Cloud and Precipitation Prediction within the GFDL SHiELD Model

Lily Johnston

Environmental Studies and Science Program

Colorado College

Supervisor

Dr. Allison Lawman, Dr. Linjiong Zhou

In partial fulfillment of the requirements for the degree of

Bachelor of Arts in Environmental Sciences

May 19, 2024

Acknowledgements

I would like to thank Dr. Linjiong Zhou and Dr. Allison Lawman for their support in this research and in the writing of this thesis. A special thank you to the Cooperative Institute for Modeling the Earth System (CIMES) internship program for giving me the opportunity to conduct research at the NOAA Geophysical Fluid Dynamics Laboratory and Princeton University. Thank you very much to the FV3 group at NOAA GFDL for allowing me to research the SHiELD model. I would also like to thank the Environmental Science Department at Colorado College for their continued support.

Abstract

Cloud and precipitation prediction poses considerable challenges in many numerical models. The global 13-km resolution System for High-resolution prediction on the Earth-to-Local Domains (SHiELD), a Unified Forecast System (UFS) prototype atmospheric model developed at the Geophysical Fluid Dynamics Laboratory (GFDL), is used to evaluate the prediction accuracy of clouds and precipitation. This research compares cloud and precipitation predictions from the 13-km SHiELD prediction system to observations. In particular, the geographic distribution, probability density function, diurnal cycle, error growth, and quantitative precipitation forecasts are carefully evaluated. In addition, to understand the extent of the prediction accuracy and its dependency on the model's horizontal resolution, I compare the 13-km SHiELD with its 25-km and 6.5-km versions.

The 13-km SHiELD shows excellent performance in ice water path, geographic mean precipitation, the global probability density function (PDF) for light to medium precipitation, PDF over land's extreme precipitation, and peak precipitation time over the land. However, the complexities of cloud and precipitation prediction have resulted in noticeable biases in predicting the geographic distribution of precipitation, precipitation diurnal cycle, ice and liquid water path, and cloud fraction in SHiELD.

We find that the SHiELD prediction system exhibits the potential for improving cloud and precipitation prediction using finer horizontal resolutions. Degradation in cloud fraction, ice water path, precipitation error, root mean square error, equitable threat score, area under the curve and fractions skill score occurs when the model's resolution is reduced to 25 km. These comparisons help uncover and understand the biases in the SHiELD system, with the goal of proposing solutions for improving cloud and precipitation prediction.

Plain Language Summary

Clouds and precipitation are difficult to predict and are a large source of error in weather prediction. This research uses a weather model called System for High-resolution prediction on the Earth-to-Local Domains (SHiELD) to evaluate if cloud and precipitation predictions are improved by doubling the resolution - the degree of refinement of a weather model's grid. Here we find that SHiELD has potential for improving cloud and precipitation prediction through the use of higher resolutions. The identification of significant differences between different model resolutions informs proposed solutions for improving cloud and precipitation prediction in SHiELD.

Contents

Introduction

The accurate prediction of clouds and precipitation is important to our daily activities and decision-making processes. However, cloud and precipitation prediction poses considerable challenges in many numerical models [1]. Numerical models struggle to accurately simulate clouds and precipitation due to factors such as terrain bias, cloud microphysics, and cloud-radiation-aerosol interactions [2].

GFDL SHiELD is the System for High-resolution prediction on the Earth-to-Local Domains developed by the National Oceanic and Atmospheric Administration (NOAA) Geophysical Fluid Dynamics Laboratory (GFDL). Previous studies [2–9], SHiELD show the power of a unified prediction system [2] across a variety of temporal and spatial scales designed for a wide array of applications such as severe weather prediction, subseasonal-to-seasonal prediction and tropical convection. It also shows the abilities of the Finite-Volume Cubed-Sphere Dynamical Core (FV3), especially its flexible nonhydrostatic dynamics, variable-resolution capabilities, and integrated physics, coupled with the elegance of the Flexible Modeling System (FMS) framework. Within SHiELD there are several domains for different regional and temporal applications: X-SHiELD, S-SHIELD, C-SHiELD, T-SHiELD, Tele-SHiELD and SHiELD (Figure 1).

Previous work demonstrates that the 13-km SHiELD model improves large-scale prediction skill (how closely model simulations resemble observations) [6]. This study finds that S-SHiELD, which is used for subseasonal to seasonal prediction, has a significantly higher prediction skill of diurnal cycle when compared to other climate models [6]. Furthermore, S-SHiELD successfully predicts the Madden-Julian Oscillation for a 28 day observational period which is an example of SHiELD's capabilities of sub-seasonal tropical climate variability [6].

Furthermore, Harris et al. (2020) suggest that T-SHiELD (the global-to-regional nested configuration) and C-SHiELD (CONUS - Contiguous United States) demonstrate large improvements in hurricane structure and exhibit the potential for the prediction of convective storms. Chen et al. (2019) find that SHiELD (previously called fvGFS), is superior in the prediction of tropical cyclone tracts over the northern Atlantic basin and northern Pacific Ocean compared to existing Global Forecasting Systems (GFS) [4]. Furthermore, the fvGFS's updated GFDL 6-category cloud microphysics scheme improved the prediction of tropical cyclone intensity [4].

Magnusson et al. (2022) compares the skill of medium-range forecast models in the Different Models, Same Initial Conditions (DIMOSIC) project studied how SHiELD and other models prediction compare [10]. DIMOSIC found that SHiELD has the strongest discrepancy between the modeled and observed precipitation over the Indian ocean resulting in high precipitation rates in 10-day forecasts [10]. Magnusson et al. (2022) indicate the potential for improvement in SHiELD through the usage of European Centre for Medium-Range Weather Forecasts (ECMWF) analyses [10].

Increasingly, numerical weather prediction is trending towards higher resolution, but it remains uncertain how resolution impacts prediction [11]. Studies, such as Mass et al. (2002), that have detailed the affects of horizontal resolution on forecast accuracy of resolutions less than 10 km found that the structure of forecasts improved, but fewer studies have indicated that higher resolutions result in enhanced skill scores [11]. One of these studies is by Gallus (1999), which found that finer horizontal resolution in weather prediction systems generally improves quantitative precipitation forecasting (QPF) and traditional skill scores [12]. This motivates our study which also uses the QPF metric to identify if changing the horizontal resolution in SHiELD improves model performance. When increasing the horizontal resolution in numerical models, it is important to understand the mechanism of resolution and why it impacts accuracy – increased horizontal resolution facilitates more vertical movement in the model, resulting in greater moisture transport and increased peak precipitation values [12].

Christopoulous (2021) found the diurnal cycle amplitude to be more realistic in higher-resolution models; thus indicating that resolution is a pathway that can be explored to improve the precipitation diurnal cycle in SHiELD [1]. A study done by Mass et al. (2002) found that increasing resolution improves the realism of the simulation but does not necessarily improve the skill of the forecasts [11]; suggesting that increasing horizontal resolution may have benefits within SHiELD but there are limitations to how significantly scored accuracy can be increased.

Studies, such as Gallus (1999), have found that the correlation between subgrid-scale and grid-resolvable scale is of utmost importance to the resulting precipitation forecasts, providing further reason for this research [12]. Generally, there is disagreement in the scientific community as to whether higher-resolution numerical weather prediction produces improvement in forecasts [11, 12]. This research aims to assess and identify the cloud and precipitation prediction errors and biases in SHiELD. A comparison between the 13-km and 6.5-km SHiELD and between the 13-km and 25-km SHiELD is performed to uncover and understand the biases, with the ultimate goal of proposing solutions for improving cloud and precipitation prediction. Model data comparison helps to reduce the uncertainty by validating the model simulations with observations. Previous studies focusing on the GFDL SHiELD did not carefully evaluate cloud and precipitation prediction [2–9]. This research aims to add to the existing scientific research on the benefits and limitations of high-resolution modeling. We evaluate how spatial resolution impacts SHiELD cloud and precipitation prediction with the intention of improving SHiELD and providing a framework for assessing the correlation between resolution and accuracy in other prediction systems.

Model and Methods

SHiELD is a Unified Forecast System (UFS) prototype atmosphere model that shows the power of a unified prediction system across a variety of time and space scales designed for a wide array of applications [6]. In this research, the global 13-km SHiELD is

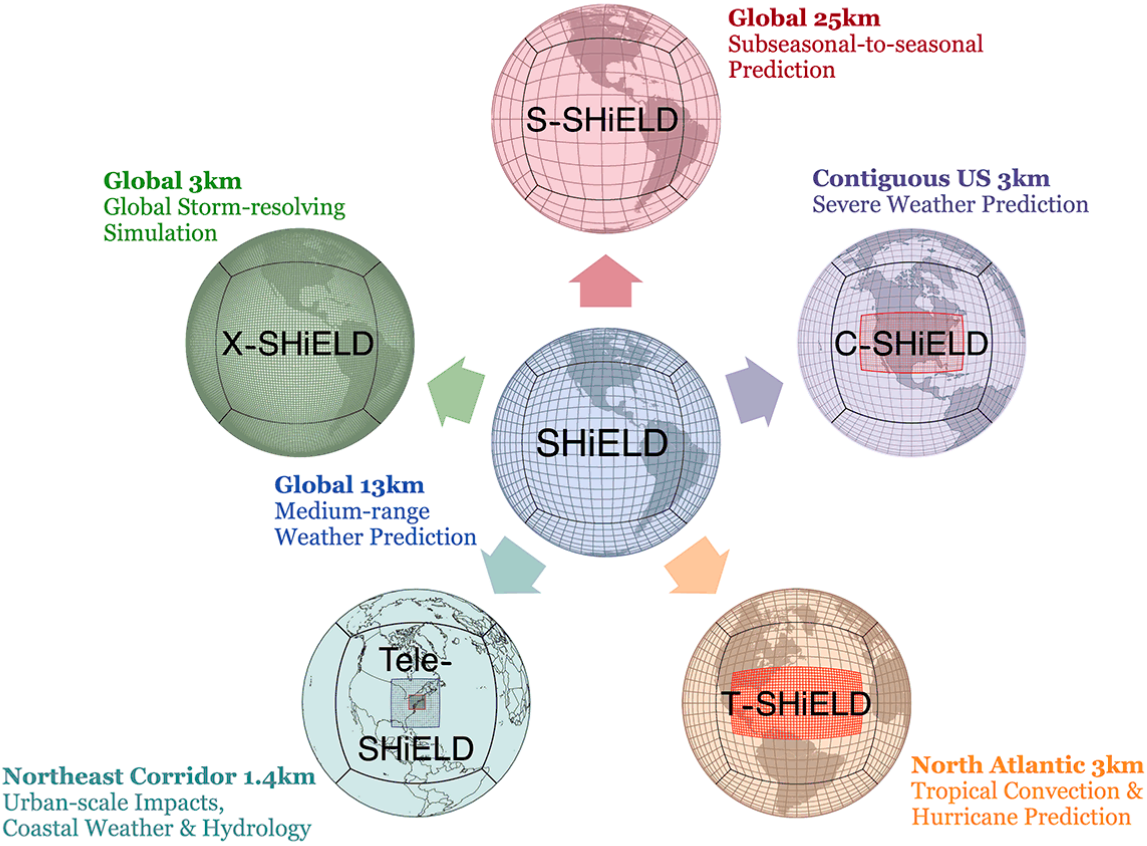


Figure 1 SHiELD Unified Forecast System. In this research, the GFDL SHiELD version 2023 is used at C384 (25 km), C768 (13 km) and C1536 (6.5 km) resolutions.

used to study cloud and precipitation prediction. GFDL SHiELD 13 km Medium-range Weather Prediction is used to simulate 21 cases from March 2021 to present day. During post-processing, the cases are averaged and regridded to 1 degree and 13 km horizontal resolution and the hindcasts are displayed in all of the figures. This research compares the model's output with the observed data at the same spatial (1 degree and 13 km) and temporal (3 hourly) resolution to assess the accuracy of the cloud and precipitation prediction. Clouds and precipitation were evaluated over the global, CONUS, Maritime Continent, South America, and the Tropics domains, including land-only and ocean-only masks. Here, we focus on the no-masked global domain. In addition to the 13-km SHiELD, the model is also reconfigured to have a horizontal resolution of 6.5 km and 25 km to study the effect of resolution on prediction.

SHiELD uses MSWEP [13], StageIV [14], CERES [15], ERA5 [16], MRMS [17], and the RTMA [18] observational products to simulate clouds, precipitation, tropical convection, etc. in the domains of SHiELD. This research focuses on the MSWEP dataset for geographic mean precipitation, zonal mean precipitation, peak precipitation time, precipitation diurnal cycle, prediction error growth and quantitative precipitation forecasts [13]. The CERES dataset is used for cloud fraction, ice water path and liquid water path [15]. The SHiELD system produces an average forecast of the 21 cases over a 10 day observational period. The forecasts are compared to the observations at the same spatiotemporal scale. Various statistical methods were used to quantitatively evaluate cloud and precipitation prediction in comparison with various observations. Statistical analyses include root mean square error (RMSE), bias, forecast error growth, diurnal cycle, probability density function (PDF), and probability density function (QPF). All significance tests are implemented at a 95% confidence.

Results

Geographic mean precipitation, zonal mean precipitation, peak precipitation time, precipitation diurnal cycle, cloud fraction, liquid water path, ice water path, prediction error growth and quantitative precipitation forecasts were analyzed in SHiELD. The simulations were compared to the observations to identify regions with significant biases. For all variables except for cloud fraction the MSWEP dataset was used [13]. For cloud fraction the CERES dataset was compared to the modeled cloud fraction [15]. SHiELD's preformance was evaluated by comparing the simulations to the observational products. In this and subsequent sections, positive bias in SHiELD indicates that SHiELD over predicts while a negative bias is representative of SHiELD's under prediction of a variable.

Geographic Mean Precipitation

25 km vs. 13 km

The geographic distribution of precipitation from the MSWEP observation and model in both 25 km and 13 km resolutions is shown in Figure 2. SHiELD generally accurately predicts geographic mean precipitation for 25 km (Figure 2c) and 13 km (Figure 2d) resolutions when compared to the MSWEP observation [13]. When comparing the geographic mean precipitation between the model resolutions, similar biases are found (Figure 2e and f). SHiELD has a strong positive bias in the tropics, especially at the South Pacific Convergence Zone (SPCZ) and over Southern Asia. There is a negative bias in SHiELD over the ITCZ as indicated in the bias panels of Figure 2e-f and Figure 2e-f.

In SHiELD, there is a significant positive bias in geographic mean precipitation over the SPCZ and a negative bias over the ITCZ. Generally, there is not a significant difference between the 13 km and 25 km resolutions as seen in Figure 2b. There are significant differences in geographic mean precipitation between the resolutions over the Arctic; however, since the geographic mean precipitation over the Arctic overall does not have a strong bias, this difference is not particularly noteworthy. The positive bias in geographic mean precipitation could be due to the updraft which brings moisture up into the atmosphere. Improvement in SHiELD as resolution is doubled is seen in South China, the South China Sea and the Philippine Sea (Figure 2d).

While there are areas of significant difference between the 13 and 25 km resolutions, overall, similar biases exist among the resolutions, thus indicating that there is not a significant improvement by doubling the resolution for geographic mean precipitation. The bias and root mean square error (RMSE) are larger when the resolution is decreased from 13 km to 25 km, which implies potential degradation of geographic mean precipitation prediction at lower resolutions (Figure 2).

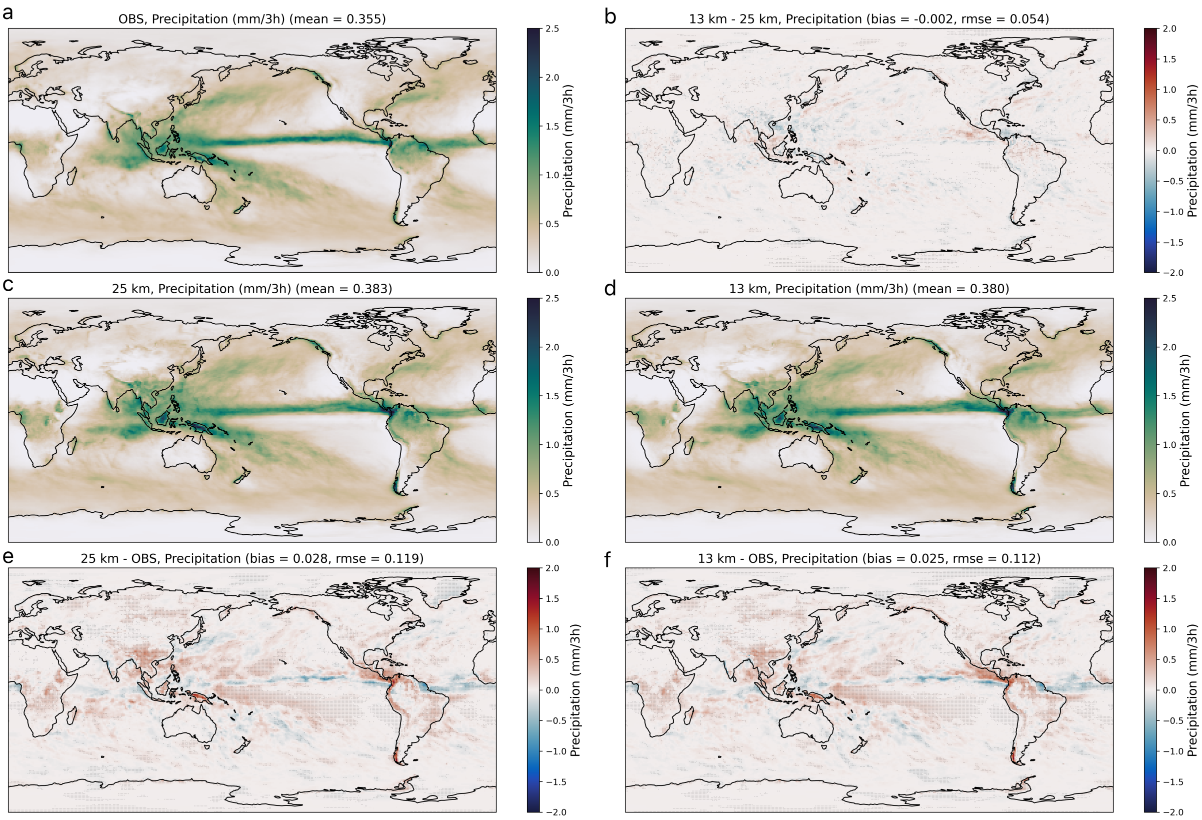


Figure 2 Geographic mean precipitation (units: mm/3h) of a) MSWEP precipitation observation, b) difference between 13-km and 25-km SHiELD, c) 25-km SHiELD, d) 13-km SHiELD, e) difference between 25-km SHiELD and MSWEP observation, f) difference between 13-km SHiELD and MSWEP. 95% significantly different areas are marked dotted.

13 km vs. 6.5 km

When comparing the model with the observations, significantly more precipitation can be found over the subtropical oceans and over most of the land in SHiELD. The highest precipitation rate is associated with the ITCZ. Negative precipitation bias is found north of the Equator. There is not much of a difference between the two resolutions, as seen in Figure 3b. Precipitation values are similar between resolutions (Figure 3c and d), and the biases (Figure 3e-f) are similar, except that the global mean bias is slightly smaller (0.380 mm/3h vs. 0.372 mm/3h) in the 6.5 km resolution model (Figure 3).

One region where the difference is noticeable is off the western coast of South America. Notably, off the western coast of South America, the 13-km SHiELD bias has significant biases (Figure 3d). In the 6.5-km SHiELD, there are fewer areas where the difference is significant; therefore, the 6.5-km SHiELD does a better job of predicting precipitation over the west coast of South America. This point is further demonstrated in Figure 3b, which shows that the difference between 6.5-km and 13-km SHiELD is significant. The negative bias in geographic mean precipitation is decreased when doubling the horizontal resolution. As seen in Figure 3b, the 13 km resolution subtracted from the 6.5-km resolution, results in a negative bias over the ITCZ, meaning that the 13-km SHiELD has a stronger bias over the ITCZ; thus suggesting that the 6.5-km SHiELD is able to improve the prediction of geographic mean precipitation over the ITCZ.

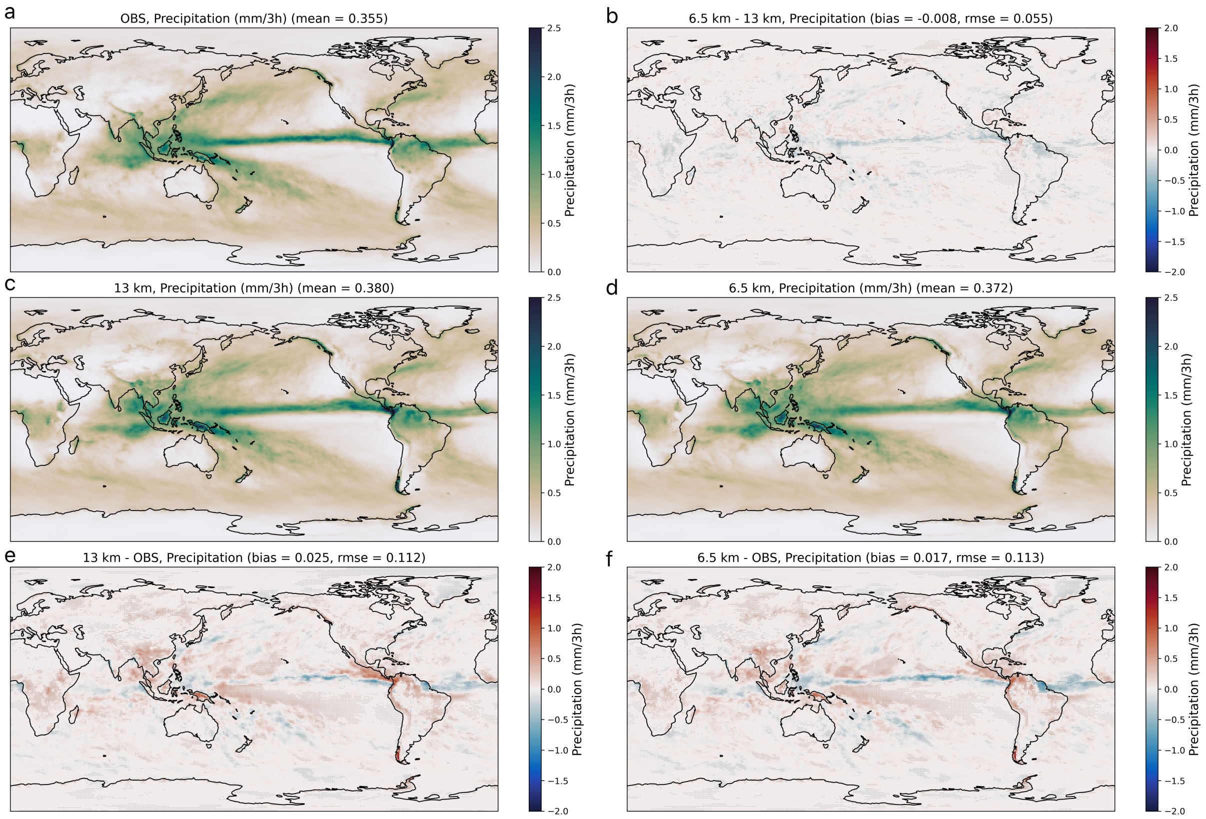


Figure 3 The same as Figure 2, but for the 13-km and 6.5-km SHiELD.

Zonal Mean Precipitation

25 km vs. 13 km

Zonal mean precipitation is compared between the two resolutions of SHiELD. In Figure 4, The zonal mean precipitation of 25-km SHiELD, 13-km SHiELD and observation are shown in the Figure 4a and the the bias and difference in resolutions is in Figure 4b. Shading represents the significance intervals of the difference between the model and observation (blue, green), and between the two configurations of the model (purple). When the observation (black) is outside the significance interval (Figure 4a), the difference between model and observation, is significant, as seen from 20°S to 0°S and 10°N to 30°N, the difference between the two configurations of models is significant. When the observation line is within the significance intervals, it means that the bias or difference between resolutions is not significant. Doubling the horizontal resolution does not necessarily increase prediction accuracy, as seen in Figure 4b with the 25 km - 13 km line staying around zero for most latitudes except for a narrow subtropical regions between 15°S- 10°S and 15°N-30°N.

13 km vs. 6.5 km

Globally, the highest zonal mean precipitation values are associated with the ITCZ (Figure 5a), consistent with the Figure 3c-d. As seen in Figure 5, the difference between the model and observation is significant north of the Equator, the subtropics of both hemispheres, and mid-latitude of the northern hemisphere. Given this bias's significance, this area could be focused on in future model updates. There is a significant difference between the two model resolutions in the tropics 5b. In the global ocean region, the zonal mean precipitation differences are quite similar to the global region, indicating that the largest contributor of bias in the global region is due to the ocean. The only exception is from the mid-latitudes of the Northern Hemisphere, where the global bias comes from the land region (figures not shown).

Over the land (Figure 6), peak zonal mean precipitation is found in the tropics and in the mid-latitudes 45°S-60°S, with a secondary peak at 45°N-60°N. The precipitation peaks at 45S-60°S and 45-60°N correspond to terrain precipitation over mountains (Figure

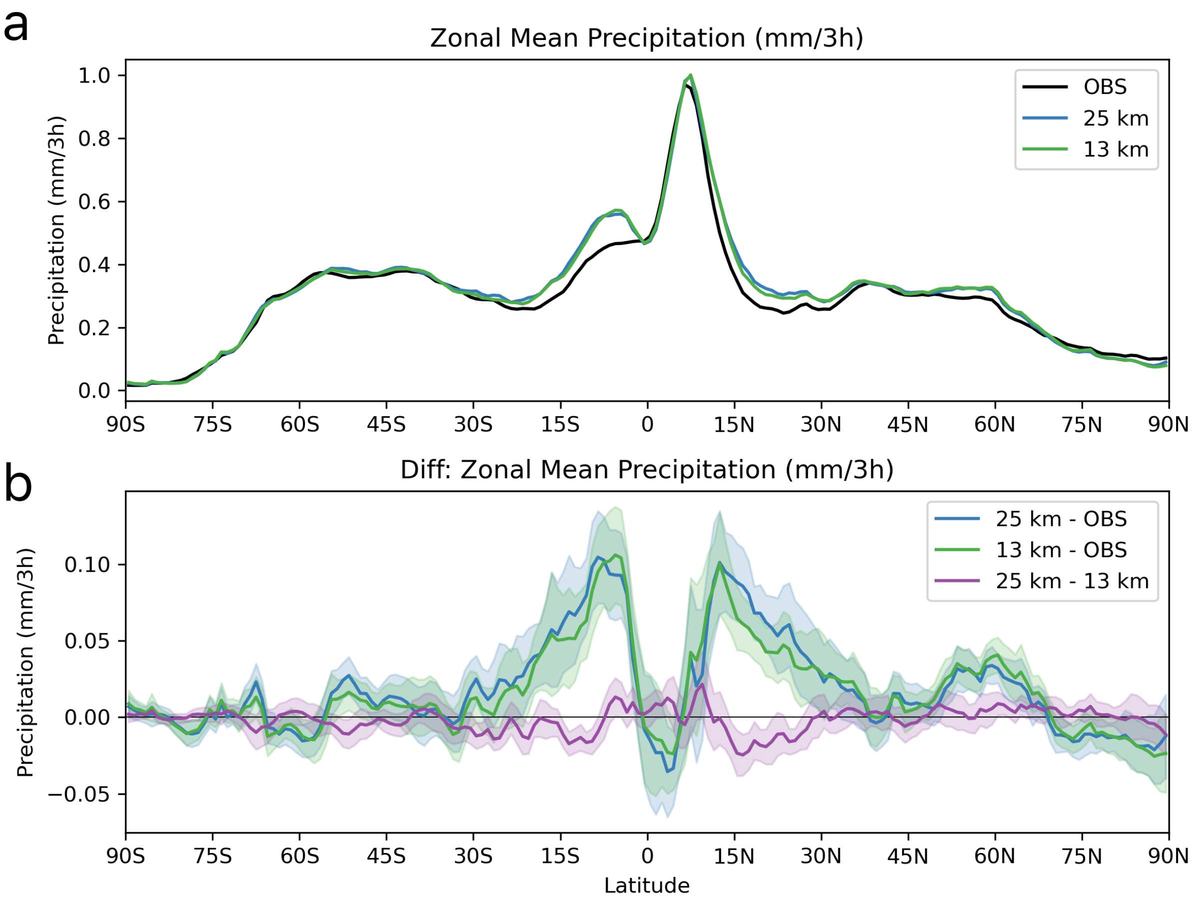


Figure 4 Zonal Mean Precipitation (units: mm/3h) a) Zonal mean precipitation of MSWEP observation (black), 25-km SHiELD (blue), and 13-km SHiELD (green). b) zonal mean precipitation difference between models and observation (blue and green), and between 13-km SHiELD and 25-km SHiELD (purple). Blue and green shadings represent the 95% significant interval between model and observation. Purple shading represents the 95% significant interval between the 13-km SHiELD and 25-km SHiELD.

6). The Andes Mountains (18S-55S) show a large terrain precipitation bias, as shown by the orange and blue lines far above the black observation line (Figure 6a). Additionally, over the tropics, the simulated precipitation is higher than observations, meaning that the model over predicts precipitation. This model bias is significant as the observation is outside of the model significance intervals. In Figure 5, doubling the resolution did not necessarily increase the skill of SHiELD, as the significance intervals for both resolutions are larger than their difference. Furthermore, over the ITCZ, doubling the resolution actually decreased the accuracy of zonal mean precipitation, as seen with a strong negative difference in zonal mean precipitation (Figure 5b).

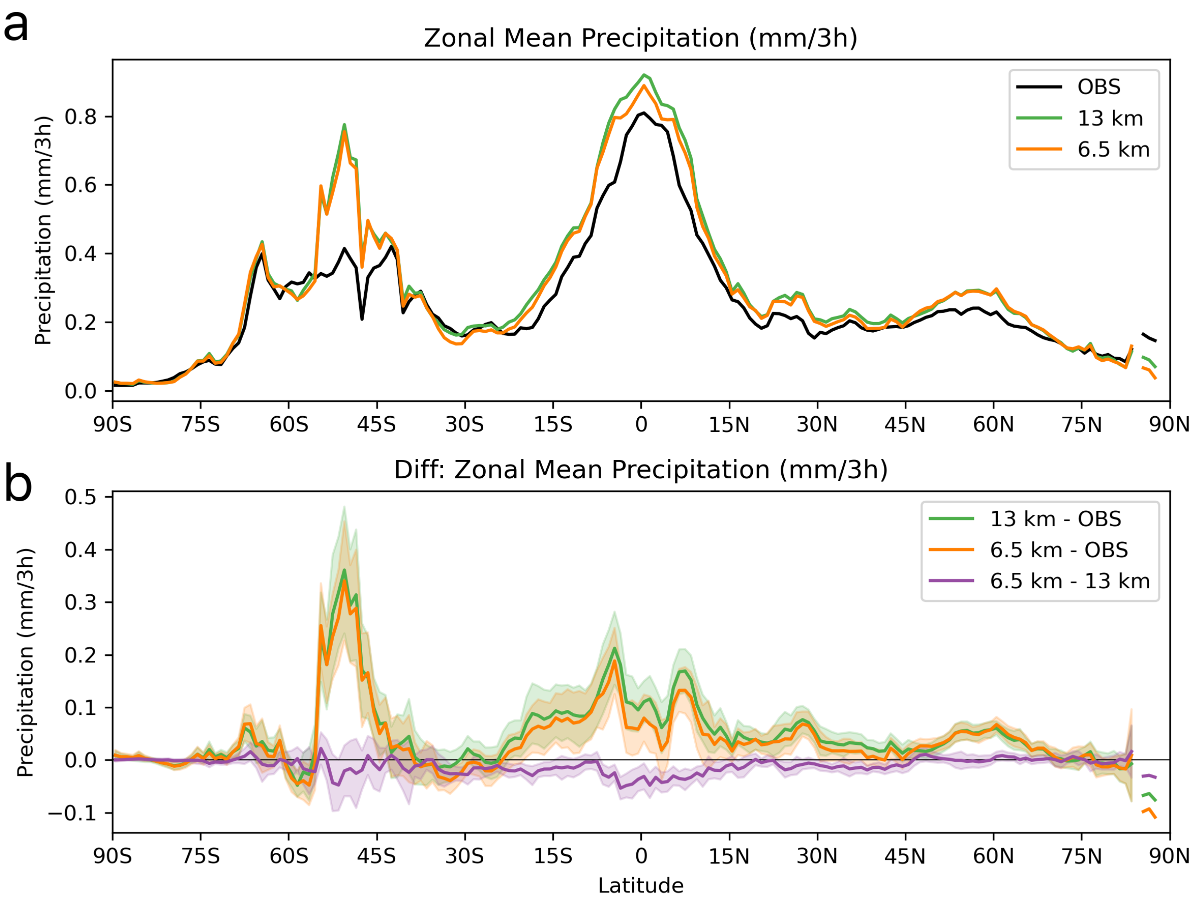


Figure 5 The same as Figure 4, but for the 13-km and 6.5-km SHiELD.

Probability Density Function (PDF)

25 km vs. 13 km

The probability density function (PDF) is useful in understanding the precipitation bias as a function of precipitation intensity. Precipitation rate spans from 0 mm/3 hours to 7 mm/3hours, which is extreme precipitation (Figure 7 x-axis). Light to medium precipitation (0-3 mm/3hours) is accurately predicted by SHiELD (Figure 4), whereas SHiELD tends to over predict heavy precipitation > 4.5 mm/h. While SHiELD over predicts precipitation rate, the observations are within the significance intervals for both resolutions, suggesting that this model bias is insignificant. Additionally, there is an insignificant difference between the two resolutions of model as the significance intervals for all precipitation intensities overlap.

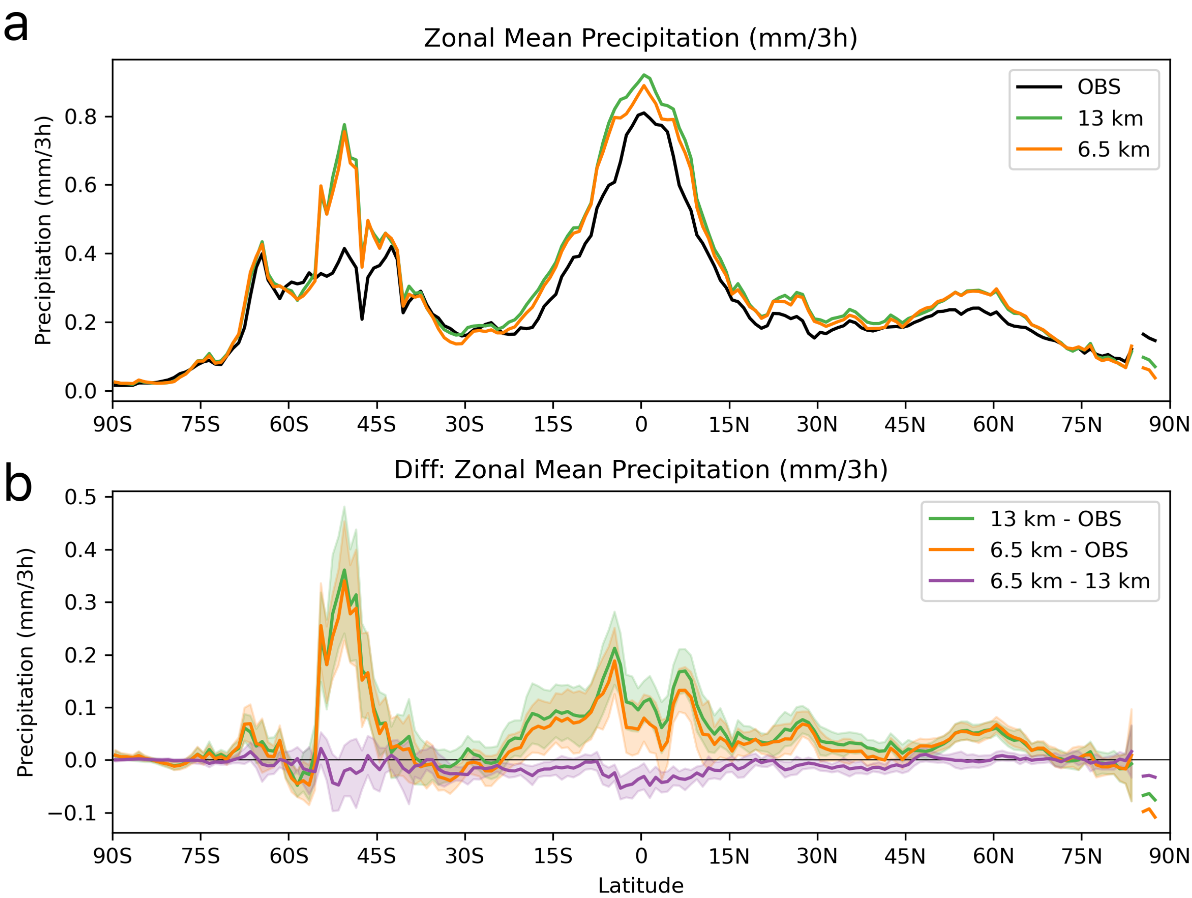


Figure 6 Land only Zonal Mean Precipitation 13 vs 6.5 km.

13 km vs. 6.5 km

Similarly, the global PDF shows that SHiELD tends to produce more heavy precipitation (> 3 mm/3h) than observation in the comparison between 13-km and 6.5-km SHiELD (Figure 8). This bias, however, is insignificant since the observation is within the significance interval. The global PDF bias is largely contributed by the ocean (Figure 10), which shows similar bias. SHiELD significantly over predicts light to medium precipitation (1-3 mm/3h) over land (Figure 9. In Figure 9 for extreme precipitation over land, there is not a large difference between the model and the observation. SHiELD tends to over predict extreme precipitation over the ocean (Figure 10); however, this difference is insignificant and should likely not be prioritized in future model revisions.

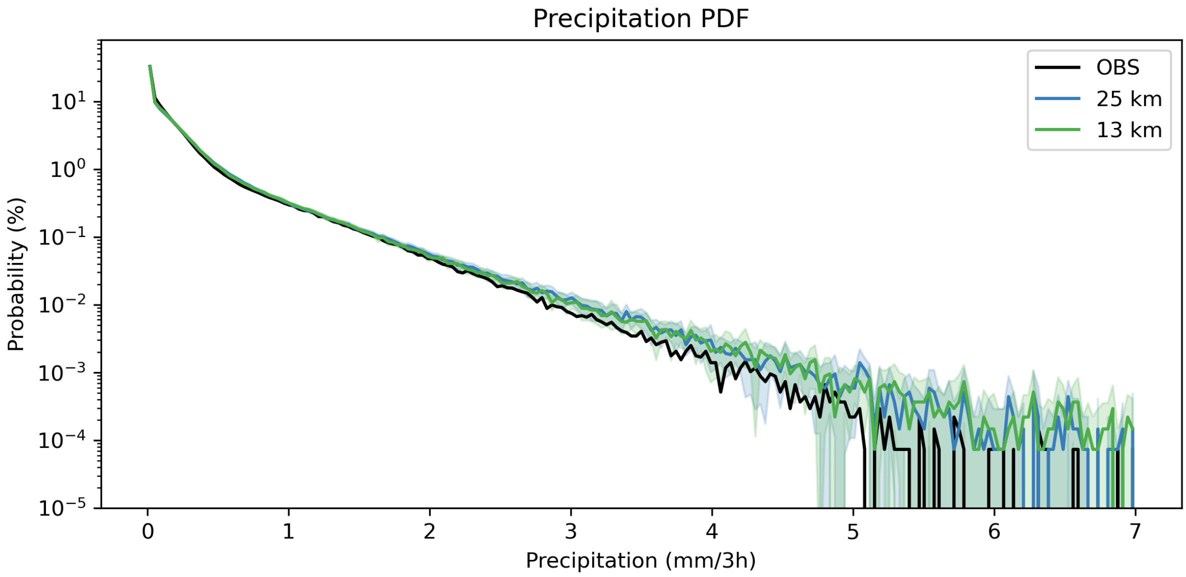


Figure 7 Precipitation probability density function (PDF; units: %) of MSWEP observation, 25-km SHiELD, and 13-km SHiELD. Shadings represent the 95% significant interval between model and observation.

Peak Precipitation Time

Precipitation tends to have two peaks, the first in the morning over the ocean and the second during the afternoon over the land. Peak precipitation time is useful for diagnosing the location and timing of maximum precipitation intensity in a 24 hour time period. The peak in precipitation over the land during the morning can be attributed to an increased relative humidity over the sea in the morning as a result of radiative cooling at night [19]. In the afternoon, peak precipitation is a result of the difference in the heat capacity of land and ocean. The differential cooling of land and ocean results in a sea breeze from the ocean that carries moisture and results in the afternoon peak precipitation over the land [19]. The Coupled Model Intercomparison Project (CMIP) found that models tend to predict precipitation too early over the ocean due to a misrepresentation of atmospheric physical properties, a result consistent with SHiELD's simulations of peak precipitation time [1].

25 km vs. 13 km

The 25 vs. 13 km peak precipitation time in SHiELD indicates that there is a morning and afternoon peak (Figure 11). The morning peak over the ocean is predicted too early,

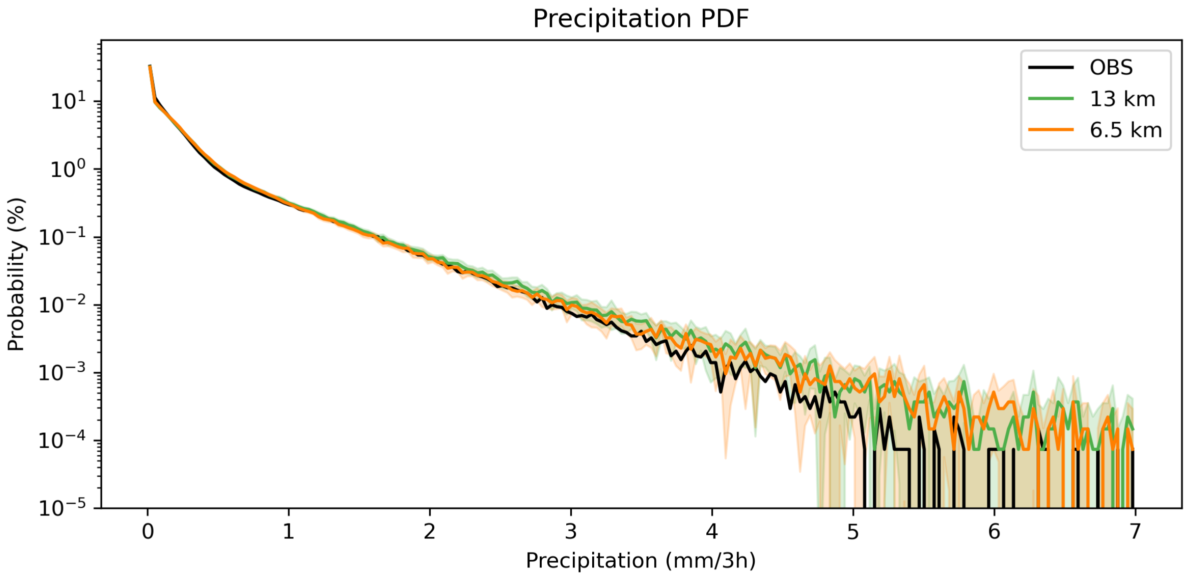


Figure 8 The same as Figure 7, but for the 13-km and 6.5-km SHiELD.

as seen with the orange color over the oceans, which is a result of bias in the SHiELD model. The afternoon peak over the land is accurately predicted with the exception of the Amazon Rainforest where the peak is predicted too early. As seen in Figure 11, there are similar biases in peak precipitation time between the resolutions. Doubling the resolution did not significantly impact the accuracy of peak precipitation time prediction as both resolutions had similar biases.

13 km vs. 6.5 km

Similarly to Figure 12, there are broadly two peak precipitation times, one in the morning and the other in the afternoon to evening. Over the majority of ocean regions, precipitation peaks in the morning (around 6-9 am); however the model predicts this peak too early (around 3-6 local time). Over land, precipitation typically peaks during the afternoon to evening (around 3-9 pm), which is generally well-represented in the SHiELD model. The one notable exception is over the Amazon Rainforest where the observed peak occurs from 3-4 pm but the model predicts the peak from noon-3 pm. Doubling the resolution for peak precipitation time did not appreciably improve the accuracy of the model.

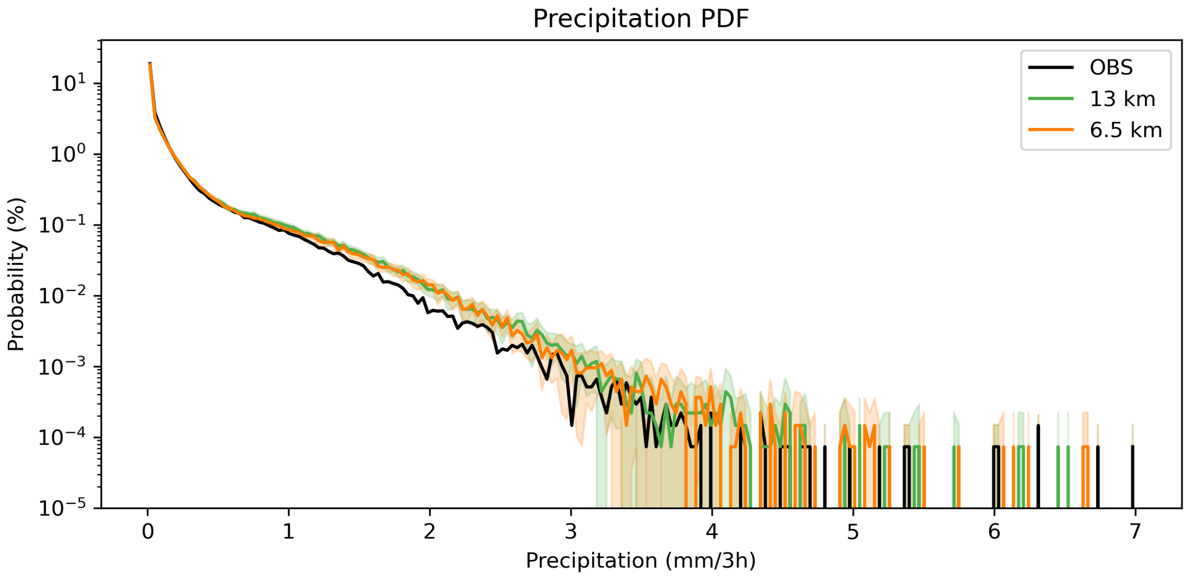


Figure 9 The same as Figure 8, but for the land.

Precipitation Diurnal Cycle

25 km vs. 13 km

The precipitation diurnal cycle describes how the intensity of precipitation changes over 24 hours. When doubling the horizontal resolution from 25 km to 13 km, there is a significant difference in the magnitude of precipitation; however in predicting the diurnal cycle both resolutions follow the same trends and biases. 13-km SHiELD has lower precipitation values and more closely resembles observations, which is consistent with Gallus (1999) which indicates that as resolution improves the magnitude of peak simulated precipitation decreases [12]. The observed morning peak over the ocean is closer in magnitude to the 13 km resolution, however the time at which it is predicted is too early in Figure 13. In the afternoon over the land, the model accurately predicts the peak. Given that the magnitude in precipitation time is more accurate in the 13 km resolution in the morning and more accurate in the afternoon using the 25 km resolution and the conclusion that both resolutions have similar biases; we can conclude that there is not a significant difference between the resolutions is insignificant.

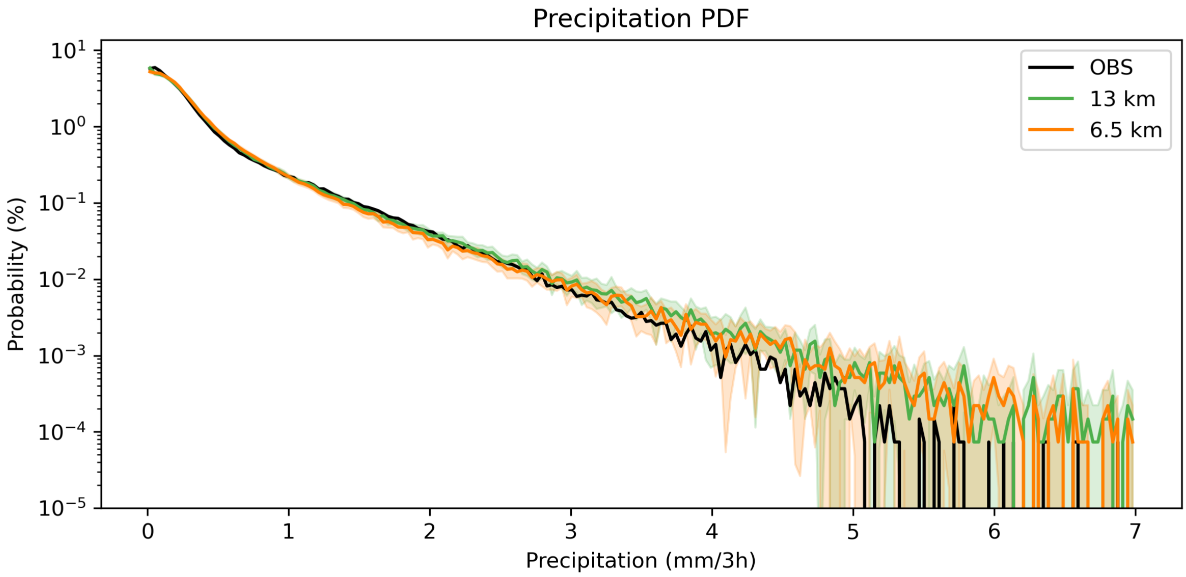


Figure 10 The same as Figure 8, but for the ocean.

13 km vs. 6.5 km

Here the diurnal cycle of precipitation is analyzed globally (Figure 14), over land (Figure 15) and over the ocean (Figure 16). In the global region, there are two peaks–one in the morning and one in the afternoon to evening; these peaks are consistent with the peaks seen in Figure 12. It is evident that peak precipitation happens in the morning over the ocean and in the afternoon to evening over the land. The model predicts the morning peaks too early (about 3-6 am) as seen in Figure 14. Additionally, the model predicts a greater magnitude of precipitation than the observed diurnal precipitation cycle (Figure 14). The significance intervals are not overlapping in Figure 14 for the morning and afternoon peaks meaning that the magnitudes are significantly different even though in this case the peaks for 13-km and 6.5-km SHiELD occur at the same time. Across all regions, the biases in the model are significant and thus an area that should be improved upon in SHiELD.

Cloud Fraction

25 km vs. 13 km

This section describes simulated cloud fraction from SHiELD compared to the CERES dataset [15]. Generally there are zonally asymmetric variations in cloud fraction. Glob-

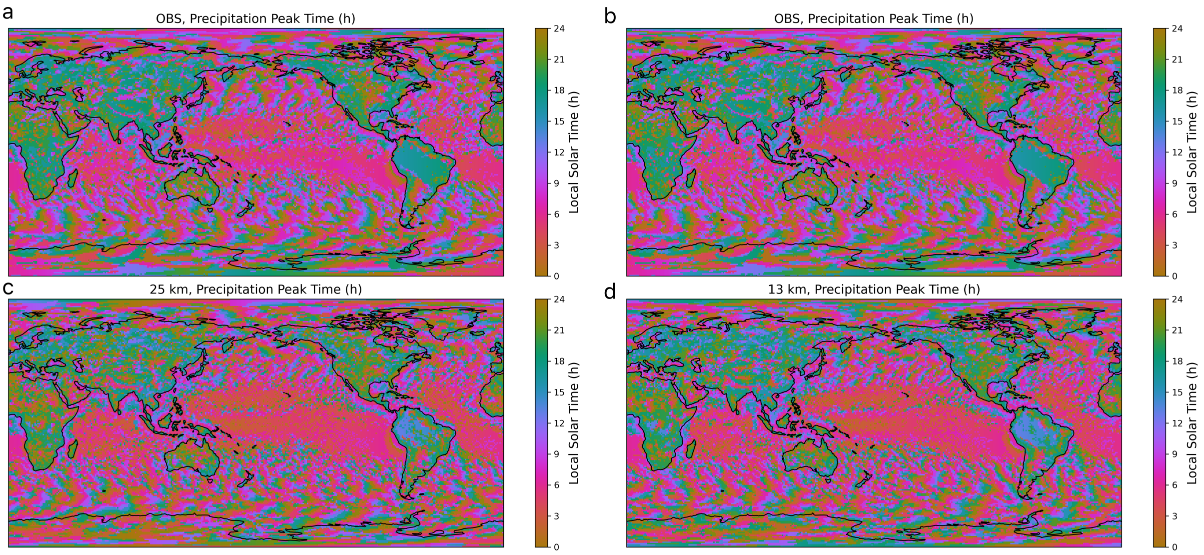


Figure 11 Peak precipitation time (units: local solar time (h)) a) MSWEP observation, b) MSWEP observation, c) 25-km SHiELD, d) 13-km SHiELD.

ally, there is a strong negative bias with the exception of Antarctica and over the tropics (Figure 17). For example, there is a positive bias for the ITCZ, SPCZ, and South America, suggesting this could be a good target for further model development. When comparing the two resolutions of the model, the 13-km SHiELD is able to reduce the positive cloud fraction bias over the tropics, as seen with the negative bias (Figure 17b). Furthermore, the 13-km SHiELD decreased the amount of CLDFR bias globally, as seen with the positive bias in (Figure 17b). The positive bias in cloud fraction is consistent with the positive bias in geographic mean precipitation over the tropics (Figure 2).

Land has lower cloud cover compared to the ocean because there is not as much water vapor in the atmosphere (Figure 17). The high cloud fraction over the tropics is due to the atmospheric updraft which brings moisture and condensation up the atmosphere, forming clouds. At 30°N, there is a downdraft, moving the dry air down from the upper to lower atmosphere. This downdraft results in a lower cloud fraction over desert areas. Land and sea distributions result in zonal variation in cloud fraction. The cloud fraction in Figure 17 is consistent with the updraft/downdraft of the atmosphere.

13 km vs. 6.5 km

As seen in Figure 18e-f, the model significantly underpredicts cloud fraction for most of the globe, with the exception of Antarctica, where the model significantly overpredicts cloud fraction. When comparing the biases for the 6.5-km and 13-km SHiELD, the higher

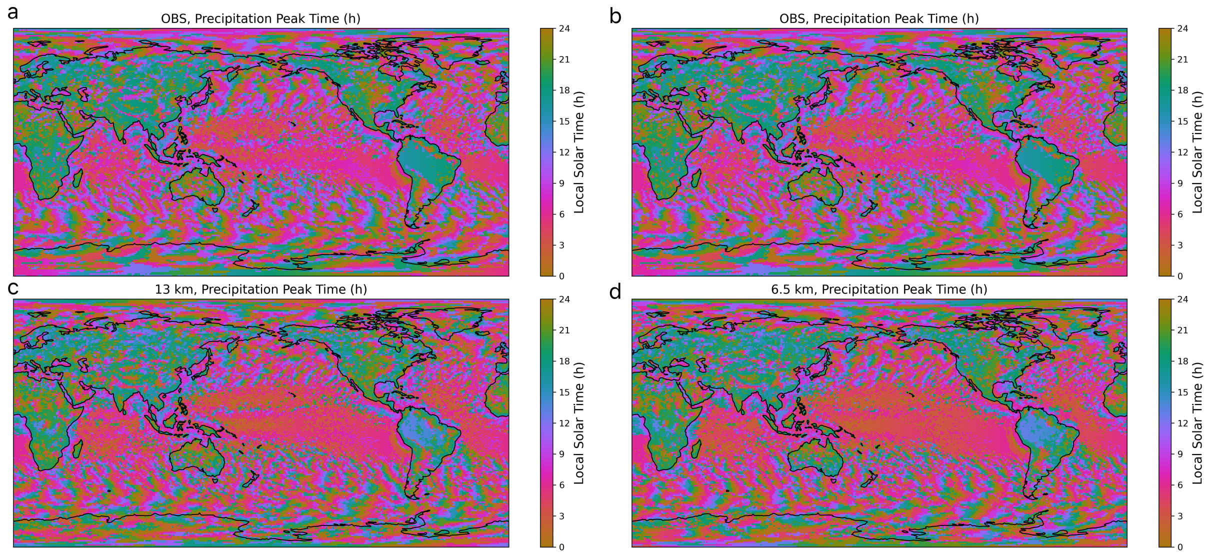


Figure 12 The same as Figure 11, but for the 13-km and 6.5-km SHiELD.

resolution actually does a better job at predicting cloud fraction. The ability to better predict cloud fraction by doubling the resolution is evident in the difference between the 6.5 km and 13 km resolution cloud fraction (Figure 18b), where the bias is positive over the ITCZ, SPCZ and tropics rather than negative. Furthermore, Figure 18b shows that globally the negative bias is significantly reduced suggesting an improvement in cloud fraction prediction at higher resolution. While the 6.5-km SHiELD is more accurate in predicting cloud fraction than the coarser resolutions, there is still much work to do to improve the accuracy of the prediction.

Liquid Water Path

25 km vs. 13 km

In Figure 19, the Liquid Water Path (LWP) is concentrated in the mid-latitude for the Figure 19c-d, but the model tends to over-predict the LWP at the subtropics, Maritime Continent, and the Amazon as seen in the bias plots with the positive bias centered around the tropics area (convective area).

This section focuses on LWP, which represents low-level clouds or liquid water clouds. By analyzing both liquid water path and ice water path, we are able to analyze SHiELD's prediction of different phases of water in clouds. The CERES dataset is able to easily identify IWP and LWP, but it is difficult to distinguish between rain and graupel, which

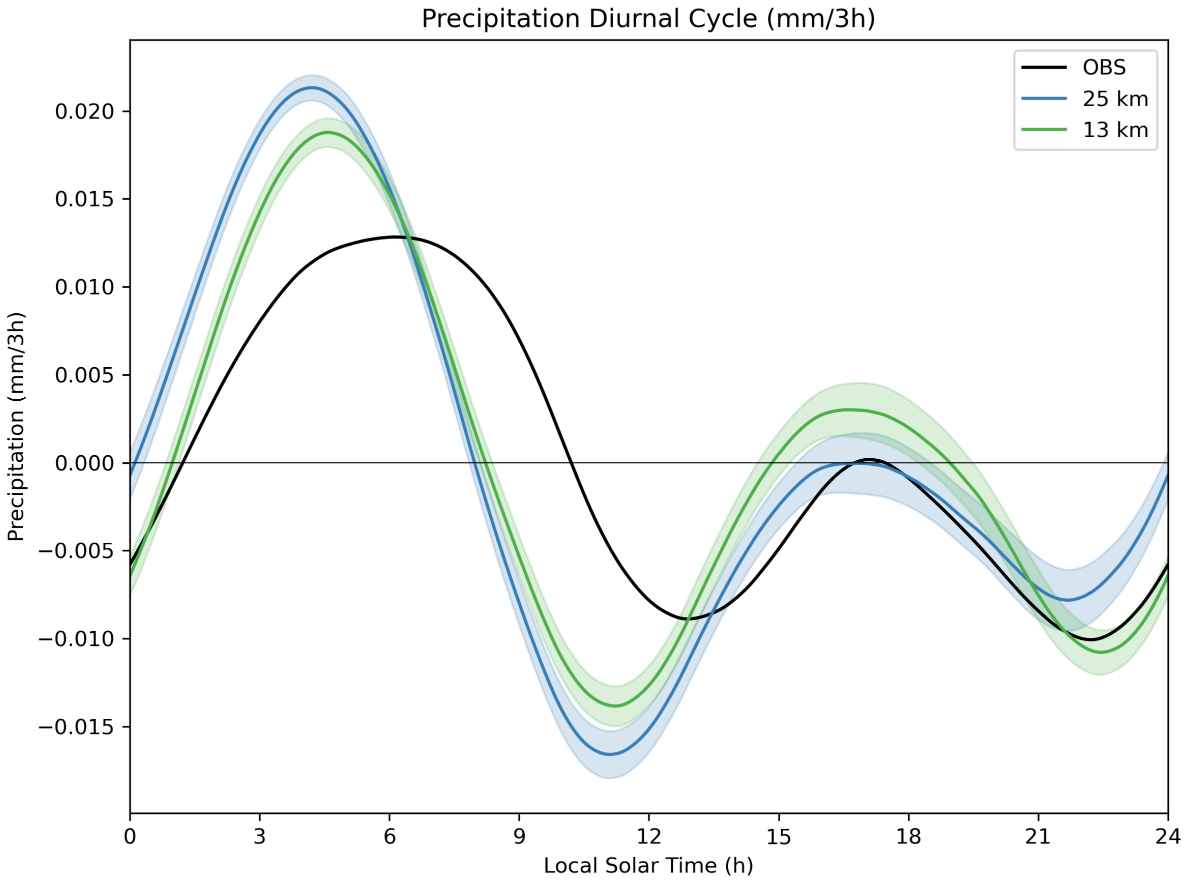


Figure 13 Precipitation diurnal cycle (units: mm/3h) of MSWEP observation, 25-km SHiELD, and 13-km SHiELD.

can result in biases in the dataset which in turn results in errors in SHiELD. In Figure 19e-f, the modeled liquid water path is significantly different from the observations. Generally, the SHiELD model shares many of the same biases, and there is not a significant difference in prediction accuracy between the two resolutions of SHiELD. One notable difference between the two resolutions of SHiELD is over Antarctica (Figure 19b). High latitude liquid water path in the polar areas are actually questionable for this dataset, so we normally do not compare the liquid cloud fraction in the polar areas.

13 km vs. 6.5 km

The geographic distribution and biases of LWP are similar for 13 km and 6.5 km so there is not a large difference in terms of the accuracy of liquid cloud prediction (Figure 20c-d). There is a lot of liquid water at high latitudes in the observation, and this is a result of bias in the dataset over the high latitudes (Figure 20c-d). When comparing the mid-latitudes, the observation has lower values for cloud water, but the

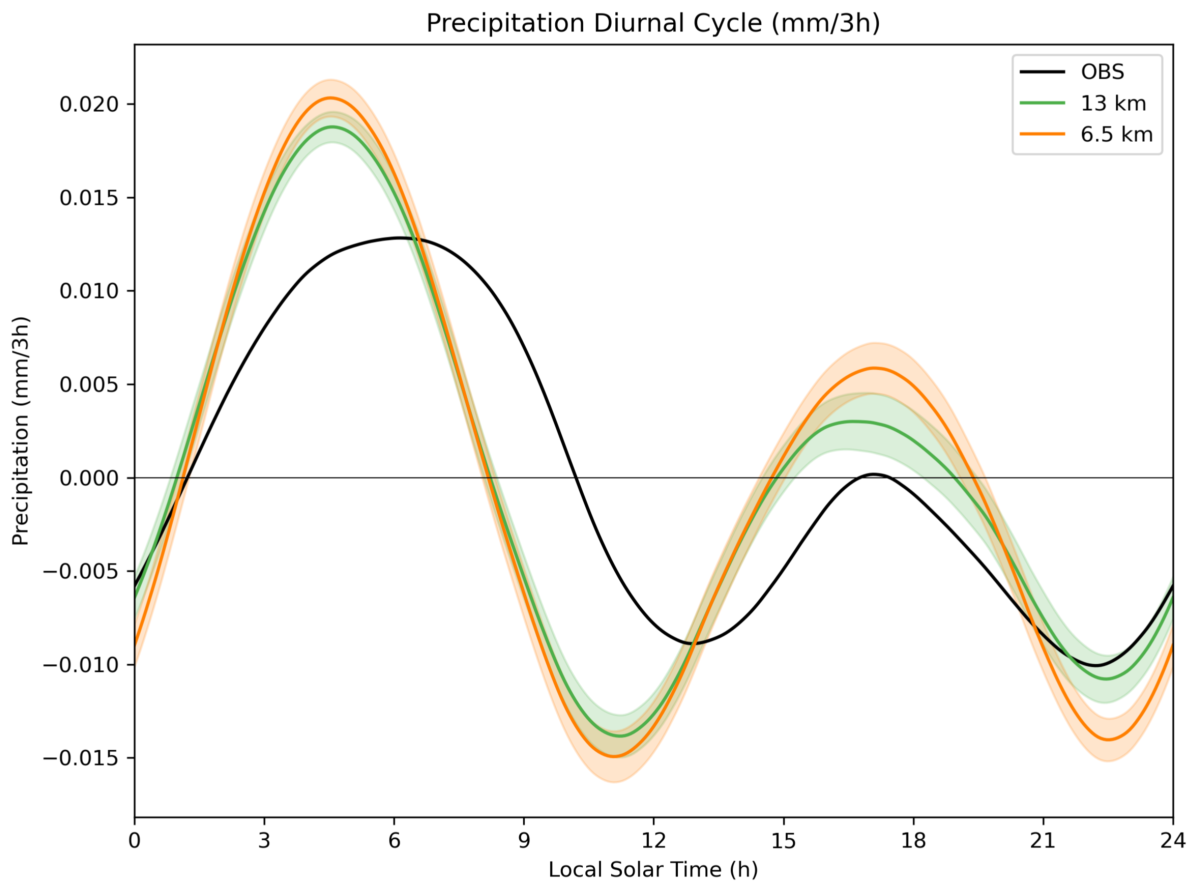


Figure 14 The same as Figure 13, but for the 13-km and 6.5-km SHiELD.

model over-predicts liquid water, resulting in a positive bias (Figure 20e-f). It was found that SHiELD tends to over-predict low-level clouds in the maritime continent area and the Amazon rainforest and that there is not a significant difference between the resolutions for liquid clouds (Figure 20b,e-f).

Ice Water Path

25 km vs. 13 km

This section will focus on the prediction of ice water path (IWP). There are significant differences between 25-km and 13-km SHiELD (Figure 21b). There is generally a negative bias of IWP in SHiELD, especially over Antarctica and the ITCZ. In the Southern Ocean and mid-latitudes, there is a positive bias in IWP. The 13-km SHiELD is able to enhance the prediction globally as seen in Figure 21b where there is a slight positive bias globally indicating that the global negative bias is reduced. Furthermore, the 13 km resolution

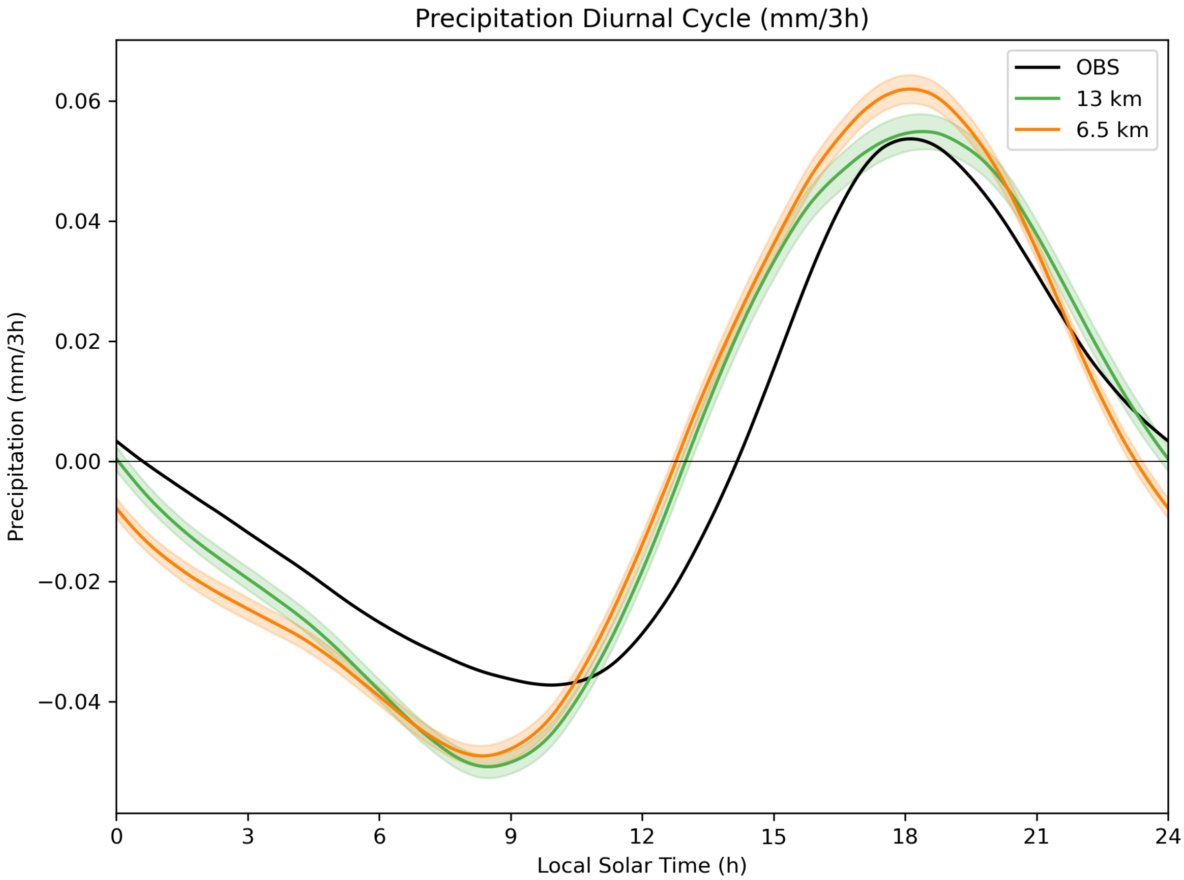


Figure 15 The same as Figure 14, but for land.

is able to significantly reduce positive bias in the Southern Ocean as seen in Figure 21b. The prediction of the ITCZ is improved in 13 km-SHiELD as seen in Figure 21b. These significant improvements suggest that increasing the horizontal resolution can be a method utilized to improve ice cloud prediction.

13 km vs. 6.5 km

Similar to the comparison between 25-km SHiELD and 13-km SHiELD, Figure 22c-d suggests that the model tends to underpredict high level clouds in the maritime continent and over predict these clouds in the mid-latitudes, especially over the ocean. While there were significant differences in IWP when doubling the horizontal resolution from 25 km to 13 km, there is not a significant difference between 13-km and 6.5-km SHiELD (Figure 22b). As seen in the bias panels of Figure 22e-f, the IWP biases in SHiELD are significant globally and thus an area that can be improved upon in model revisions.

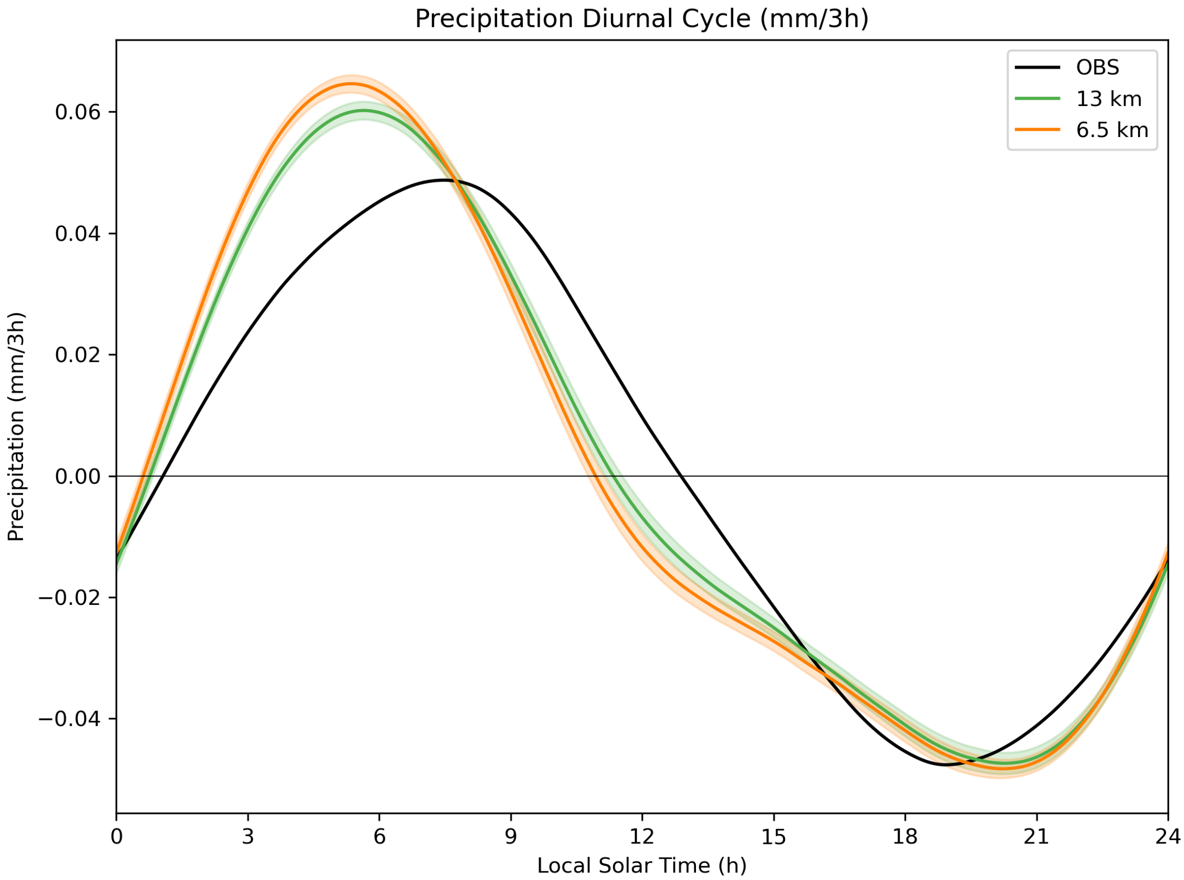


Figure 16 The same as Figure 14, but for the ocean.

Prediction Error Growth

25 km vs. 13 km

This section analyzed the error growth of precipitation prediction. The model's prediction of precipitation is significantly overestimated, as seen with the observation being outside of the significance intervals of the model (Figure 23a). The 25-km SHiELD has a stronger positive bias as seen with the negative difference line (Figure 23b). The root mean square error (RMSE) is higher for 25-km SHiELD, but as the forecast gets closer to 10 days, the difference becomes insignificant (Figure 23c). While biases were identified, this suggests that doubling the resolution does not necessarily mean improving prediction accuracy, and thus, another method should be implemented to reduce biases in SHiElD (Figure 23b).

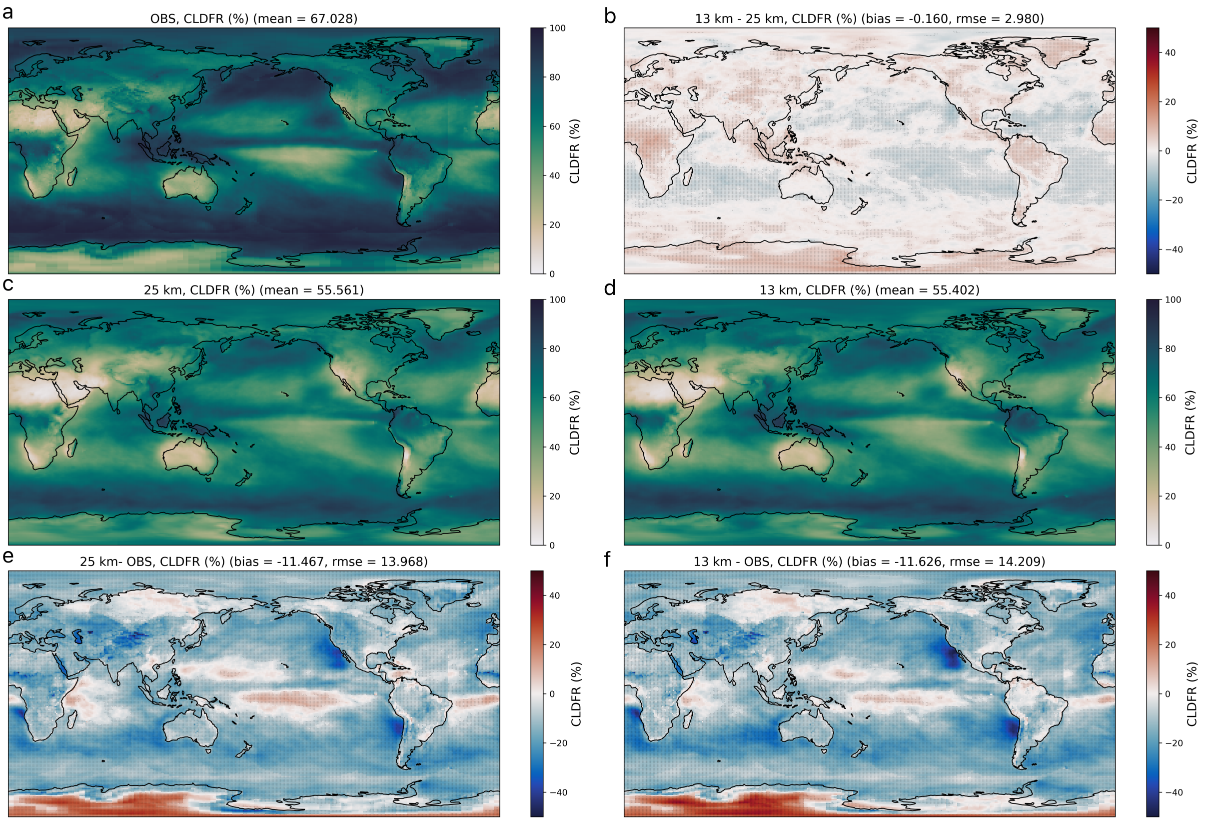


Figure 17 Geographic cloud fraction (CLDFR; units: %) of a) CERES cloud fraction observation, b) difference between 13-km and 25-km SHiELD, c) 25-km SHiELD, d) 13-km SHiELD, e) difference between 25-km SHiELD and CERES observation, f) difference between 13-km SHiELD and CERES. 95% significantly different areas are marked dotted.

13 km vs. 6.5 km

The 6.5-km SHiELD mean precipitation is closer to the observation and its bias is closer to zero, therefore its precipitation prediction is more accurate (Figure 24a). (Figure 24b) suggests that the difference between 6.5-km and 13-km SHiELD is significant since the significance interval is smaller than their difference, indicating that doubling the model resolution can significantly improve the accuracy of SHiELD. While the RMSE follows similar trends for both resolutions, the difference is significant as the significance interval is larger than their difference (Figure 24c). It is interesting to note that the bias of the 6.5-km SHiELD (Figure 24b)is smaller, but its RMSE is larger than the 13 km (Figure 24c). This suggests that high-resolution models can reduce the global mean precipitation but have difficulty predicting the precipitation pattern or distribution.

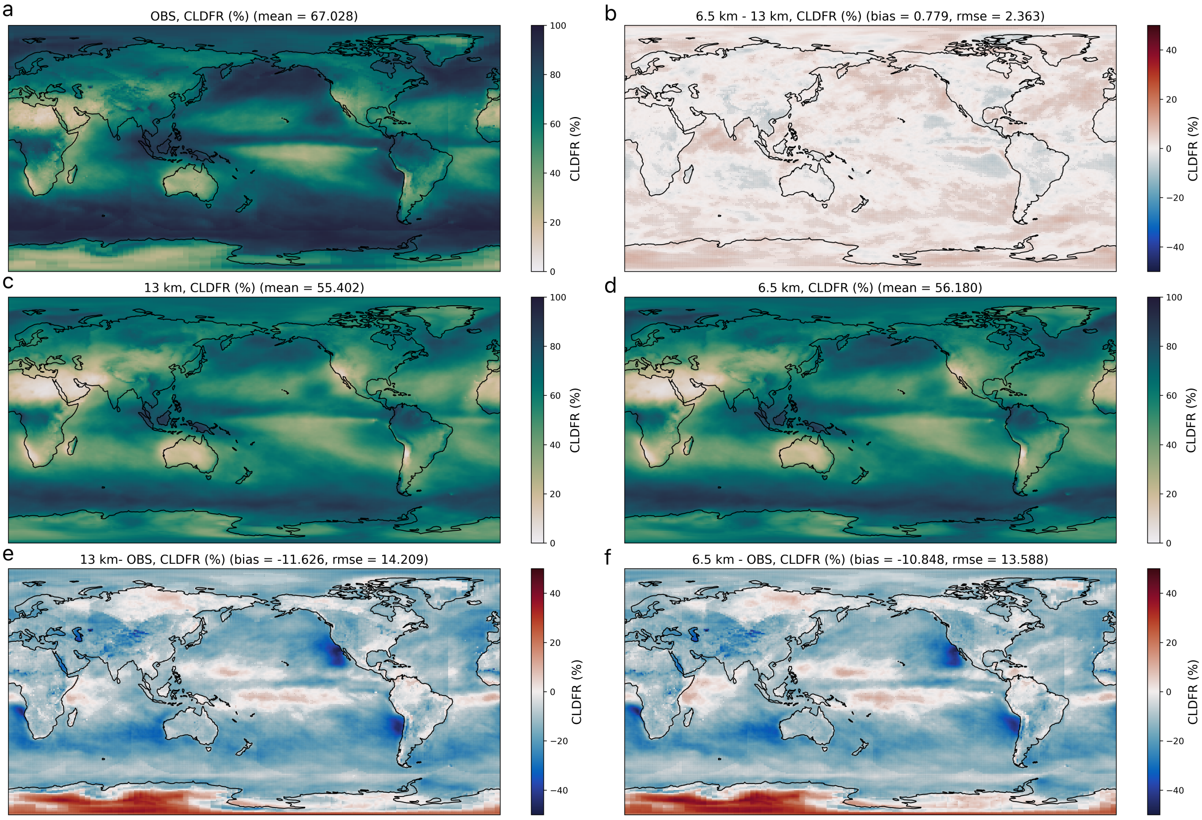


Figure 18 The same as Figure 17, but for the 13-km and 6.5-km SHiELD.

Quantitative Precipitation Forecasts (QPF)

25 km vs. 13 km

QPF (reference: Model Evaluation Tools (MET) Ch.33 Appendix C Verification Measures) is a statistic that quantitatively indicates the skill of the forecast. The Equitable Threat Score (ETS) is based on the Critical Success Index (CSI) and a larger ETS means that the model predicts the observation more accurately. BIAS is the ratio of the number of forecasts of an event to the total number of events and the model is more accurate when the BIAS is close to 1. Compared with the 25-km SHiELD, the 13-km SHiELD has a higher ETS for all days and thresholds as seen with the positive difference. This indicates that the 13-km SHiELD does have the potential to improve prediction in some areas. Both resolutions were found to have similar biases with the exception of the negative bias for less than 2 mm/3h indicating that the 25-km SHiELD's bias is great for those thresholds. The Area Under the Curve (AUC) is higher for 13-km SHiELD, further indicating an improvement in prediction. The Fractions Skill Score is high for both resolu-

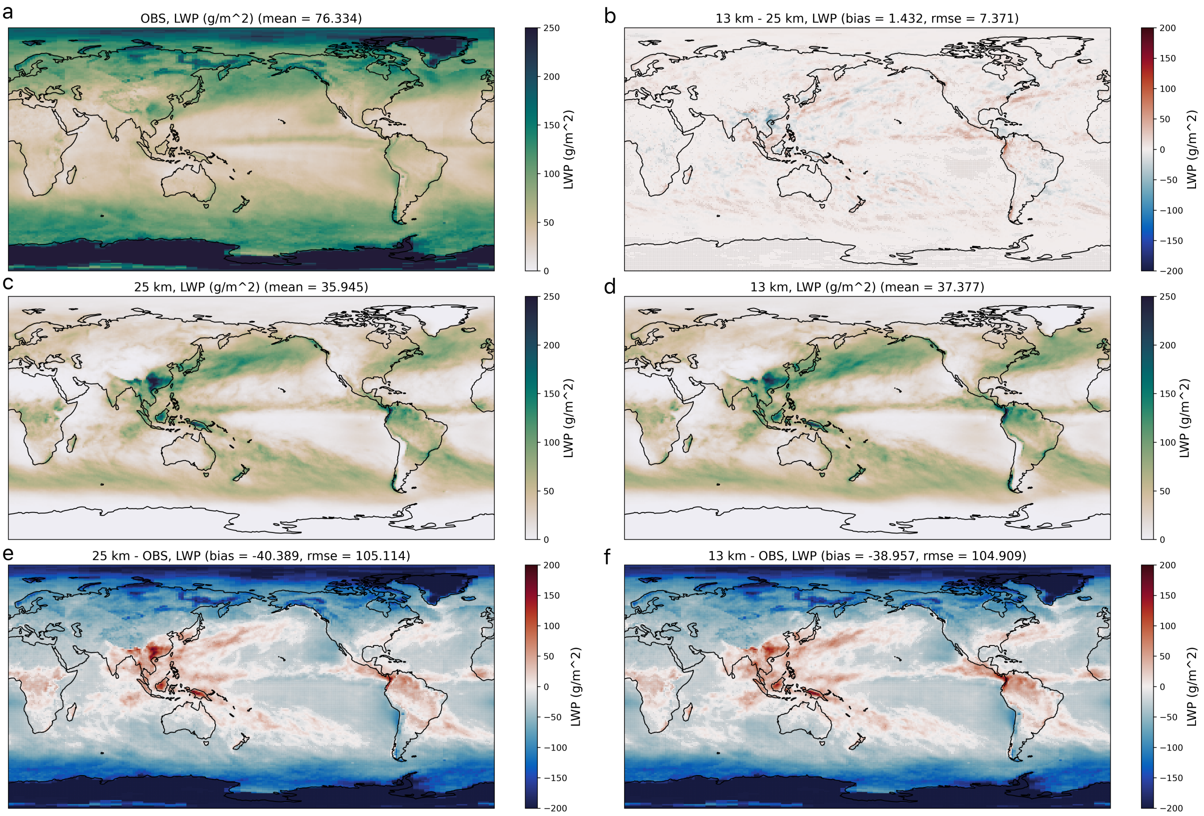


Figure 19 Geographic liquid water path (LWP; units: g/m^2) of a) CERES LWP observation, b) difference between 13-km and 25-km SHiELD, c) 25-km SHiELD, d) 13-km SHiELD, e) difference between 25-km SHiELD and CERES observation, f) difference between 13-km SHiELD and CERES. 95% significantly different areas are marked dotted.

tions indicating that the SHiELD generally accurately predicts clouds and precipitation. The Fractions Skill Score is significantly higher in the 13-km SHiELD than the 25-km SHiELD which indicates an improvement of prediction when doubling the resolution.

13 km vs. 6.5 km

This section found that the precipitation skill is higher for light precipitation and short forecast time while lower for heavy precipitation and long forecast time regarding the ETS. However, regarding the BIAS, the model has the highest forecast skill at light precipitation and medium precipitation and there is little change in 10-day forecast. There are some significant differences between the 6.5-km and 13-km SHiELD. The skill is significantly higher for light to medium precipitation and slightly lower for extreme precipitation in the 6.5-km SHiELD. The 6.5-km SHiELD was found to have a high bias for light rain but low bias for heavy rain.

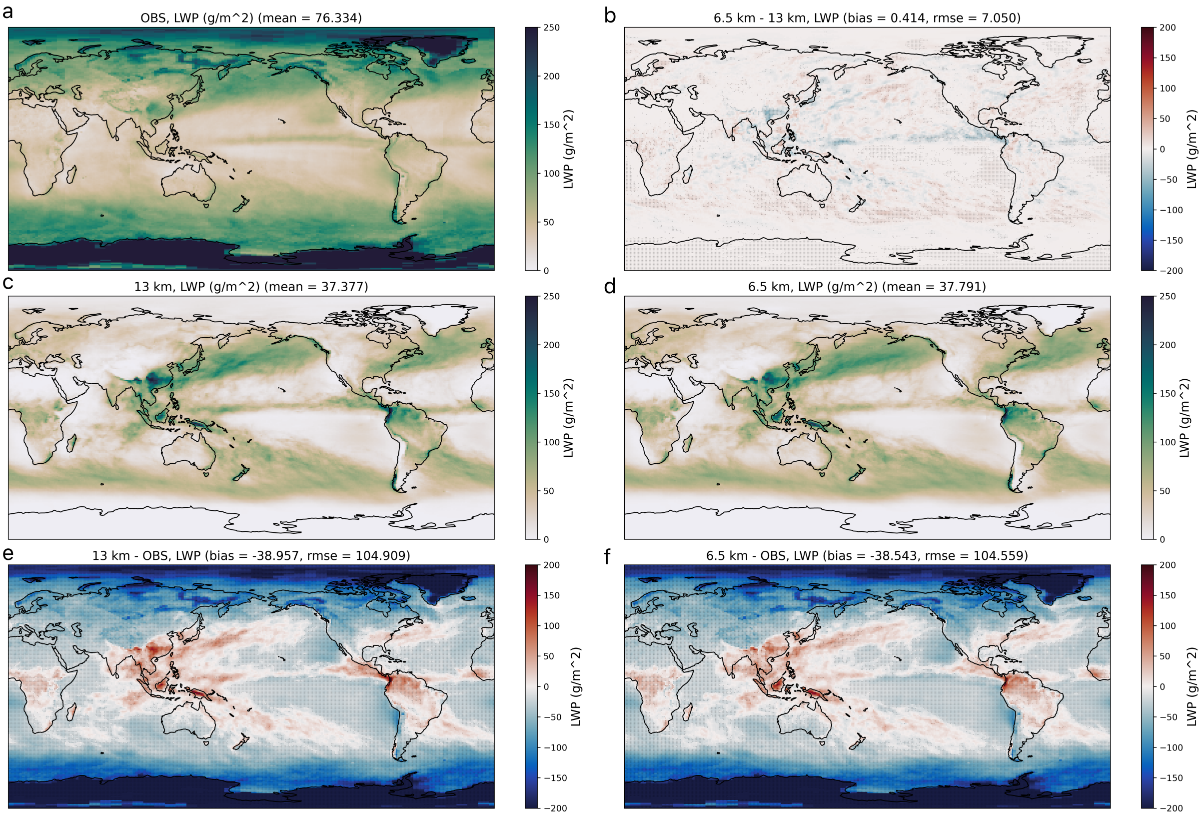


Figure 20 The same as Figure 19, but for the 13-km and 6.5-km SHiELD.

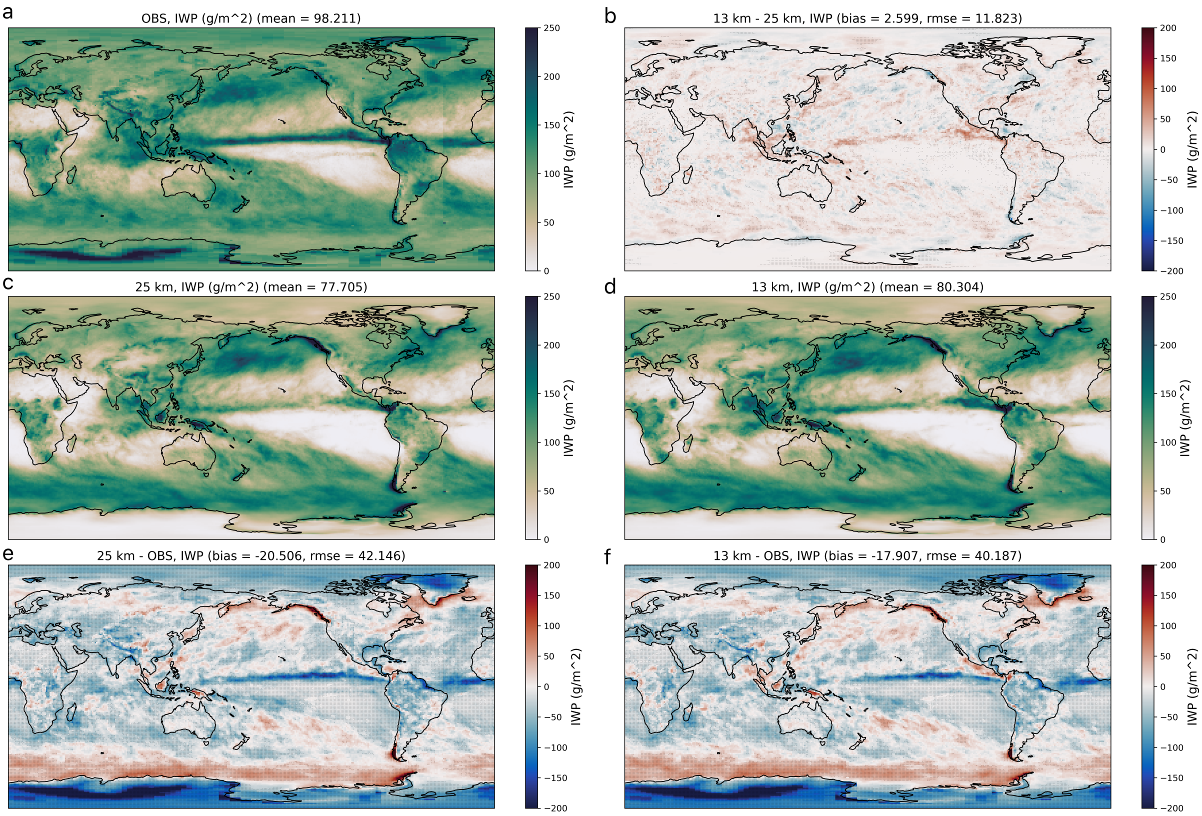


Figure 21 Geographic ice water path (IWP; units: g/m^2) of a) CERES IWP observation, b) difference between 13-km and 25-km SHiELD, c) 25-km SHiELD, d) 13-km SHiELD, e) difference between 25-km SHiELD and CERES observation, f) difference between 13-km SHiELD and CERES. 95% significantly different areas are marked dotted.

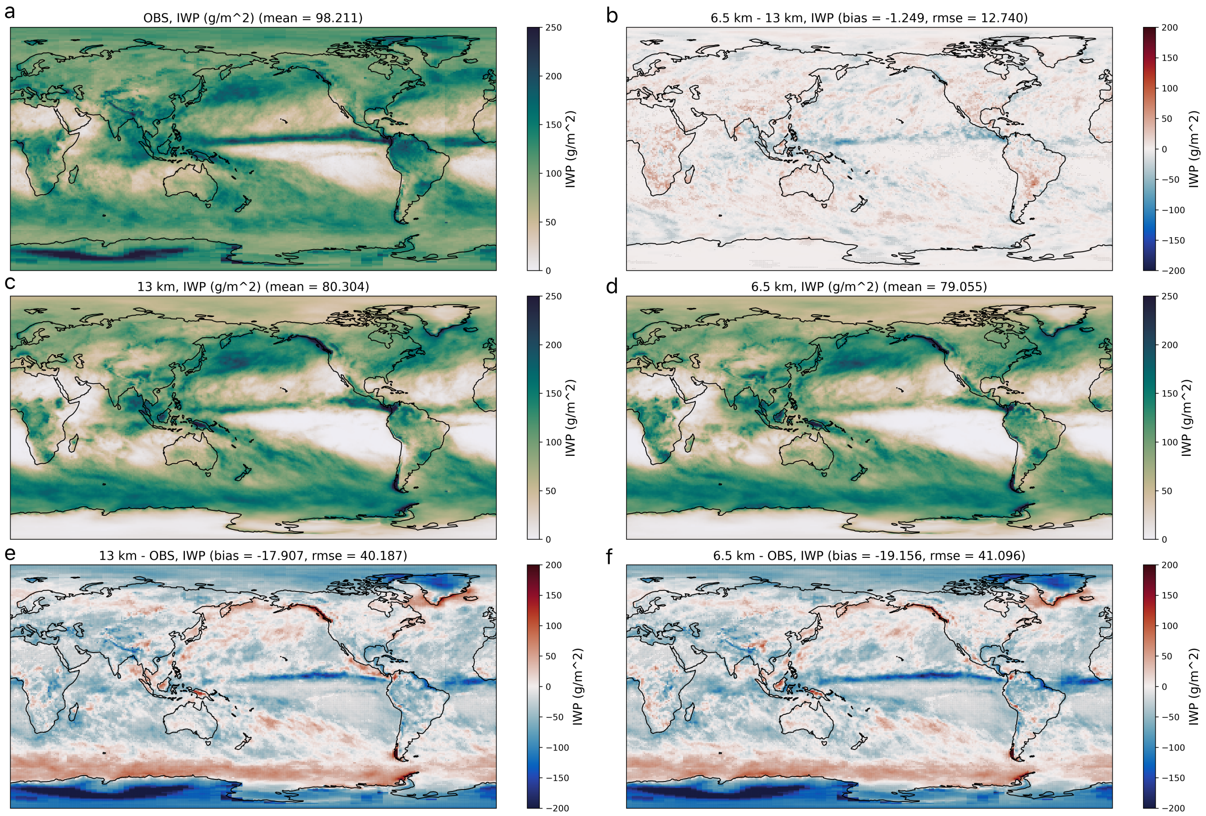


Figure 22 The same as Figure 20, but for ice water path (IWP).

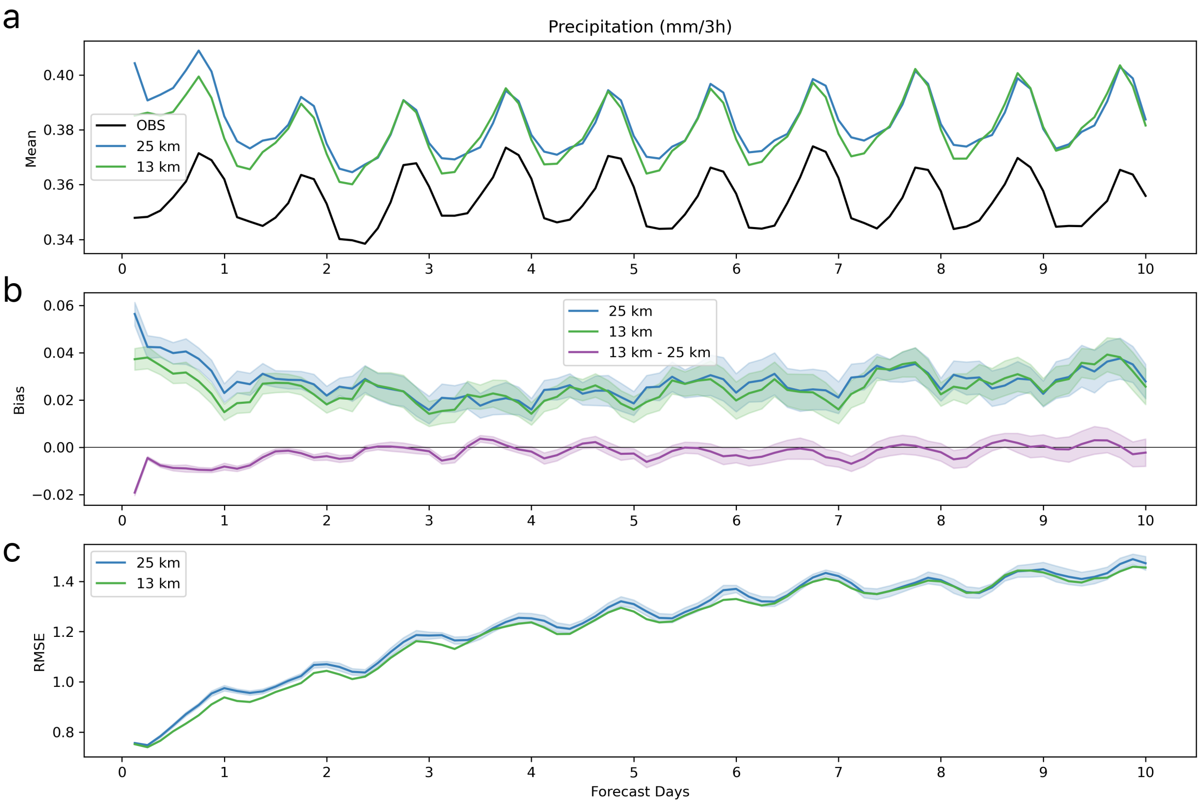


Figure 23 a) 10-day evolution of global mean precipitation of MSWEP observation (black), 25-km SHiELD (blue), and 13-km SHiELD (green). b) 10-day evolution of global bias of 25-km SHiELD (blue) and 13-km SHiELD (green), compared with MSWEP observation (black). Difference between 13-km SHiELD and 25-km SHiELD (purple). Shading areas represents the 95% significant intervals. c) 10-day evolution of global RMSE of 25-km SHiELD (blue) and 13-km SHiELD (green), with respect to MSWEP observation (black). Shading areas represents the 95% significant intervals.

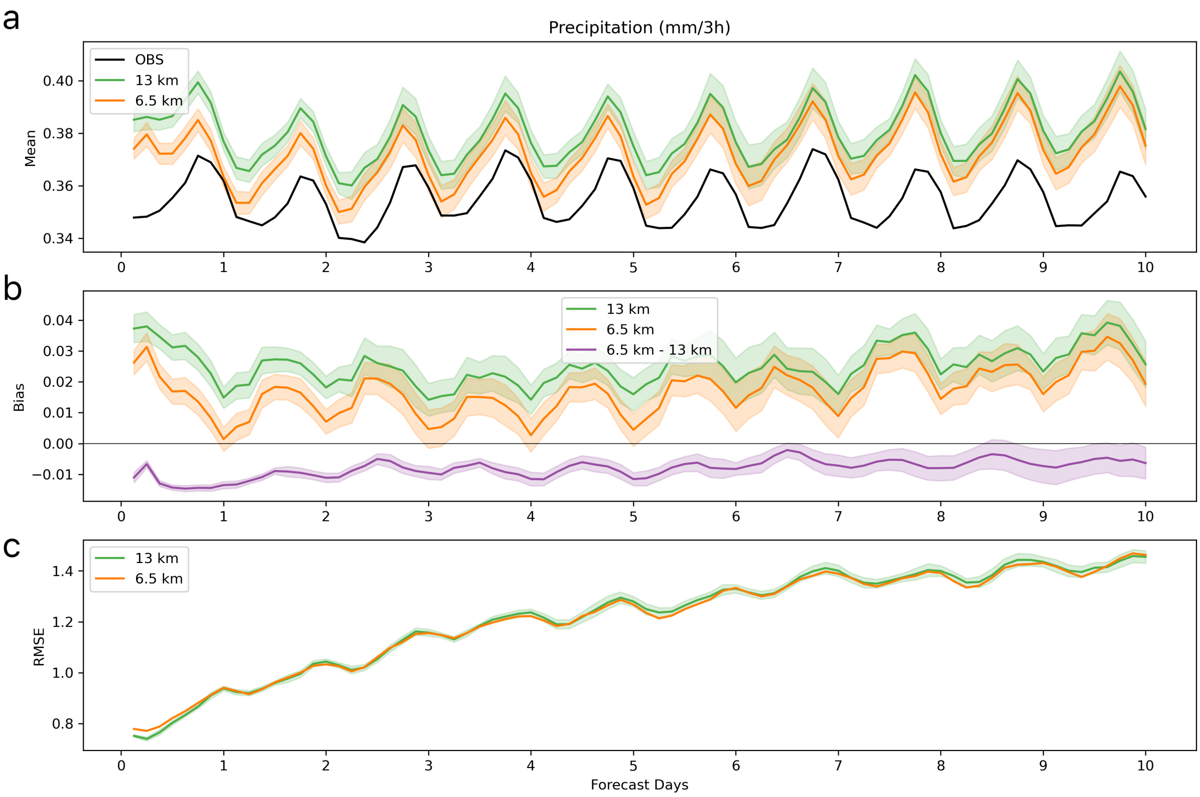
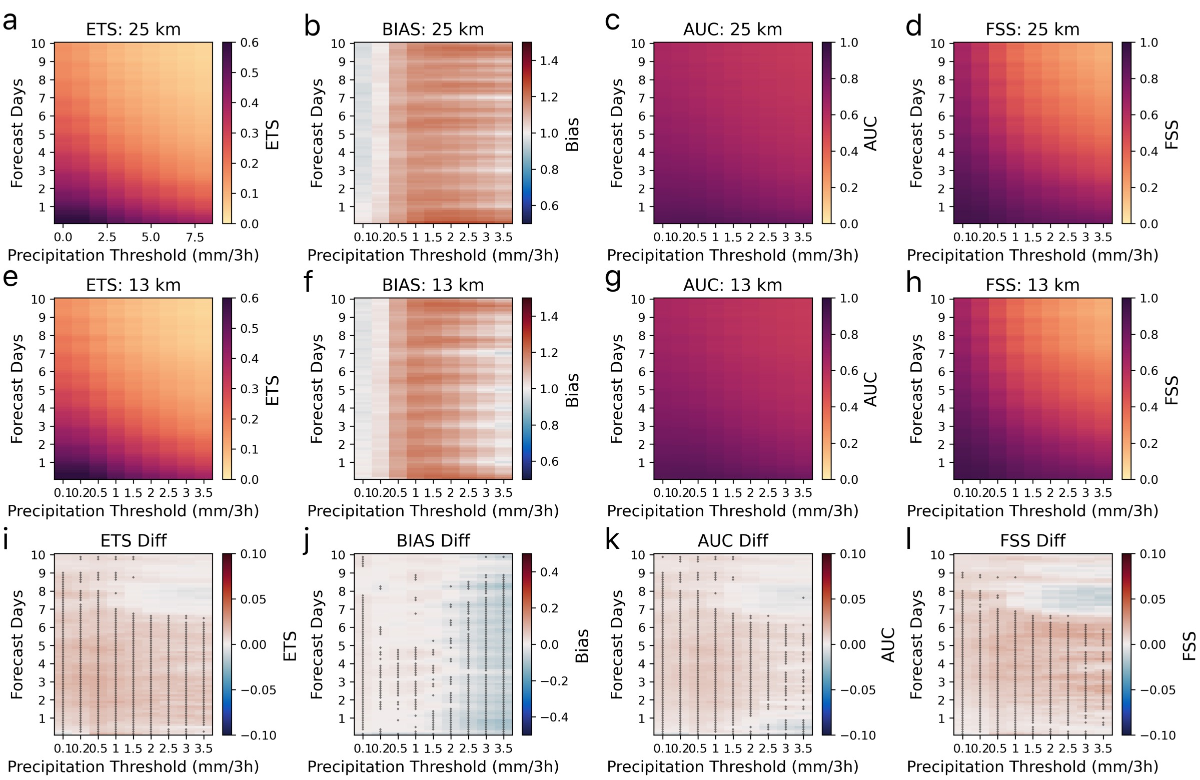


Figure 24 The same as Figure 23, but for the 13-km and 6.5-km SHiELD.



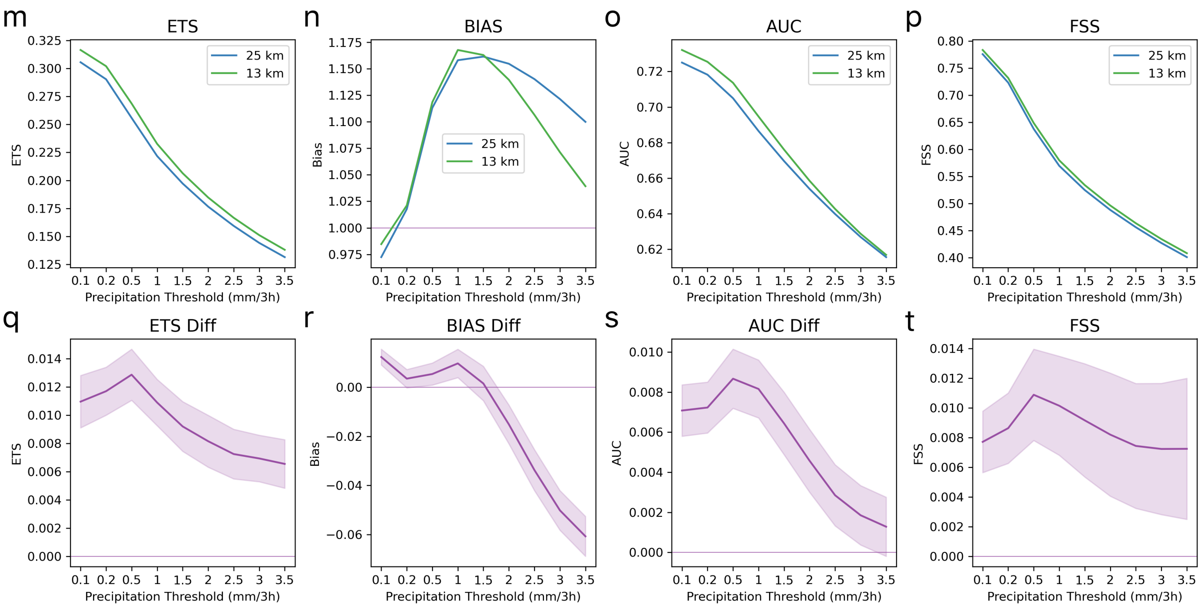
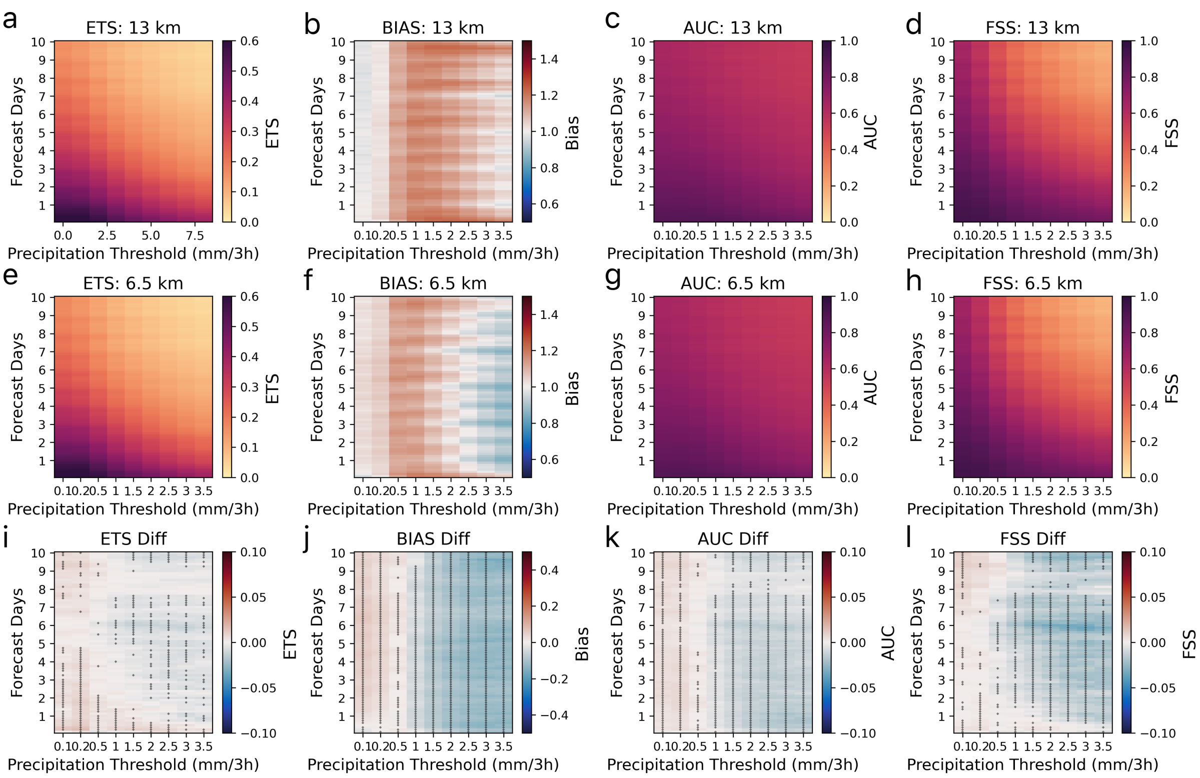


Figure 25 Quantitative Precipitation Forecasts. a) 25 km equitable threat score (ETS), b) 25 km bias, c) 25 km area under the receiving operating characteristic (ROC) curve, d) 25 km fractions skill score, e) 13 km equitable threat score (ETS), f) 13 km bias, g) 13 km area under the receiving operating characteristic (ROC) curve, h) 13 km fractions skill score, i) difference in equitable threat score (ETS), j) difference in bias, k) difference in area under the receiving operating characteristic (ROC) curve, l) difference in fractions skill score, m) 25 vs 13 km Equitable Threat Score (ETS), n) 25 vs 13 km bias, o) 25 vs 13 km Area Under the Curve, p) 25 vs 13 km fractions skill score, q) difference in equitable threat score, r) difference in bias, s) difference in area under the curve, t) difference in fractions skill score.



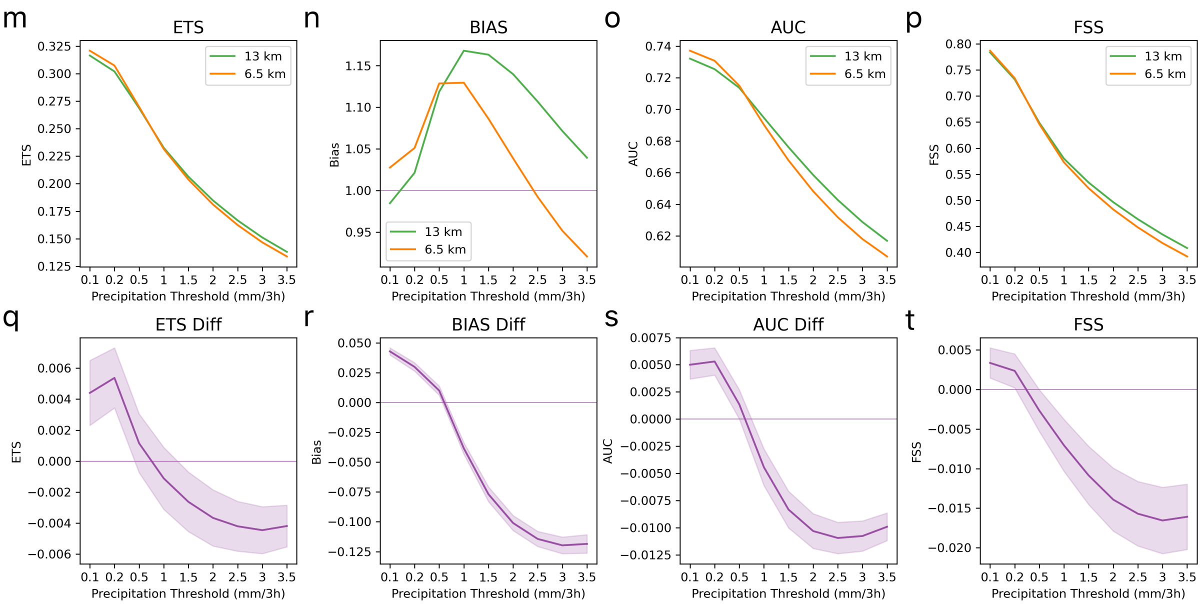


Figure 26 The same as Figure 25, but for the 13-km and 6.5-km SHiELD.

Discussion and Conclusions

Cumulus parameterization and how cloud droplets phase change to precipitation has been found to strongly influence the diurnal cycle in models and could be used with varying resolution to optimize model revisions [1]. Furthermore, entrainment and detrainment rates affecting cumulus mixing have been shown to impact ocean precipitation rates in the diurnal cycle but with little significance in geographic mean precipitation; thus, this is another factor that must be accounted for in SHiELD [1]. Low cloud depth and amount has proven to be significant in liquid water path, ice water path, cloud fraction and diurnal cycle and the correlation between cloud depth and diurnal cycle is another area that can be addressed to improve prediction [1].

Previous studies have found that the influence of varying resolution on precipitation forecasts and QPF is highly dependent on the choice of convective parameterization; thus, further research can address how increasing resolution in tandem with convective parameterization or other methods can advance prediction.

GFDL SHiELD version 2023 accurately predicts clouds and precipitation; however, many variables exhibit potential for improvement such as liquid water path, ice water path, cloud fraction, quantitative prediction forecasts, and zonal mean precipitation. Terrain precipitation can be improved upon in zonal mean precipitation forecasts. Cloud fraction and geographic mean precipitation can be more accurately predicted with higher resolutions, but there is still much to be done to increase the skill of SHiELD. Terrain bias and convective parameterization were found to have significant impacts on zonal mean precipitation and diurnal cycle respectively, and can thus be improved upon in the future.

While horizontal resolution is one method to increase prediction accuracy, it should be used in tandem with other methods to improve the forecast capabilities. Enhancing the prediction accuracy can be accomplished by increasing resolution; however, it should not be seen as the sole method of improving forecasts. Increasing horizontal resolution is by no means perfect in its ability to heighten the skill and should be paired with convective parametrization to improve forecasts. As spatial resolution is increased in many climate and weather systems, it is important to understand how simulations are impacted and how to assess significant biases in the prediction systems.

In this research, we assume that changing the horizontal resolution in SHiELD occurred in isolation and that SHiELD did not undergo other model revisions while the resolution changed. This assumption allowed us to use the 25 km, 13 km and 6.5 km resolutions as controlled variables that could then be compared to assess how changing only the horizontal resolution impacts prediction. Some uncertainties lie in how varying resolution compares to other methods used to improve prediction accuracy.

Various resolutions can be used for different purposes; for example, T-SHiELD has a 3 km resolution for studying tropical convection and hurricane prediction [6]. One resolution may significantly raise the accuracy for a certain model that may not impact the accuracy of another Unified Forecast System; thus, while the 6.5 km resolution was able to advance prediction accuracy for some aspects of SHiELD, it may not have similar results in other models. This research is not intended to identify a singular resolution that should be applied to all numerical weather prediction; rather, it is meant to propose a method for improving forecasting and how it may be applied to other models.

Here we vary horizontal resolution in order to assess how resolution impacts the skill of numerical weather prediction and if it may be used as a tool to enhance prediction. To uncover and understand the biases in SHiELD, significance tests are implemented to evaluate if differences between model resolutions and observations are statistically significant. The analyses provide targets for future model development and present methods for assessing the model revisions and an assessment of the implementation of 25-km SHiELD and 6.5-km SHiELD.

Accurate predictions of clouds and precipitation are important to our daily lives and to decision makers in preparing a weather ready nation. Weather prediction can help minimize the impact of weather hazards on communities and initiate preparedness and responses to extreme weather. This research provides a framework for improving the accuracy of the SHiELD model by enhancing horizontal resolution and an assessment tool for how effective model revisions and resolution are in augmenting prediction capabilities.

Future research can be done to find the "ideal" resolution for predicting clouds and precipitation. Here, 6.5 km resolution was found to have the highest skill; however, future work could evaluate whether further increasing the resolution heightens the model skill. As models are increasingly being moved towards higher resolutions, it is important to understand how resolution impacts the accuracy of the variables that are being predicted. This research addressed how increasing the resolution impacts prediction with the hope of providing a framework for identifying biases and evaluating forecast skills as the resolution is augmented. Since higher resolutions have increased computational costs, it is important to analyze cost-benefit at which point increasing resolution no longer improves prediction accuracy.

References

(1) Christopoulos, C.; Schneider, T. *Geophysical Research Letters* **2021**, *48*, DOI: 10.1029/2021gl093017.

(2) Zhou, L.; Harris, L.; Chen, J.-H.; Gao, K.; Guo, H.; Xiang, B.; Tong, M.; Huff, J. J.; Morin, M. *Journal of Advances in Modeling Earth Systems* **2022**, *14*, DOI: 10.1029/2021ms002971.

(3) Zhou, L.; Lin, S.-J.; Chen, J.-H.; Harris, L. M.; Chen, X.; Rees, S. L. *Bulletin of the American Meteorological Society* **2019**, *100*, 1225–1243.

(4) Chen, J.-H.; Lin, S.-J.; Magnusson, L.; Bender, M.; Chen, X.; Zhou, L.; Xiang, B.; Rees, S.; Morin, M.; Harris, L. *Geophysical Research Letters* **2019**, *46*, 4495–4501.

(5) Harris, L. M.; Rees, S. L.; Morin, M.; Zhou, L.; Stern, W. F. *Journal of Advances in Modeling Earth Systems* **2019**, *11*, 1847–1869.

(6) Harris, L. et al. *Journal of Advances in Modeling Earth Systems* **2020**, *12*, DOI: 10.1029/2020ms002223.

(7) Harris, L.; Zhou, L.; Kaltenbaugh, A.; Clark, S.; Cheng, K.-Y.; Bretherton, C. *Journal of Geophysical Research: Atmospheres* **2023**, *128*, DOI: 10.1029/2022jd037823.

(8) Gao, K.; Harris, L.; Zhou, L.; Bender, M.; Morin, M. *Journal of the Atmospheric Sciences* **2021**, *78*, 3007–3021.

(9) Gao, K.; Harris, L.; Bender, M.; Chen, J.-H.; Zhou, L.; Knutson, T. *Geophysical Research Letters* **2023**, *50*, DOI: 10.1029/2023gl103329.

(10) Magnusson, L. et al. *Bulletin of the American Meteorological Society* **2022**, DOI: 10.1175/bams-d-21-0234.1.

(11) Mass, C. F.; Ovens, D.; Westrick, K.; Colle, B. A. *Bulletin of the American Meteorological Society* **2002**, *83*, 407–430.

(12) Gallus, W. A. *Weather and Forecasting* **1999**, *14*, 405–426.

(13) Beck, H. E.; Wood, E. F.; Pan, M.; Fisher, C. K.; Miralles, D. G.; van Dijk, A. I. J. M.; McVicar, T. R.; Adler, R. F. *Bulletin of the American Meteorological Society* **2019**, *100*, 473–500.

(14) Du, J. *UCAR/NCAR - Earth Observing Laboratory* **2011**, DOI: https://doi.org/10.5065/D6PG1QDD.

(15) Wielicki, B. A.; Barkstrom, B. R.; Harrison, E. F.; Lee, R. B.; Louis Smith, G.; Cooper, J. E. *Bulletin of the American Meteorological Society* **1996**, *77*, 853–868.

(16) Hersbach, H. et al. *Quarterly Journal of the Royal Meteorological Society* **2020**, *146*, 1999–2049.

(17) Zhang, J.; Howard, K.; Langston, C.; Kaney, B.; Qi, Y.; Tang, L.; Grams, H.; Wang, Y.; Cocks, S.; Martinaitis, S.; Arthur, A.; Cooper, K.; Brogden, J.; Kitzmiller, D. *Bulletin of the American Meteorological Society* **2016**, *97*, 621–638.

(18) De Pondeca, M. S. F. V.; Manikin, G. S.; DiMego, G.; Benjamin, S. G.; Parrish, D. F.; Purser, R. J.; Wu, W.-S.; Horel, J. D.; Myrick, D. T.; Lin, Y.; Aune, R. M.; Keyser, D.; Colman, B.; Mann, G.; Vavra, J. *Weather and Forecasting* **2011**, *26*, 593–612.

(19) Tang, X.; Cai, Q.; Fang, J.; Tan, Z.-M. *Journal of Geophysical Research: Atmospheres* **2019**, *124*, 12010–12021.

(20) Jensen, T.; Prestopnik, J.; Soh, H.; Goodrich, L.; Brown, B.; Bullock, R.; Gotway, J. H.; Newman, K.; Opatz, J. *The MET Version 11.1.0 User's Guide*; tech. rep.; Developmental Testbed Center, 2023.