Weekly Updates (09/05/24)

Evaluating rainfall data ⇲

Satellite vs Ground Observations

PRESENTED BY:
Rafa Tagulao

content

datasets ↘

my scope of study: 2013 to 2018

Weather Philippines Foundation (WPF)	Multi-Source Weighted-Ensemble Precipitation (MSWEP)	Integrated Multi-satellitE Retrievals for GPM (IMERG)
2012 to 2019 (missing oct-dec 2012, and **sept-dec 2014)**	2013 to 2020 (missing **2019 data)**	2013 to 2019
10-minute precipitation	3-hourly precipitation	30-minute precipitation

WPF Data ↘

- contains multiple columns, but only extracted **‘Datetime’, ‘station_id’, ‘Rain’, ‘FLAG’**
  - **‘FLAG’** for rainfall are A2 and A3
    - A2: Extreme test: >25 mm/ 10 min (150mm/hr) rainfall
    - A3: Variability Test: difference between a data point and the data point before/after it is more than eight (8) times the standard deviation
  - Dropped rows with **A2 and A3 Flag**
- Dropped the days with at least >20% missing (10-minute) data
- Aggregated per day for the remaining stations
  - 767 Stations

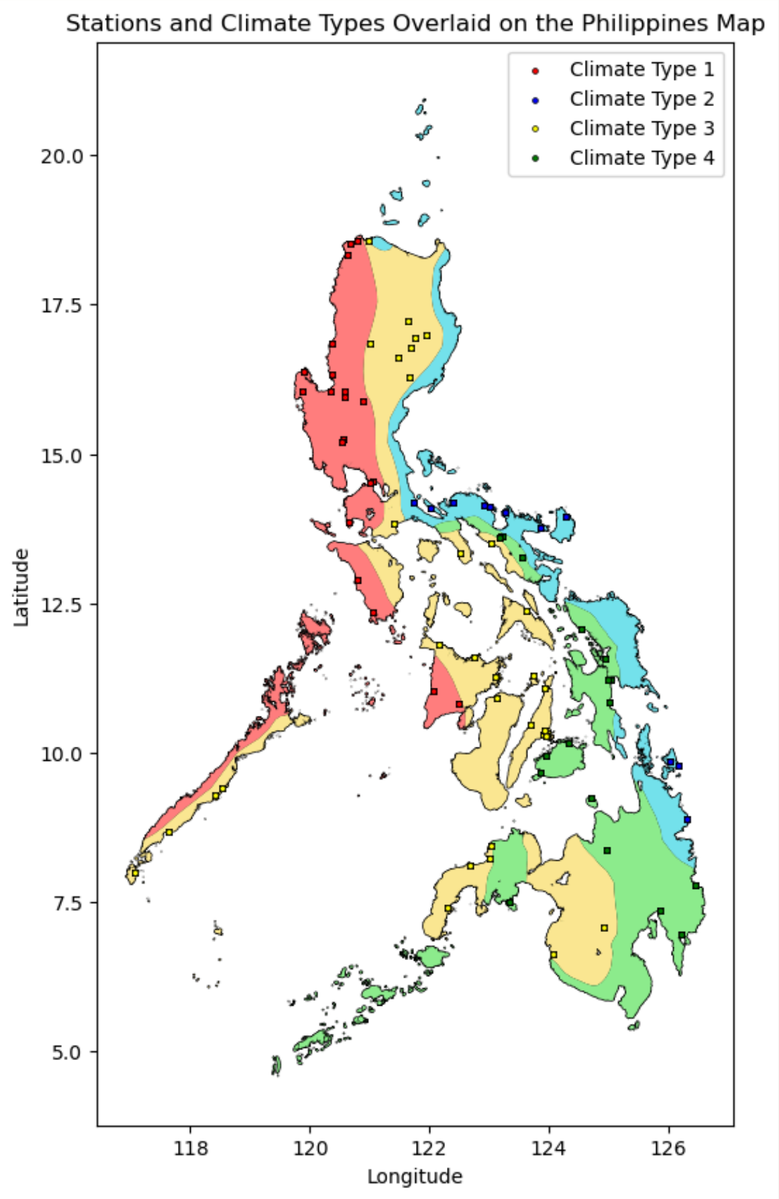
WPF Data ↘

	station_id	year	month	day	Rain	CLIM_TYPE
0	980001	2013	2	15	0.0	1.0
1	980001	2013	2	16	0.0	1.0
2	980001	2013	2	21	0.0	1.0
3	980001	2013	2	22	4.0	1.0
4	980001	2013	2	23	0.0	1.0
...	...	...	...	...	...	...
757865	980989	2018	8	19	10.6	1.0
757866	980989	2018	8	20	2.2	1.0
757867	980989	2018	8	21	0.0	1.0
757868	980989	2018	8	22	0.0	1.0
757869	980989	2018	8	23	0.0	1.0

- Some stations had no coordinates (from the csv file from Eco)
  - 688 stations remain
    - Type 1: 223
    - Type 2: 121
    - Type 3: 222
    - Type 4: 122

WPF Data ↘

- More quality control for data
  - Dropped stations with >20% missing data for at least 1 year
  - Further reduced the number of **years** and **stations**
    - **years:** no more remaining stations for 2013 and 2014
    - **stations:** only 82 left
      - Type 1: 20
      - Type 2: 11
      - Type 3: 33
      - Type 4: 18



IMERG Data ↘

- Identified the closest coordinates in IMERG that matches with WPF station data

	station_id	lat	lon	CLIM_TYPE
0	980088	16.278100	121.6582	3.0
364	980133	7.400000	122.3000	3.0
728	980183	10.292700	123.9019	3.0
2133	980369	16.937100	121.7692	3.0
3573	980371	13.620500	123.1899	4.0
...	...	...	...	...
90915	980949	10.268300	123.9600	3.0
91561	980952	9.283400	118.4222	3.0
92234	980953	8.677431	117.6346	3.0
92900	980955	18.526837	120.6872	1.0
93577	980987	7.370087	125.8621	4.0

IMERG Data

	station_id	CLIM_TYPE	year	month	day	imerg_daily_Pr
0	980088.0	3.0	2013	1	1	0.00
1	980088.0	3.0	2013	1	2	0.00
2	980088.0	3.0	2013	1	3	0.59
3	980088.0	3.0	2013	1	4	3.38
4	980088.0	3.0	2013	1	5	0.00
...	...	...	...	...	...	...
179739	980987.0	4.0	2018	12	28	0.88
179740	980987.0	4.0	2018	12	29	0.80
179741	980987.0	4.0	2018	12	30	7.42
179742	980987.0	4.0	2018	12	31	0.76
179743	980987.0	4.0	2019	1	1	1.71

Comparison ↘

- Merged the daily data from WPF and IMERG

	station_id	year	month	day	Rain	CLIM_TYPE	imerg_daily_Pr
0	980088	2015	1	2	0.00	3.0	0.00
1	980088	2015	1	3	0.00	3.0	0.00
2	980088	2015	1	4	0.00	3.0	0.00
3	980088	2015	1	5	1.29	3.0	0.00
4	980088	2015	1	6	5.42	3.0	0.00
...	...	...	...	...	...	...	...
93886	980987	2018	12	27	0.00	4.0	0.95
93887	980987	2018	12	28	0.00	4.0	0.88
93888	980987	2018	12	29	0.00	4.0	0.80
93889	980987	2018	12	30	7.20	4.0	7.42
93890	980987	2018	12	31	0.00	4.0	0.76

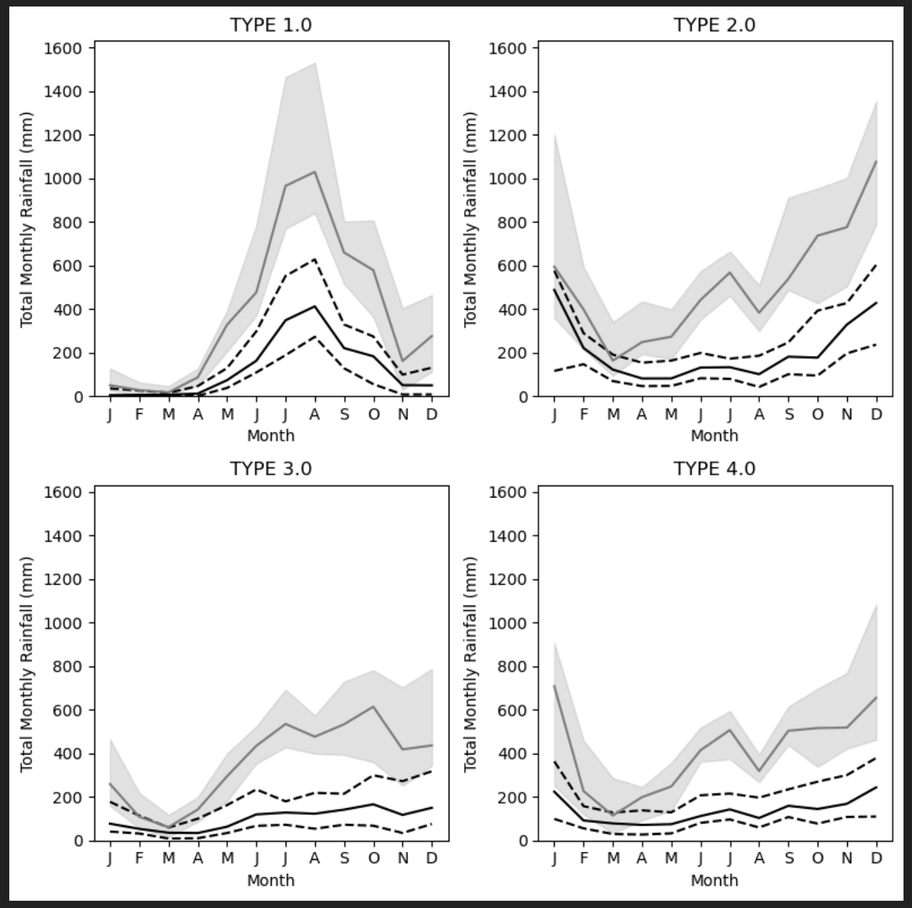
Comparison ↘

- Used both daily and monthly/seasonal aggregates
  - Removed stations with at least 5 missing days per month for the aggregates
- Categorized the rainfall values for extreme rainfall indices

Event category	Rainfall range
Dry days	<1 mm
Light-to-moderate rain days	1–10 mm
Heavy rain days	10–20 mm
Very heavy rain days	>20 mm

Comparison ↘

monthly percentiles for each climate type



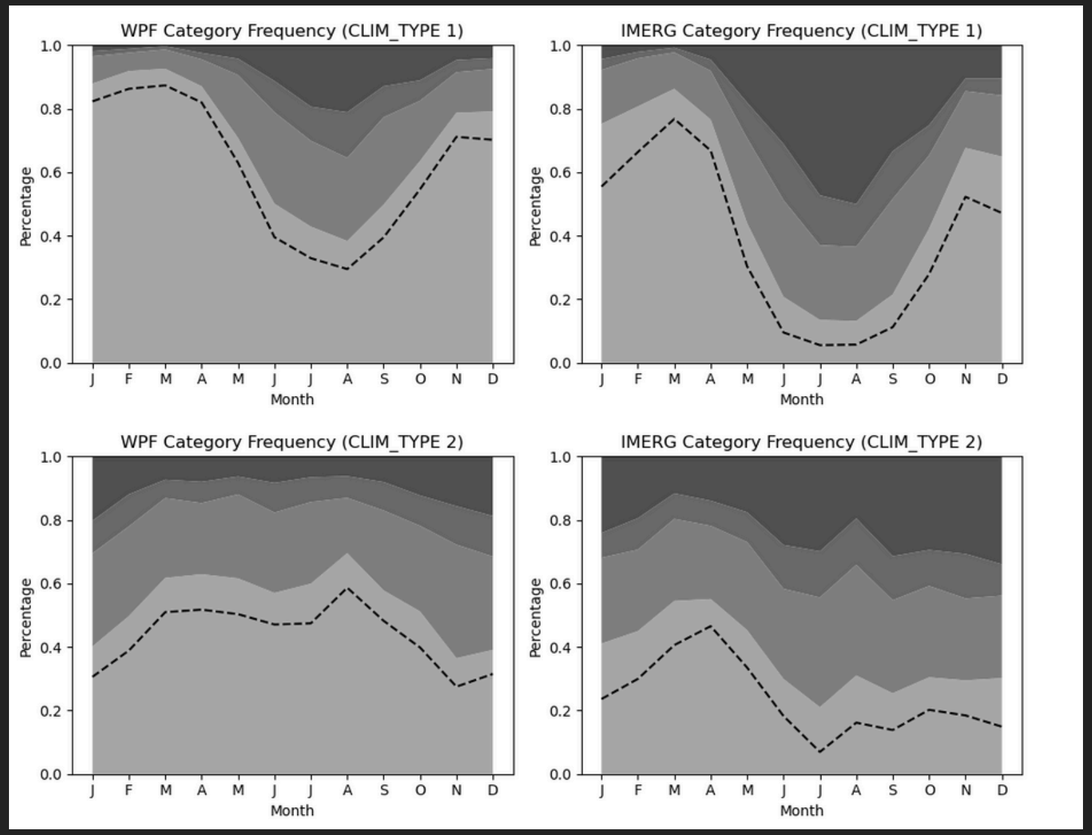
- Solid black line: median of WPF;
- solid gray line: median of IMERG
- Broken black lines: P25 and P75 of WPF;
- gray area: P25 and P75 of IMERG

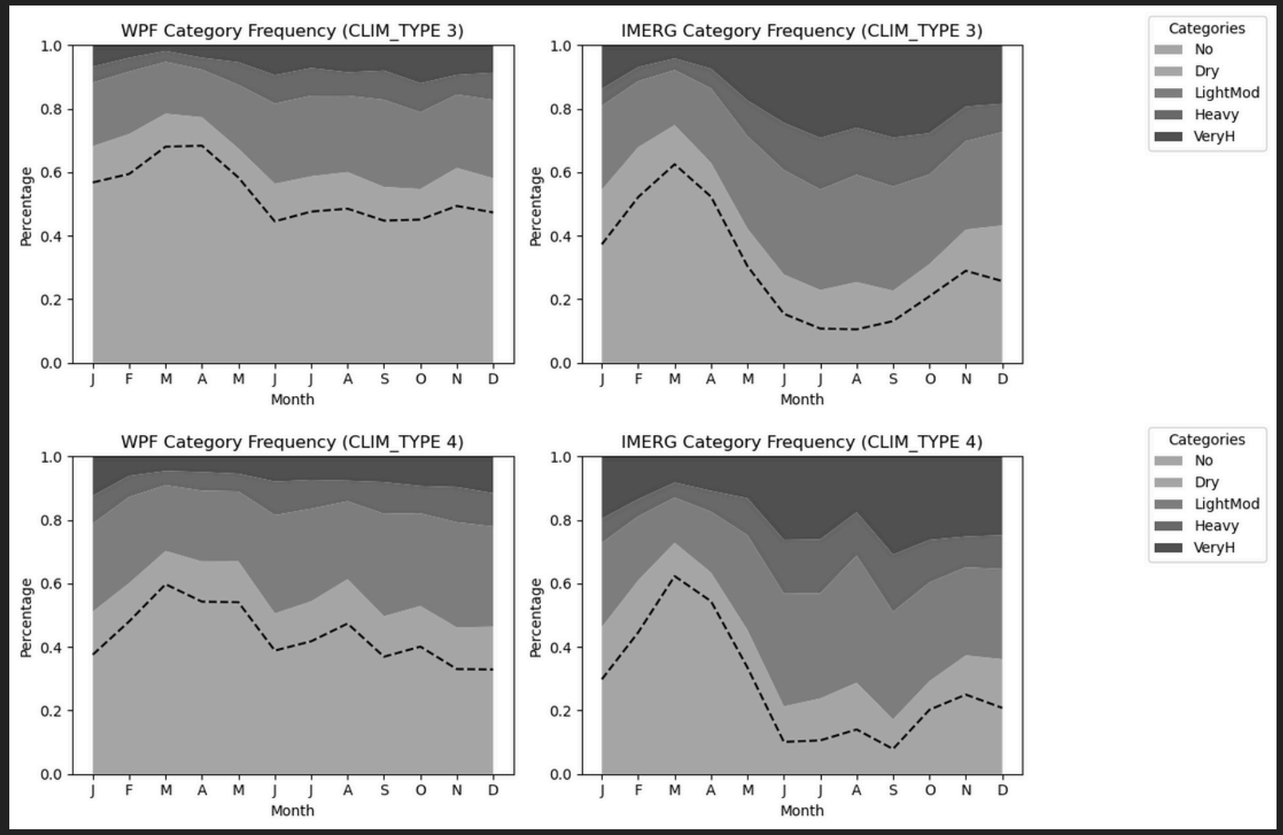
For all climate types, monthly spread is not well-represented by IMERG

- IMERG tends to **overestimate median monthly rainfall for all types**, with huge deviations in the months beyond April.
- Overall, monthly statistics does not fit well with that of WPF's as IMERG tends to **overestimate**

Comparison ↘

monthly category frequency





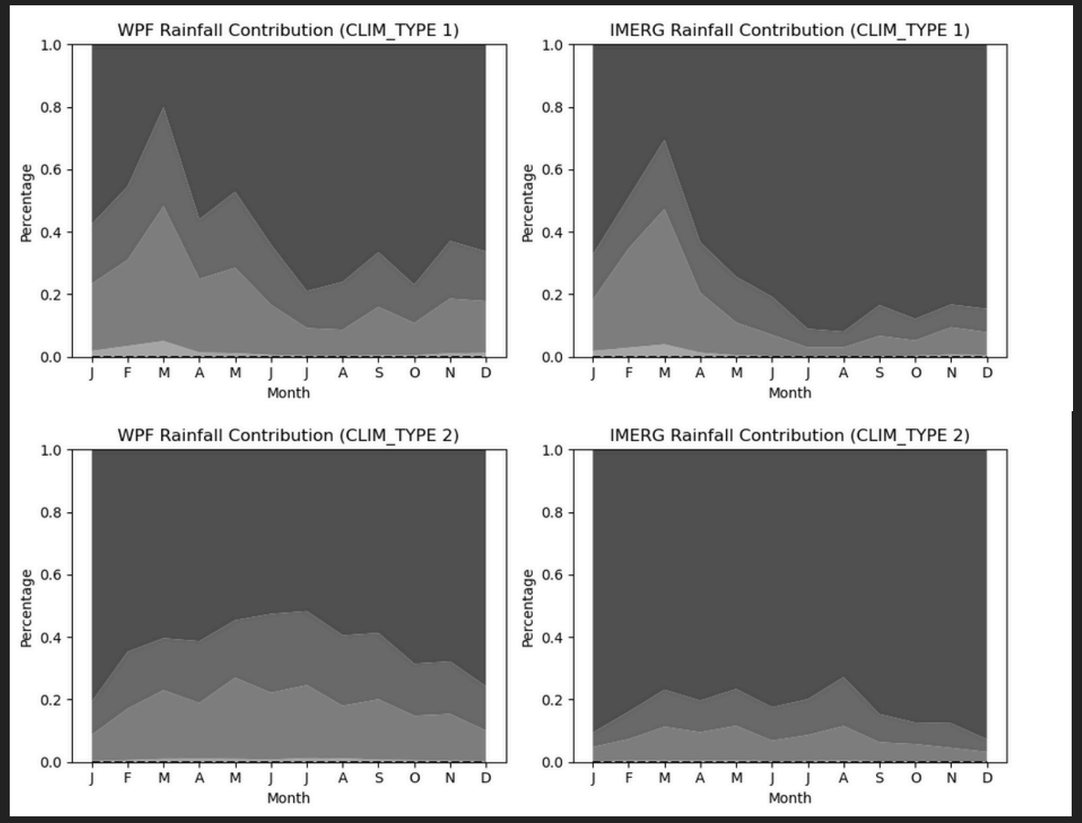
Comparison ↘

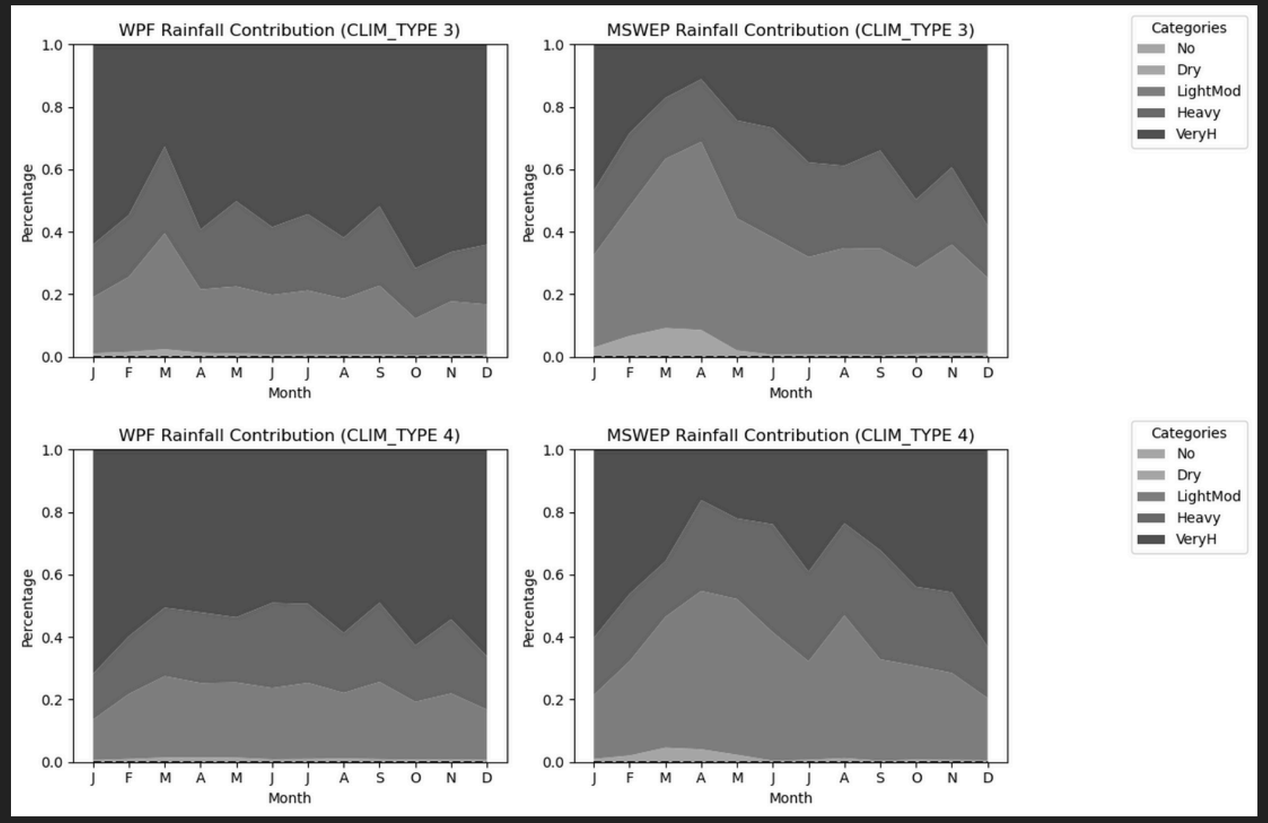
monthly category frequency

- For all Climate Types, IMERG **underrepresents "No Rain" Days and "Dry" Days** during MJJAS.
- Does relatively well in representing **"Light to Moderate" and "Heavy" Rain** Days
- Largely **overrepresents "Very Heavy"** days during **MJJAS** for Climates 1 and 3, and all throughout for Climate Types 2 and 4

Comparison ↘

monthly category contribution





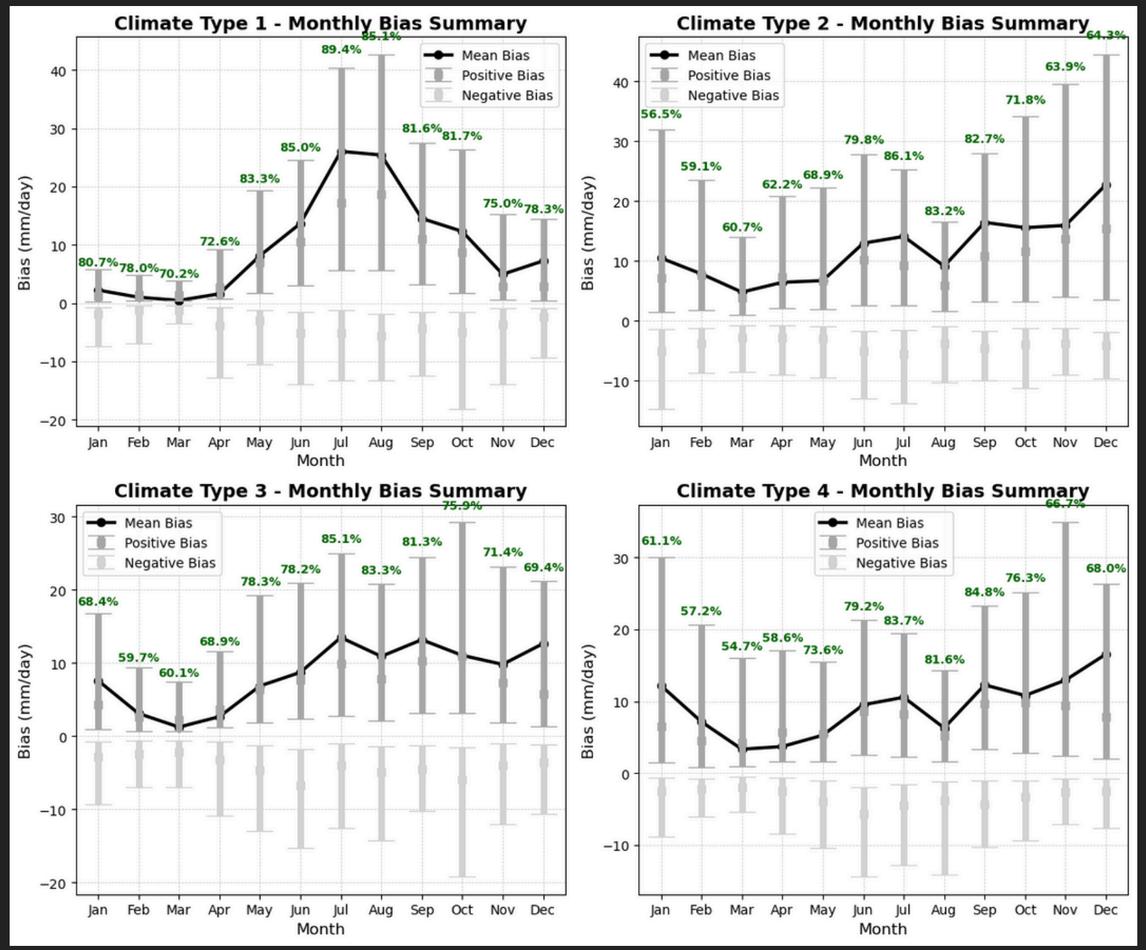
Comparison ↘

monthly category contribution

- IMERG underrepresents rainfall contribution for **“Light to Moderate” and “Heavy” days from March onwards** for Climate Type 1
- Overrepresents **“Light to Moderate”,** underestimates **“Very Heavy” Days** for Climate Type 2
- Little underrepresentation for **“Light to moderate”** and **“Heavy“** days for Climate Type 3
- But overrepresentation for **“Light to moderate”** and **“Heavy“** days for Climate Type 4

Comparison ↘

monthly statistics on bias



For all months, IMERG to overestimate the daily rainfall

- Very large positive biases for all months (with mean bias has large as 20mm/day)
- Majority of biases are all positive

Comparison ↘

Statistical Metrics for rainfall

	Climate Type	Aggregation	RMSE	MAE	R2	KS_Statistic	KS_P_Value
0	1.0	Daily	30.036577	12.296692	0.448346	0.238691	0.000000e+00
1	1.0	Monthly	481.948104	327.482416	0.570190	0.356098	1.465362e-46
2	1.0	Seasonal	1234.184808	883.361142	0.563954	0.371237	1.003472e-18
3	2.0	Daily	37.656870	16.542251	0.319377	0.204901	4.131646e-252
4	2.0	Monthly	505.024173	394.486177	0.137349	0.515012	1.026700e-52
5	2.0	Seasonal	1337.827145	1070.880610	0.100926	0.574194	5.990653e-24
6	3.0	Daily	28.500256	12.543467	0.149435	0.245010	0.000000e+00
7	3.0	Monthly	384.414461	300.729945	0.061908	0.491589	1.086604e-117
8	3.0	Seasonal	994.449370	805.216618	0.033220	0.552083	3.878001e-54
9	4.0	Daily	29.818381	12.758140	0.307476	0.184985	2.783493e-299
10	4.0	Monthly	374.402778	300.886201	0.217011	0.518868	1.769646e-78
11	4.0	Seasonal	982.317709	817.340458	0.187027	0.599119	3.267916e-38

Really bad metrics!

- **Very high values of RMSE and MAE** as aggregation increases
- **low R^2 values** in general except for relatively high values for Climate Type 1
- **KS Statistics** have p<0.001 for all, indicating that their distribution significantly differ

Concerns ↘

- IMERG involves both land and ocean data, so I might have to remove stations that are near water bodies and check if results will change