Evaluating Convection-Permitting Ensemble Forecasts of Precipitation over Southeast Asia

SAMANTHA FERRETT,[a] THOMAS H. A. FRAME,[a] JOHN METHVEN,[a] CHRISTOPHER E. HOLLOWAY,[a] STUART WEBSTER,[b] THORWALD H. M. STEIN,[a] AND CARLO CAFARO[a]

[a] *Department of Meteorology, University of Reading, Reading, United Kingdom*
[b] *Met Office, Exeter, United Kingdom*

Outline

1. Introduction
2. Data
3. Methods
4. Results
5. Summary and conclusions

1. Introduction

a. *The characteristics of deep convection in Southeast Asia*

- It can cause ***extreme rainfall intensity*** and result in devastating impacts (flooding, landslide).
- It plays an active part in the ***dynamics*** of the larger-scale atmospheric phenomena in this region.
- It can be influenced by the systems across a range of time and space scales. (El Nino-southern oscillation, Madden-Julian oscillation, cold surges, equatorial waves, tropical cyclones, and land-sea breeze circulations)

1. Introduction

b. The issues of NWP models

- Global models which rely on ***convection parameterizations*** cannot forecast tropical rainfall features accurately (e.g. diurnal cycle, MJO, and equatorial waves).
- Increasing model ***resolution*** can simulate deep convection explicitly.
- The current operational "***convection permitting (CP)***" model (resolution from 1-10 km) can only partially resolve deep convection.
- CP models perform better than convection parameterizations in convection initiation, ***diurnal cycle***, and large-scale modes.
- There is a ***trade-off*** between resolution, model domain size, and ensemble size within real-time operational NWP routine.

1. Introduction

c. The issues of ensemble forecasts

- Forecast ***uncertainties*** are associated with model physics, initial conditions and boundary conditions.
- ***Ensemble forecasts*** can account for model uncertainty.
- Several studies have shown the benefits of ensemble forecasts to high impact weather in the ***extratropics***. However, there are fewer studies examining the extreme rainfall in the ***tropics***, especially in Southeast Asia.
- Porson et al. (2019) and Sun et al. (2020) evaluate the CP ensemble forecasts around Singapore but in a relatively small area (400 km × 400 km).

1. Introduction

d. The aim of this study

- ***Quantify the usefulness*** of CP ensemble forecasts in Southeast Asia.
- Three domains are examined, including peninsular Malaysia, Java and the Philippines.
- The issues include the ***precipitation***, the ***scale dependence*** of forecast skill, and the role of ***diurnal cycle*** in forecast skill.
- Besides, the ***spread of the skill*** between the ensemble members are also verified.

2. Data

a. Ensemble forecasts – Met Office Unified Model (MetUM)

- ***18 ensemble members*** perturbed from:

  ensemble transform Kalman filter (ETKF) for IC

  stochastic parameterization scheme

- Initialized twice daily (***00Z***, 12Z) and forecast 120 hours

  from Oct. 2018 to Mar. 2019

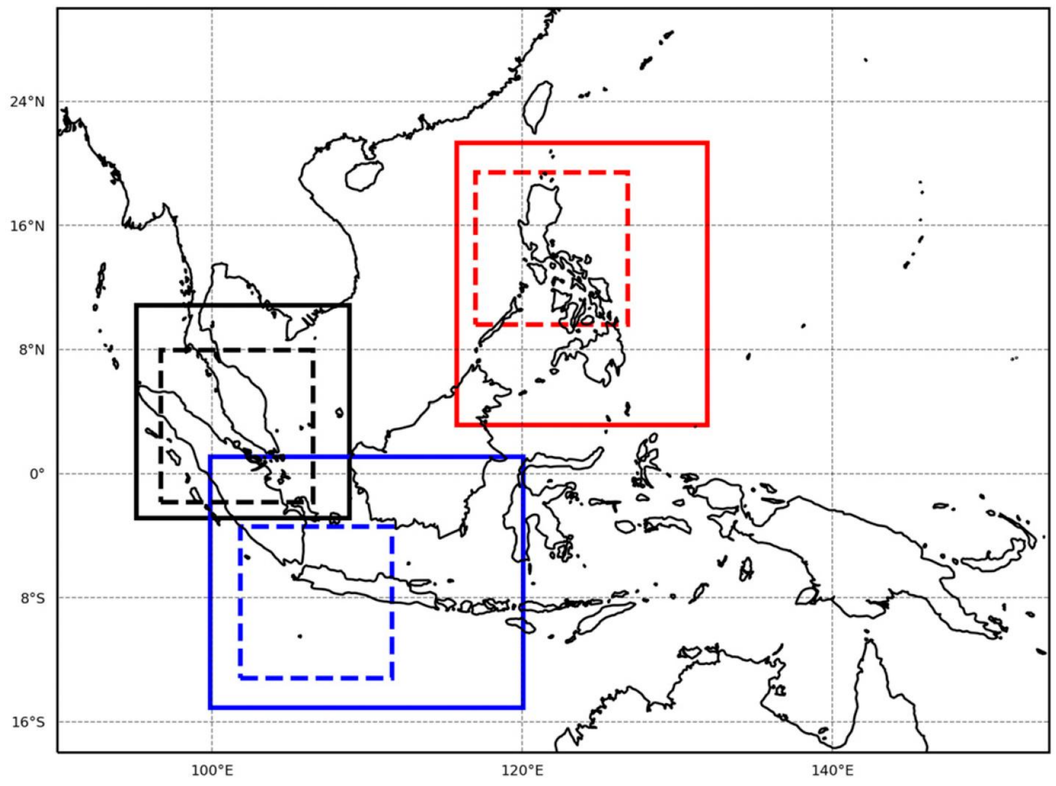
- ***3 resolutions***: 2.2 km, 4.5 km, and 8.8 km

- The 8.8 km forecasts include ***parameterized run (GA)***

- Regrid to a common 9-km grid before analysis

- ***3 domains:***

  Malaysia*, *Indonesia*, and *the Philippines



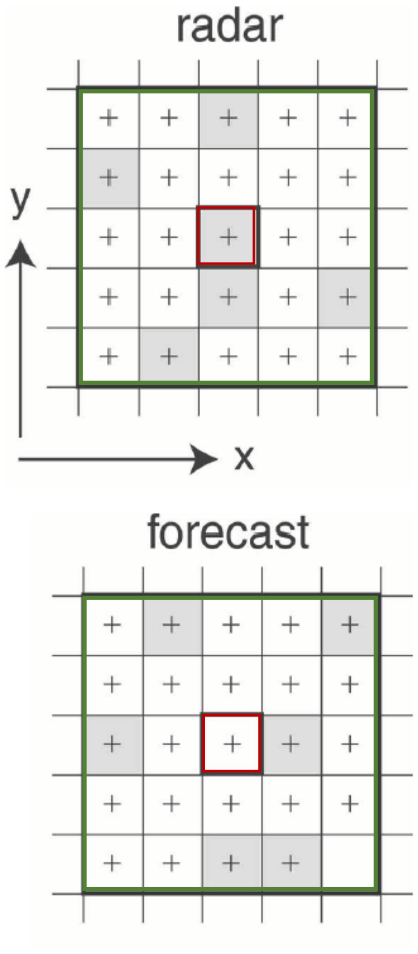
2. Data

b. Observation – GPM-IMERG

- GPM-IMERG: Integrated Multisatellite Retrievals for Global Precipitation Mission
- The product used is at the resolution of ***0.1 °*** in space and ***half hour*** in time which combines precipitation estimates from GPM satellites and Global Precipitation Climatology Centre (GPCC) rain gauges.
- The dataset are abbreviated as "*GPM*" in this study.
- The GPM precipitation is also interpolated to a common 9-km grid.
- N. De Silva et al. (2021) finds that GPM estimated precipitation is similar to local rain gauge for ***percentiles between 85^{th} and 95^{th}*** .

3. Methods

a. Fractions skill score (FSS, Roberts and Lean, 2008)



$(N = 1, N = 5)$

- ***Concept:*** Compare two gridded fields and measure the degree of correspondence as a function of spatial scale.

- For each grid point:

  1. Select a rainfall ***threshold*** "q"

  2. Select a ***spatial scale*** "N"

  3. Calculate the portion exceeding threshold "q" within $N \times N$ ***neighborhood***, *called **fraction.***

  4. After the calculation of each grid point, the ***fraction fields*** $O(n)$ and $M(n)$ are calculated.

(Roberts, N. and H. Lean, 2008)

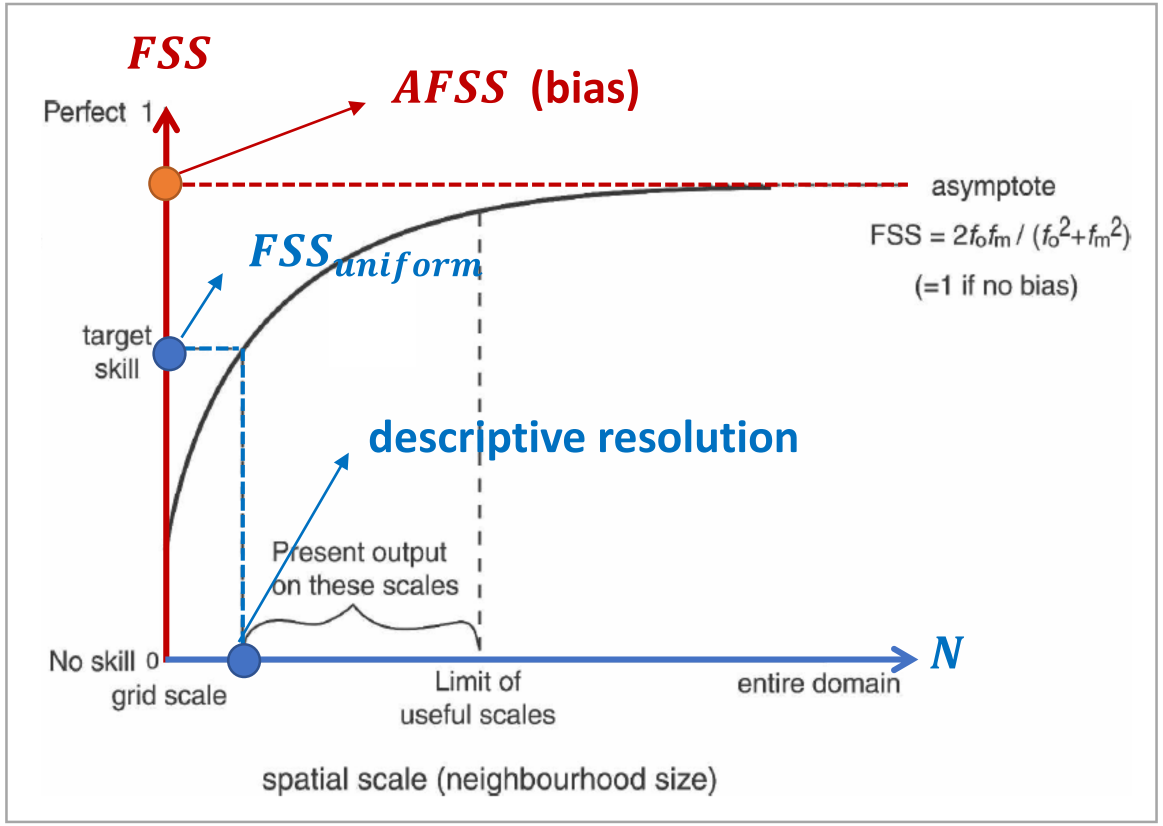
3. Methods

a. Fractions skill score (FSS, Roberts and Lean, 2008)

$$MSE_{(N)} = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} [O - M]^2$$

$$MSE_{(N)ref} = \frac{1}{N_x N_y} \left[\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} O^2 + \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} M^2 \right]$$

$$FSS_{(N)} = 1 - \frac{MSE_{(N)}}{MSE_{(N)ref}}$$



(Roberts, N. and H. Lean, 2008)

3. Methods

a. Fractions skill score (FSS, Roberts and Lean, 2008)

➢ ***Three types of FSS:***

1. eFSS: ensemble-aggregated FSS (Dey et al. 2014)

The ***purpose*** is to summarize the performance of ensemble members ***as a whole***.

$$\underbrace{MSE_{(N)} = \frac{1}{N_x N_y}\sum_{i=1}^{N_x}\sum_{j=1}^{N_y}[\boldsymbol{O}-\boldsymbol{M}]^2 \qquad MSE_{(N)ref} = \frac{1}{N_x N_y}\left[\sum_{i=1}^{N_x}\sum_{j=1}^{N_y}\boldsymbol{O}^2+\sum_{i=1}^{N_x}\sum_{j=1}^{N_y}\boldsymbol{M}^2\right]}_{\text{Average across all forecasts and ensemble members}}$$

$$\boldsymbol{eFSS_{(N)} = 1 - \frac{MSE_{(N)}}{MSE_{(N)ref}}}$$

3. Methods

a. Fractions skill score (FSS, Roberts and Lean, 2008)

➢ ***Three types of FSS:***

2. dFSS: dispersion FSS (Rezacova et al. 2009; Dey et al. 2014)

The ***purpose*** is to evaluate the ensemble ***spread*** of performance.

Replace the observation by a ***control member*** of the ensemble. (Replace the "***O***" in the formula.)

$$MSE_{(N)} = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} [\boldsymbol{O} - \boldsymbol{M}]^2$$

$$\boldsymbol{dFSS_{(N)} = 1 - \frac{MSE_{(N)}}{MSE_{(N)ref}}}$$

$$MSE_{(N)ref} = \frac{1}{N_x N_y} \left[\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \boldsymbol{O}^2 + \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \boldsymbol{M}^2 \right]$$

Conditions	Meaning
$dFSS > eFSS$	Underspread
$dFSS < eFSS$	Overspread

3. Methods

a. Fractions skill score (FSS, Roberts and Lean, 2008)

➢ ***Three types of FSS:***

3. LFSS: localized FSS (Woodhams et al. 2018)

The ***purpose*** is to evaluate the ***spatial distribution*** of FSS.

$$MSE_{(N)} = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} [\boldsymbol{O} - \boldsymbol{M}]^2 \qquad MSE_{(N)ref} = \frac{1}{N_x N_y} \left[\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \boldsymbol{O}^2 + \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \boldsymbol{M}^2 \right]$$

$$MSE_{(N)} = \frac{1}{\boldsymbol{N_t}} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} [\boldsymbol{O} - \boldsymbol{M}]^2 \qquad MSE_{(N)ref} = \frac{1}{\boldsymbol{N_t}} \left[\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \boldsymbol{O}^2 + \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \boldsymbol{M}^2 \right]$$

$$\boldsymbol{LFSS_{(N)} = 1 - \frac{MSE_{(N)}}{MSE_{(N)ref}}}$$

3. Methods

b. *Persistence forecast and shifted forecast*

➢ ***Three variations of forecast*** over Oct. 2018 ~ Mar. 2019

1. Standard CP ensemble forecasts:

Observation	Day 0	Day 1	Day 2	Day 3	Day 4	Day 5
Forecast		Day 1	Day 2	Day 3	Day 4	Day 5

2. Shifted ensemble forecasts:

- The forecast is verified against observations that occur ***a day later*** than the actual forecast verification times.

- To test how much potential predictability comes from the similarity of the observed ***diurnal cycle*** from one day to the next.

Observation	Day 0	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6
Forecast			Day 1	Day 2	Day 3	Day 4	Day 5

3. Methods

b. Persistence forecast and shifted forecast

- ***Three variations of forecast*** over Oct. 2018 ~ Mar. 2019

  3. **Persistence forecasts:**

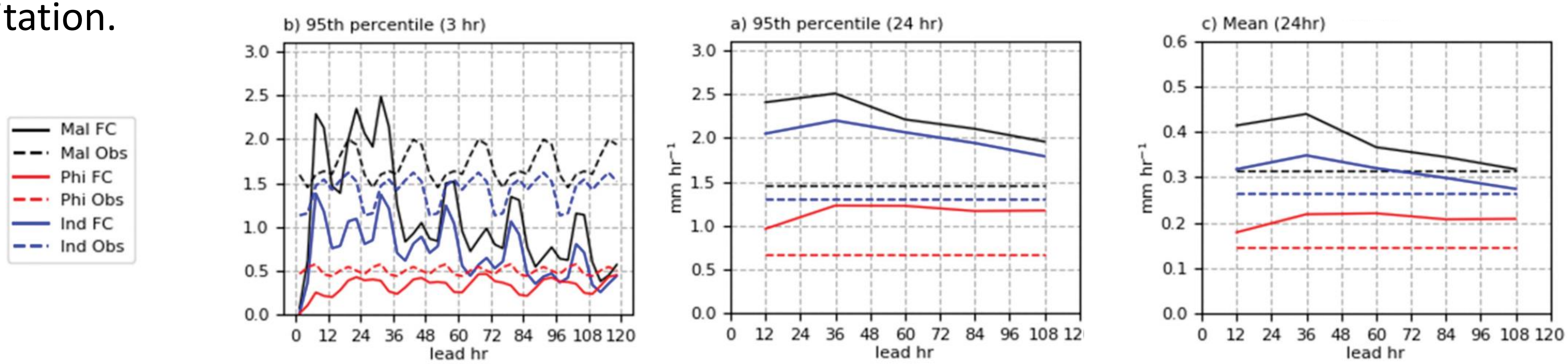
  - ***Use no model data*** but instead use GPM precipitation from the day prior to forecast initialization and replicate for every 24 hours in the 120-h forecast period.
  - ***To act as a benchmark*** for how skillful the ensemble forecasts are.

Observation	Day 0	Day 1	Day 2	Day 3	Day 4	Day 5
Forecast		Day 0	Day 0	Day 0	Day 0	Day 0

4. Results

a. Rainfall climatology and 4.5-km forecast bias

- Pronounced ***diurnal variation*** are observed in Malaysia and Indonesia.
- Each domain has ***2 peaks*** in rainfall: 08Z and 20Z (03~04 LST and 15~16 LST)
- The peaks correspond to the rainfall over ***land*** in the evening and over ***sea*** in the morning.
- The ***differences between two peaks*** in model are larger than the GPM observation, which is caused by the underestimation of rainfall over ocean.
- The forecasts tend to ***underestimate*** the 3-hour rainfall but ***overestimate*** the 24-hour precipitation.



4. Results

a. *Rainfall climatology and 4.5-km forecast bias*

- The ensemble members tend to ***underestimate*** the 95th rainfall but ***overestimate*** the 99th rainfall for the 3-h precipitation.

- Other studies also find that the CP model underestimate low rainfall intensities and overestimate high intensities in other regions.

  (Woodhams et al. 2018; Kendon et al. 2012)

- GPM is known to underestimate heavy rainfall events, which may also cause the "overestimation" of model rainfall.

  (Tan and Duan 2017; Sunikumar et al. 2019)

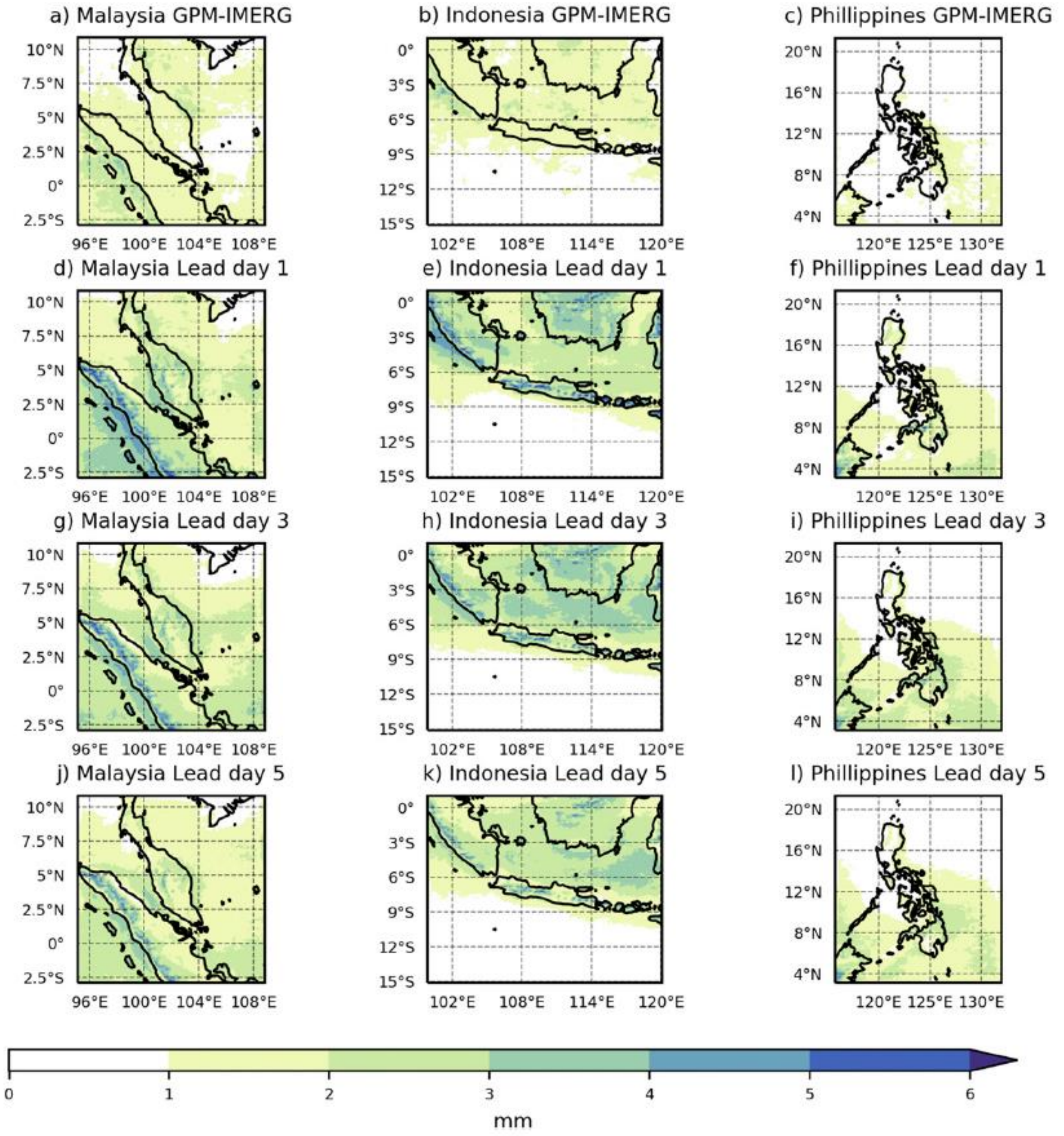
Region	GPM/model resolution	95th pc	99th pc
Malaysia	GPM	1.85	5.62
	2.2	1.49	7.13
	4.5	0.78	6.45
	8.8	0.62	6.04
	8.8 (GA)	1.41	2.78
Indonesia	GPM	1.35	4.76
	2.2	0.52	5.69
	4.5	0.48	5.87
	8.8	0.12	4.21
	8.8 (GA)	1.23	2.57
Philippines	GPM	0.34	2.66
	2.2	0.18	3.21
	4.5	0.27	4.34
	8.8	0.11	2.93
	8.8 (GA)	0.58	2.19

4. Results

a. *Rainfall climatology and 4.5-km forecast bias*

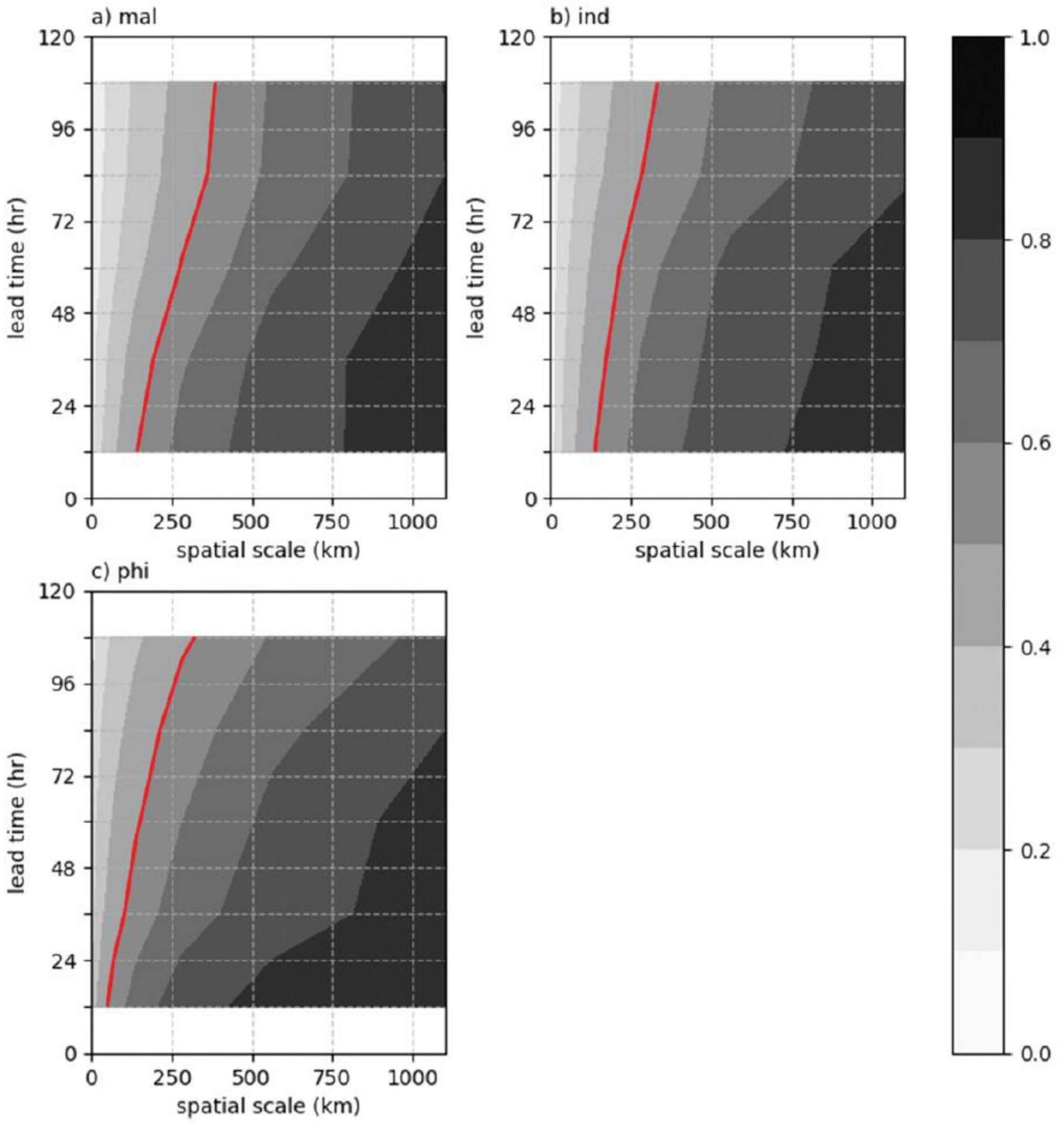
95th percentile for 24-hour precipitation between GPM and 4.5-km forecast:

- ➢ The highest rainfall in GPM occur over ***ocean***.
- ➢ The highest rainfall in model occur over ***mountainous regions*** of Sumatra and Java.
- ➢ The rainfall off the west coast of Sumatra decrease rapidly as the ***lead time*** increases (fig. d, g, j).
- ➢ The rainfall amount in the Philippines domain is well forecasted compared to the other two.



4. Results

b. Skill of 4.5-km ensemble forecasts of daily precipitation accumulations



<u>eFSS of 24-hour precipitation exceeding 95th percentile in 4.5-km forecasts:</u>

- The forecasts are considered skillful if ***eFSS > 0.5*** (on the right of the red line).
- Skill decreases as ***lead time*** increases.
- The skill in the Philippines domain is higher than the other two.
- The skill is likely to vary with both ***location (land, sea)*** and ***time of a day (diurnal cycle)***, so they are further examined in the following.

4. Results

c. *The role of the diurnal cycle in 4.5-km forecast skill*

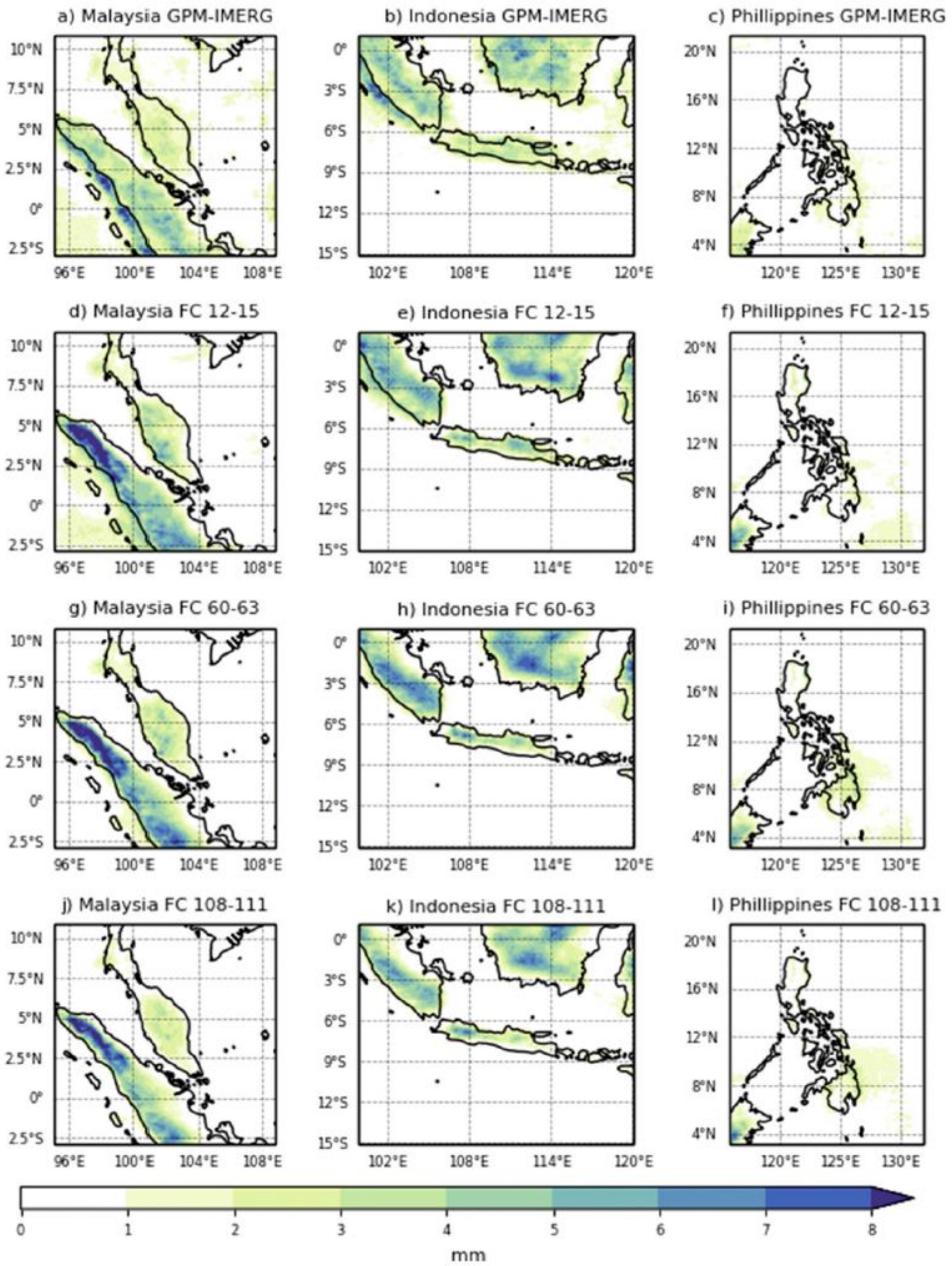
95th percentile of 12-15 UTC 3-hour rainfall (local evening):

➢ **GPM:**

1. The highest values are over Sumatra and Borneo.

2. The rainfall over Sumatra is located at the mountains and the west coast.

3. The rainfall over Java, peninsular Malaysia, and the Philippines are weaker.



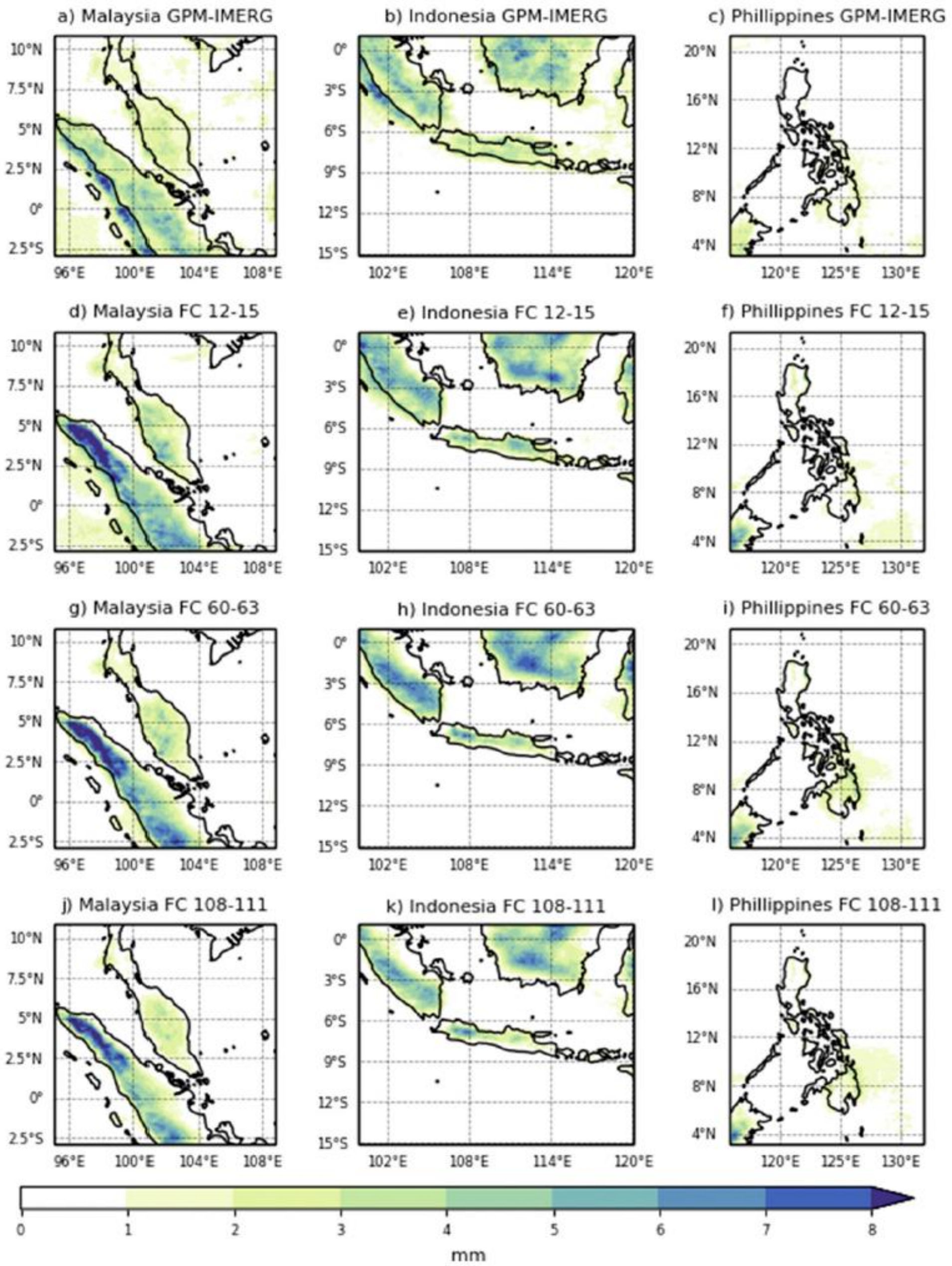
4. Results

c. *The role of the diurnal cycle in 4.5-km forecast skill*

95th percentile of 12-15 UTC 3-hour rainfall (local evening):

- **Model:**

1. The rainfall over Sumatra is concentrated at the northwest and the precipitation off the west coast is not captured.
2. Rainfall over the ocean decreases more apparent than that over the land as the lead time increases.
3. The overall spatial pattern of the precipitation remains the same as the lead time increases.

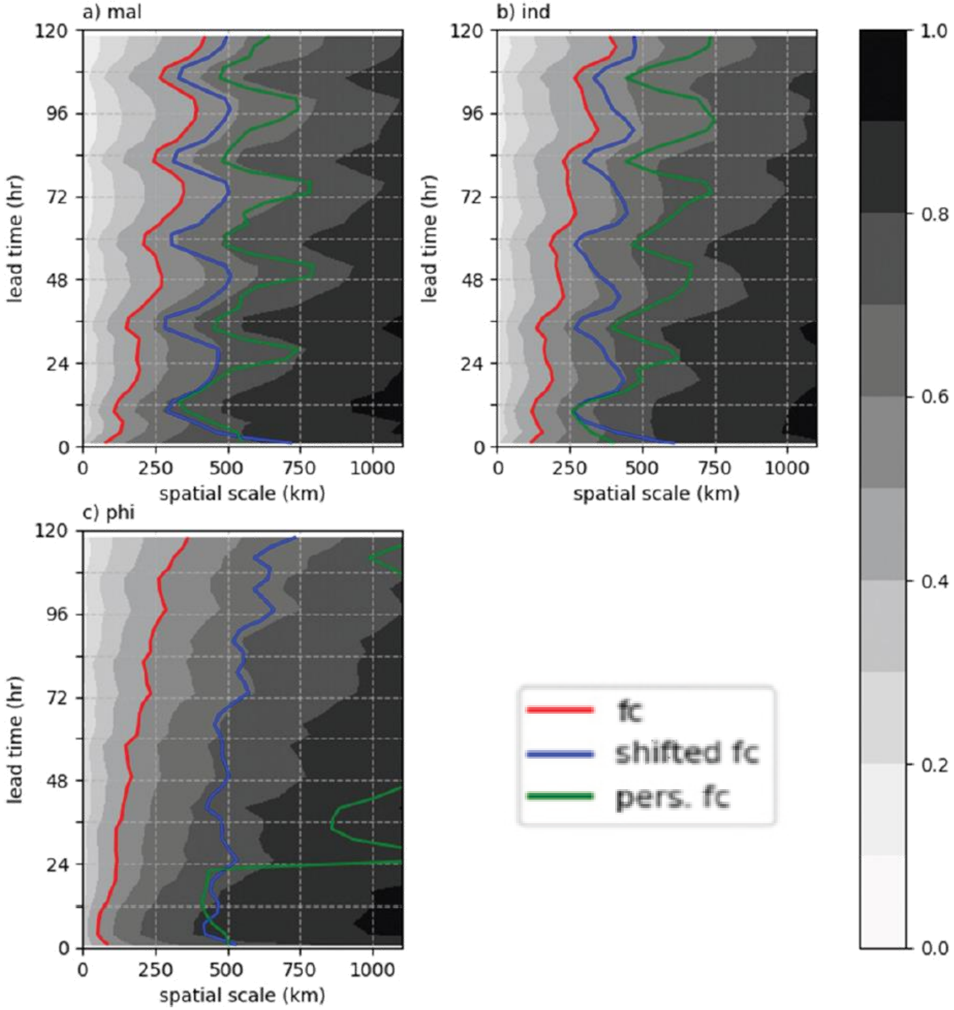


4. Results

c. The role of the diurnal cycle in 4.5-km forecast skill

<u>eFSS of 3-hour accumulated rainfall exceeding 95th percentile:</u>

- The shading is the eFSS for standard forecast, but the skillful standards (***eFSS=0.5***) are shown for all three forecast variations.
- For the ***red line (standard forecast):***
  1. The skill is strongly tied to the ***diurnal cycle*** in the Malaysia (fig. a) and Indonesia (fig. b) domain.
  2. The skill tends to be largest in the daytime when rainfall is over land and smallest at night when precipitation is offshore.
  3. The skill decrease as the ***lead time*** increases.



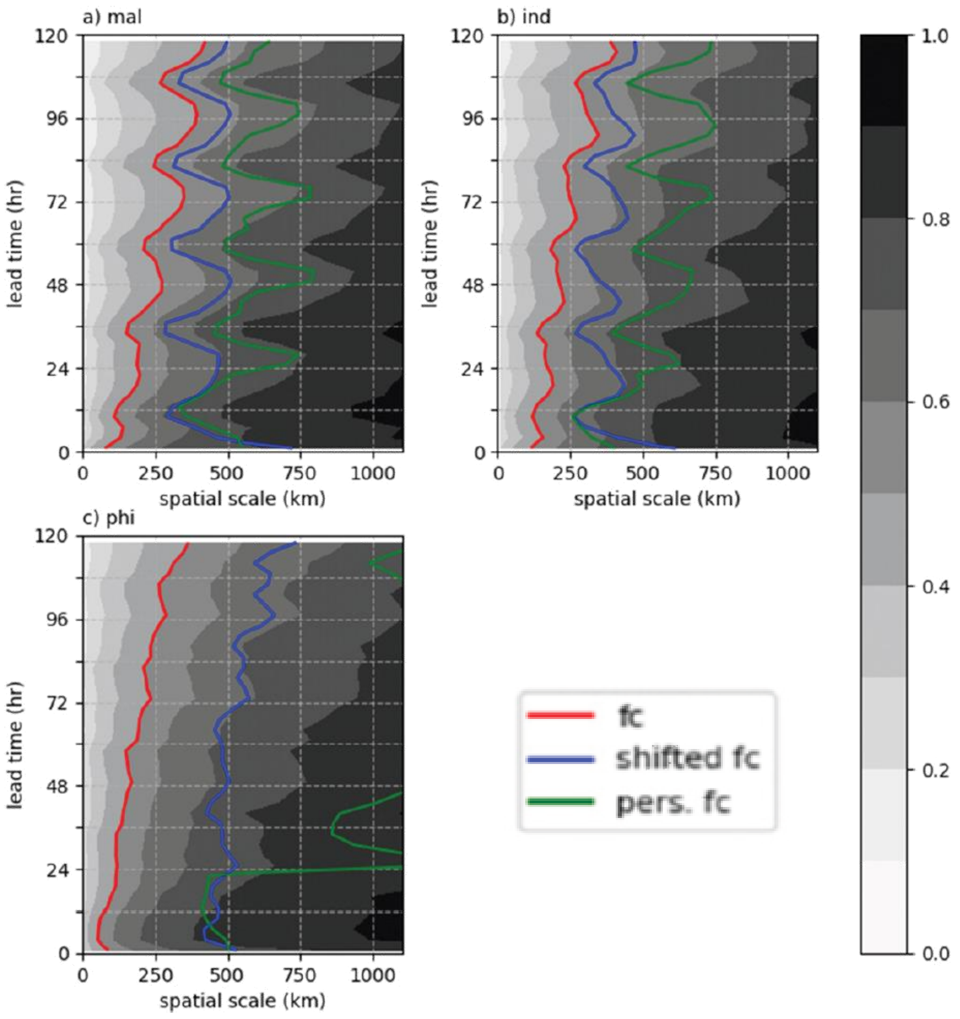
4. Results

c. *The role of the diurnal cycle in 4.5-km forecast skill*

<u>eFSS of 3-hour accumulated rainfall exceeding 95th percentile:</u>

➢ For the ***green line (persistent forecast):***

1. To test how much skill is driven simply by diurnal variations and how much is added by dynamical ensemble forecast (red line).
2. The standard ensemble forecast (red line) is more skillful than the persistence forecast (green line)
3. The ensemble forecast contains more information about the weather occurring in the future.



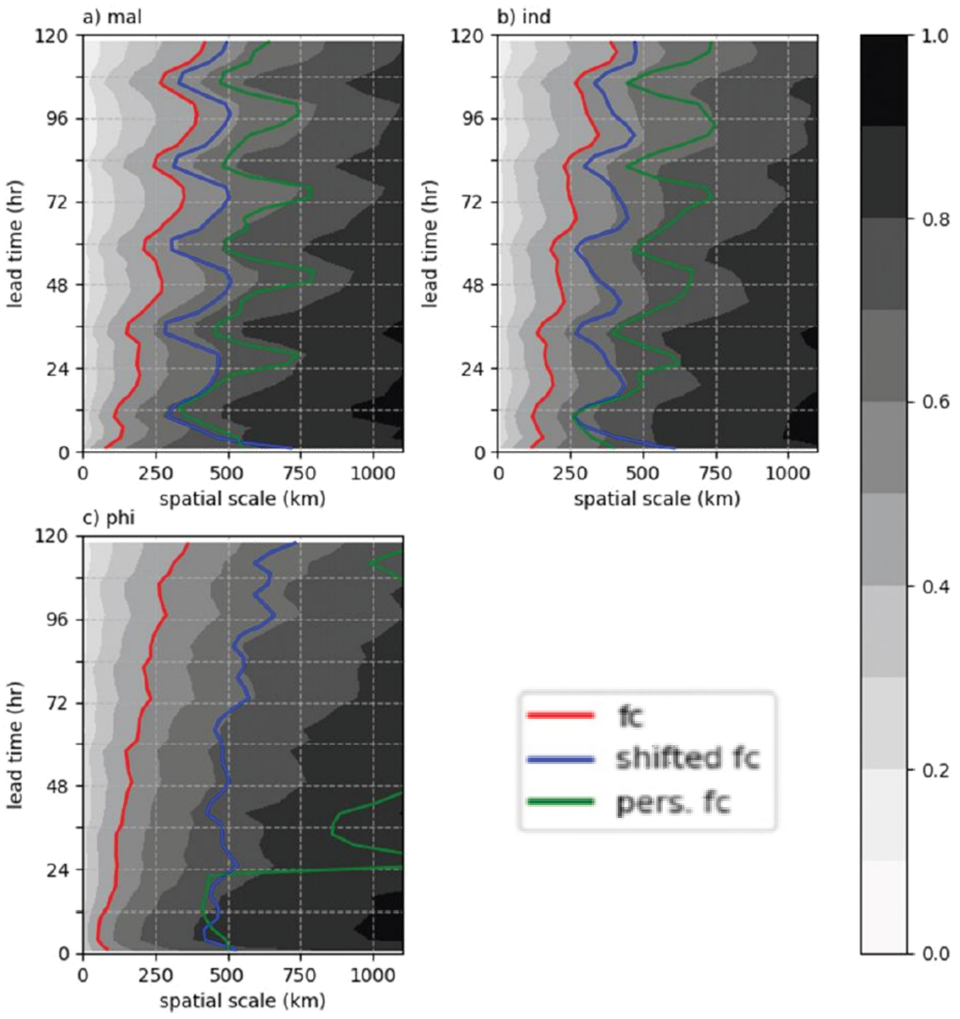
4. Results

c. *The role of the diurnal cycle in 4.5-km forecast skill*

eFSS of 3-hour accumulated rainfall exceeding 95th percentile:

- For the ***blue line (shifted forecast):***

  1. To test how much predictability comes from the diurnal cycle in the current flow regime.
  2. The standard forecast performs better, which indicates some forecast skill is associated with the phenomena in multi-day time scale captured by the model.
  3. The shifted forecast (blue line) tends to move toward standard forecast (red line) as the lead time increase.

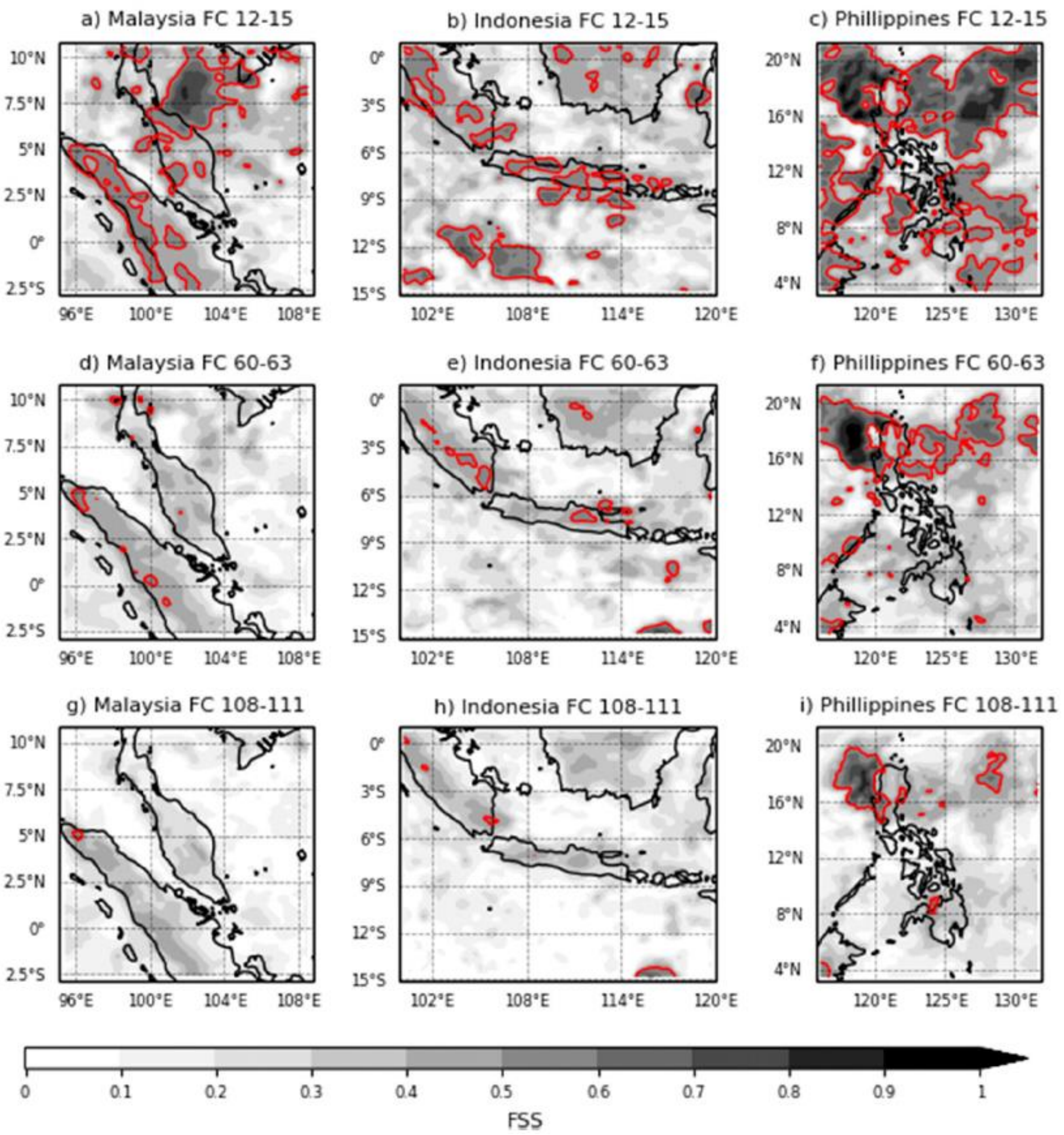


4. Results

c. *The role of the diurnal cycle in 4.5-km forecast skill*

LFSS of 3-hour accumulated rainfall exceeding 95th percentile in the evening (N=72km):

- The aim is to find the ***spatial distribution*** of skill across the full domain.
- Skill tends to be located over land where the forecast has most precipitation during the evening for the Malaysia and Indonesia domains.
- The Philippines domain shows higher skill over ocean due to the stronger impact from ***synoptic scale variability*** and weaker ***diurnal cycle***.

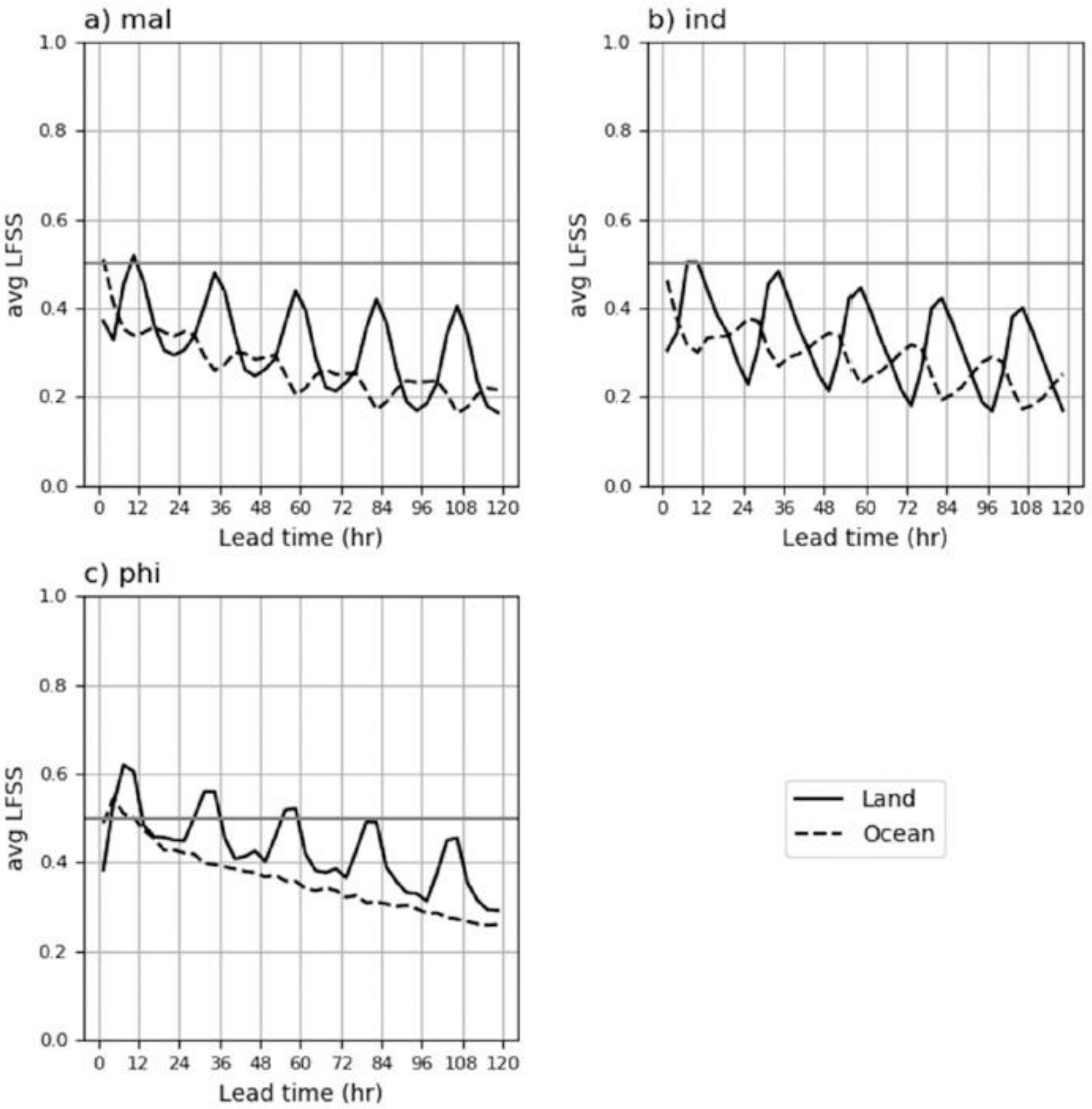


4. Results

c. *The role of the diurnal cycle in 4.5-km forecast skill*

<u>LFSS of 3-hour accumulated rainfall exceeding 95th percentile over the land and ocean (N=72km):</u>

- Peaks in skill occur at the time when the rainfall is heavier in the ocean or the land.
- There is much less skill over ocean in the Malaysia and Indonesia domains.
- Most skill comes from precipitation that is constrained by topography.



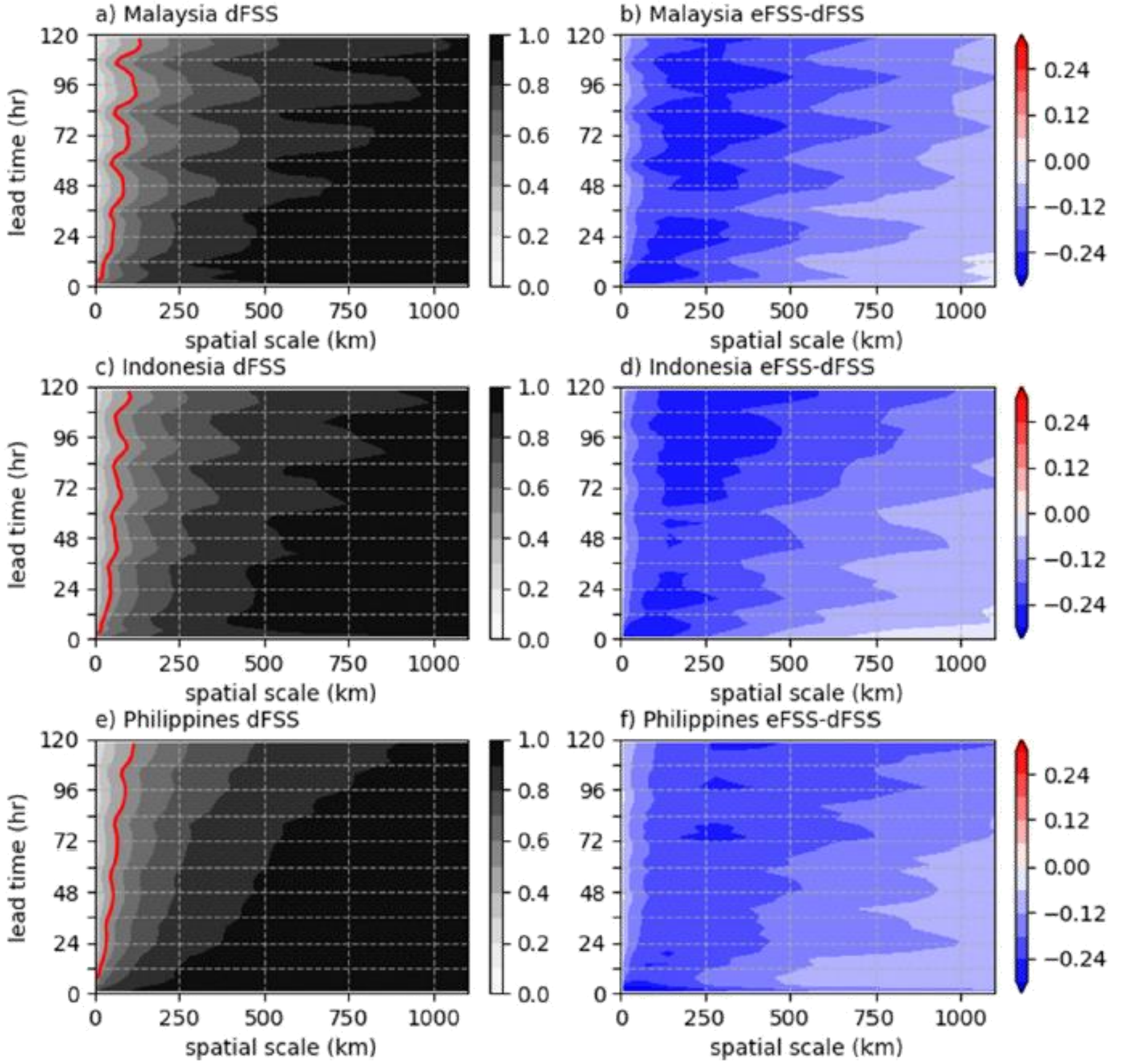
4. Results

d. Ensemble spread-skill relationship for 4.5-km forecasts

- ***dFSS*** (Dey et al. 2014) is used to assess skill differences between ensemble members, which is "***ensemble spread***" of forecast skill.

Conditions	Meaning
$dFSS > eFSS$	Underspread
$dFSS < eFSS$	Overspread

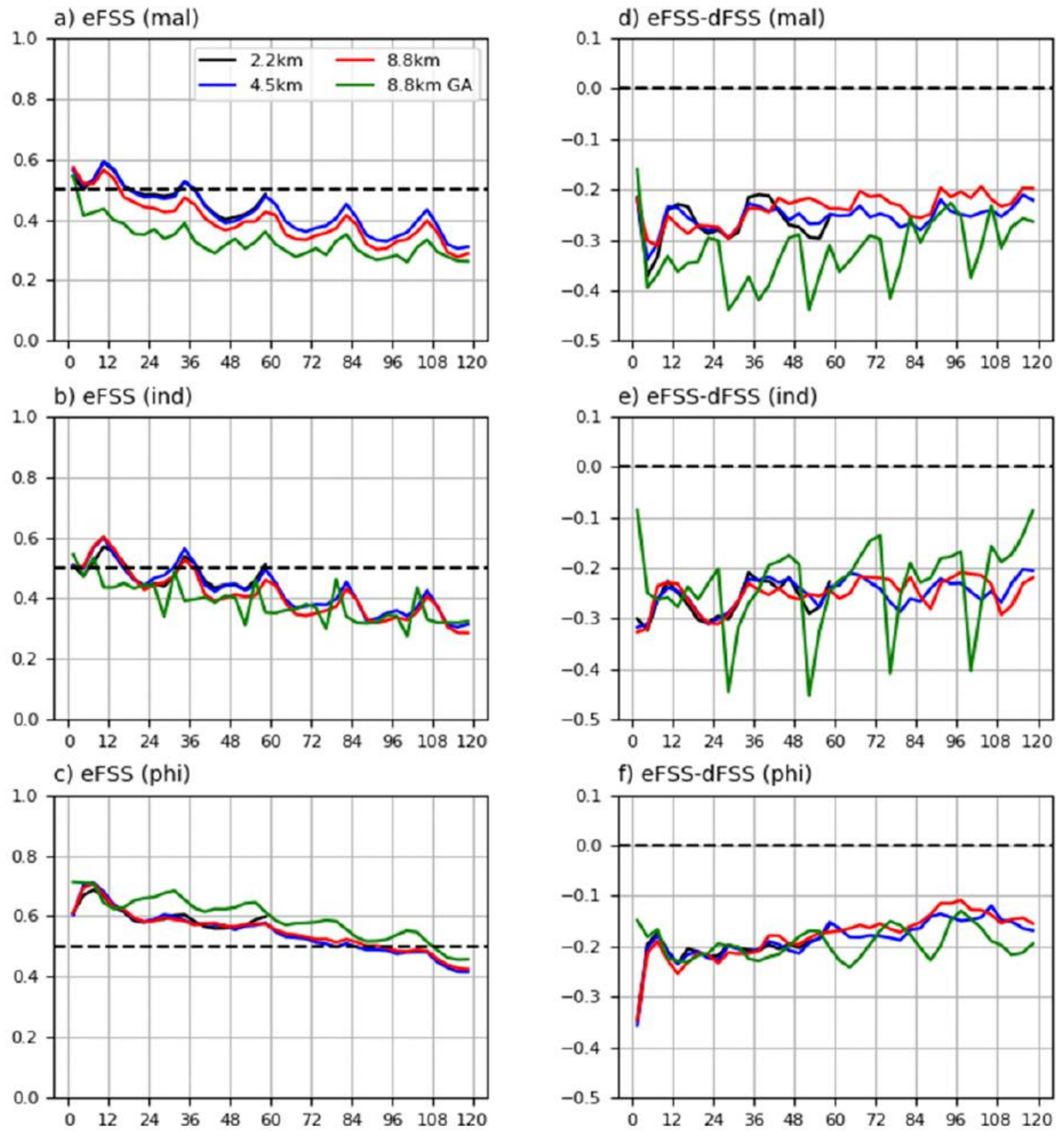
- $\boldsymbol{dFSS > eFSS}$ for all domains, which indicates the ensemble members are too similar to one another.
- The spread also varies with diurnal cycle and is smallest (largest dFSS) in the local evening.



4. Results

e. *The role of resolution and convection parameterization in forecast skill and spread*

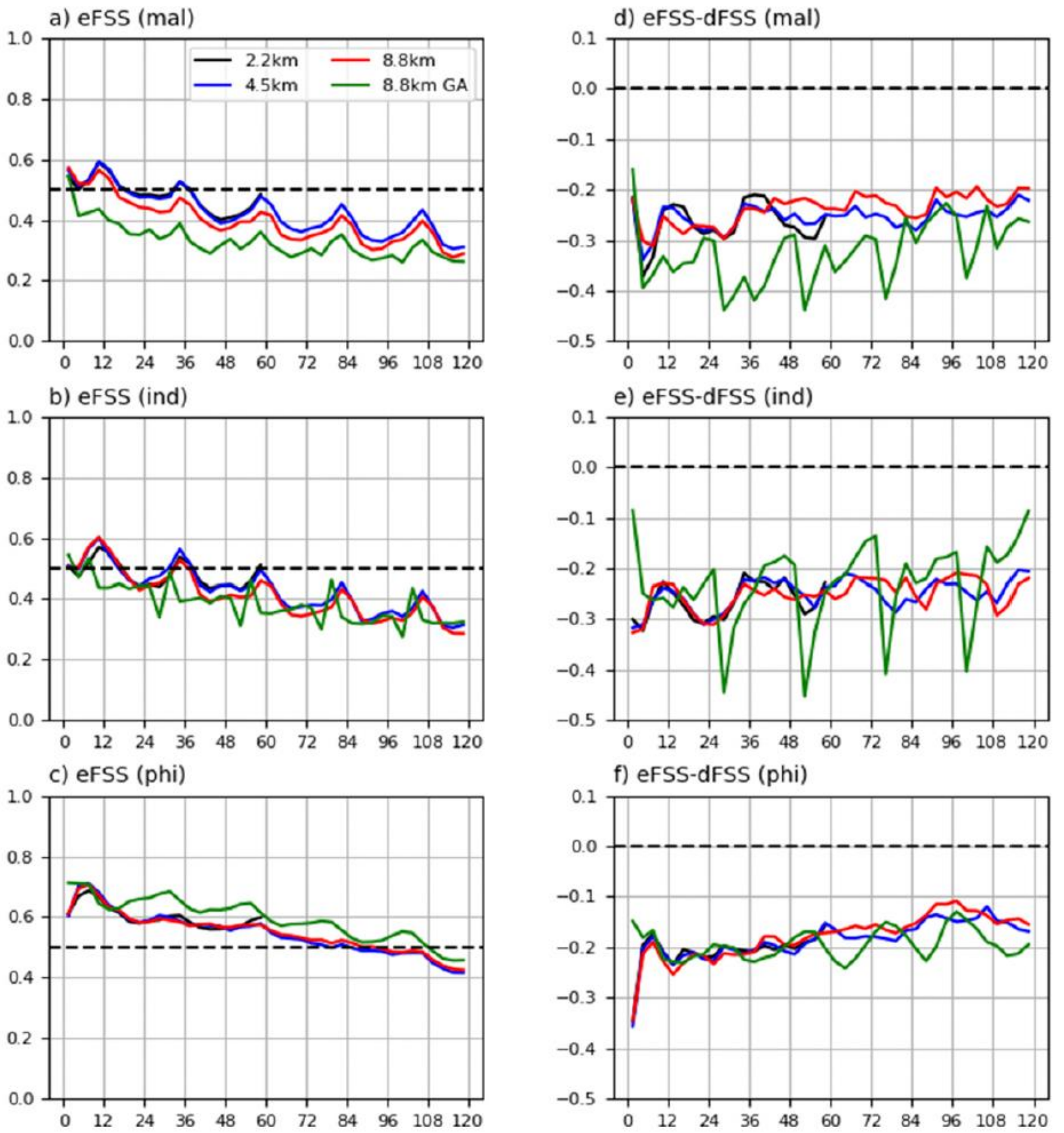
- ***Resolution*** tends to play a minor role in forecast skill and ensemble spread.
- GA ensemble shows less skill in Malaysia and Indonesia domains, in which the rainfall rely more on the ***diurnal cycle***.
- The Indonesia rainfall peaks in GA ensemble occur a few hours earlier, which cause the skill decreasing and followed by a increasing in local evening.
- The rainfall in coarser resolution ensemble (8.8 km) tend to be smaller, which cannot be evaluated by the ***percentile-based*** FSS evaluation.



4. Results

e. The role of resolution and convection parameterization in forecast skill and spread

- The GA ensemble performs better than CP ensemble in the Philippines domain.
- The precipitation patterns at the Philippines are influenced by ***larger-scale phenomena***.
- The ensemble spread is smaller in the local evening for both CP and GA ensembles.
- The spread in the evening is smaller in GA ensemble, which may be caused by the ***parameterization*** used.



5. Summary and conclusions

1. In this study, the skill of ***convection permitting (CP) ensemble forecasts*** is evaluated in the domains covering Malaysia, Indonesia, and the Philippines by ***fractions skill score (FSS)***.
2. This study mainly focuses on the impacts of ***diurnal cycle*** on the ***skill and spread*** of the ensemble rainfall forecast.
3. The ***diurnal cycle*** plays an important role in the domains of Malaysia and Indonesia.
4. The forecasts perform better over the ***land*** than over the ***ocean***.
5. The comparisons between the ***standard forecasts***, the ***shifted forecasts*** and the ***persistence forecasts*** conclude that the ensemble forecast systems can capture the scopes beyond the diurnal cycle.

5. Summary and conclusions

6. The skill decreases as the ***lead time*** increases.
7. The ***resolution*** plays a fairly small role in the skill of ensemble rainfall forecasts.
8. The convection permitting (CP) ensemble performs better than the ***parameterized ensemble (GA ensemble)*** in the precipitation that is highly associated with diurnal cycle.
9. Although the forecast skill evaluated by FSS are similar across all resolutions, the ***percentile-based*** method can only describe the ***spatial patterns*** of the rainfall areas.
10. The ***ensemble spread (dFSS)*** is 59%, 61%, and 33% less than the ***ensemble mean forecast error (eFSS)*** on average for Malaysia, Indonesia, and the Philippines domains, respectively.