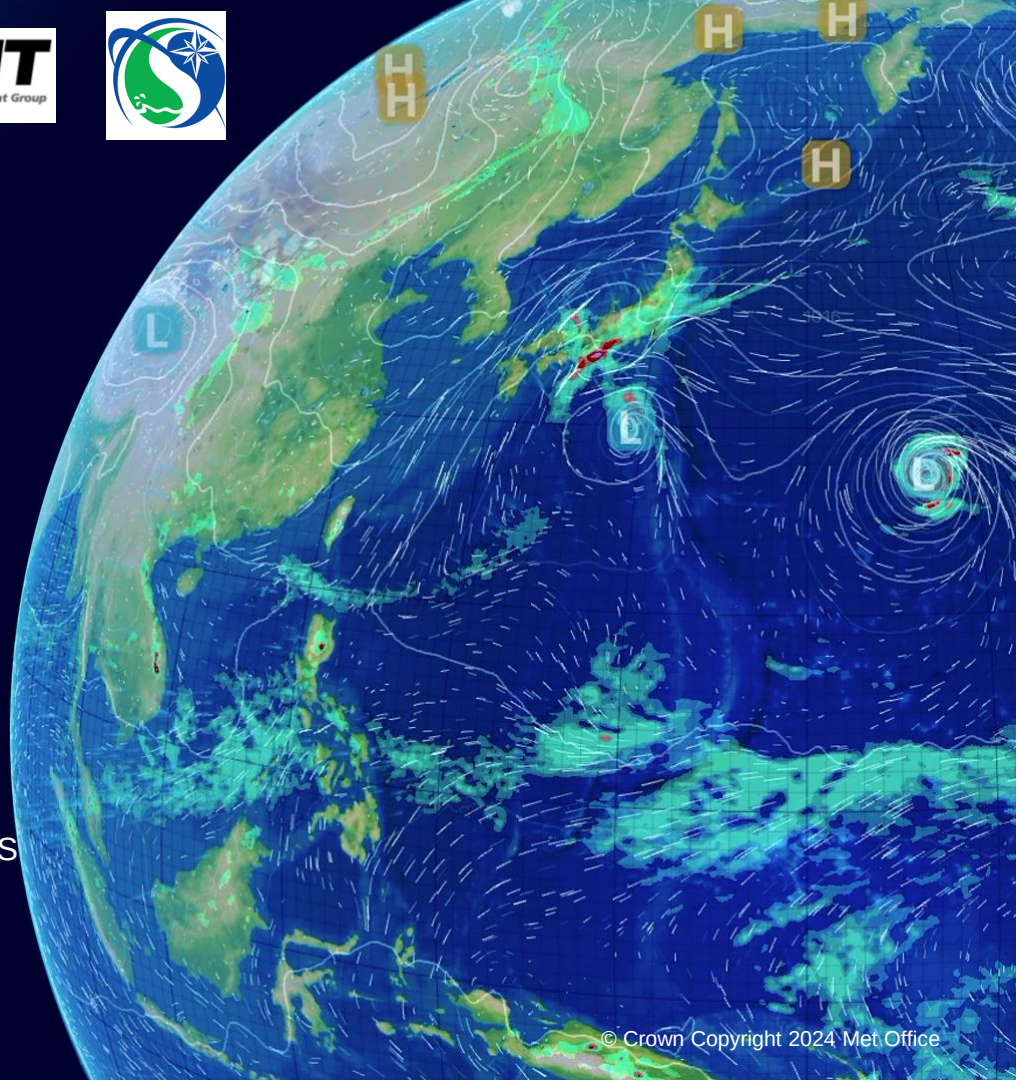
High score ≠ skill: A study over the Maritime Continent using a new derivative of the FSS

Marion Mittermaier

9th international verification methods workshop

Cape Town, May 2024

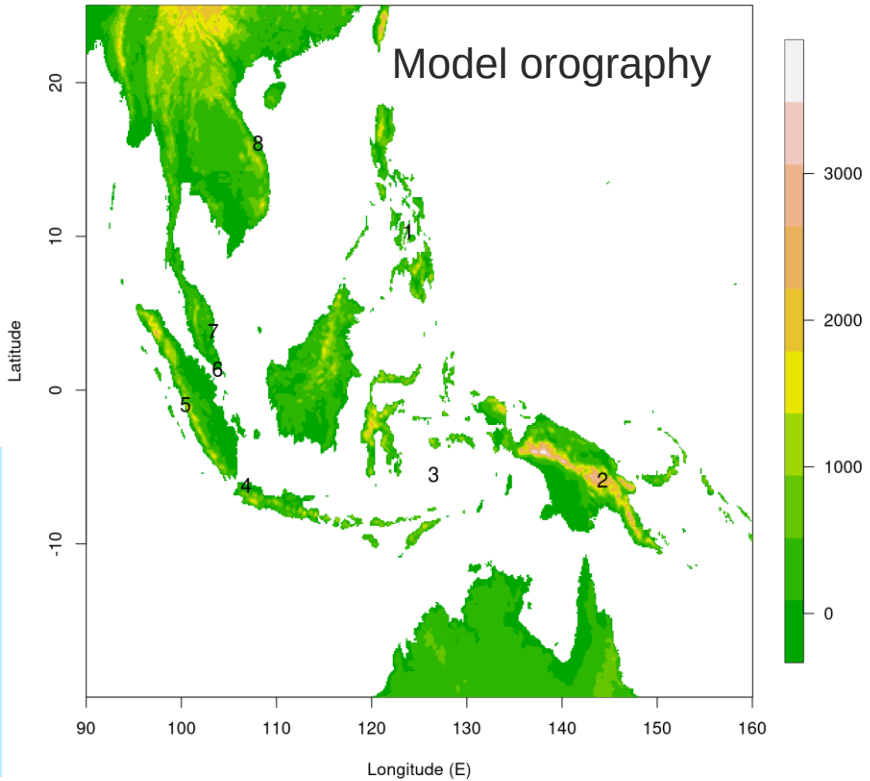
Mittermaier, 2024, accepted subject to minor revisions, QJRMS



The Maritime Continent

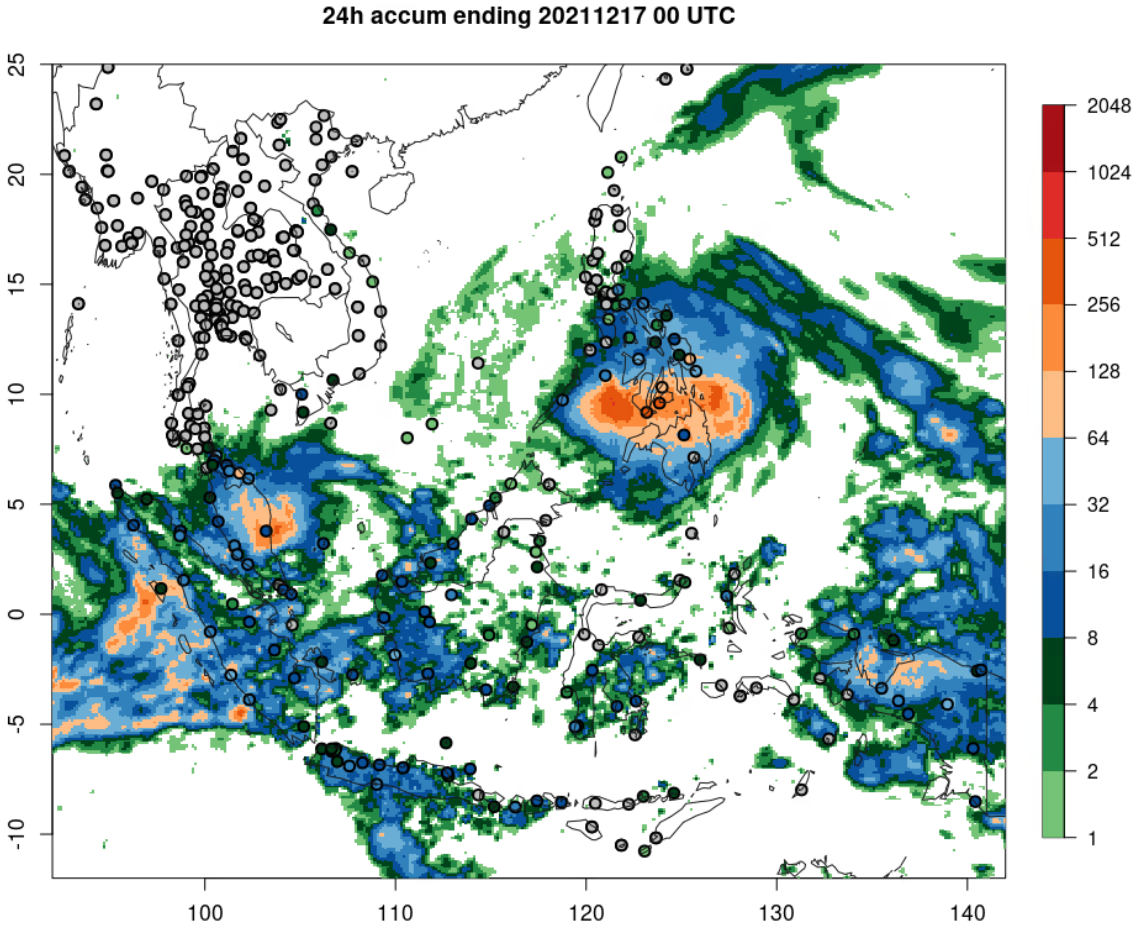
- Using Global UM operational forecasts from December 2021 (GA7, N1280),
- GPM IMERG
- This is a region where it is often difficult to demonstrate forecast skill…. Is there a way to discriminate?

1=Cebu City (Philippines, 159 m)
2=Mount Hagen (Papua New Guinea, 2051 m)
3=Banda Sea (0 m)
4=Central Jakarta (Indonesia, Java, 33 m)
5=Padang (Indonesia, Sumatra, 558 m)
6=Singapore (12 m)
7=Kuantan (Malaysia, 32 m)
8=Da Nang (Vietnam, 309 m).



How good is GPM?

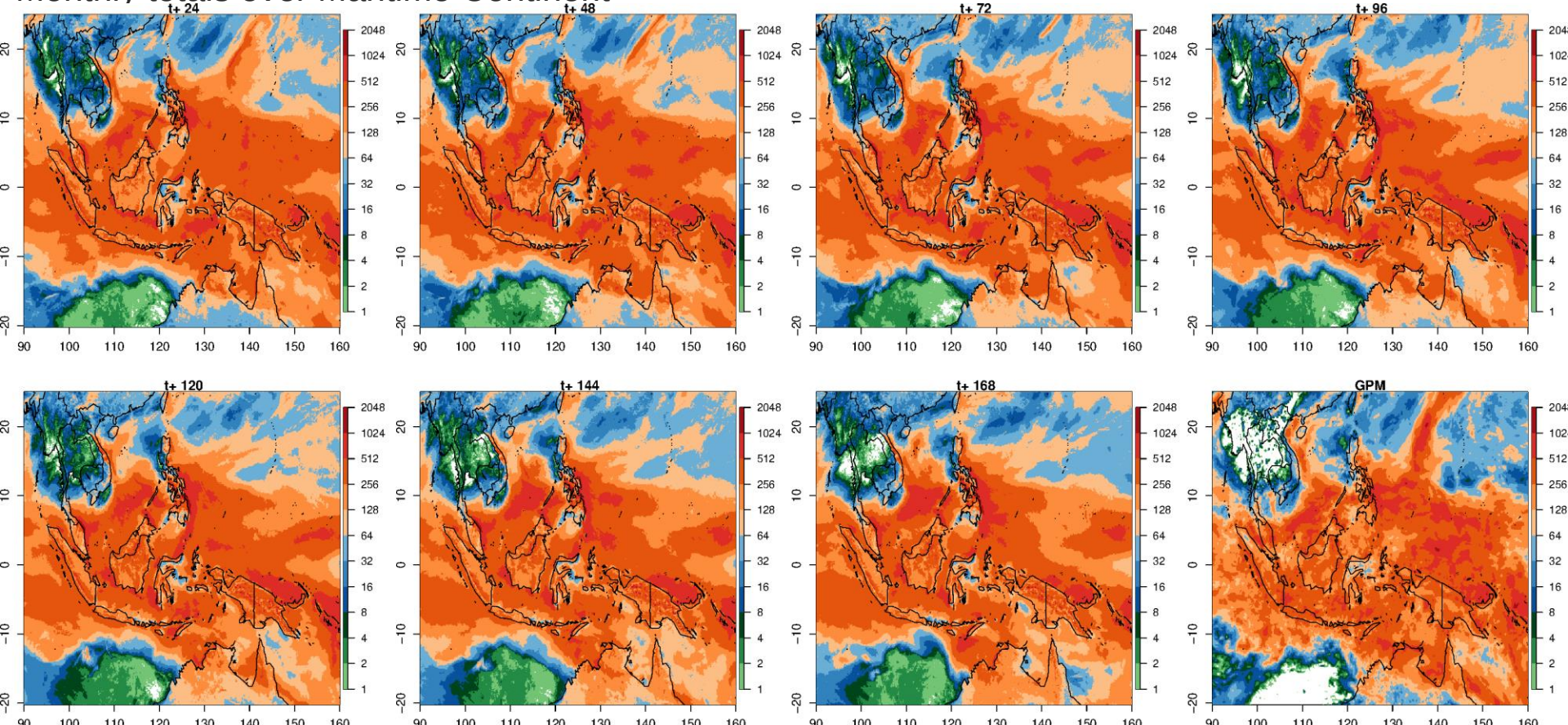
- Not perfect, but acceptable over the region.
- Example of comparison of daily accumulations.
- Some large mismatches are possible.
- When comparing the same forecast to these two observation types, one should expect the results to be different!



Met Office BMKG MiGHT Malaysian Industry-Government Group for High Technology Department for Science, Innovation & Technology

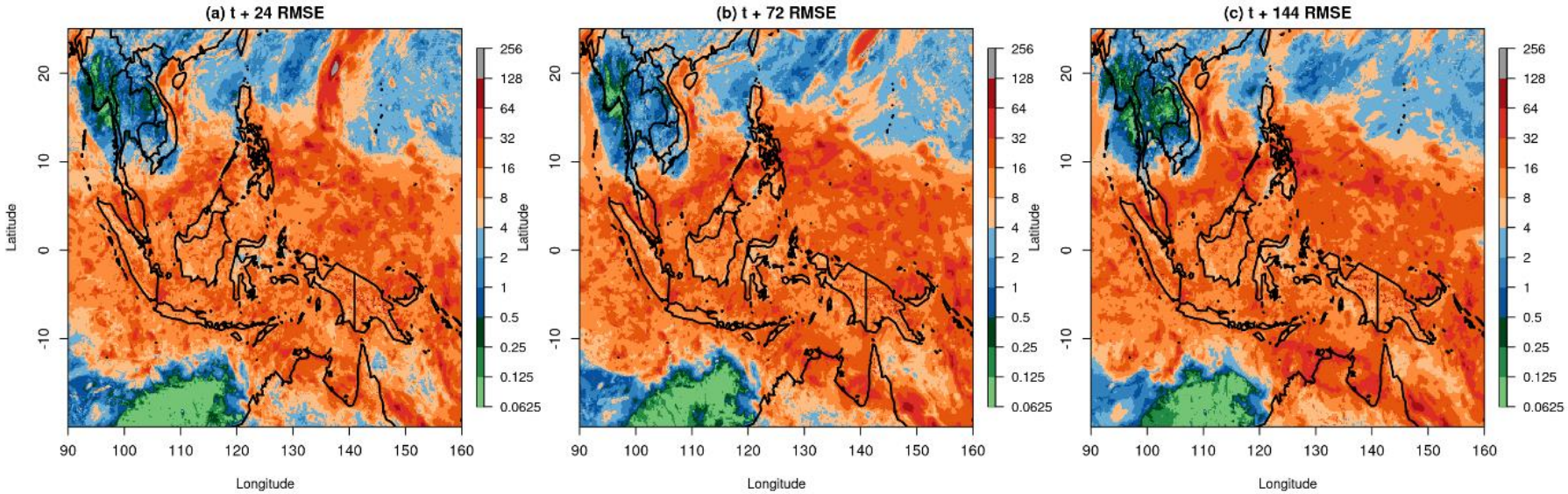
Dec 2021, GA7

Monthly totals over Maritime Continent

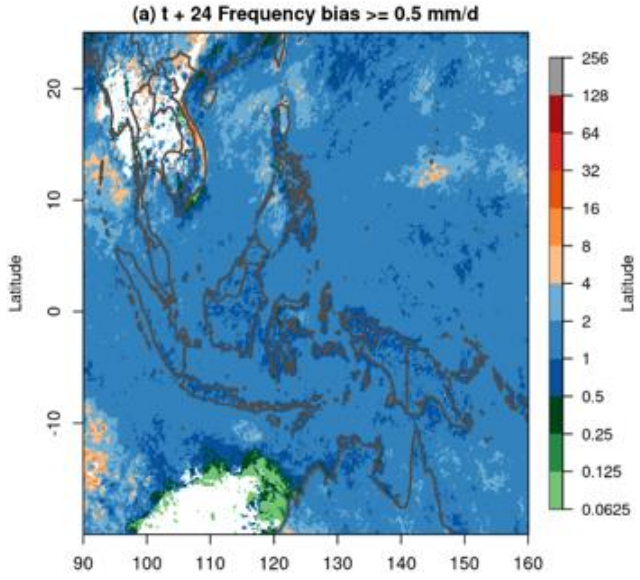


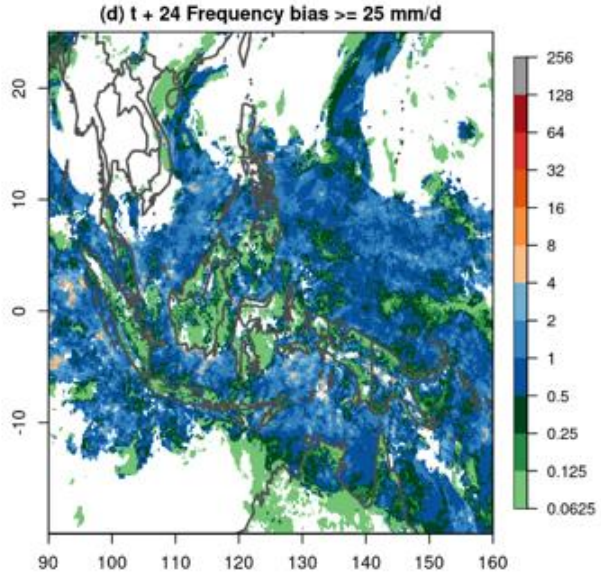
Met Office BMKG MiGHT Malaysian Industry-Government Group for High Technology Department for Science, Innovation & Technology

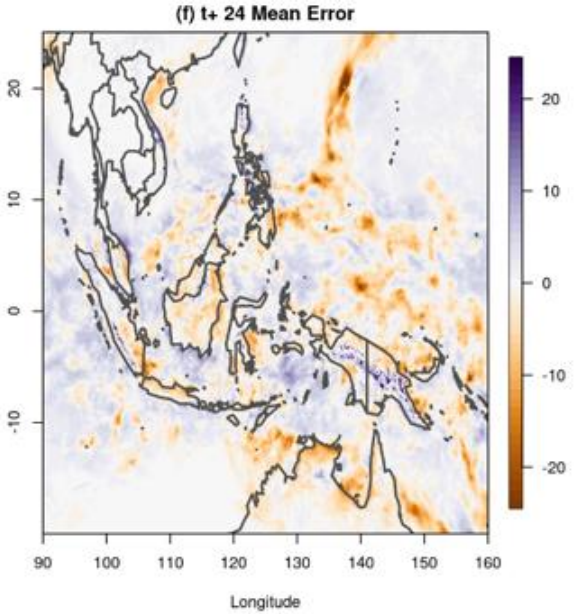
… and the rmse is… pretty uniform!



The biases aren’t terrible….



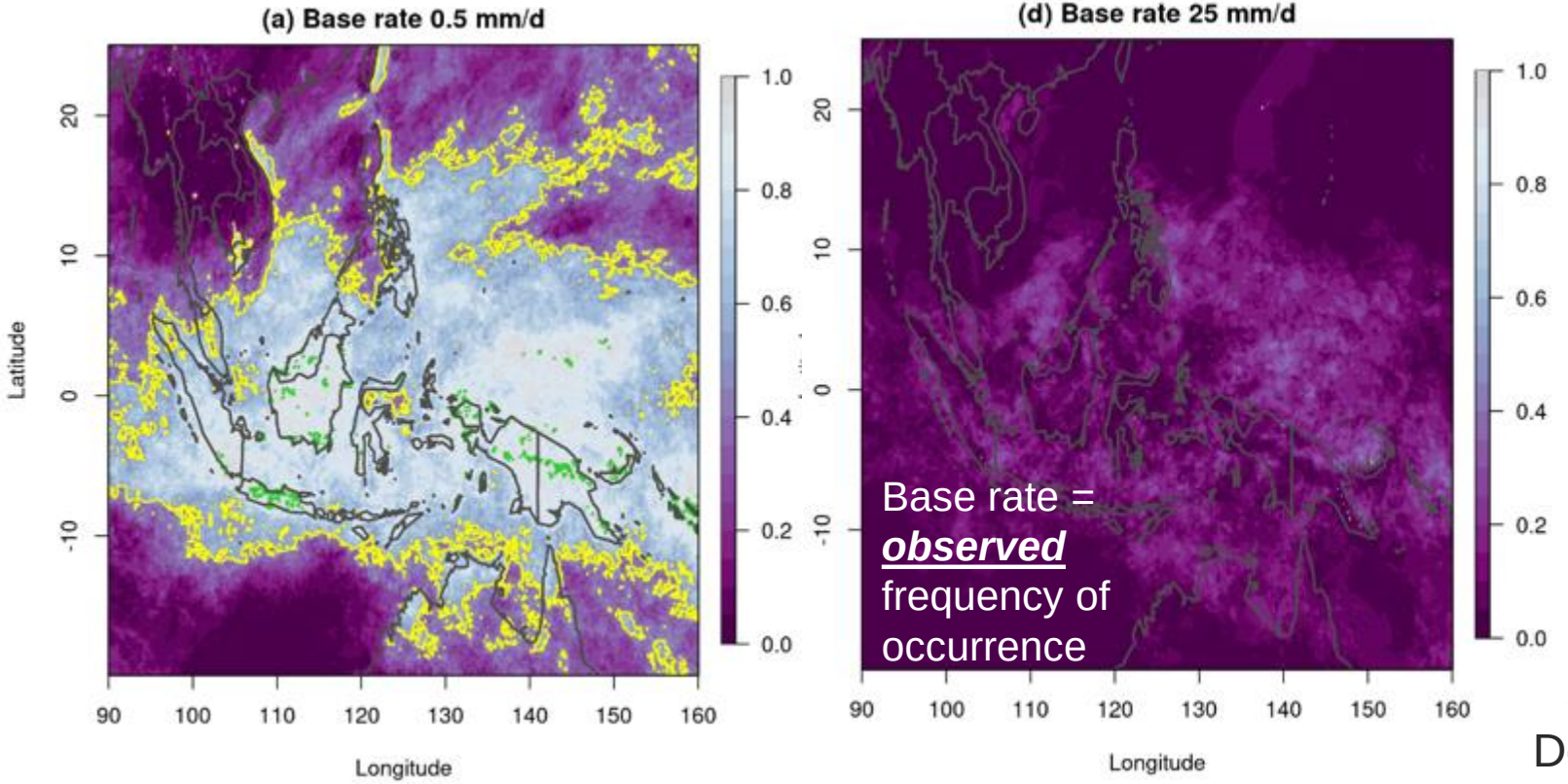




The base rate f_0 is interesting however

Green contour = f_0 = 1

Yellow contour = f_0 > 0.5



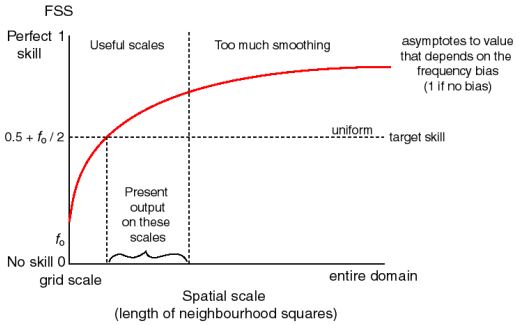
Dec 2021

The *Localised* FSS (LFSS). What is it?

- **FSS on the grid** (see Woodhams et al. 2018) enables one to maintain the geographical location whilst casting the net ever wider to compensate for potential double penalty errors.
- It is “localised” because the **location of the central grid point remains fixed** and the scores for that location (and neighbourhoods centred on this location) are aggregated over time.

Why is this useful and/or important?

- Local effects are accounted for
- Local skill is assessed taking the impact of any double penalty into account
- **Systematic patterns** in spatial skill can be identified
- Grid points with **vastly different climatologies are not aggregated in space** (this is especially relevant when FSS -or any score- is computed over large domains)
- **Vastly different model performance across large domains is not “hidden” or cancelled out, especially if there are systematic regional signals**



“Factorisations” of the FSS…

- **Skilful Spatial Scale (SSS)** defined as L(FSS > 0.5 + f_0/2) has been around since Roberts and Lean (2008). Here the FSS is framed in terms of a displacement and the useful or skilful spatial scales.
- ***NEW* Lead Time Potential (LTP)** defined as maximum lead time where FSS > 0.5 + f_0/2. Here the FSS is framed in terms of how long a forecast is skilful for at a given location and in terms of measuring improvement whether forecast performance is staying higher for longer (in a new model version), i.e. whether hours or days are gained in terms of skill.

- Both of these can also be computed in a localised sense.

“Factorisations” of the FSS…

- **Skilful**
since R
displac
- ***NEW L***
0.5 + f_0
a given
perform
hours c

We often ignore the $f_0/2$ bit, but it is REALLY influential in determining whether you have skill or not.

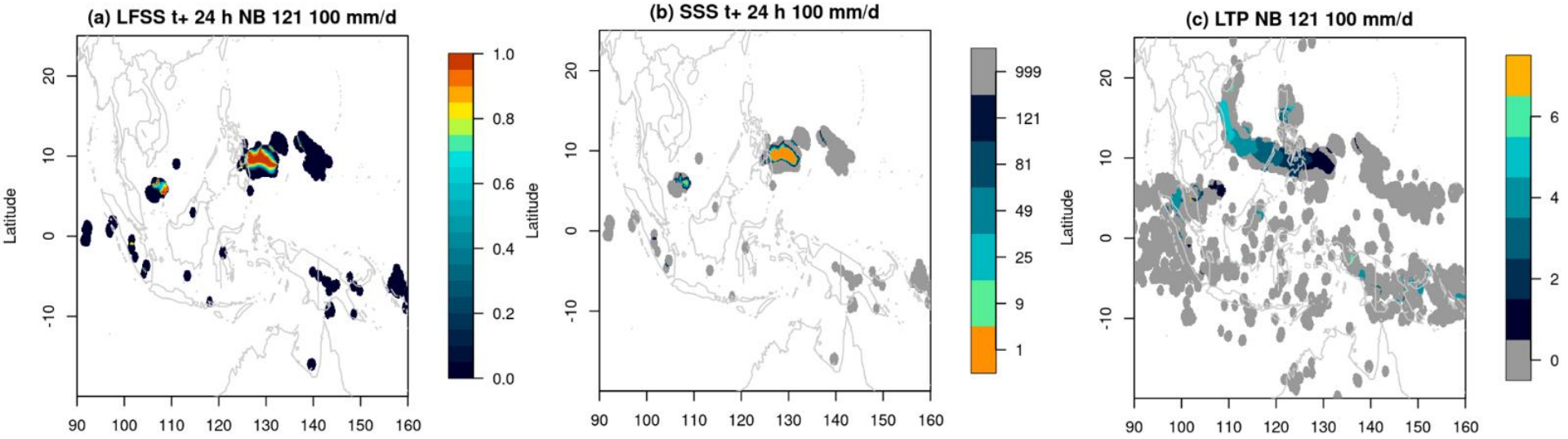
Let’s demonstrate!

- Both of these can also be computed in a localised sense.

Mittermaier (QJ, 2024, accepted subject to minor revisions)

Performance for a single forecast

Landfall of Rai



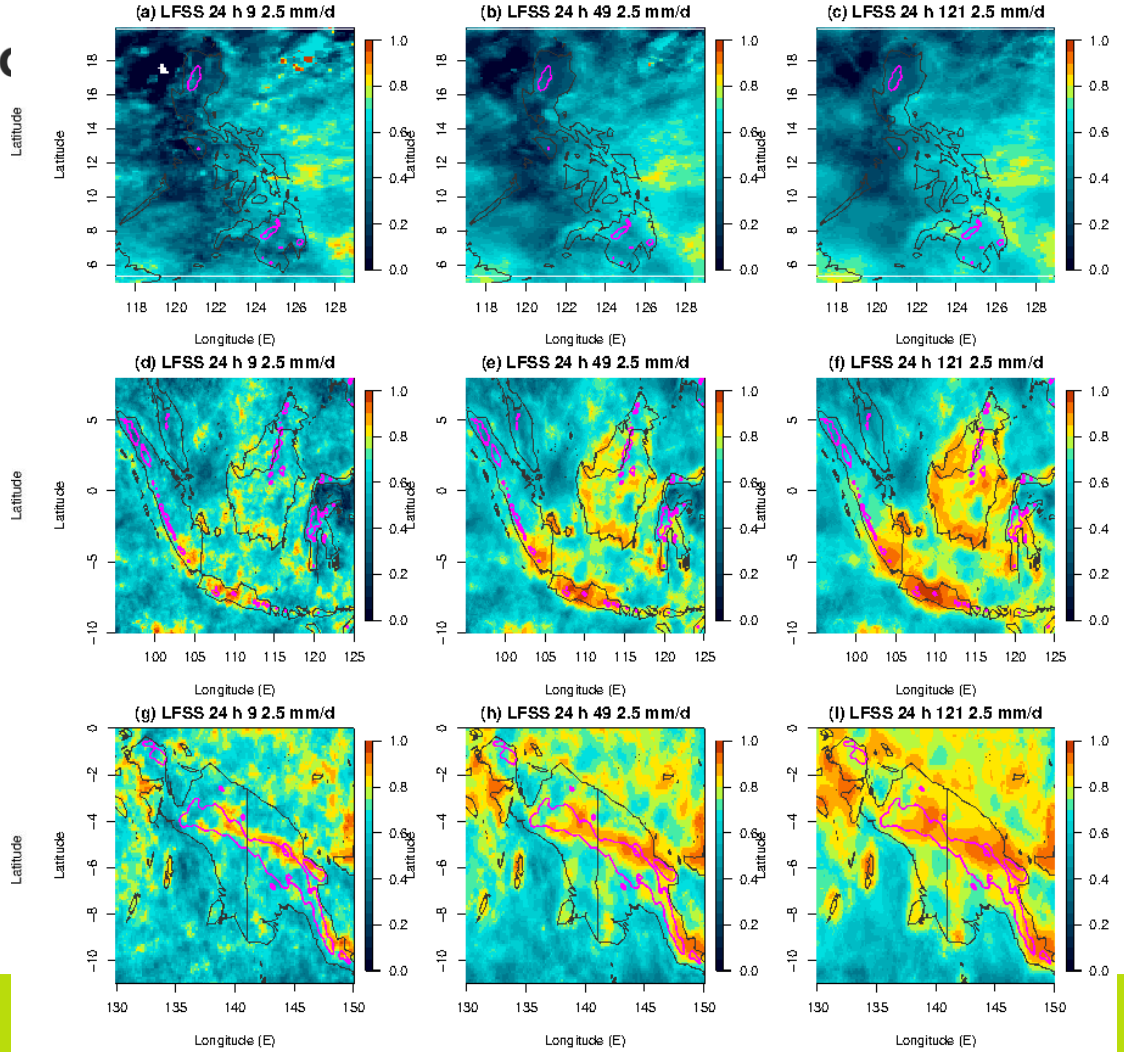
- Scores can be really high. Can also compute the LFSS for a GPM-based **persistence forecast to see what real value NWP forecasts add in this region.**

Mittermaier (QJ, 2024, accepted subject to minor revisions)

LFSS maps for Dec 2021

Maps as a function of threshold & n'hood size.

Provides rich detail to unpack the performance over high ground, coastal zones, just off-shore etc.



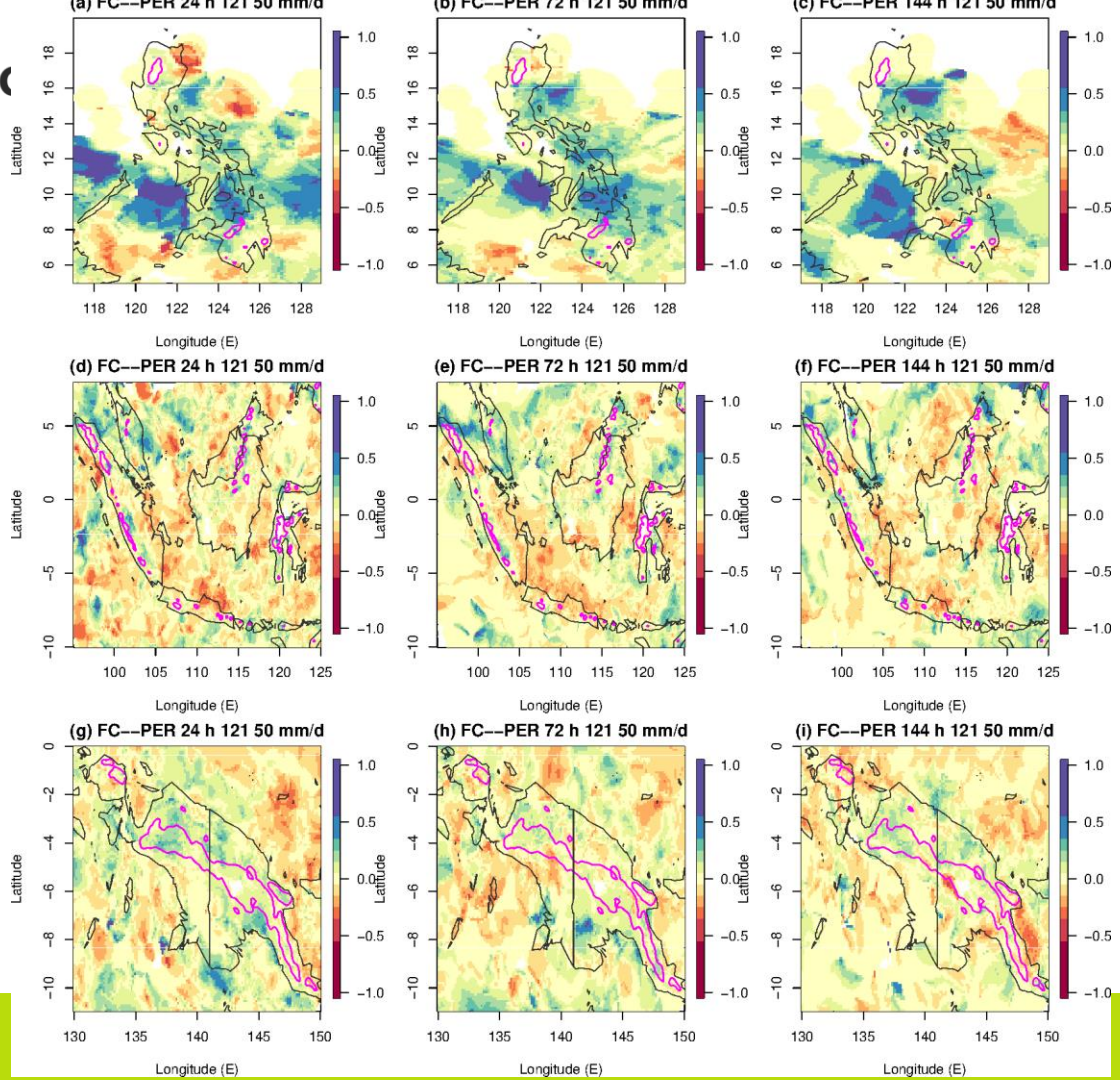
Comparison to persistence shows that NWP model adds *some* skill over persistence, with larger positives for higher thresholds and for tropical cyclones.

Mittermaier (QJ, 2024)

LFSS maps for Dec 2021

Maps as a function of threshold & n'hood size.

Provides rich detail to unpack the performance over high ground, coastal zones, just off-shore etc.

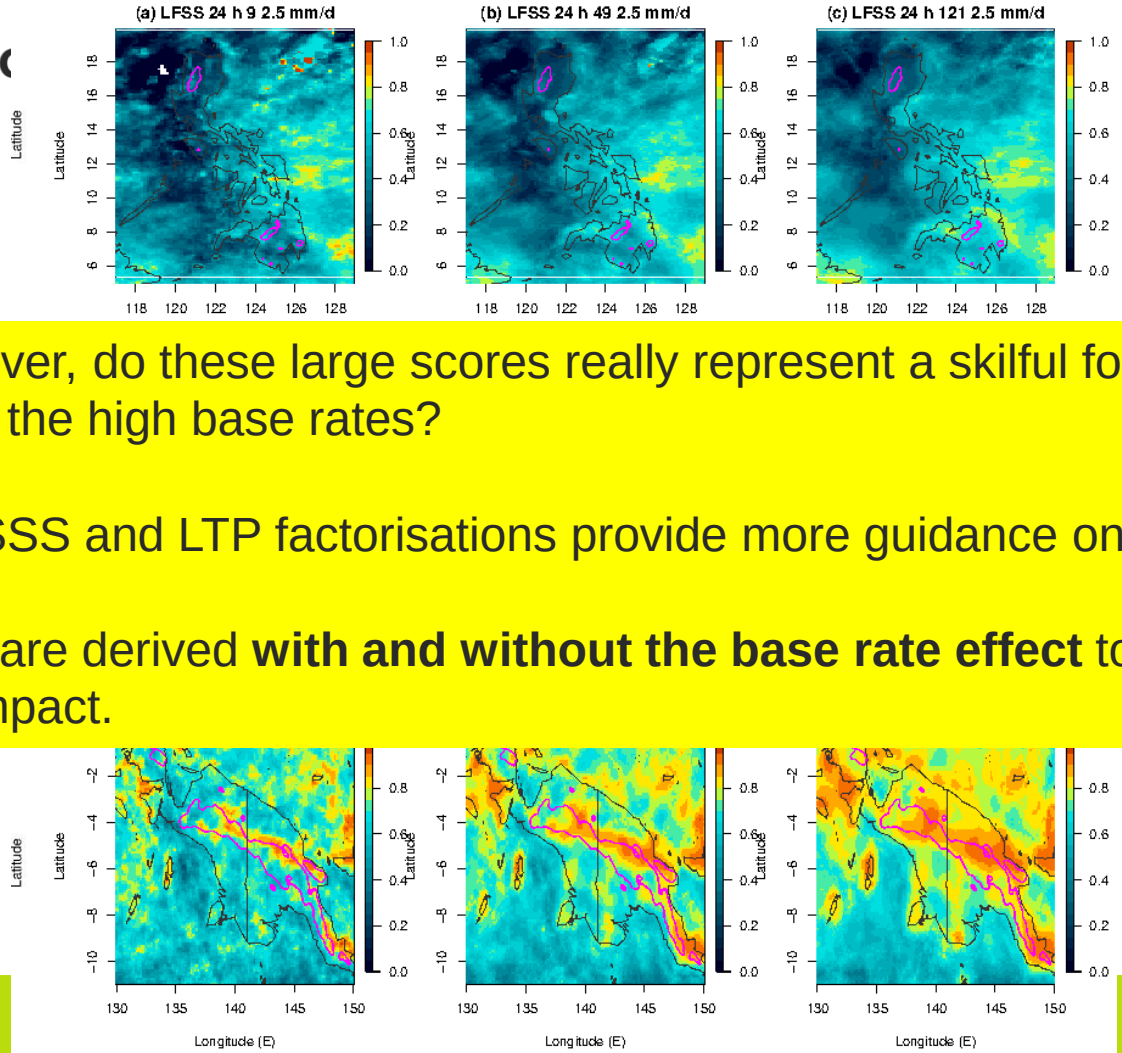


Department for Science, Innovation & Technology

Comparison to persistence shows that NWP model adds *some* skill over persistence, with larger positives for higher thresholds and for tropical cyclones.

Mittermaier (QJ, 2024)

LFSS maps for Dec 2021



Maps as function o threshold n'hood si

Provides detail to unpack th performa over high ground, coastal zones, just off-shore etc.

However, do these large scores really represent a skilful forecast, given the high base rates?

The SSS and LTP factorisations provide more guidance on skill.

They are derived **with and without the base rate effect** to show the impact.

Comparison to persistence shows that P model s *some* skill

istence, larger tives for er sholds and opical cyclones.

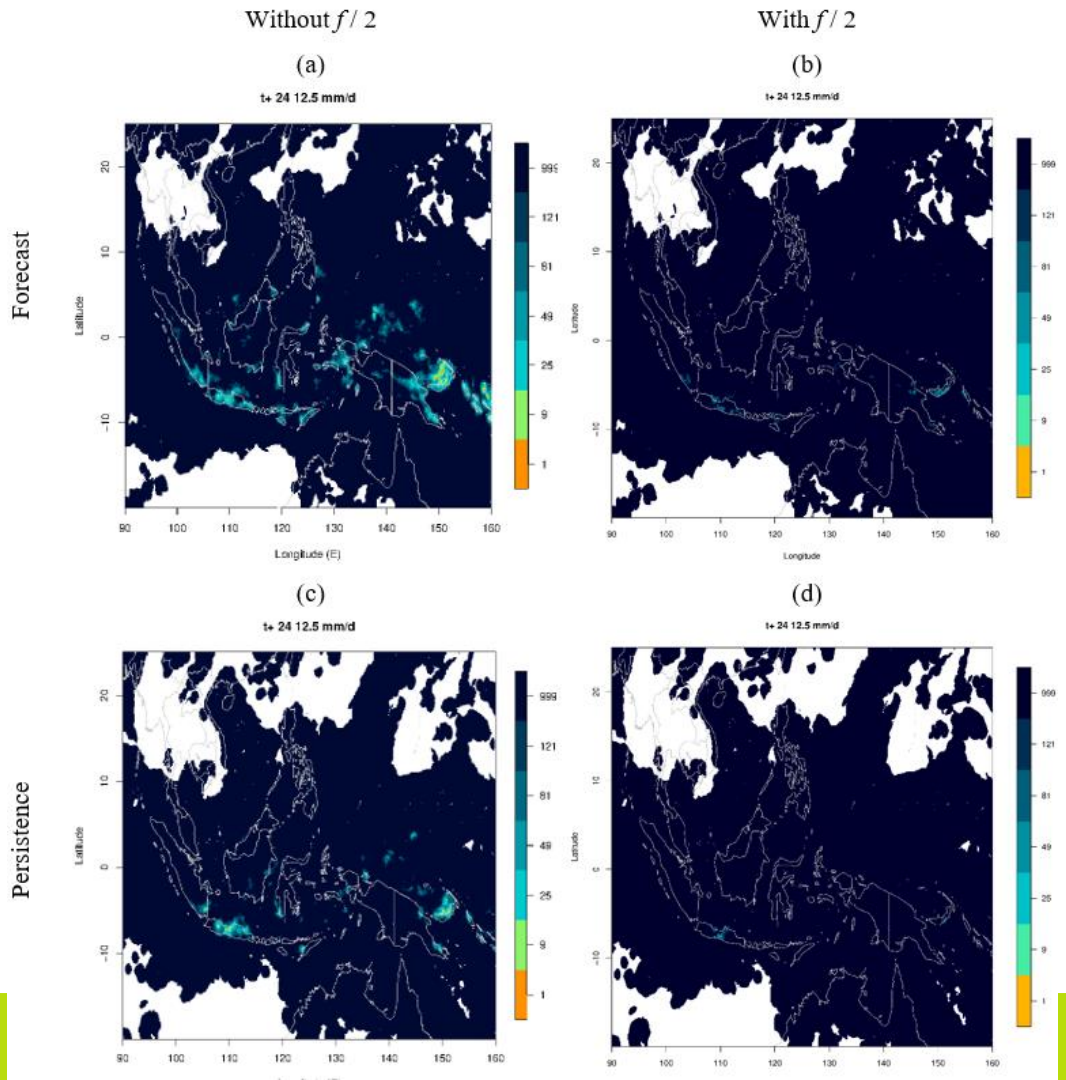
Mittermaier (QJ, 2024)

Met Office

Skilful Spatial Scale

Smaller SSS is better.

Including the base rate effect reduces the benefit of the forecast over persistence.



Department for Science, Innovation & Technology

The picture looks quite promising when *f* is disregarded (left column).

Once *f* is included (right column) many of the locations with low SSS disappear.

This holds for both the forecast (top) and persistence (bottom).

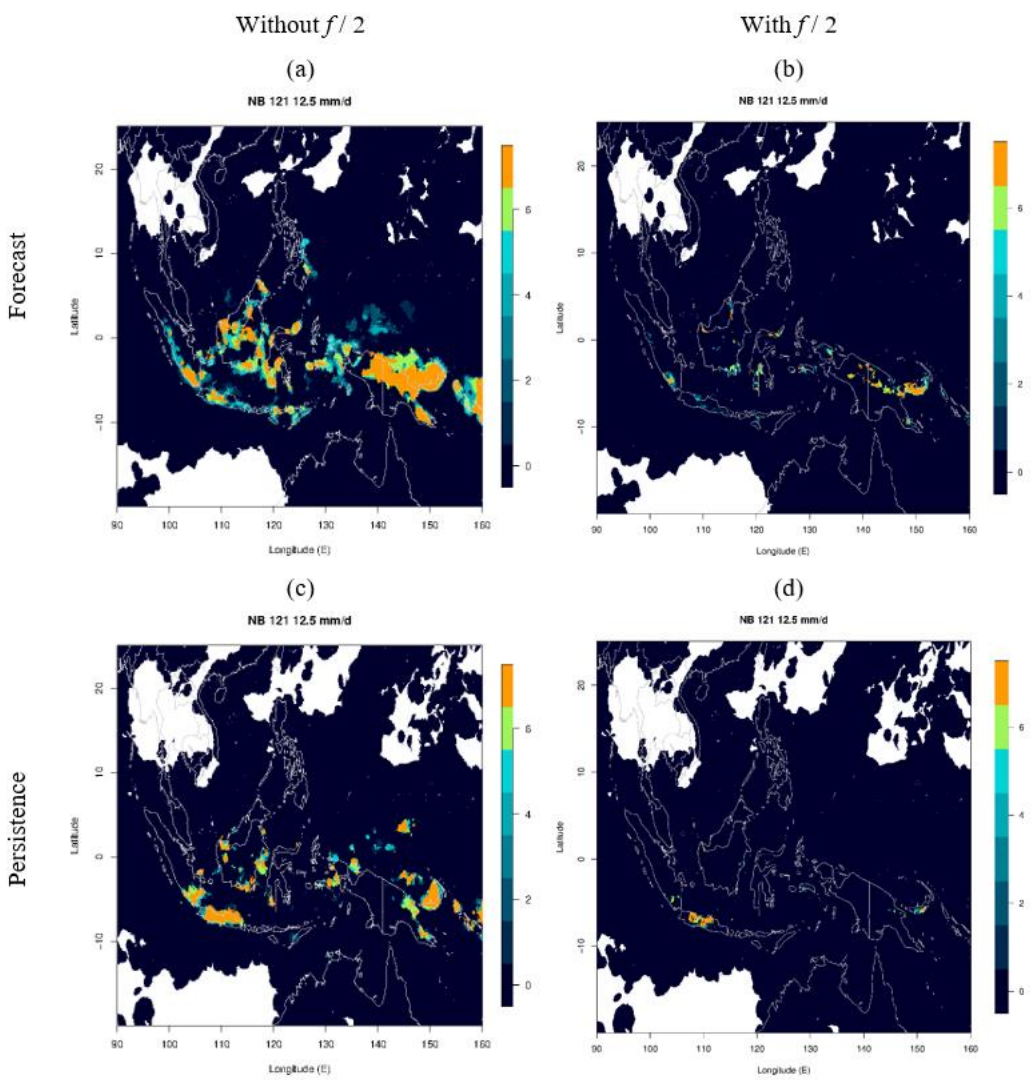
For Dec 2021

Mittermaier (QJ, 2024)

LTP

Bigger LTP is better (in days)

Including the base rate effect reduces the benefit of the forecast over persistence.



As for SSS, LTP is much reduced when *f* is included (right).

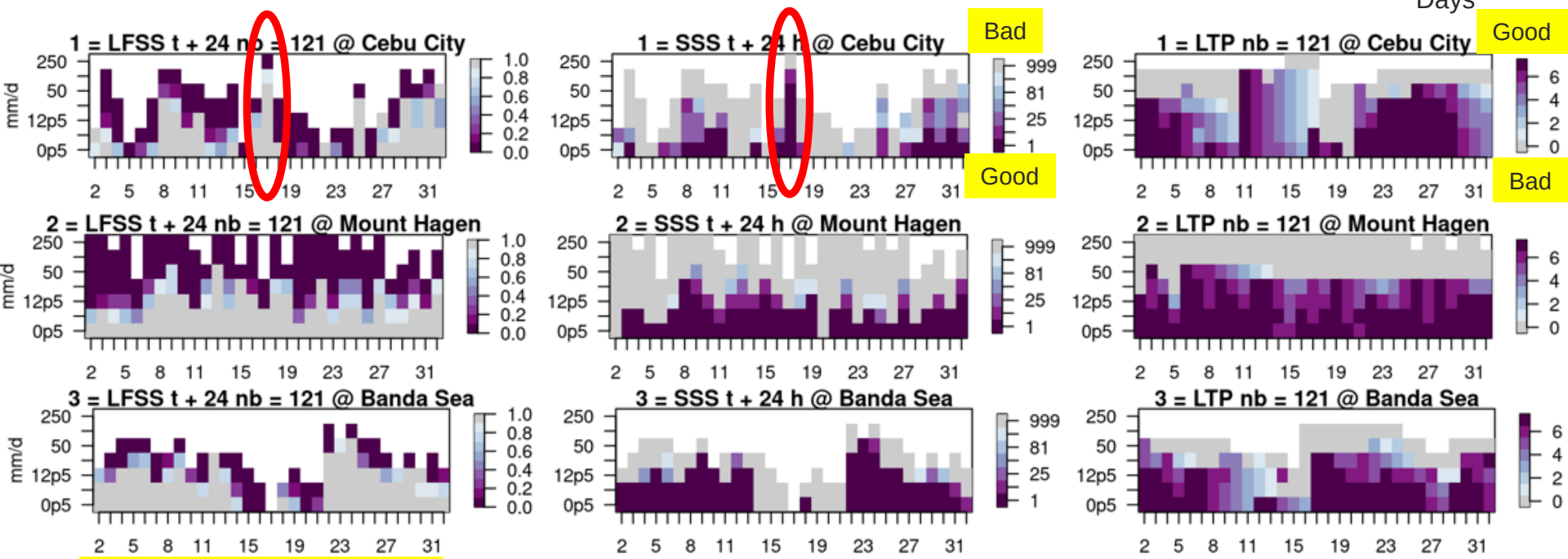
Forecasts (top) pick out some different locations to persistence (bottom).

Forecasts provide more skilful guidance further ahead, so that the persistence forecasts are impacted more at longer lead times, except for the most persistent rainy spots.

For Dec 2021

Mittermaier (QJ, 2024)

Performance at a location (including f_0)



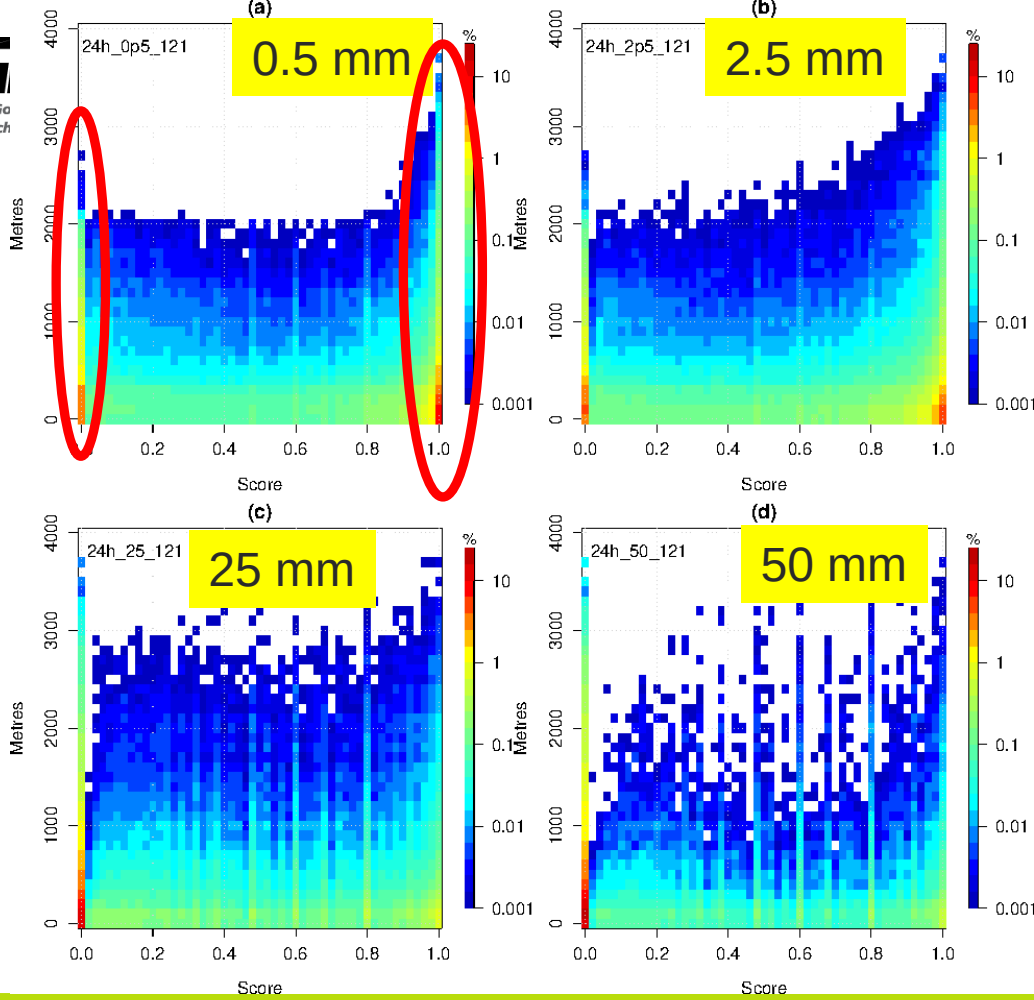
Score maxes out frequently for modest thresholds

Both sea points and Mount Hagen persistently low SSS and high LTP

Mittermaier (QJ, 2024)

Stratification by height

- LFSS values stratified using GM orography for different thresholds and the largest neighbourhood size (121)
- High ground affects predictability and skill of daily precipitation (as seen for Mount Hagen).
- Provides valuable input on the variability in skill over land and how to aggregate results to ensure systematic signals are preserved.

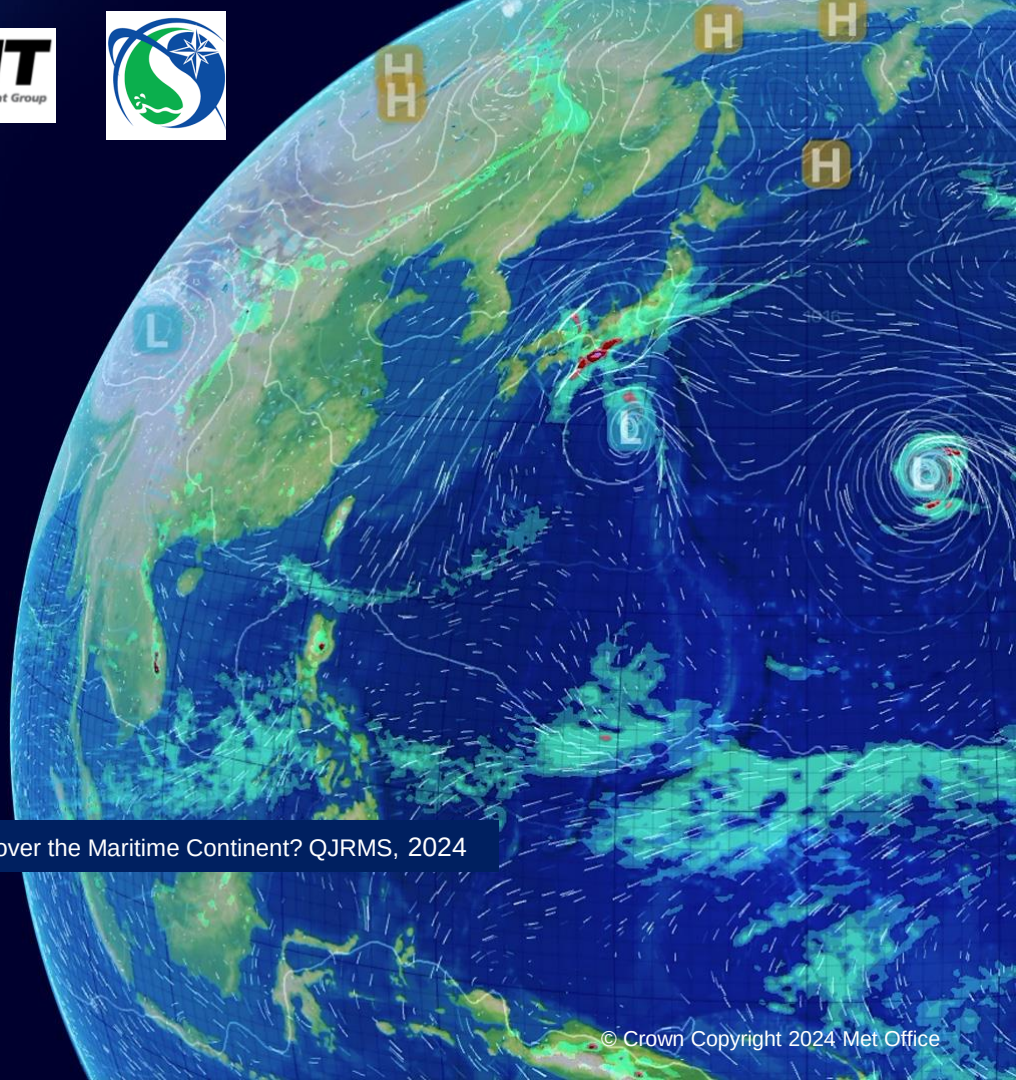


Mittermaier (QJ, 2024)

Summary

- Though the FSS is a “skill” score, it does not measure skill relative to a credible reference (Mittermaier 2021). When the FSS is computed *for* a credible reference forecast such as persistence, the true skill becomes clearer.
- Even then, **high FSS values**, as seen over the MC (for both forecast and persistence), **do not necessarily indicate *useful skill* without having an awareness of the observed frequency of occurrence or base rate**. In cases where $f \rightarrow 1$, FSS → 1 may not provide “useful” skill. In cases where $f \sim 0.8$, FSS > 0.9 must be achieved.
- Locally (or at the domain scale), ***f* could *only* be considered negligible for very high centiles**, with the caveat that using quantile mapping means some of the day-to-day thresholds could be very negligible too.
- Changing the perspective (e.g. to the space or time dimension) illustrates this very well, i.e. **when the base rate is included** and when it is not. **Much of the perceived useful skill disappears even though the scores are very high**.
- **These facets are hidden to some degree when computing the classic FSS** where the fractions are aggregated over the domain, but that does not mean they are not there!
- Considering forecast performance in space provides useful detail which is lost when aggregating over an area.

Thanks for listening!
Questions?



Mittermaier, M.P., 2024: Is there any skill in daily global precipitation forecasts over the Maritime Continent? QJRMS, 2024