Data dictionary

Original data:

The rainfall data (2B 31 V 7) for the region with latitude range 21°N to 31°N and longitude range 84°E to 90°E, has been downloaded from http://trmm.gsfc.nasa.gov/ (with the data locator: http://disc.gsfc.nasa.gov/datacollection/2B31.V7.html). These data are available in .HDF format. We converted the data to .csv format. We merge the data arising from all satellite passes of a monsoon month and save them as monthly data for each of the year over 2002-2012.

The following are description of the data files

1) The folder "rf_data" contains twelve subfolder. Each subfolder contains boundary contours of the rainfall regions of the monsoon months of each of the year 2002-2012. The following is the description of the variables:

M6_cluster_points: represents the co-ordinates of the boundary of a rainfall regions, M6_cluster_lengths: number of boundary point of a rainfall region, M6_cen_spell: index of regions which are censored (not completely observed), M6_rainfall_rate: rainfall rate (positive) in each of the grids of a rainfall region.

2) The folder "satellite_passes_2012" contains four files. Each file contains rainfall data corresponding to a monsoon month (June, July, August and September) of the year 2012. Each of the csv file contains eight variables (columns) such as long (longitude), lat (latitude), rainfall rate (rrsurf), year, day, hour, minute, second.

Processed data:

We provided processed data for the entire period (2002-2012) considered in the manuscript.

1) The folder "starhull_complete_contours" contains twelve files (named according to corresponding year) each containing the star-hulls of rainfall regions for each of the years 2002-2012. A year with n rainfall regions consists of a 1000xn matrix with each column representing the star-hull contour of a rainfall region (i.e., 1000 equi-spaced point from the boundary).

2) The folder "area_yearwise" contains twelve files (named according to the corresponding year). Each file contains a single column containing area of the rainfall regions for a particular year.

3) The folder "censor_status_yearwise" contains twelve files (named according to corresponding year). Each file contains a single column containing the status of the rainfall regions: whether the rainfall region is censored or completely observed.

4) The folder “area_status_combined” contains two files. These two files contain area and censor status of all monsoon rainfalls regions over the years 2002-2012.

5) The file “map.txt” contain the map co-ordinates (geographical reference) of the region of interest of the paper.