A Global High Quality Gridded Dataset of Daily Precipitation from 1950

Steefan Contractor, Lisa Alexander, Markus Donat, Zak Bailie

Climate Change Research Centre, ARC Centre of Excellence for Climate Systems Science

Corresponding author: **s.contractor@unsw.edu.au**

Never Stand Still | Faculty of Science | School of Biological Earth & Env. Science

Abstract

- Existing global gridded datasets of daily precipitation to investigate variability and extremes are lacking due to limited availability of high quality precipitation records
- Here we present our efforts to create a **spatially dense**, **high quality** in situ daily precipitation data for interpolation, by combining major data archives and applying an automated quality controlled process

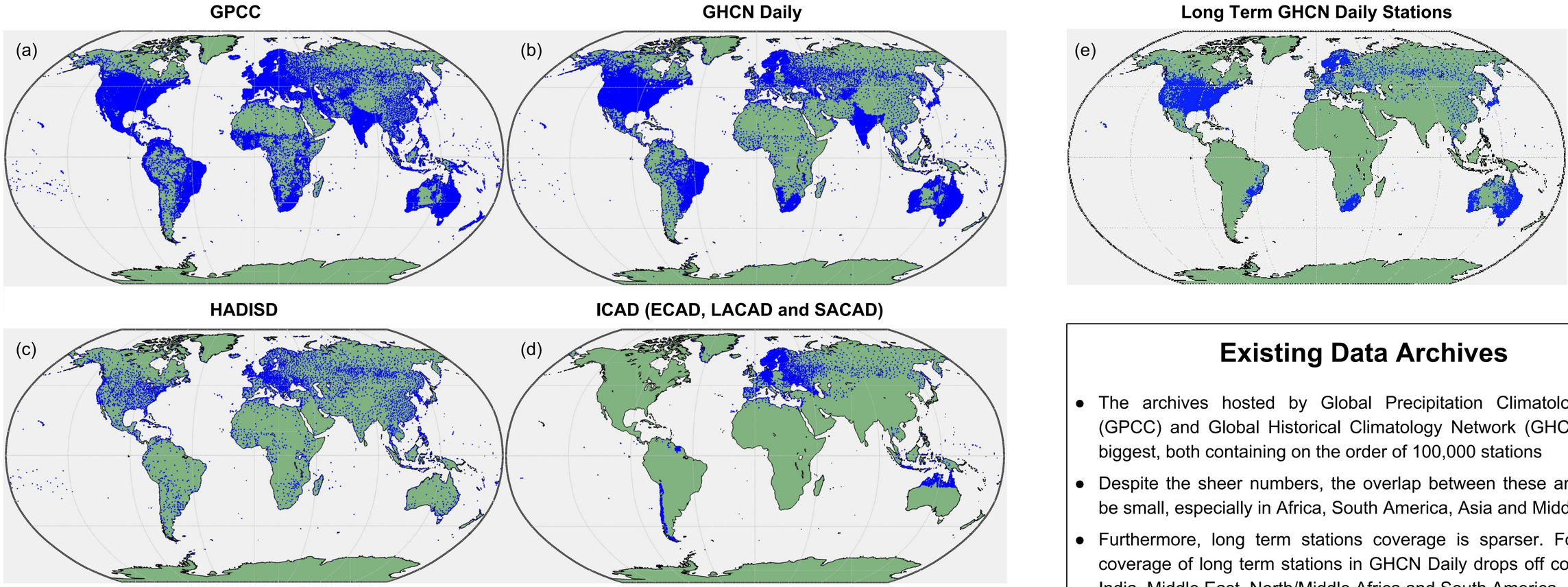
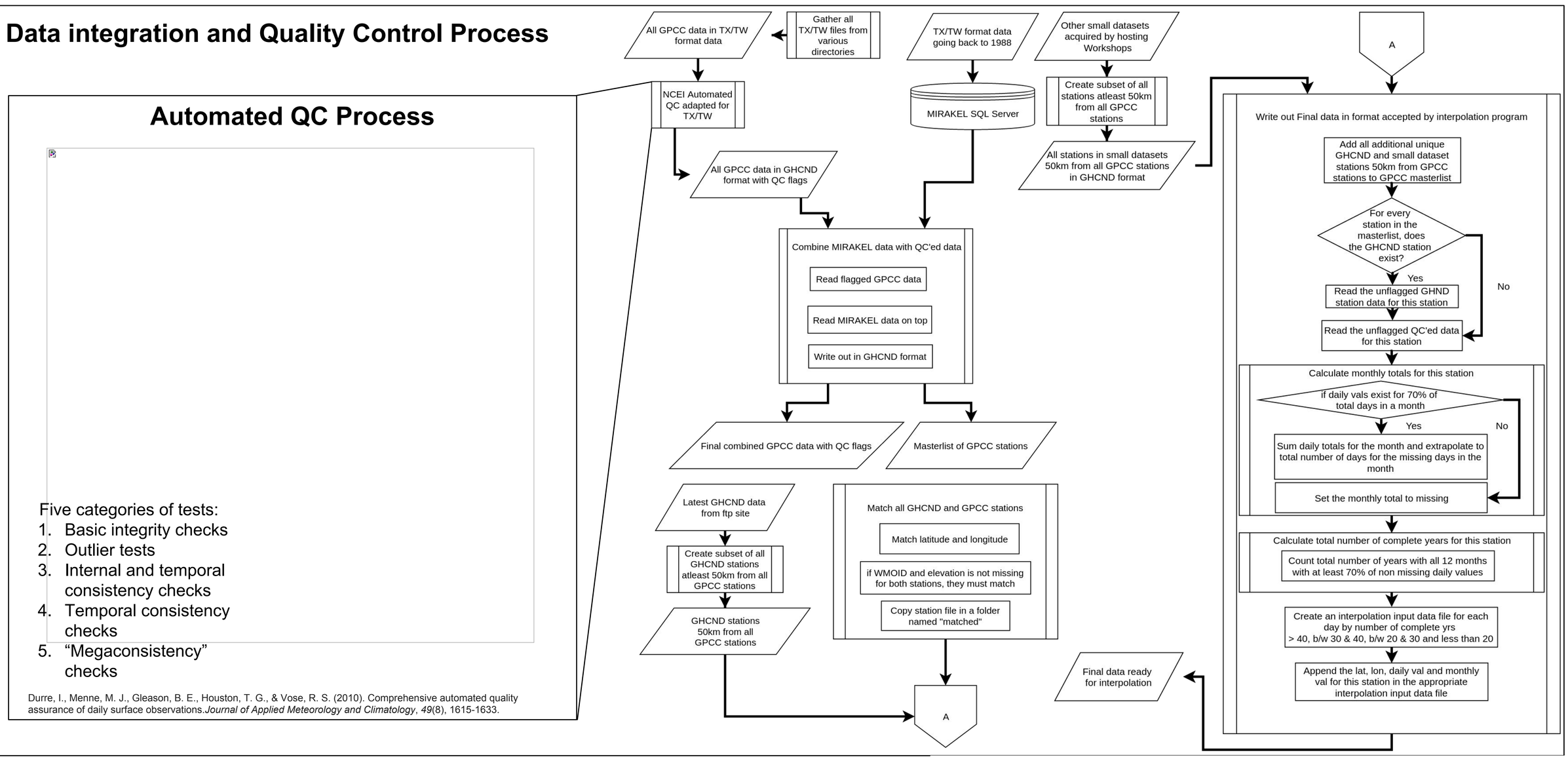


Figure 1: (a)-(d)The four biggest archives of in situ daily/sub-daily precipitation data. (e) Only the stations from GHCN-Daily containing 40 complete years of data.

Existing Data Archives

- The archives hosted by Global Precipitation Climatology Centre (GPCC) and Global Historical Climatology Network (GHCN) are the biggest, both containing on the order of 100,000 stations
- Despite the sheer numbers, the overlap between these archives can be small, especially in Africa, South America, Asia and Middle East.
- Furthermore, long term stations coverage is sparser. For e.g. the coverage of long term stations in GHCN Daily drops off completely in India, Middle East, North/Middle Africa and South America

Data integration and Quality Control Process



Automated QC Process

Five categories of tests:

1. Basic integrity checks
2. Outlier tests
3. Internal and temporal consistency checks
4. Temporal consistency checks
5. "Megaconsistency" checks

Durre, I., Menne, M. J., Gleason, B. E., Houston, T. G., & Vose, R. S. (2010). Comprehensive automated quality assurance of daily surface observations. *Journal of Applied Meteorology and Climatology*, *49*(8), 1615-1633.

Conclusions and Future Work

- The largest data archives were combined to create a raw dataset with the highest achievable spatio-temporal density
- We applied strict quality control to this dataset before using only the long term stations for interpolation
- In future, we will provide various measures of uncertainties and metadata, such as number of stations per grid cell used in interpolation, to assist users with the sensible use of this dataset.

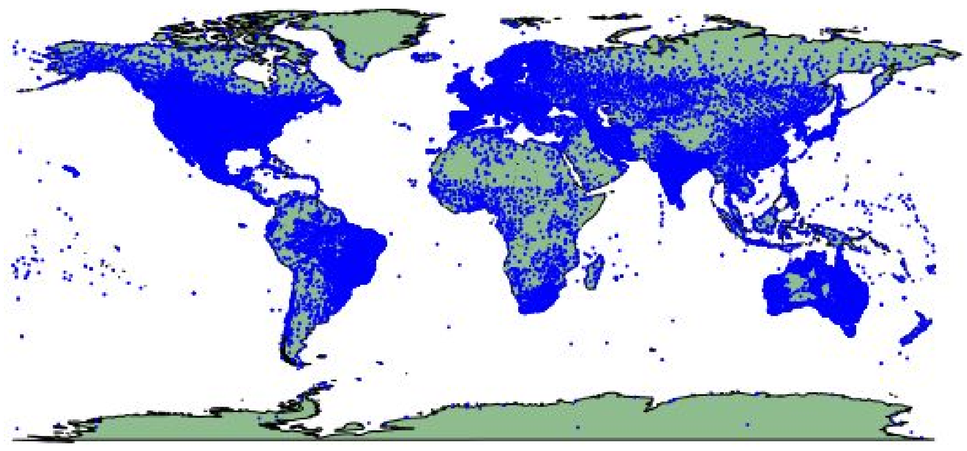


Figure 2: The locations of the final list of stations used for interpolation. There are 94,004 stations.

ARC CENTRE OF EXCELLENCE FOR CLIMATE SYSTEM SCIENCE

Climate Change Research Centre