

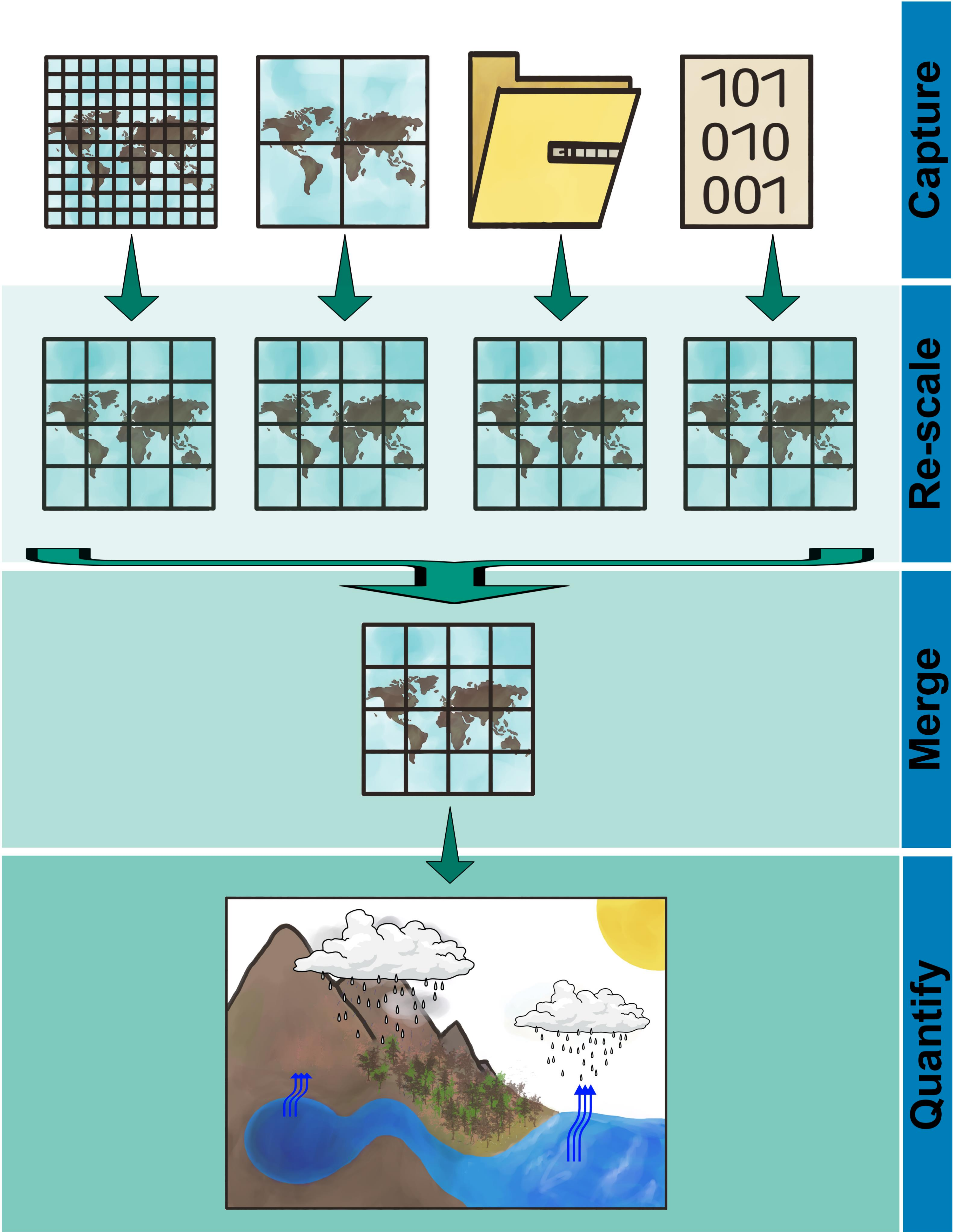
## Abstract

The knowledge of global precipitation is of crucial importance to the study of climate dynamics and the global water cycle in general [8]. Although global precipitation climatologies have existed for some time, and their understanding has improved dramatically due to the vast amount of different data sources, their information has not been comprehensive enough due to precipitation spatial-temporal variability. Thus, ground station reports are, in some cases, not representative of the surrounding areas. Remote sensing data and model simulations complemented the traditional surface measurements and offered unprecedented coverage on a global scale. It is important to note that satellite data records are now of sufficient time frame lengths and with methods “mature” enough to develop meaningful precipitation climatologies that are able to provide information on precipitation patterns and intensities on a global scale. While data (and in some cases exploration/visualization tools as well) are widely available, each dataset comes with different spatial resolution, temporal resolution, and biases.

Consequently, this unique opportunity to obtain a robust quantification of global precipitation has been hindered by the uncertainty, already revealed in the first attempts of the unification of different data products. Herein, we present a multi-source quantification of global precipitation, focusing on the description of the underlying uncertainties. Our approach combines station (CRU, GHCN-M, PRECL, UDEL, and CPC Global), remote sensing (PERSIANN, PERSIANN-CCS, PERSIANN-CDR, GPCP, GPCP\_PEN\_v2.2, CMAP, and CPC-Global) and reanalysis (NCEP1, NCEP2, and 20CRv2) data products, providing an updated overview of the role of precipitation in global water cycle.

## References

- Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P.-P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., et al. (2003). The version-2 global precipitation climatology project (gpcp) monthly precipitation analysis (1979–present). *Journal of hydrometeorology*, 4(6):1147–1167.
- Ashouri, H., Hsu, K.-L., Sorooshian, S., Braithwaite, D. K., Knapp, K. R., Cecil, L. D., Nelson, B. R., and Prat, O. P. (2015). Persiann-cdr: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bulletin of the American Meteorological Society*, 96(1):69–83.
- Chen, M., Xie, P., Janowiak, J. E., and Arkin, P. A. (2002). Global land precipitation: A 50-yr monthly analysis based on gauge observations. *Journal of Hydrometeorology*, 3(3):249–266.
- Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., et al. (2011). The twentieth century reanalysis project. *Quarterly Journal of the Royal Meteorological Society*, 137(654):1–28.
- Harris, I., Jones, P. D., Osborn, T. J., and Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations—the cru ts3. 10 dataset. *International journal of climatology*, 34(3):623–642.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., et al. (1996). The ncep/ncar 40-year reanalysis project. *Bulletin of the American meteorological Society*, 77(3):437–472.
- Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.-K., Hnilo, J., Fiorino, M., and Potter, G. (2002). Ncep–doe amip–ii reanalysis (r-2). *Bulletin of the American Meteorological Society*, 83(11):1631–1644.
- Markonis, Y., Papalexiou, S. M., Martinkova, M., & Hanel, M. (2019). Assessment of Water Cycle Intensification Over Land using a Multisource Global Gridded Precipitation DataSet. *Journal of Geophysical Research: Atmospheres*, 124(21), 11175–11187.
- Markonis, Y., & Strnad, F. (2019). Representation of European hydroclimatic patterns with self-organizing maps. *The Holocene*, 0959683620913924.
- Peterson, T. C. and Vose, R. S. (1997). An overview of the global historical climatology network temperature database. *Bulletin of the American Meteorological Society*, 78(12):2837–2850.
- Sorooshian, S., Hsu, K.-L., Gao, X., Gupta, H. V., Imam, B., and Braithwaite, D. (2000). Evaluation of persiann system satellite-based estimates of tropical rainfall. *Bulletin of the American Meteorological Society*, 81(9):2035–2046.
- Willmott, C. J. and Matsuura, K. (1995). Smart interpolation of annually averaged air temperature in the united states. *Journal of Applied Meteorology*, 34(12):2577–2586.
- Xie, P. and Arkin, P. A. (1997). Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bulletin of the American Meteorological Society*, 78(11):2539–2558.
- Xie, P., Janowiak, J. E., Arkin, P. A., Adler, R., Gruber, A., Ferraro, R., Huffman, G. J., and Curtis, S. (2003). Gpcp pentad precipitation analyses: An experimental dataset based on gauge observations and satellite estimates. *Journal of Climate*, 16(13):2197–2214.
- Xie, P., Chen, M., and Shi, W. (2010). Cpc unified gauge-based analysis of global daily precipitation. In *Preprints, 24th Conf. on Hydrology*, Atlanta, GA, Amer. Meteor. Soc, volume 2.



## Data Sets

NAME	SPATIAL RESOLUTION	TEMPORAL RESOLUTION	RECORD LENGTH	REFERENCE
CRU	0.5°x0.5°	Monthly	1901-2018	[5]
GHCN-N	5°x5°	Monthly	1900-2015	[10]
PRECL	0.5°x0.5°	Monthly	1948-2020	[3]
UDEL	0.5°x0.5°	Monthly	1900-2017	[12]
CPC Global	0.5°x0.5°	Daily	1979-2020	[15]
PERSIANN-CCS	0.04°x0.04°	Monthly	2003-2020	[11]
PERSIANN CDR	0.25°x0.25°	Monthly	1983-2020	[2]
GPCP	2.5°x2.5°	Monthly	1979-2020	[1]
GPCP_PEN_v2.2	2.5°x2.5°	5 Days	1979-2017	[14]
CMAP	2.5°x2.5°	Monthly	1979-2020	[13, 14]
CPC-Global	0.5°x0.5°	Daily	1979-2020	[15]
NCEP1	T62	Monthly	1948-2020	[6]
NCEP2	T62	Monthly	1979-2020	[7]
20CRv2	1°x1°	Monthly	1836-2015	[4]

## Methods

- Capture: To download the different data sets. As previously mentioned and evidenced by the above table, each data set comes with different spatiotemporal resolutions. Furthermore, each data provider has its storage and naming systems, thus, adding different file formats into the mix as well.
- Re-scale: To transform the individual data sets into a common spatiotemporal resolution because we do need all the data to be in the same spatiotemporal resolution before we move on [9]. In addition, we have to guarantee that any alteration of the spatiotemporal resolution will have a minor impact on the statistical properties of the original data.
- Merge: To combine the re-scaled data sets into a single one by means of a weighted average. The weight of a given data set will be inversely proportional to its difference to the mean of all data sets. Basic unbiasing will tackle the time periods in which not all data sets overlap at the same time (e.g. Despite CRU and CPC-Global having a similar spatiotemporal resolution, CRU data record starts in 1901, whereas CPC-Global record starts in 1979).
- Quantify: To this point we would have produced a single data set to be used in the quantification of precipitation in the global water cycle, and with these global time series we will be able to analyze the climatology of global precipitation.

A necessary component of the estimation of precipitation in the global water cycle is uncertainty quantification and validation analysis. Through the above steps we will keep track of the errors, their propagation, and possible creation by our processing. Several statistical metrics like RMSE, FAR, and correlation coefficient to mention some of them, will be adopted to quantify uncertainty and validate different data sets among themselves as well as versus the new data set generated by the merge.