Assessment and Construction of a Precipitation Dataset for the Yalong River Basin

Haicheng Liu
GISt group, TUD
H.Liu-6@tudelft.nl

TU Delft

Outline

- Background
- Initial data analysis
- Data interpolation approaches (direct + indirect)
- Cross validation
- Conclusions

Background

- Complete daily rainfall input with high accuracy for the Yalong River real time hydrologic forecast system

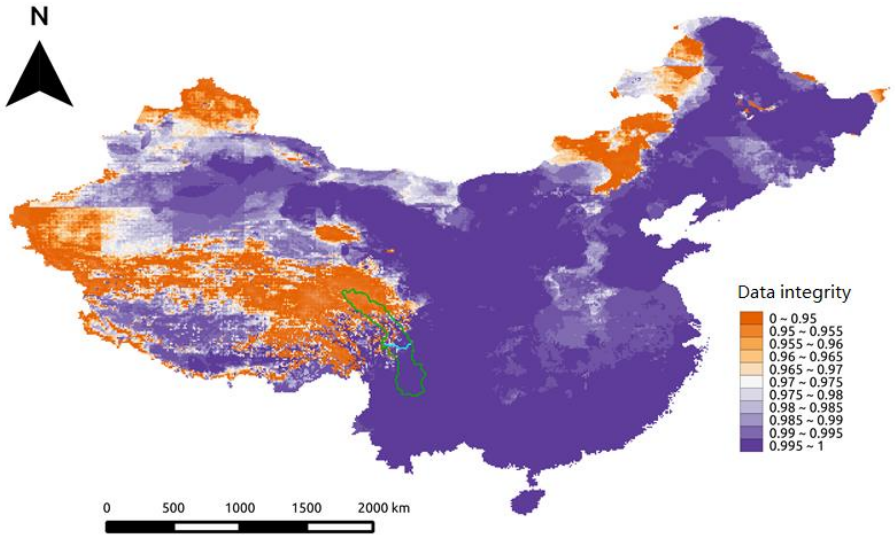
Three precipitation datasets from China national meteorological agency

Precipitation products	Temporal resolution	Area coverage	Duration	Data sources	Construction methods
Hourly 0.1° real time precipitation dataset	Hourly	70 ～ 140°E 15 ～ 59°N	Jan. 2008 ～ current	more than 30,000 automatic gauges + CMORPH precipitation product	PDF matching + Optimum interpolation
Daily 0.25° real time precipitation dataset	Daily	72 ~ 136°E 18 ～ 54°N	Apr. 2008 ～ current	2419 gauges	Climatological optimal interpolation
Daily 0.5° precipitation dataset	Daily	72 ~ 136°E 18 ～ 54°N	Jan. 1961 ～ current	2472 gauges + GTOPO30 DEM	Thin-plate spline interpolation

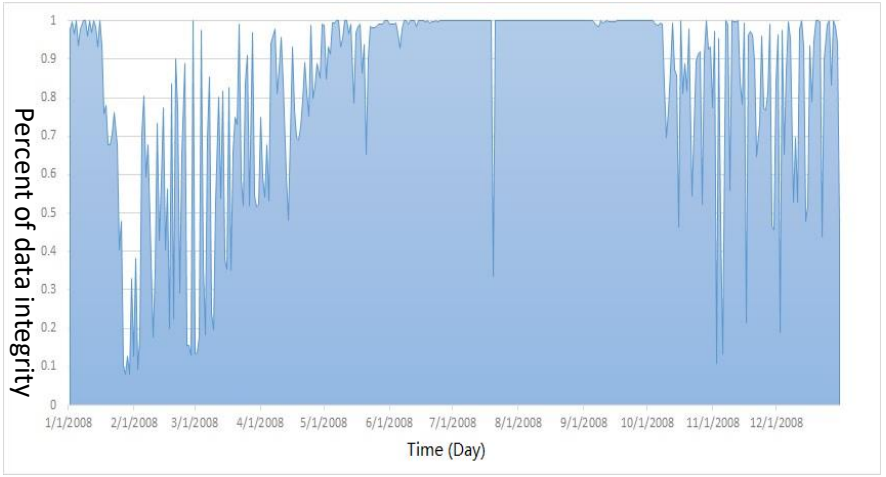
- More gauges (accuracy)+satellite observation (spatial distribution) -> 0.1° hourly data is the most accurate

- But a lot of nodata values especially in the Qinghai-Tibet Plateau due to missing data in the original CMORPH observations

Missing data problem



Integrity map of the 0.1° integrated daily precipitation dataset in 2008.
For every cell, value=number of nodata hours/366 x 24



Time series of data integrity in the upper-stream area of the Yalong River in 2008.

$$value = \frac{1}{24}\sum_{0}^{24} \text{Number of nodata cells/number of all cells}$$

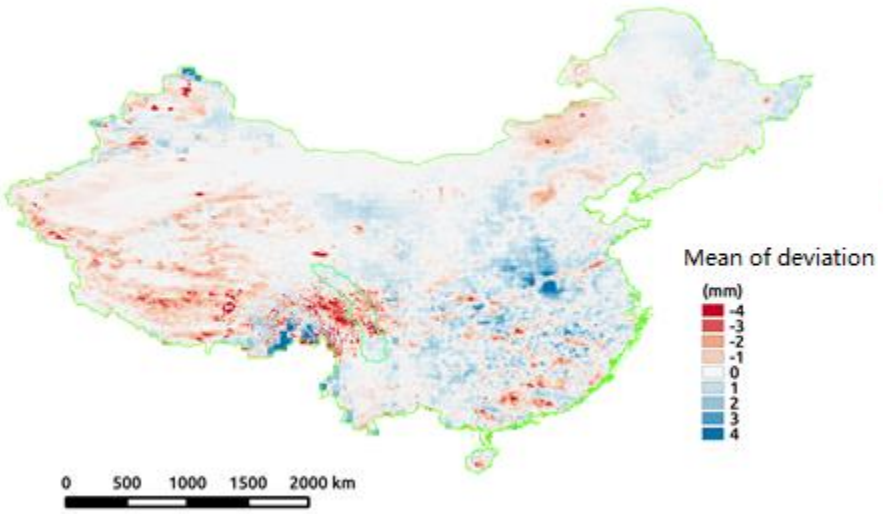
Solution 1 – direct interpolation

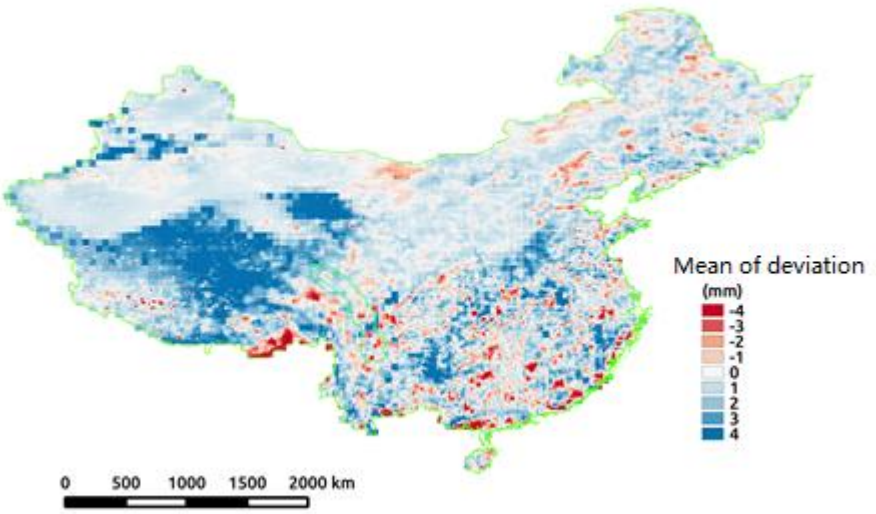
- Kriging method with linear semivariogram model -> better spatial distribution
- Calculation grid consists of 9 x 9 neighboring cells -> efficiency and accuracy
- At least 10 cell values are above 1, otherwise, do averaging -> higher accuracy
- To improve accuracy, kriging for hourly gird, and then aggregate to daily data

Solution 2 – indirect interpolation

- Alternative data sources, 2 products from metrological agency, 7 from satellites
  - 0.5° -> more gauges, longer time span
  - TMPA-v7 -> highest accuracy by literature
- 0.5° data and v7 have various deviation patterns in different seasons and regions
- Interpolation workflow
  - 0.5° : subtraction grid -> Inverse distance weighted (IDW) interpolation -> aggregation
  - v7: same procedure but use 0.1° for north area because of no data.

Comparison between 0.5° and 0.1° data

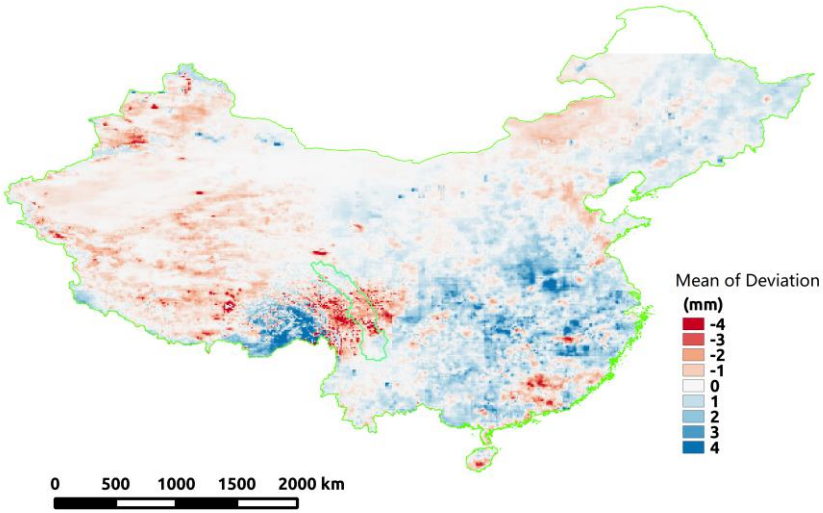


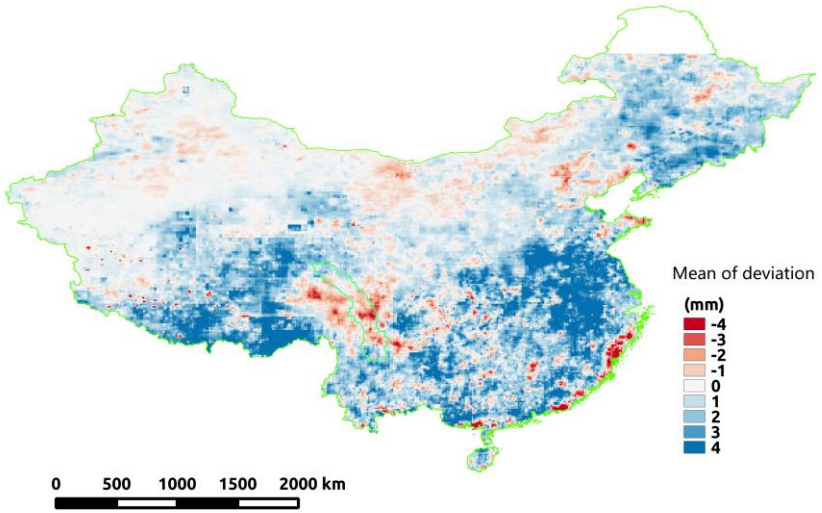


April, 2008 **July, 2008**

Monthly mean of deviations

Comparison between v7 and 0.1° data



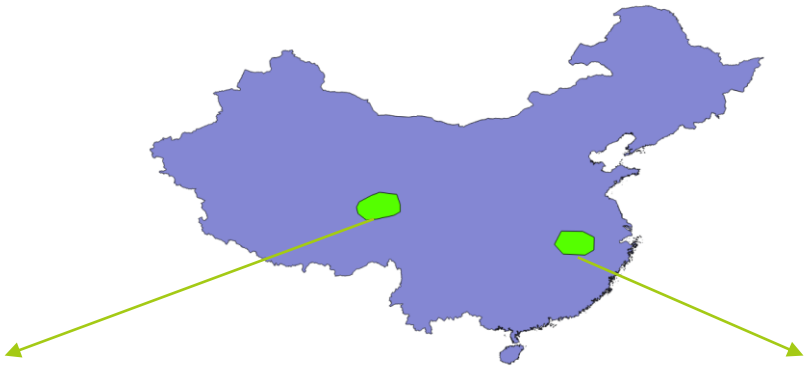


April, 2008 **July, 2008**

Monthly mean of deviations

Cross validation

- 3 approaches comparison in 2 regions, 4 seasons (4 single days)

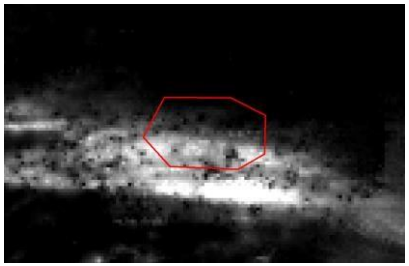


902 cells in Tibet (Yalong River), 770 cells in downstream area of the Yangtze River

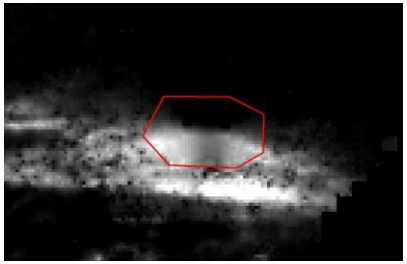
- 4 criteria including avg, sdv, largest error, R
- 4 scenarios for the direct method
  - 1,5,10,24 hours data missing

Cross validation

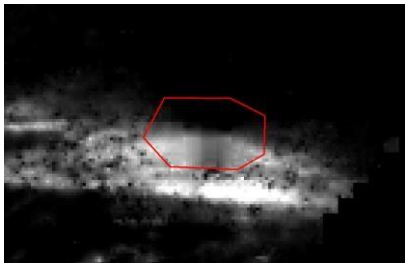
- If only several hours' data is missing, direct method is the best among the 3 interpolations in both quantity and spatial distribution
- Effect of 0.5° data is better than v7, key reason is the temporal span definition
- Sometimes R of interpolation results can be high due to simple spatial distribution



Original data



Daily direct interpolation



0.5° indirect interpolation

10th, April, 2008

Conclusions

- Accuracy order is direct Kriging > indirect 0.5° > indirect v7
- Climatological features for qualitative comparative analysis
- Seasonal effect and temporal correlation of precipitation data may also be utilized in the interpolation plan
- All fail when spatial pattern is complex in nodata area. Thus, more sensors are needed. But data is later utilized in research group's work

Thanks