Weekly Updates (08/28/24)

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)**	yet to explore :)
10-minute precipitation	3-hourly 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

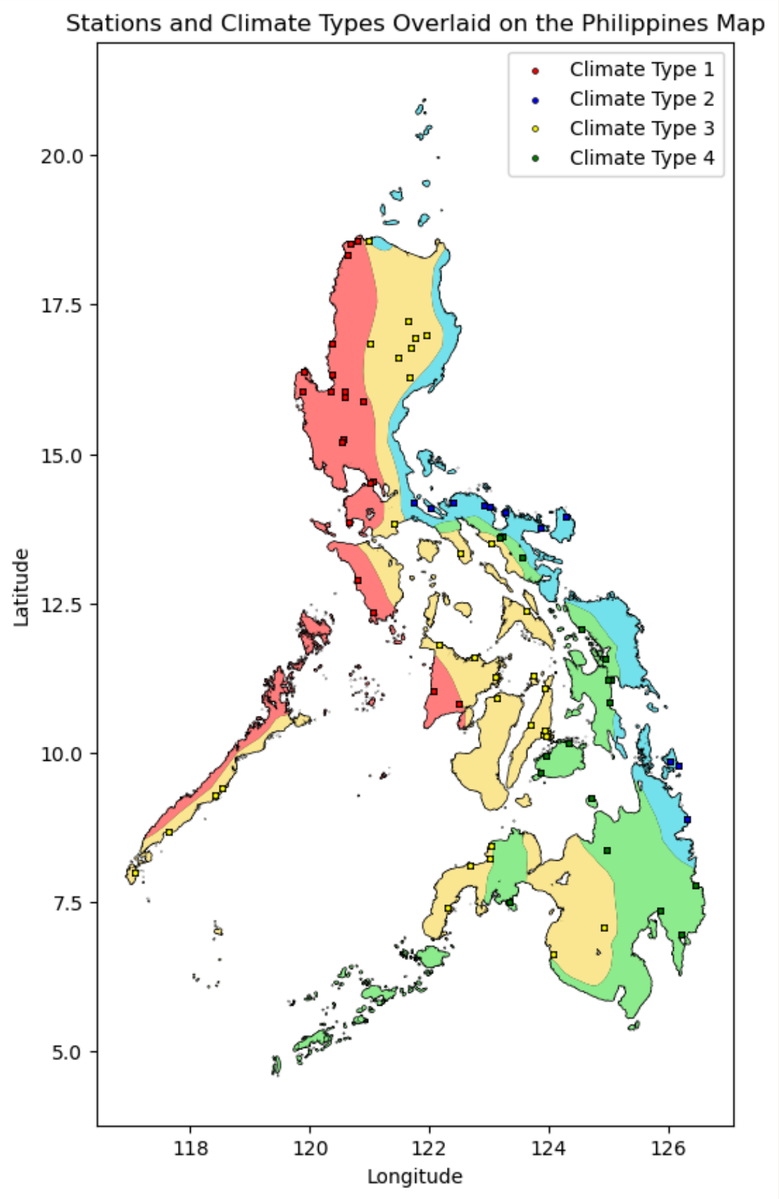
WPF Data ↘

	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

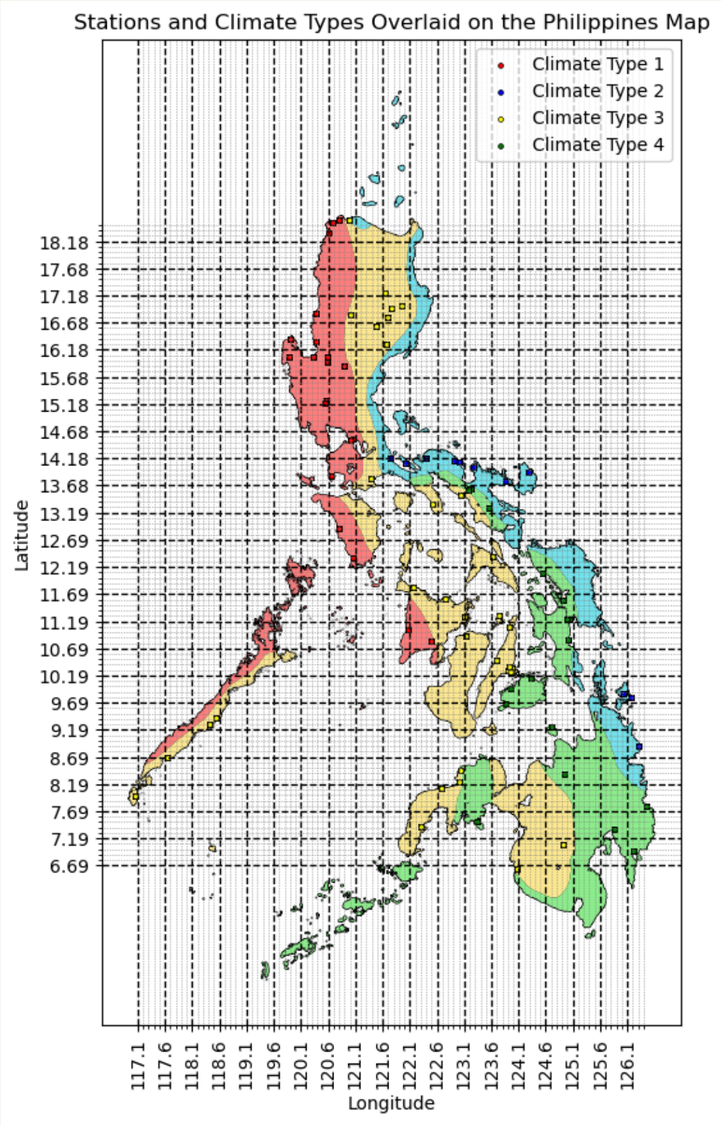
WPF Data ↘

- 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



MSWEP Data ↘

- Overlaid grid lines for MSWEP



MSWEP Data ↘

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

	station_id	lat_MSWEP	lon_MSWEP	CLIM_TYPE
0	980088.0	16.249996	121.650009	3.0
1	980133.0	7.350001	122.250015	3.0
2	980183.0	10.249995	123.949997	3.0
3	980369.0	16.950001	121.750015	3.0
4	980371.0	13.649997	123.150009	4.0
...	...	...	...	...
77	980949.0	10.249995	123.949997	3.0
78	980952.0	9.249995	118.449997	3.0
79	980953.0	8.649997	117.650009	3.0
80	980955.0	18.549999	120.650009	1.0
81	980987.0	7.350001	125.850021	4.0

MSWEP Data ↘

	station_id	CLIM_TYPE	year	month	day	mswep_daily_Pr
0	980088.0	3.0	2013	1	1	0.5625
1	980088.0	3.0	2013	1	2	1.1250
2	980088.0	3.0	2013	1	3	2.8125
3	980088.0	3.0	2013	1	4	10.9375
4	980088.0	3.0	2013	1	5	2.8125
...	...	...	...	...	...	...
179739	980987.0	4.0	2018	12	28	0.0000
179740	980987.0	4.0	2018	12	29	0.7500
179741	980987.0	4.0	2018	12	30	4.6250
179742	980987.0	4.0	2018	12	31	1.7500
179743	980987.0	4.0	2019	1	1	3.6250

- Aggregated per day to obtain daily data

Comparison ↘

- Merged the daily data from WPF and MSWEP

	station_id	year	month	day	wpf_daily_Pr	CLIM_TYPE	mswep_daily_Pr
0	980088	2015	1	2	0.00	3.0	0.6250
1	980088	2015	1	3	0.00	3.0	0.5625
2	980088	2015	1	4	0.00	3.0	0.2500
3	980088	2015	1	5	1.29	3.0	0.9375
4	980088	2015	1	6	5.42	3.0	1.6875
...	...	...	...	...	...	...	...
93886	980987	2018	12	27	0.00	4.0	0.0000
93887	980987	2018	12	28	0.00	4.0	0.0000
93888	980987	2018	12	29	0.00	4.0	0.7500
93889	980987	2018	12	30	7.20	4.0	4.6250
93890	980987	2018	12	31	0.00	4.0	1.7500

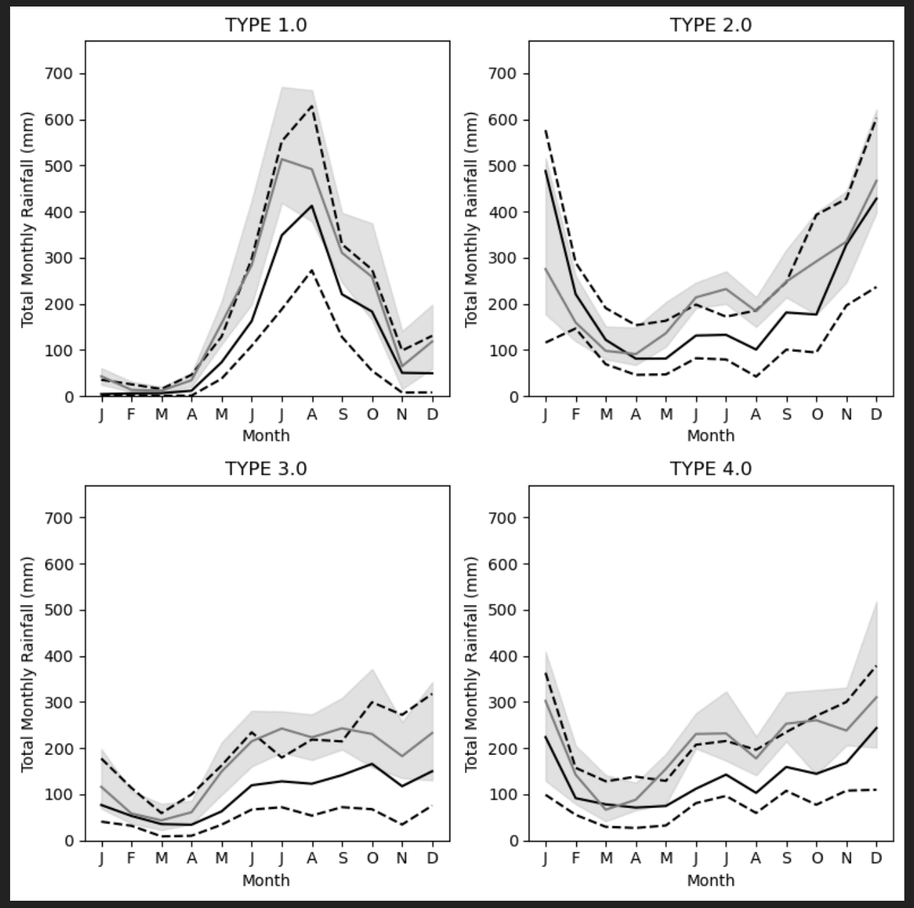
Comparison ↘

- 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



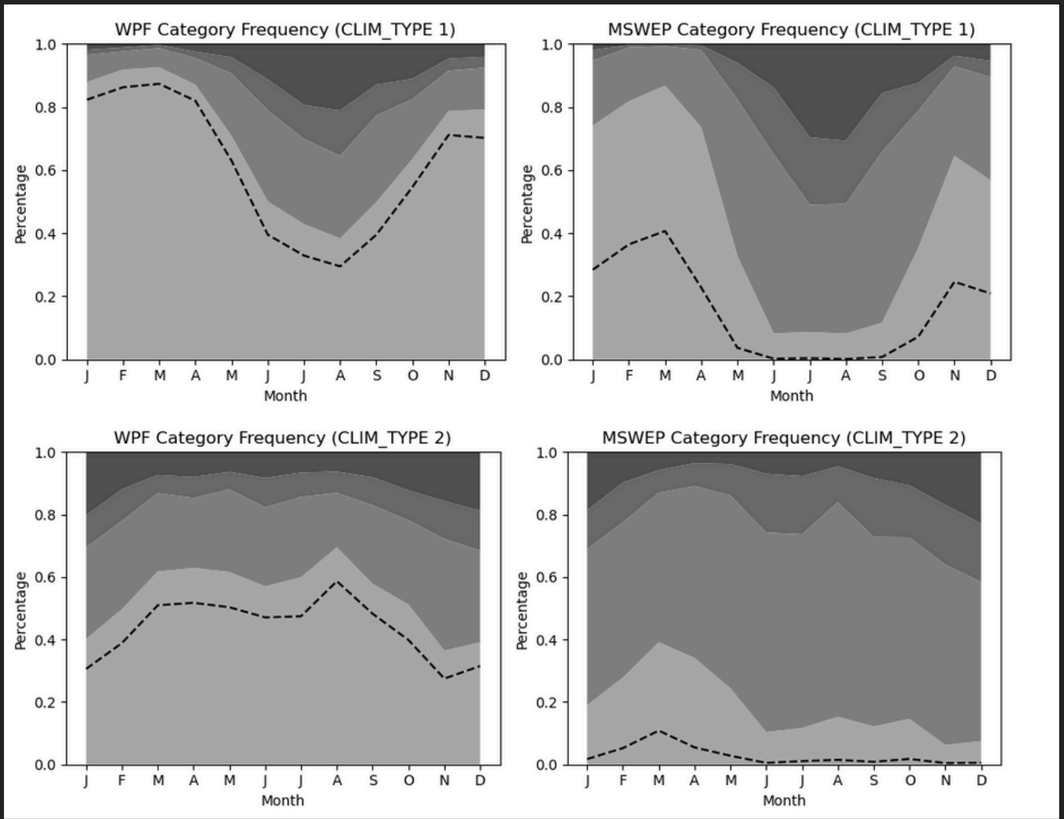
- Solid black line: median of WPF;
- solid gray line: median of MSWEP
- Broken black lines: P25 and P75 of WPF;
- gray area: P25 and P75 of MSWEP

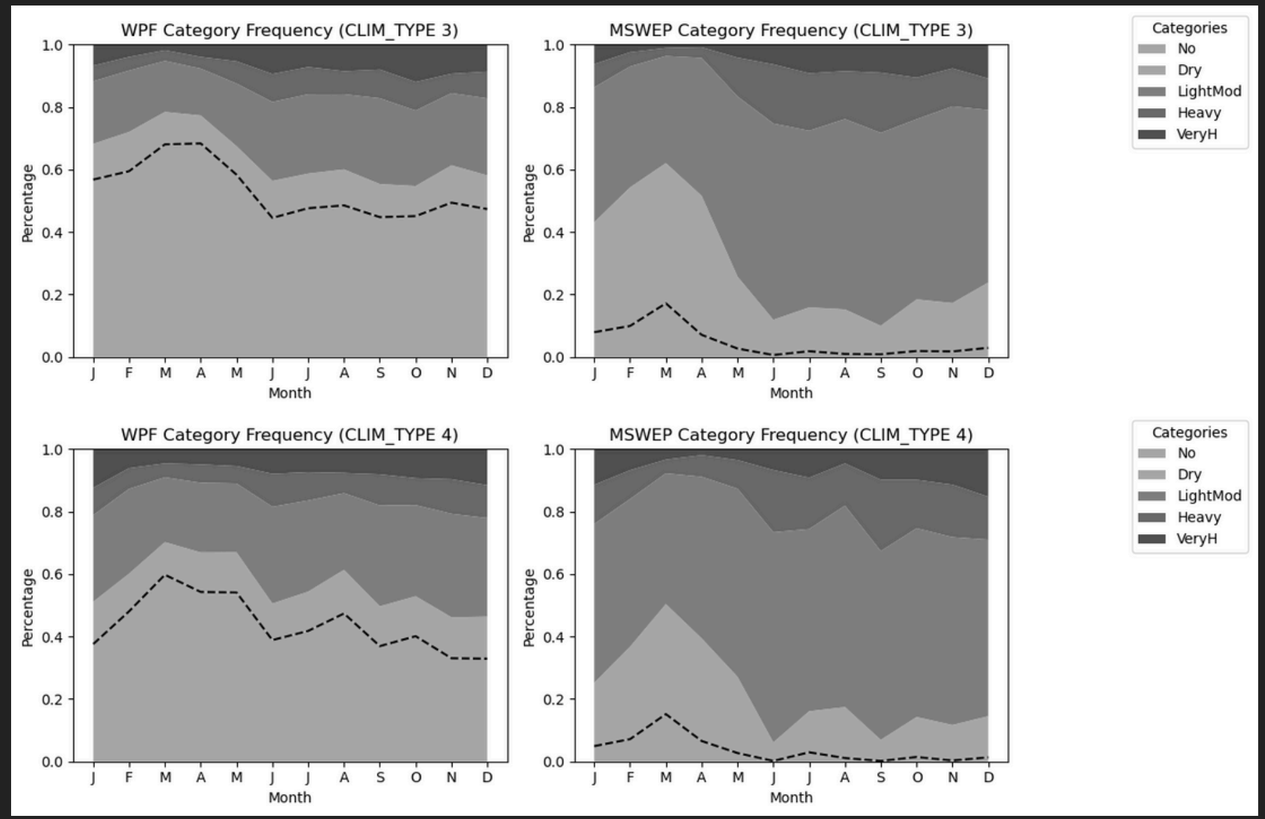
For all climate types, monthly spread is not well-represented by MSWEP

- MSWEP tends to **overestimate median monthly rainfall for all types**, with huge deviations in the months beyond April
- P25 for MSWEP is usually higher than median of WPF for all climate types

Comparison ↘

monthly category frequency





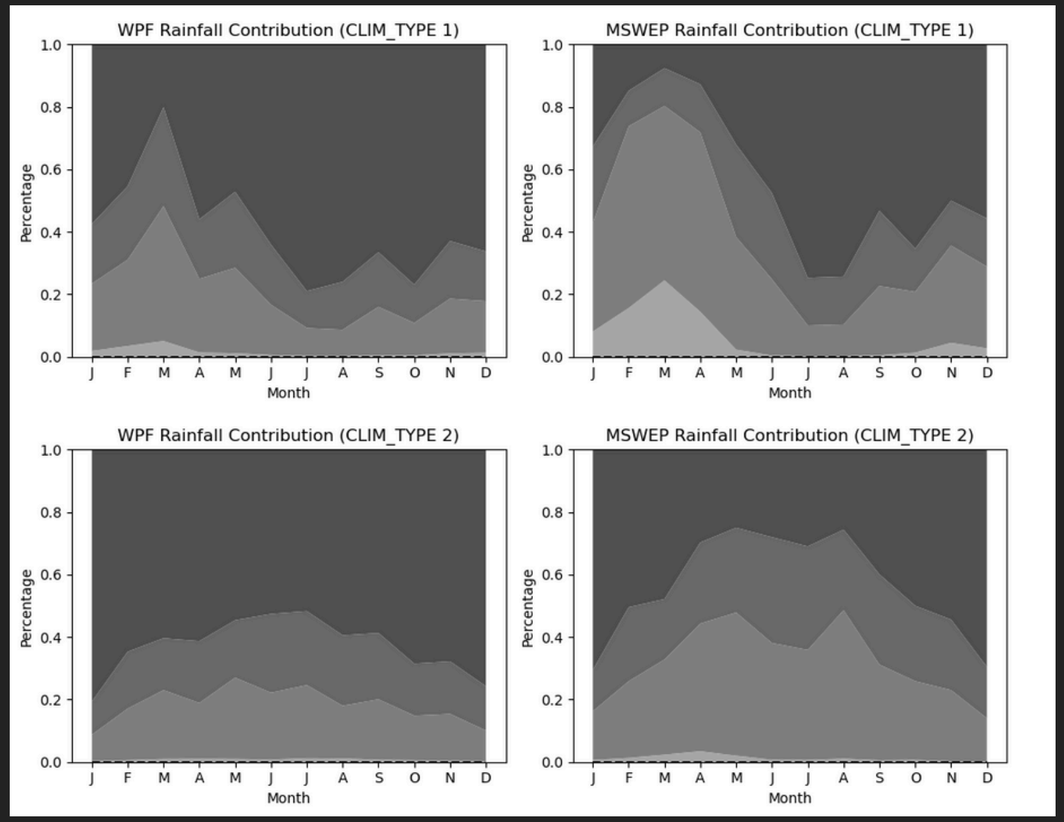
Comparison ↘

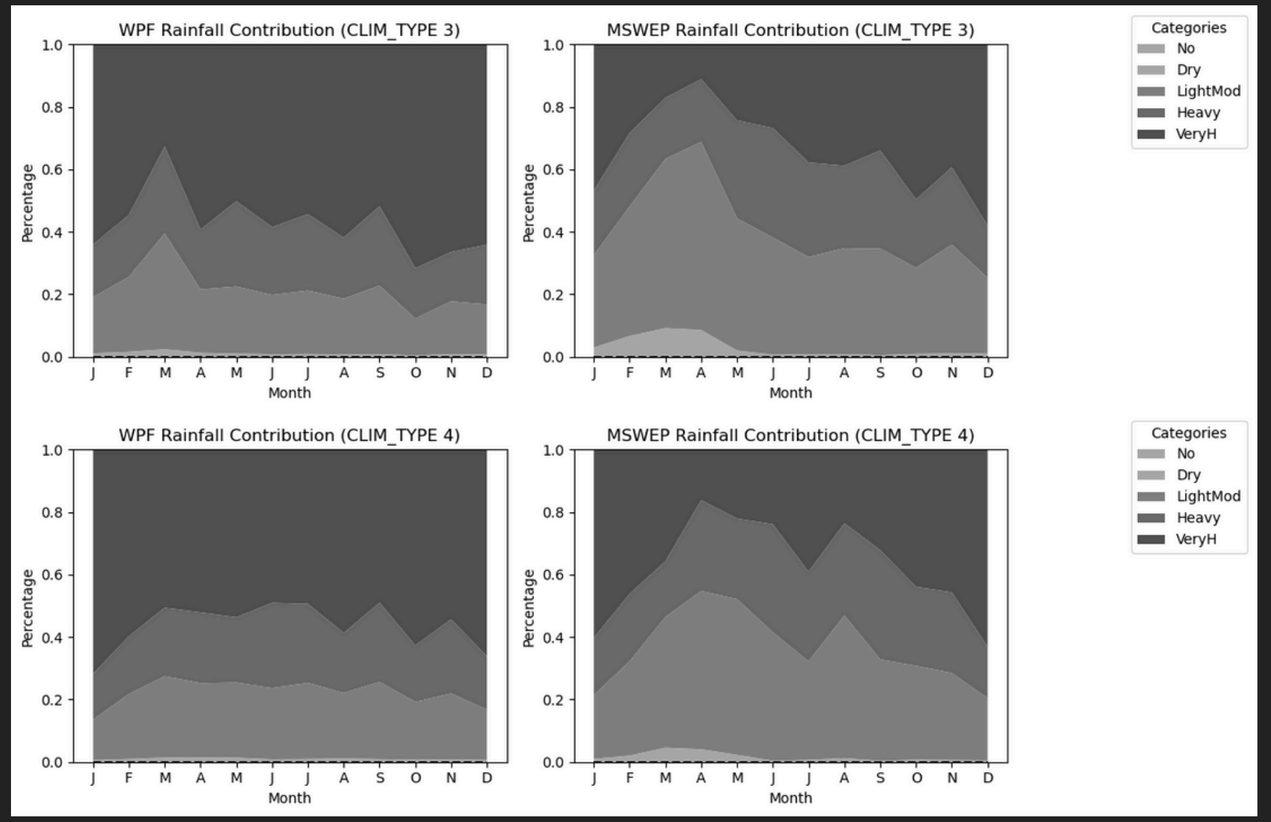
monthly category frequency

- MSWEP consistently **underrepresents** “No Rain” and “Dry” days especially for Climate Types 2, 3, and 4
- **Overrepresents** “Light to Moderate” Days
- Does **relatively well** for “Heavy” and “Very Heavy” Days

Comparison ↘

monthly category contribution





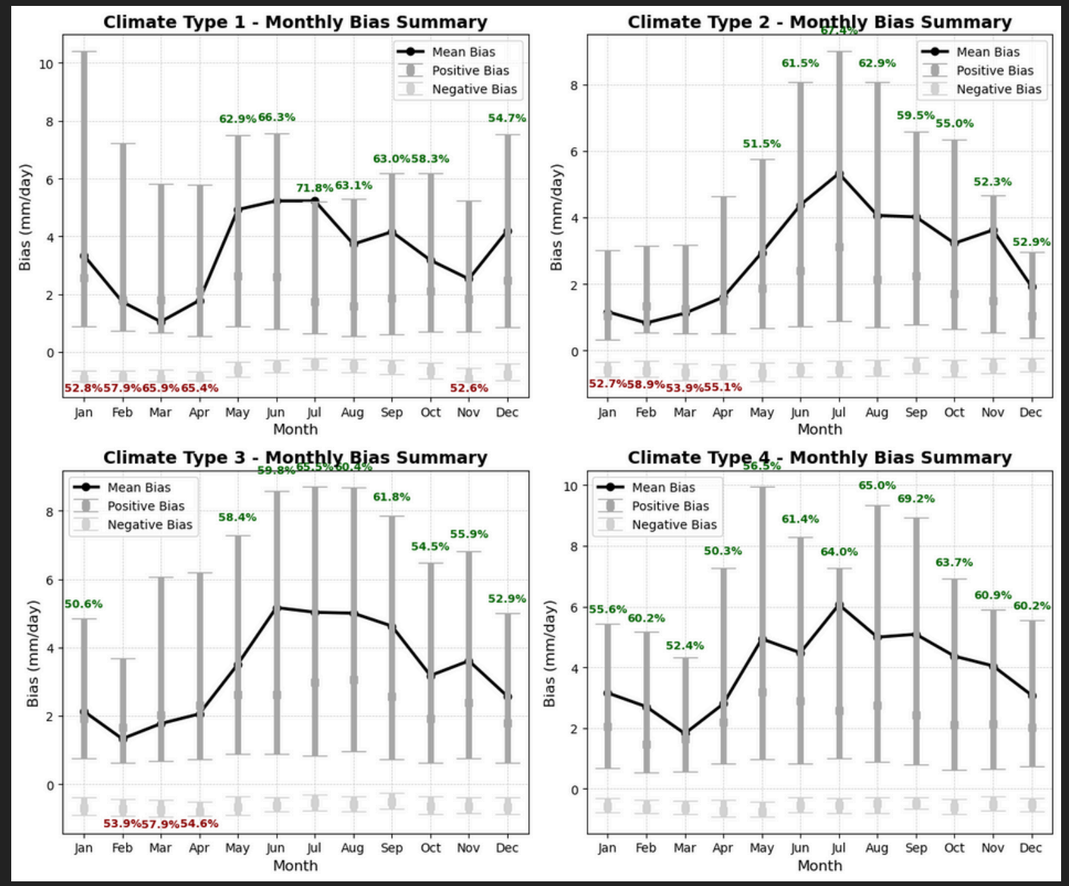
Comparison ↘

monthly category contribution

- MSWEP underrepresents rainfall contribution for **“Very heavy” days,** and overrepresents for **“Light to Moderate” days**
- **Overrepresents** rainfall contribution for “Light to Moderate” rains
- Does fairly well on **“Heavy” days** and **“Dry” days** (with some overestimation on November to May

Comparison ↘

monthly statistics on bias



For all months, MSWEP tends to overestimate the daily rainfall

- Larger positive biases than negative biases; number of positive biases are also more than the negative
- Positive **Mean** Bias for all months
- Majority of biases are negative for first few months (Climate types 1, 2 and 3)

Comparison ↘

Statistical Metrics for rainfall

Climate Type	Aggregation	RMSE	MAE	R2	KS_Statistic
1.0	Daily	13.800165	6.049857	0.413946	0.457001
1.0	Monthly	178.157622	112.867189	0.565224	0.210976
1.0	Seasonal	431.050968	280.724348	0.588977	0.240803
2.0	Daily	15.541377	7.607774	0.311802	0.417965
2.0	Monthly	207.489331	146.259706	0.170327	0.235566
2.0	Seasonal	521.186971	359.460435	0.139280	0.316129
3.0	Daily	12.959300	6.298090	0.172008	0.486076
3.0	Monthly	175.941342	123.127124	0.106410	0.244860
3.0	Seasonal	438.202825	314.725371	0.099880	0.281250
4.0	Daily	12.979356	6.518519	0.252683	0.413435
4.0	Monthly	162.423991	118.167983	0.177468	0.265723
4.0	Seasonal	401.902260	305.625606	0.164116	0.325991

Really bad metrics!

- (still have to wait for other data from datasets to compare which have better values)
- **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 ↘

- Results don’t seem too positive for MSWEP
  - Will try to also run the script for IMERG to see if there are issues, if there are, I might have to check my code again
- Consider increasing threshold for completeness so that data will be more reliable.