtain the best precision? To answer this question we investigated the existing modeling possibilities and classified them by physical, functional, and stochastic aspects. For the best possible precision we focused on a static network approach with dual-frequency receivers with which (both code and) phase observables are acquired.

physical modeling

The hydrostatic part of the delay can be modeled with high accuracy if the surface pressure is known, but the more variable wet delay cannot be modeled with enough precision based on meteorological parameters. Both delays can be approximated as a zenith value multiplied by a mapping function. The mapping functions differ for both parts because they depend on their respective layer thicknesses. The wet layer thickness is smaller because of the temperature-dependent saturation pressure value of water vapor. Several different mapping functions were derived in the past decades. They rely on mean atmospheric conditions or surface meteorological parameters and do not account for anisotropy.

functional modeling

In the troposphere-fixed models, a-priori corrections are applied for the tropospheric delays. These models are rank deficient, but by repeated reparameterizations full-rank (undifferenced) models can be obtained. To prevent errors in the GPS heights by inadequate a-priori models, it is necessary to parameterize the tropospheric delay. One can parameterize Zenith Tropospheric Delays (ZTDs), gradients, and residual Slant Tropospheric Delays (STDs). The presence of satellite clock errors causes near rank deficiencies in the ZTDs, coordinates, and gradients. In the troposphere-float model, ZTDs are the most common parameters; one ZTD per station is estimated for every batch of epochs, but sometimes the ZTDs of one station are fixed because of this near rank deficiency. This

does however introduce biases in the heights, and should therefore not be done. Gradients or residual STDs can refine the parameterization. The latter parameters loosen the isotropy assumption in the mapping. Each slant delay is parameterized, but an equal amount of constraints is applied to prevent the model from becoming underdetermined. Residual STDs are temporal (time-varying) parameters.

stochastic modeling

A theoretic stochastic model for the constraints on residual STDs was derived by Emardson and Jarlemark (2001). Their polynomial turbulence model turned out to be indefinite, but a summation model proved to be a positive-definite alternative. This was found by tests with these covariance matrices on several locations around the world using a full day of satellite configurations. The models imply zenith angle dependent weighting, as well as correlations between different observations of the same epoch, which seems more realistic than diagonal covariance matrices.

The observation model could further be strengthened by relative constraints on the ZTDs (troposphere-weighted model). Based on the assumption of Kolmogorov turbulence, the Zenith Wet Delay can be considered a power-law process. For practical implementation of relative constraints, a random-walk process can be assumed.

filtering

With recursive filtering techniques, not only the parameters can become available in real time, but also the impact of observation time can be analyzed. It is an appealing technique because parameters can be treated when they are available. New ambiguities can be introduced when new satellites rise, and old ambiguities expired when they become dormant; ambiguity transformations are needed to change pivot satellites. Two filters are considered: the Kalman Filter and the Square Root Information Filter (SRIF). The latter needs reordering transformations in the expiring of parameters. The Kalman Filter turned out to be much faster than the SRIF. Both filters benefit from Cholesky-factor transformations and pre-elimination of temporal nuisance parameters, such as clock and ionosphere parameters, and residual STDs. Two pre-elimination methods exist: one uses a projector on the orthogonal subspace of the temporal parameters and one uses the orthogonal subspace itself. The latter method is faster and implies double-differencing, linear inter-frequency combinations, and full covariance matrices for the original (untransformed) observations.

simulations

The filter finally built in simulation software is a Kalman–Cholesky Filter with the latter pre-elimination method. The simulation software was used to analyze the sensitivities of baseline heights to different model components. Error sources that involve a-priori corrections were not considered in the simulations.

Because the wet and hydrostatic part of the tropospheric delay are mapped to the zenith differently, and because they can cause biases if we map with the wrong function, we assume that a good a-priori model for the Zenith Hydrostatic Delay is applied, which is possible with good surface pressure values. Only the more variable Zenith Wet Delay is mapped this way.

recommendations

Based on the simulation results we come to the following recommendations for GPS leveling:

- Do not use spatiotemporal constraints; the troposphere-float model is preferred.
- Only use troposphere fixing of all stations for distances $< \pm$ 1–2 km.

- Do not fix Zenith Total Delays (ZTDs) of a reference station; estimate absolute ZTDs.

- Use small batches.

And further conclude that:

- Possible improvement is obtainable by modeling residual Slant Tropospheric Delays (STDs; 10–30%). The model should be proved however, because if it turns out to be incorrect it can also deteriorate the precision.

- Increasing the zenith cut-off angle ($\pm$ 5%/degree in the range of 70–80 degrees) and ambiguity fixing (15–20%) have a relatively large contribution to the precision.

formal precision

A formal precision of 5 mm is certainly obtainable for 24-hour data sets. How much faster this precision can be obtained depends on the precision of the observations. Several error sources other than the tropospheric delay contribute to a lower precision. Figure 32.1 shows (approximate) baseline height precision based on simulations for standard deviations of the phase observations of 2 mm (realistic observation noise) and 10 mm (scaled value because of unmodeled error sources) and for zenith angle dependent weighting and modeling residual STDs. These scenarios relate to the conclusions above (troposphere-float model, absolute ZTDs estimated, batch size of 1 epoch and ambiguity fixing). Note that the actual precision may be lower when there are systematic errors.
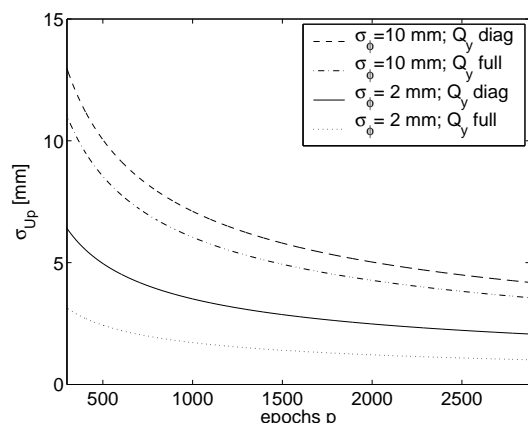
Figure 32.1   Approximated formal baseline height precision, using $\sigma_{Up}^2 = s_0^2/p$ (Eq. (30.5)), for the troposphere-float model with ambiguity fixing and absolute ZTDs estimated; epochs 300–2880. Dashed line: Zenith angle dependent weighting, standard deviation phase observables 10 mm ($s_0 = 225$ mm). Dash-dot line: Residual STD estimation, standard deviation phase observables 10 mm ($s_0 = 191$ mm). Solid line: Zenith angle dependent weighting, standard deviation phase observables 2 mm ($s_0 = 111$ mm). Dotted line: Residual STD estimation, standard deviation phase observables 2 mm ($s_0 = 54$ mm). Other options according to defaults of Table 27.3.

further research

Because a good description of the observation precision is important for the decision on the observation time to be used, quantifying the error sources remains an important issue. Reducing these errors will of course also stay an ongoing task.

In recent developments external information from Numerical Weather Models are considered, mainly to reduce the error in the mapping of STDs to the zenith. Further improvement can be achieved when large zenith cut-off angles are used. Because of the impact of the observations to low-elevation satellites,

especially the modeling of error sources for these observations leaves room for further improvement.

The modeling of residual STDs showed possibilities of improving the precision of not only heights but also of horizontal coordinates and ZTDs. Validation of this model can therefore be considered a task for further research. Points of concern are the scaling of the covariance matrix and the assumption of height independence of the refractivity structure function.

Acknowledgments

Curriculum Vitae

June 5, 1968	Born in Rotterdam, the Netherlands
1987–1994	M.Sc. Geodesy at Delft University of Technology
1994	Military service
1995	Surveyor
1995–1998	Ministry of Transport, Public Works and Water Management
1998–2003	Ph.D. research at Delft University of Technology
2004	Harvard-Smithsonian Center for Astrophysics

Luck is the residue of design.

Niles Crane.