

Comparison of probabilistic forecasts of extreme precipitation for a global and convection-permitting ensemble and hybrid statistical-dynamical method based on equatorial wave information

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Wolf, G., Ferrett, S. ORCID: <https://orcid.org/0000-0003-4726-847X>, Methven, J. ORCID: <https://orcid.org/0000-0002-7636-6872>, Frame, T. H. A. ORCID: <https://orcid.org/0000-0001-6542-2173>, Holloway, C. E. ORCID: <https://orcid.org/0000-0001-9903-8989>, Martinez-Alvarado, O. ORCID: <https://orcid.org/0000-0002-5285-0379> and Woolnough, S. J. ORCID: <https://orcid.org/0000-0003-0500-8514> (2024)
Comparison of probabilistic forecasts of extreme precipitation for a global and convection-permitting ensemble and hybrid statistical-dynamical method based on equatorial wave information. Quarterly Journal of the Royal Meteorological Society, 150 (759). pp. 877-896. ISSN 1477-870X doi: 10.1002/qj.4627 Available at <https://centaur.reading.ac.uk/114232/>

work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1002/qj.4627>

Publisher: Royal Meteorological Society

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur


CentAUR

Central Archive at the University of Reading

Reading's research outputs online

RESEARCH ARTICLE

Comparison of probabilistic forecasts of extreme precipitation for a global and convection-permitting ensemble and hybrid statistical–dynamical method based on equatorial wave information

G. Wolf¹  | S. Ferrett²  | J. Methven¹ | T.H.A. Frame¹ | C.E. Holloway¹  |
O. Martinez-Alvarado¹  | S.J. Woolnough¹

¹Department of Meteorology, University of Reading, Reading, United Kingdom

²Department of Mathematics and Statistics, University of Exeter, Exeter, United Kingdom

Correspondence

G. Wolf, Department of Meteorology, Earley Gate, University of Reading, P.O. Box 243, Reading, Berkshire RG6 6BB, United Kingdom.

Email: g.a.wolf@reading.ac.uk

Funding information

NERC National Capability International Programmes Award, Grant/Award Number: NE/X006263/1; Newton Fund WCSP project Forecasting for Southeast Asia (FORSEA)

Abstract

Recent work has demonstrated that skilful hybrid statistical–dynamical forecasts of heavy rainfall events in Southeast Asia can be made by combining model forecasts of the phases and amplitudes of Kelvin, Rossby, and westward-moving Rossby gravity waves with climatological rainfall statistics conditioned on these waves. This study explores the sensitivity of this hybrid forecast to its parameter choices and compares its skill in forecasting extreme rainfall events in the Philippines, Malaysia, Indonesia, and Vietnam to that of the Met Office Global and Regional Ensemble Prediction System (MOGREPS). The hybrid forecast is found to outperform both the global and convection-permitting ensemble in some regions when forecasting the most extreme events; however, for less extreme events, the ensemble is found more skilful. A weighted blend of the MOGREPS forecasts and the hybrid forecast was found to have the highest skill of all for almost all definitions of extreme event and in most regions. To quantify the influence of errors in the predicted wave state on the skill of the hybrid forecast, the skill of a hypothetical best-case forecast was also calculated using reanalysis data to specify the wave amplitudes and phases. This best-case forecast indicates that errors in the forecasts of all wave types reduce the skill of hybrid forecast; however, the reduction in skill is largest for Kelvin waves. The skill in convection-permitting models is greater than for global models in the regions where Kelvin waves dominate, but the added value of limited-area high-resolution forecasts is hampered by the poor representation of Kelvin waves in the parent global model.

KEYWORDS

convection-permitting models, ensemble forecasts, equatorial waves, high-impact weather, probabilistic forecasts

1 | INTRODUCTION

Southeast Asia is a region of frequent heavy precipitation events leading to devastating societal impacts through flooding or landslides (Kirschbaum et al., 2015). Therefore, it is crucial to better understand the occurrence of heavy precipitation events and to what extent models are able to predict them, keeping in mind that there are different ways to identify such events (e.g., Ferrett et al., 2020, 2021; Schlueter et al., 2019). Some modes of variability shown to be linked to heavy rainfall are the Madden–Julian oscillation, Borneo vortices, cold surges, and equatorial waves (Chang et al., 2005; Juneng et al., 2007; Tangang et al., 2008; Xavier et al., 2014; van der Linden et al., 2016; Ferrett et al., 2020). Because of the range of temporal and spatial scales involved, forecasting of rainfall in the Maritime Continent is a great challenge. However, large-scale systems, such as equatorial waves, are expected to have longer range predictability than individual convective weather systems and their associated rainfall (e.g., Judt, 2020). For southeast Asia, Ferrett et al. (2020) have established that there is a strong observed statistical dependence of heavy precipitation on the amplitude and phase of different equatorial wave modes, but this connection is not limited to Southeast Asia (e.g., Schlueter et al., 2019).

Equatorial waves can be identified in global forecast data, and there is appreciable skill in the forecast of wave amplitude and phase out to 4–5 days for Kelvin waves and 6 days for the westward Rossby (R1, R2) and westward-moving mixed Rossby gravity (WMRG) waves, as shown for deterministic global Met Office forecasts by Yang et al. (2021). For an observed strong connection between a difficult-to-forecast quantity (precipitation, sea breeze front, etc.) and an easier-to-predict large-scale condition (such as equatorial waves in this case), a statistical–dynamical model can be expected to produce skilful forecasts (Cafaro et al., 2019; Ferrett et al., 2023; Maier-Gerber et al., 2021). Such statistical approaches can be used as a useful benchmark test for the numerical weather prediction (NWP) models (Walz et al., 2021). To exploit the skill in the Met Office forecasts of equatorial wave activity, a prototype hybrid dynamical–statistical forecast model for the probability of widespread heavy precipitation, given only predicted wave information, has been introduced by Ferrett et al. (2023). By using forecasts of the waves, it might be expected to extend skill of probabilistic forecasts of heavy rainfall (occurring anywhere within subregions of countries) into the medium range. Ferrett et al. (2023) were able to show that their hybrid model did indeed lead to better predictions of heavy precipitation than raw precipitation forecasts from the Met Office Global and Regional Ensemble Prediction System

(MOGREPS-G) ensemble did, at least for some tropical regions and particular seasons.

As mentioned, global models have difficulties in producing skilful precipitation forecasts within the Tropics, suggesting the necessity of higher resolution convection-permitting (CP) models or statistical–dynamical models to generate skill beyond the climatology forecasts (Vogel et al., 2020). Although introducing a CP model can lead to an increase in forecast skill compared with global models that use parametrised convection (Woodhams et al., 2018), more evaluation of such models for a variety of regions within the Tropics seems necessary to better understand their potential. The UK Met Office have developed higher resolution limited-area CP ensemble forecasts over 6 months (October 2018 to March 2019) in a number of regions of southeast Asia: Malaysia, Indonesia, and the Philippines. The forecasts are available at a range of resolutions and yield skill in precipitation forecasts over those from NWP models with parametrised convection. However, skill is strongly dependent on spatial scale and also on local time of day. In general, CP forecasts of precipitation over land during the day yield skill in Malaysia, Indonesia, and the Philippines (Ferrett et al., 2021) on scales of 70 km or greater. Though it is established that CP forecasts are an improvement over global NWP forecasts with parametrised convection in Malaysia and Indonesia, it is still not clear how the skill of the CP ensemble depends on equatorial waves and how the prediction of the probability of high-impact weather (HIW) compares with the previously developed hybrid forecasting method (Ferrett et al., 2023).

The aim of this article is to analyse the hybrid wave-based model and investigate to what extent skill in precipitation from ensemble forecasts can be explained by the prediction of waves (using their average relationships with precipitation) and where within southeast Asia, and under which circumstances, high-resolution CP ensembles have skill beyond the hybrid model. If there is skill beyond the wave hybrid model, what atmospheric features contribute to this skill? Although the use of a wave-based statistical–dynamical hybrid model shows promising results, forecast skill of equatorial waves is still somewhat limited despite their large-scale features (Bengtsson et al., 2019; Dias et al., 2018). For example, Yang et al. (2021) and Ferrett et al. (2023) find that in global Met Office NWP forecasts, westward-moving waves, R1 and WMRG, have good predictability up to 6 days, but Kelvin waves tend to have predictability to around 4 days, thus placing a limit on the effectiveness of the wave-based hybrid model. Therefore, it is possible that NWP forecasts may be preferred over the hybrid model in some instances. Owing to the dynamical–statistical set-up of the hybrid model, which makes use of the connection between HIW

and equatorial wave modes, it can also be used to identify NWP model deficiencies associated with the forecast of large-scale dynamics and the subsequent consequences for the use of a high-resolution nested regional model. An important aspect of these comparisons is not only to identify the better approach, but also the possibility to improve impact forecasting by improving early warnings about HIW (Merz et al., 2020).

The remainder of this article is organized as follows. Section 2 presents the data and methods used to define HIW events, the set-up of the hybrid statistical–dynamical model, and a detailed description of the forecast metrics to evaluate and compare all models. Section 3 covers in detail the hybrid model, presenting its connection to HIW for several regions, including its seasonal dependencies, as well as the sensitivity to its most important parameters. Section 4 then investigates the performance of all individual models in predicting HIW, and how they compare with each other. The key conclusions of this article are summarized in Section 5.

2 | DATA AND METHOD

2.1 | Data

The skill of three different sets of precipitation forecasts will be compared in this work: forecasts from the global ensemble MOGREPS-G, forecasts from the CP version of the MOGREPS ensemble to estimate the added value of the increased resolution, and a hybrid statistical forecast derived from MOGREPS-G forecasts of equatorial waves (described in Section 2.2.2).

The MOGREPS-G ensemble (Bowler et al., 2008) comprises 17 members (11 members before June 2017) that are run with a grid spacing of 33 km with parametrised convection out to a lead time of 168 hr. These are routinely run as part of the Met Office operational suite, and data are used for the 2015 to 2019 period.

The high-resolution CP model is given by the limited high-resolution area simulations of the Met Office Unified Model 17 ensemble members, nested within the global ensemble of MOGREPS-G with 2.2, 4.5, and 8.8 km grid spacing with lead times out to 120 hr (60 hr for the 2.2 km resolution). Our main focus will be the regions available for the 2.2 km resolution areas (Mal-22, Phi-22, and Ind-22 in Figure 1), which are the regions for which all models are available. The evaluation period of the CP model is October 2018 to March 2019. The global MOGREPS-G ensemble will be evaluated for the same short 6-month period, but also for the longer January 2015 to March 2019 period to get results less affected by the noisy nature of HIW events.

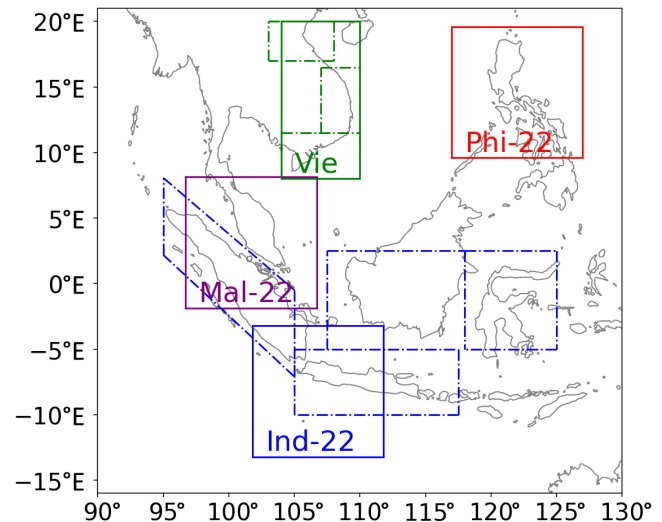


FIGURE 1 Boxed regions of interest. Main areas for Malaysia, Indonesia, Philippines, and Vietnam are given by solid lines. Subareas for the same countries are represented by dashed lines, which in the text will be referred to as Vie-S, Vie-C, and Vie-N for the southern, central, and northern parts of Vietnam, and Sumatra, Borneo, Java and Sulawesi for the blue dashed boxes, according to their large overlap with the main Indonesian islands. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/qj.4627)]

All model forecasts are used to diagnose 24 hr accumulated precipitation from 0000 UTC daily.

The statistical–dynamical model, introduced by Ferrett et al. (2023), estimates the probability of HIW events based on the forecasted wave state of the global MOGREPS-G ensemble, calculated from the horizontal wind and geopotential height at 850 hPa. Those fields are filtered in time and space and spatially projected onto the theoretical equatorial wave mode of Kelvin, Rossby (R1), and WMRG waves. See Yang et al. (2021) for details about the calculation of the wave state from real-time forecasts. The statistical relationship between the equatorial wave state and HIW events is derived from the 2000 to 2014 climatology, for which the wave state is calculated from European Centre for Medium-Range Weather Forecasts Reanalysis v5 (ERA5) (Hersbach et al., 2020). The method to produce this climatological wave dataset is described in Yang et al. (2003).

The precipitation observational dataset 2000 to 2019 is based on the Integrated Multi-satellite Retrievals for Global Precipitation Measurement (IMERG) (Huffman et al., 2015), which is not only used for the evaluation period of the model forecasts, but also to train the statistical–dynamical model. For a proper evaluation of the statistical–dynamical model it is essential to separate the full time range into a mutually exclusive training period (2000–2014) and evaluation period (2015–2019).

2.2 | Method

2.2.1 | Definition of HIW events

The focus of this work is forecasts of high-impact rainfall events. We use two different methods to convert gridded rainfall data to a binary variable that defines the occurrence/non-occurrence of a high-impact rainfall event within a given region. Both methods are calculated using only grid points that lie over land. The first, HIW_{av} , is very simple: an HIW event is said to have occurred if the area-averaged 24 hr rainfall accumulation exceeds a chosen percentile of climatology. The second method, HIW_{area} , is a two-step method: First, we define a rainfall threshold value for a given region to be the 95th percentile of 24 hr accumulated rainfall at all land points within the region pooled into a single dataset. An HIW event is then said to have occurred in the region if the area fraction of land points exceeding that rainfall threshold in a given 24 hr period exceeds a chosen value. The value is chosen according to a specific percentile, defining how extreme the HIW should be. All area thresholds for all percentiles and all regions for both MOGREPS-G and the CP model are included in Supporting Information Tables S1 and S2. The rationale for using the second definition, HIW_{area} , is that impacts and required mitigations are likely to be greater if a larger area is affected. Using this simpler definition of an HIW event can also be used as a check on the sensitivity of the results on the specific definition of a high-impact rainfall event.

To account for systematic biases in model forecast rainfall, the percentile threshold value used in the HIW definition is calculated from model data at each lead time separately. This is done only for the percentile thresholds of HIW_{av} and the 95th grid-point-based percentile of the HIW_{area} definition. The area threshold is not changed. As the model forecasts are not available for the same full climatology, the precipitation threshold derived from the observed climatology for a specific percentile might not be represented by the same percentile within the evaluation period. The bias correction of the model precipitation forecasts therefore needs an additional step. We first identify the percentile of the precipitation within the evaluation period that is closest to the climatological precipitation threshold. This can differ from the original percentile defining the HIW event, or for investigating very rare events even lead to the case that no HIW event can be identified in the evaluation period. Then, in a second step, this percentile of the evaluation period is applied to the model forecast to define the associated precipitation threshold for the given lead time in forecast model data.

2.2.2 | Statistical–dynamical hybrid model

The statistical–dynamical hybrid forecast model, introduced by Ferrett et al. (2023) and in the following referred to as “hybrid model”, makes use of the connection between the probability of heavy precipitation in a defined region conditional on the phase and amplitude of an equatorial wave mode. Three types of equatorial waves are used to construct the hybrid forecasts: the eastward-propagating Kelvin wave, the R1 wave, and the WMRG wave. The horizontal structures of these waves vary and can be seen in Yang et al. (2021, fig. 1) or Ferrett et al. (2020, fig. 2). Kelvin wave activity maxima are located on the Equator and characterised by oscillations between zonal wind convergence or divergence. R1 and WMRG waves propagate westward and are anti-symmetric about the Equator in v for R1 and divergence for WMRG. The R1 wave consists of twin vortices straddling the Equator, and the WMRG wave is dominated by rotational winds centred on the Equator. The WMRG wave winds alternate between northerly and southerly winds crossing the Equator with convergence/divergence in the Northern Hemisphere (NH) and opposing divergence/convergence in the Southern Hemisphere. These structures can be identified (e.g., Yang et al., 2021) from full horizontal wind fields and are used to construct wave phases to use in the hybrid model, as described later herein. For the Bayesian forecast methodology we are following Cafaro et al. (2019).

As in Yang et al. (2021) and Ferrett et al. (2023), the evolution of each equatorial wave type is described using two parameters, λ_1 and λ_2 , defined at a fixed longitude and latitude and calculated from the relevant wave wind fields. Therefore, these parameters relate to the local wave structure (Table 1). These are chosen as they characterise the propagation and amplification of the waves through the defined region. The longitude varies depending on the region, using the centre longitude of the specified region, but the latitude is fixed based on the horizontal structure of the wave. With a further normalisation by their temporal standard deviations σ_1 and σ_2 , we use those two

TABLE 1 Variables λ_1 and λ_2 spanning the wave phase space for the different wave types.

Wave type	λ_1	λ_2
Kelvin	u_p (0° N)	du_p/dx (0° N)
WMRG	$-u_p$ (10° S)	v_p (0° N)
R1	$-u_p$ (0° N)	v_p (8° N)

Note: WMRG, westward-moving mixed Rossby gravity; R1, mode 1 Rossby; u_p and v_p are respectively the spatial projections of the zonal and meridional winds onto the associated equatorial wave mode. x corresponds to distance along the Equator.

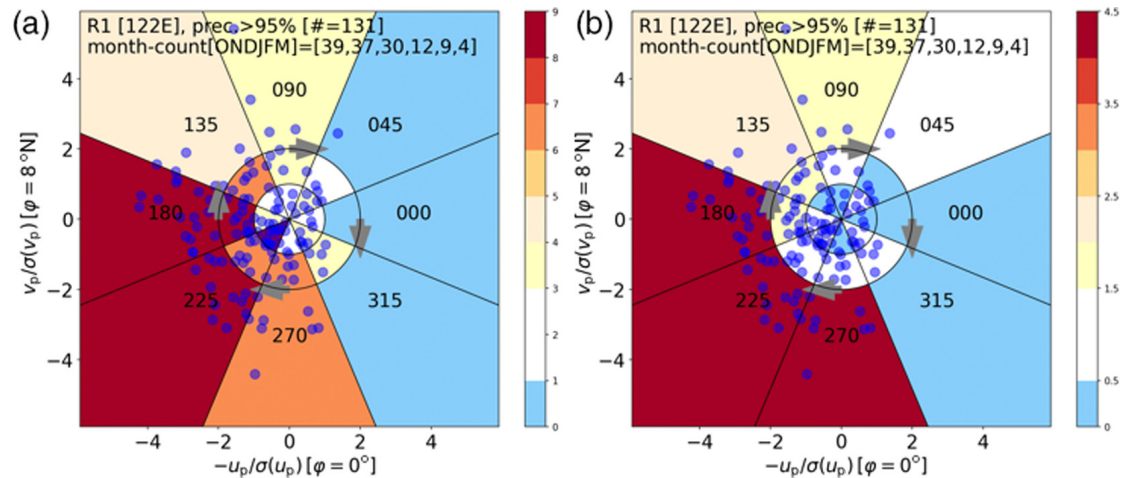


FIGURE 2 Wave-phase diagram for mode 1 Rossby (R1) waves derived for the Philippines (region Phi-22) using European Centre for Medium-Range Weather Forecasts Reanalysis v5 data for October–March (ONDJFM) within 2000 to 2015. Blue dots show the high-impact weather (HIW) events. Colours in panel (a) represent the percentage of total HIW events; Equation (2). Colours in panel (b) show the HIW conditional probability in each sector, normalised by the overall HIW probability; Equation (3). The associated values in panel (b) therefore show how much more likely (values >1) or less likely (values <1) an HIW event occurs in a particular sector compared with the general, wave-independent, HIW probability. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/qj.4627)]

normalised variables to span the local wave-phase diagram (as shown in Figure 2). The current wave state at a given time t for a specified geographical region Ω can then be defined by

$$W_{\Omega} = W_{\Omega}(A, \alpha, t), \quad (1)$$

where $A = \sqrt{(\lambda_1/\sigma_1)^2 + (\lambda_2/\sigma_2)^2}$ is defined as the equatorial wave amplitude and α is the wave phase angle in the wave-phase diagram, defined as the anticlockwise angle between the vector $(\lambda_1/\sigma_1, \lambda_2/\sigma_2)$ and $(1, 0)$.

Then, the time-dependent wave state $W_{\Omega}(A, \alpha, t)$ is linked with HIW events using an indicator variable:

$$\theta_{\Omega}(t) = \begin{cases} 1 & \text{HIW event occurs} \\ 0 & \text{no HIW event} \end{cases},$$

where t represents the validation time and Ω is the evaluation region. $\theta_{\Omega}(t)$ therefore represents a time series of HIW or no-HIW events for a specific region that can be associated with the time series of the wave state $W_{\Omega}(A, \alpha, t)$. With the projection of the wave state and the HIW events of a reference climatology into the wave phase space one can eliminate the time dependence and HIW events can then be associated with specific areas within the wave phase space $W_{\Omega}(A, \alpha)$ (HIW events in the wave phase diagram of Figure 2 shown by blue circles). In the following, we drop the index Ω as all variables within an equation are representations of the same region Ω , keeping in mind that each of the equations must be applied for each investigation region individually.

The wave-phase diagrams of Figure 2 are based on R1 waves for the Philippines during the extended NH winter season (October–March, ONDJFM). To link HIW events with specific wave states, the wave phase space is split into several sectors (delimited by black lines in Figure 2), each sector separated by 45° (different to the original method of Ferrett et al., 2023, where 90° sectors were used). The sectors are labelled according to the angle of their centre from 0° to 315° in steps of 45° . The sectors are further split into three wave-amplitude ranges, between 0 and 1, 1 and 2, and above two standard deviations where higher wave amplitudes occur further away from the centre of the diagram. The location of the HIW events within the wave-phase diagram are highlighted by the blue dots, each dot representing one HIW event. From such a diagram one can calculate the number of HIW events falling within each sector or wave state ($n_{\text{HIW}}(W)$) of the wave phase space, normalised by the total count of HIW events, n_{HIW} (Figure 2a):

$$f(W|\theta = 1) = \frac{n_{\text{HIW}}(W)}{n_{\text{HIW}}}. \quad (2)$$

The difficulty of interpreting this distribution (Figure 2a) results from the fact that higher amplitude waves (further away from the centre of the diagram) occur less often and therefore there are also fewer non-HIW events within those high-amplitude sectors. Despite this, f does show increased values within those high-amplitude sectors with a substantial percentage of all HIW events.

Bayes' rule can then be used to derive a formula for the probability of HIW occurrence conditional on the specific

wave state:

$$\pi(\theta = 1|W) = \frac{f(W|\theta = 1)\pi(\theta = 1)}{f(W|\theta = 1)\pi(\theta = 1) + f(W|\theta = 0)\pi(\theta = 0)}, \quad (3)$$

where $\pi(\theta = 1) = P$ is the unconditioned probability of HIW occurrence. The reason that Equation (3) differs from the normalised count f is that it is necessary to account properly for the count of non-events in each sector of the wave phase space since these are not evenly distributed. In particular, counts of HIW events (relative to non-events) are much higher when wave amplitude is high than when wave amplitude is low. If HIW events were to occur independent of the wave state, one would expect that $\pi(\theta = 1|W) \approx \pi(\theta = 1)$. For example, if an HIW event is defined by the top 5% of strongest precipitation events within the region of investigation, this would simply mean $P = 0.05$. However, Figure 2b shows that the probability of HIW depends strongly on wave state.

Using the conditional probability of Equation (3) therefore allows the identification of specific sectors with high frequencies of HIW; for example, sectors 180° , 225° , and 270° , highlighted by the dark red colours (Figure 2b). HIW in the Philippines can therefore frequently be associated with high R1 wave amplitudes ($A > 2$) in the transition phase into the strong cyclonic flow (from $v_p(8^\circ\text{N}) < 0$ to $u_p(0^\circ\text{N}) > 0$). Ferrett et al. (2023) also found a strong connection between HIW and high-amplitude R1 waves within their Philippine regions. Periods with little or no HIW (blue colours in Figure 3) occur within and slightly after the main anticyclonic flow phase (from $u_p(0^\circ\text{N}) < 0$ to $v_p(8^\circ\text{N}) < 0$) with a sharp transition to the wave states associated with highly increased HIW probabilities.

To get a meaningful connection, the wave-phase diagram should include long enough climatological data for precipitation and the associated wave state. We used the 15-year period 2000–2014, with IMERG for precipitation and ERA5 to derive the wave state. For the investigation of a specific climatological season, the training data for the hybrid model are limited to the same season. Instead of using the model forecast of precipitation, the hybrid model uses the model forecast wave amplitude and phase combined with the climatological conditional probability of HIW events. In the following we will analyse how this hybrid model, applied to the MOGREPS-G ensemble, performs in comparison with high-resolution CP ensemble precipitation forecasts. The hybrid ensemble forecast estimate for the probability of an HIW event, π_{ens} (for n_{ens} ensemble members) is defined as the ensemble mean of the conditional probabilities estimated from the wave state of each member:

$$\pi_{\text{ens}}(\theta = 1|W) = \frac{1}{n_{\text{ens}}} \sum_{i=1}^{n_{\text{ens}}} \pi(\theta = 1|W_i). \quad (4)$$

We will use two evaluation periods, January 2015 to March 2019 and October 2018 to March 2019, both without overlap with the climatological training period. The 6-month period is used to evaluate the CP ensemble, which is only available for this shorter time period. The evaluation of the MOGREPS-G ensemble will be done for both time periods, the shorter period for a direct comparison with the CP ensemble and the longer period to check the robustness of the results and to analyse in more detail the potential of the statistical–dynamical hybrid model. When analysing the high-resolution CP model, the HIW definition (as presented in Section 2.2.1) is based on the high-resolution IMERG data, whereas for the global model the HIW definition will be based on the interpolation of the high-resolution IMERG data onto the lower resolution grid of MOGREPS-G. This is so that the HIW variable is calculated consistently in both forecasts so that skill differences are attributable due to different forecasts of the same predictand rather than differences in the definition of the predictand.

2.3 | Forecast metrics analysed to evaluate HIW forecasts

2.3.1 | Area under the receiver operating characteristic curve diagnostic applied for HIW events in the Philippines

The main forecast skill metric that we will use for assessing the relative performance of the different forecasts is the area under the receiver operating characteristic (ROC) curve (AUC). This measure of skill is designed to test whether there is a systematic relationship between variations in the forecast probability of an event and whether or not that event occurs, even if the probabilities themselves are not well calibrated. In the rest of the article this will be distilled down to a set of summary figures. However, to illustrate the detailed calculations used and to include some discussion of properties of the forecast, such as reliability, we first present a detailed example of forecasts for the October–March 2015–2019 period in the Philippines region, Phi-2.2 (Figure 1), with the hybrid model based on Rossby R1 waves.

Figure 3a shows the 3-day lead time forecast probability (restricted to the 2018–2019 season) of the occurrence HIW_{area} (95th percentile) from the MOGREPS-G forecast (green dash-dotted line) and the hybrid forecast (solid orange line). The area threshold of this HIW_{area} definition is 29.4%, meaning that at least 29.4% of grid points over land exceed the 24 hr accumulated precipitation 95th percentile threshold of 34.4 mm. In the following we will only refer to the percentiles of the HIW_{area}

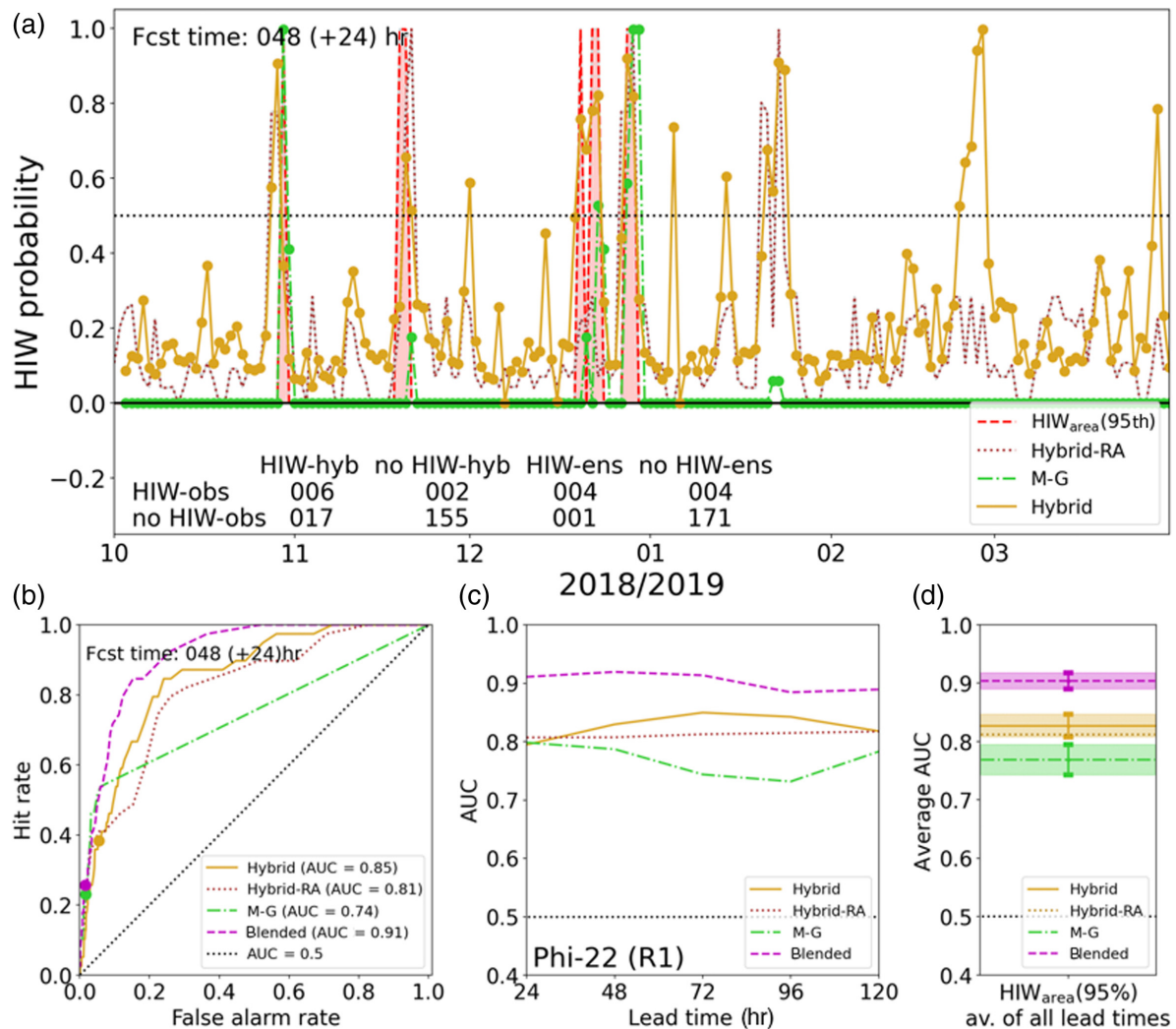


FIGURE 3 Area under the receiver operating characteristic (ROC) curve (AUC) diagnostic for Phi-22 for the 5% strongest precipitation events (HIW_{area}) in the October–March season for 2015–2019. (a) The limited time series from October 2018 to March 2019 of high-impact weather (HIW) events (red dashed), average of Met Office Global and Regional Ensemble Prediction System (MOGREPS-G) ensemble members predicting an HIW event (green line with dots, labelled “M-G”), rescaled hybrid model (based on Rossby R1 waves) applied to MOGREPS-G ensemble (orange line with dots, labelled “Hybrid”), and the hybrid model based on reanalysis data (brown dotted line, labelled “Hybrid-RA”). Available data are indicated by dots. The horizontal black dotted line represents an arbitrary choice for the ensemble forecast to be interpreted as a deterministic forecast of an extreme event, for which the contingency table in the lower part of the panel is created. (b) The associated ROC curve for the full four seasons, where the threshold to identify extreme events (as done by the horizontal black dotted line in (a) with the associated contingency table) is varied between 0 and 1 (three filled circles show values for threshold of 0.5, also included in (a), shown by horizontal black dotted line). (c) The AUCs for all lead times (panel (b) represented by the values for a 72 hr forecast). (d) The average of panel (c) for all lead times with plus/minus one standard deviation as a measure of its variability. Colours are the same as for (a) (green for M-G, orange for Hybrid, and brown for Hybrid-RA), with the additional purple-coloured line a linear combination of M-G (0.8 weighting) and the hybrid model (0.2 weighting). Black dotted line shows the diagonal within the ROC curve, representing the lower limit of a useful model forecast performance, which corresponds to a value of 0.5 for the AUC. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/terms-and-conditions)]

definition; the associated area percentages are included in Supporting Information Table S2. Additionally to the MOGREPS-G and hybrid forecasts, the binary outcome variable (red dashed line) and the hypothetical best-case hybrid forecast with wave state taken from reanalysis (brown dotted line) are shown. For ease of viewing, probabilities from the hybrid forecasts have been scaled by

the maximum value within the time-period such that in the figure they vary between zero and one, however the hybrid forecast probabilities do not exceed about 0.4 for this October–March period in this or other regions.

The forecast probabilities are sharply peaked, meaning that there are a small number of high-probability forecasts with values that exceed considerably the typical range.

This is most clearly seen for the MOGREPS-G forecasts, where all but a few forecast probabilities are zero. For the MOGREPS-G forecasts there are notable hits where very high probability is forecasted and the event occurs, but also notable misses where an HIW event is observed but a probability of zero is given by the ensemble (e.g., the events in November and the first event in December). These misses seem to be associated with timing errors, as the event is forecasted to occur a day later than it occurs in the verification data. Since they are based on climatological rainfall statistics, the hybrid forecasts very rarely produce forecast probabilities close to zero.

Values of the ensemble averages above a threshold (the horizontal black dashed line in Figure 3a) could be used to transform the probabilistic forecast into a deterministic forecast for the occurrence of an HIW event. A comparison of the predicted HIW events of the hybrid model (HIW-hyb) or the MOGREPS-G direct precipitation forecasts (HIW-ens) with the actual occurring HIW events (HIW-obs), and similarly for the non-events (no HIW-obs), could be used to create contingency tables, as included in the lower part of Figure 3a. Such a table can therefore be used to identify correct forecasts of events occurring (HIW-obs and HIW-X with X either hyb or ens) or not occurring (no HIW-obs and no HIW-X), as well as false alarms (no HIW-obs and HIW-X) and missed events (HIW-obs and no HIW-ens). However, interpreting those tables would be difficult, as they would strongly depend on the subjectively chosen threshold.

To get an objective measure of how well the models predict an HIW event we produce so-called ROC curves. To do so, we vary the applied HIW probability threshold in 0.01 increments between 0 and 1 and calculate the resulting hit and false alarm rates (Figure 3b). To help relate the curves to Figure 3a, the filled circles show the hit and false alarm rate relationship for a chosen threshold of 0.5 (horizontal black dashed line in Figure 3a). We then calculate the AUC as an evaluation metric for the HIW forecast, with better forecasts being represented by higher AUC values. A perfect forecast would have $AUC = 1$; a useful forecast must have $AUC > 0.5$, which one would get if the associated ROC curve were to lie along the diagonal.

We repeat the calculation of the AUC for all lead times of 1 to 5 days. A comparison between the hybrid and the precipitation forecast model for the chosen example reveals that the hybrid model performs better than the precipitation forecast model for all available lead times (Figure 3c). This means, based on this forecast metric and the chosen region, using the MOGREPS-G wave predictions for the hybrid model does lead to better precipitation-based HIW forecasts than using the actual precipitation forecasts of MOGREPS-G. To easily compare different set-ups for the hybrid model, such as use

of different percentiles to define an HIW event, different regions within southeast Asia, and so forth, we compress the information further by averaging over all lead time values, resulting in one value for the model performance of a region for all lead times, with the standard deviation of all AUC values for the different lead times giving an uncertainty range (Figure 3d). This diagnostic of the compressed information for the 1 to 5 days lead time will be referred to as the “AUC diagnostic”. Using this compressed information also means that this study is not particularly focusing on the tendency of the forecast performance with increasing lead time, but rather on the general model skill within this lead time range. This procedure will allow us to draw conclusions about the general model performances and possible model issues.

In general, one would expect the Hybrid-RA line to show better performances than the Hybrid line, as it is based on the actual reanalysis wave state and not the forecasted one. However, this is not necessarily the case, as both models evaluate HIW events in a time range different to the climatological training period; that is, there is no guarantee that the distribution of the observed small number of HIW events of the evaluation period within the wave-phase diagram is identical to the distribution obtained from the climatological training data (although we expect them to be very close). Nevertheless, a comparison of Hybrid and Hybrid-RA can still be used to identify possible model deficiencies in forecasting the wave state. The differences in the example (Figure 3d) are, however, very small and within the error margin, meaning that one would conclude that MOGREPS-G has no systematic difficulty in representing the R1 wave state correctly for the Philippines.

In Figure 3b–d we showed an additional line (purple) apart from the hybrid model (orange and brown) and the precipitation forecast model (green). This line, and therefore the associated high model performance, can be achieved if both models, the hybrid and MOGREPS-G, are blended linearly with a weighting of 0.2 and 0.8. This linear combination (blended forecast) is outperforming both individual models. Simple testing indicates that higher performances can be achieved by putting more weight on the direct precipitation forecasts. A possible reason for this might already be visible in the ROC curve shown (Figure 3a,b), which indicates that, in general, the model predicts several HIW events fairly well but completely misses several other HIW events. This behaviour can be the result of an underdispersive ensemble. The exact choice of the weighting does not seem to matter too much, as long as more weighting is put on the direct precipitation forecast. The blended forecast is used here to emphasise the additional gained insight and improved forecast that can be achieved by incorporating the hybrid

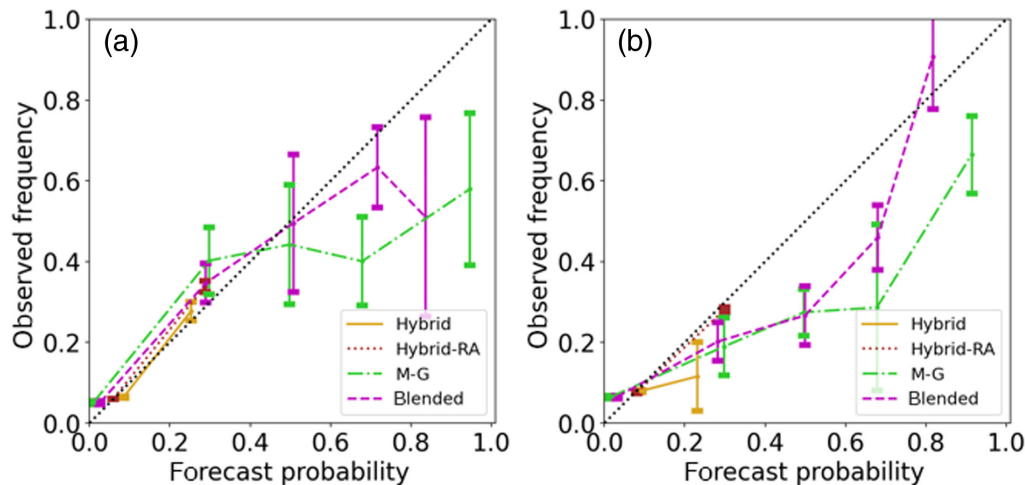


FIGURE 4 Forecast reliability for (a) Phi-22 and (b) Mal-22 with the hybrid model based on (a) R1 and (b) Kelvin waves. Forecasted probabilities separated into five bins with bin edges given by values on the x-axis. The values shown represent the average of all lead times (1 to 5 days), with error bars representing the standard deviation of the averaged values. Coloured lines represent the same models as in Figure 3. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

model. Therefore, no in-depth analysis was done to identify the best weighting split, as this would likely depend on the specific ensemble forecast system being used and other details of the data.

2.3.2 | Comparison of AUC diagnostic with other evaluation metrics

Good model performance according to the AUC diagnostic does not necessarily mean the model can be used as a good HIW forecast. The AUC diagnostic allows us to identify if there is a link between the model forecast of HIW events and observed HIW events. The hybrid model in its unmodified form cannot be expected to deliver reliable forecasts of high probabilities, because the associated maximum HIW probabilities are in general below 50% (as the highest percentage of HIW events within one sector of Figure 2). The reliability diagrams of the Phi-22 and Mal-22 regions (Figure 4) show this limitation of the hybrid model, which only has forecast probabilities for the 0 to 0.2 and 0.2 to 0.4 probability bins. However, for those bins, the hybrid model, based on the real wave state (brown dotted line), is quite reliable, and it should also be noted that for higher probabilities the MOGREPS-G forecast does strongly deviate from the diagonal; so, although the MOGREPS-G is able to produce higher forecast probabilities, those high values seem to be associated with overforecasting the frequency of the HIW events.

However, although reliability of the hybrid model is quite good for the examples shown, the resolution is rather poor, which will also affect the Brier skill score (BSS) negatively. Going back to our example region of the Philippines

with the hybrid model based on R1, one can see that MOGREPS-G does have a BSS of above 0.3 for the first 24 hr lead time but then drops to a low value (green dash-dotted line in Figure 5). The hybrid model (orange solid line) does not reach a similarly high BSS, but it outperforms MOGREPS-G for all lead times exceeding 24 hr. Comparing the hybrid model forecast with the hybrid model based on reanalysis data we can see that one cannot expect much larger BSS values with a BSS expectation of about 0.1 if the model is able to predict the wave state correctly. If all lead time values are averaged, one can get to the same compressed information as for the AUC diagnostic (Figure 5b). As was the case for the AUC diagnostic, the combined method of MOGREPS-G and the hybrid model does lead to an improved BSS, highlighting again the additional benefit of the hybrid model. A notable feature of this compressed data is the large standard deviation of MOGREPS-G compared with the hybrid model, which is of course a result of the strong decline of the forecast metric with increasing lead time. This should be a general message from this compressed information plot; a large error range is expected to indicate a strong decrease of model performance for increasing lead time.

In the following, we will mainly focus on the AUC diagnostic. Including the BSS and reliability metric here was useful to highlight that the hybrid model does also give useful performance in other metrics, but also to show the limitations of the hybrid model, which can best be seen in the reliability diagram. This limitation of producing only relatively small probabilities for HIW event predictions can be seen not only for the Philippines, but all regions (not shown). This is, of course, what one should expect from a simple climatological statistical relationship between HIW

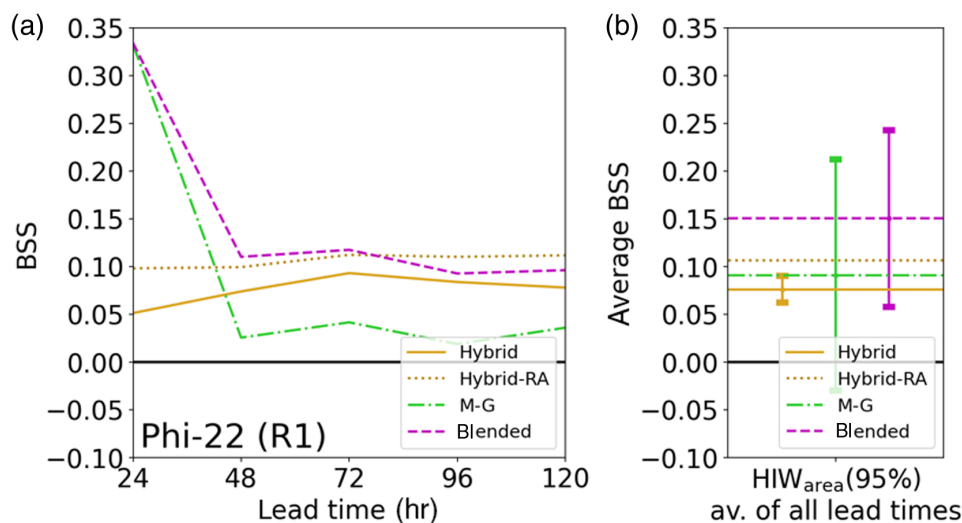


FIGURE 5 Brier skill score (BSS) for Phi-22 with hybrid model based on the R1 wave. (a) Lead-time-dependent BSS for Met Office Global and Regional Ensemble Prediction System (green dash-dotted), hybrid model (orange solid), the blended forecast (purple dashed), and the hybrid model based on reanalysis (brown dotted). (b) The average values of all those models with associated standard deviation. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/qj.4627)]

and equatorial wave modes. It seems plausible that clever weighting with increased probabilities for high values of the hybrid model while damping the lower noisy signal might lead to improved performances. However, the goal of this study is to understand the dynamical importance of the larger scale wave state in improving accuracy of forecasts of HIW, how forecast models compare in general with a much simpler dynamical–statistical model, as well as the question of whether higher resolution nested regional models can be expected to lead to further improvements. To that end, we will mainly focus on the AUC diagnostic.

3 | STATISTICAL-DYNAMICAL HYBRID FORECAST MODEL

In this section we will present the general link between HIW events and the wave state for all wave types (Kelvin, R1, and WMRG), all seasons and all regions (as given in Figure 1). Those results can be used to identify regions and seasons for which the statistical–dynamical hybrid model is likely to produce skilful forecasts. However, from a strong link between HIW events and the wave state one cannot directly conclude the associated forecast performance. The performance analysis is included in Section 4, where all models (MOGREPS-G, CP, and hybrid) are compared against each other, although only for the restricted time period of October–March. The results of this section are further used to identify for the various regions which wave type does show the strongest connection with HIW events and which is therefore being used for the hybrid model in Section 4.

The hybrid model has, however, two further relevant dependencies: the number of sectors within the wave phase space and the length of the training period. The analysis of these dependencies supports the choice of the

chosen eight wave phase-angle sectors and highlights the importance to use longer training periods, with the full 15-year training period in general showing better results than any of the individual 3-year subperiods (analysis in Section A in Supporting Information Figures S1–S3).

3.1 | General link between HIW events and the wave state for all wave types, seasons, and regions

A consistent feature of all wave-phase diagrams investigated with an underlying connection between HIW and equatorial waves is the strong anomalous values for the high-amplitude sectors (outer sectors in Figure 2, in agreement with the findings of Ferrett et al., 2023). Those high-amplitude values can therefore be used to show the seasonal variability of these connections. Using the Phi-22 region as an example, one can identify a strong connection of HIW and the R1 high-amplitude ($A > 2$) 135° to 225° labelled sectors for nearly all seasons, clearest for NH summer and autumn (Figure 6a). The high-amplitude sectors associated with the anticyclonic flow of the R1 wave (sectors 045° , 000° and 315°) indicate an absence of HIW events, represented by the strongly reduced values (blue colours).

The information is further condensed by averaging all 3-month periods related to a season (e.g., May–July, June–August, and July–September for NH summer) to make a further comparison between regions or wave types easier. The condensed information (Figure 6b) allows an easy identification of the most relevant wave sectors; for example, there is no HIW within Phi-22 for all seasons during the later stage of the anticyclonic flow phase of the R1 wave (sectors 000° and 315°). In contrast, the immediate transition into the early stage of the cyclonic flow wave phase (sectors 270° to 225°) is associated, for all seasons,

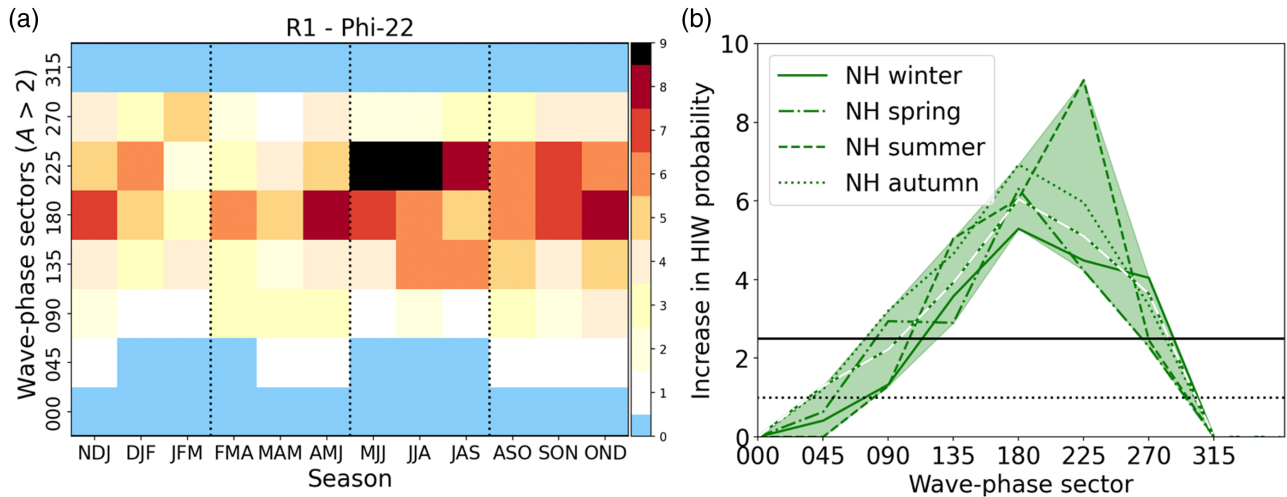


FIGURE 6 Link of high-impact weather (HIW) with R1 wave amplitudes for the Phi-22 region. HIW_{area} defined by the 95th percentile. (a) The seasonal variability of the connection between large-amplitude ($A > 2$) R1 waves and HIW. Values for wave-phase sector indicate the anticlockwise angle within the wave-phase diagram as given in Figure 2. (b) Compressed information for seasonal averages; solid line for the average of the Northern Hemisphere (NH) winter seasons (November–January [NDJ], December–February [DJF], January–March [JFM]), dash-dotted for NH spring (February–April [FMA], March–May [MAM], April–June [AMJ]), dashed for NH summer (May–July [MJJ], June–August [JJA], July–September [JAS]), and the dotted line for NH autumn (August–October [ASO], September–November [SON], October–December [OND]). Green/white dashed line shows the result for the extended 6-month training season (October–March). The horizontal dotted and solid line at a value of 1 and 2.5 separate the occurrence of HIW events into decreased (below 1), slightly to moderately increased (between 1 and 2.5), and strongly increased HIW probabilities (above 2.5). [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/qj.4627)]

with a strong increase of HIW probability by a factor of up to above 8 relative to the climatological probability of HIW events in this region.

The consecutive 3-month training seasons from October to March show a rather consistent signal (Figure 6a), although the signal changes a bit towards the end of this period. To reduce the variability of the results by increasing the available number of HIW events we will use the hybrid model trained for the October–March training season for comparing it with the CP ensemble forecasts in the next section. We therefore also include this extended 6-month training season in Figure 6b by the green/white dashed line. We use this extended 6-month season as it represents the period for which the CP data are available. For a better comparison with the results of the CP data, we used the 6-month period also for the results shown using MOGREPS-G.

The resulting connections for waves with HIW events within all available regions (Figure 1) are presented in Figure 7, which can be summarised as follows.

Peninsular Malaysia (Mal-22)

- Strong connection with convergent sector of Kelvin waves (blue area in Figure 7a), and some additional connection to WMRG waves within their transition phase into a positive vorticity (blue area in Figure 7c).

Indonesia

- Connection mainly with Kelvin waves slightly before the peak of the convergence sector (red area in Figure 7a), but also some further signal with WMRG waves within their transition phase into positive vorticity (red area in Figure 7c).
- Individual Indonesian main islands (Sumatra, Java, Borneo, and Sulawesi) all show strong connections with Kelvin waves within or close to the peak of the convergence sector, with Java showing the smallest signal (Figure 7d).
- Connection with WMRG waves within their transition phase into positive vorticity mainly linked to Java (green area in Figure 7f). This can be expected because Java is further from the Equator and therefore less influenced by Kelvin wave winds and more by WMRG structures.

Philippines

- Very strong connection with R1 and WMRG waves in a broad range of sectors in the transition sectors into and out of peak positive vorticity (green area in Figure 7b,c).
- This signal linked with westward waves is dominated by precipitation around tropical cyclones in summer and autumn (dashed and dotted lines).

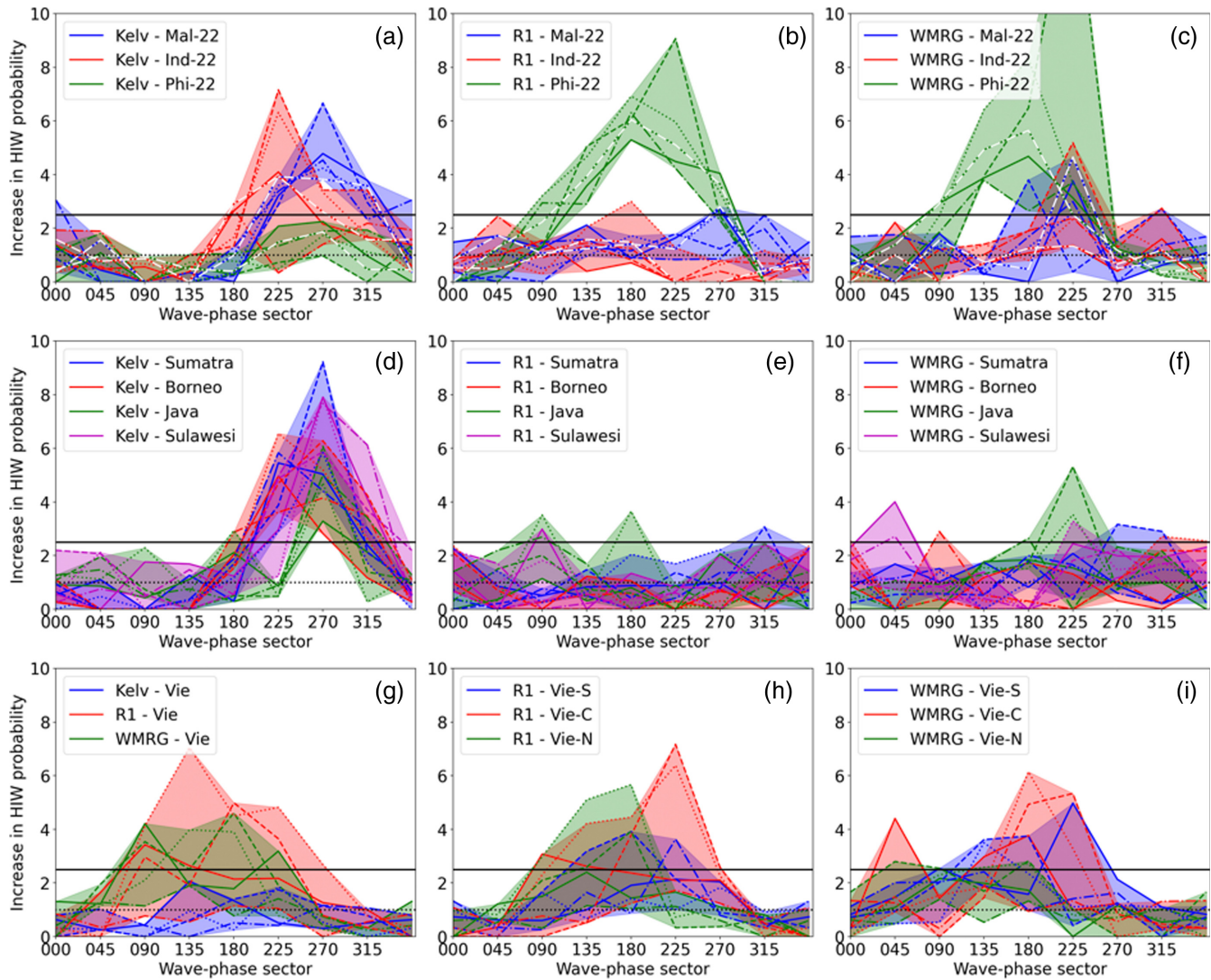


FIGURE 7 Link of HIW_{area} (95th percentile) with large-scale wave amplitudes for Malaysia, Indonesia, Philippines, and Vietnam. Values for wave-phase sector indicate the anticlockwise angle within the wave-phase diagram as given in Figure 2. Different line styles represent different seasons (solid: winter; dash-dotted: spring; dashed: summer; dotted: autumn). (a)–(c) The connection between high-impact weather (HIW) and (a) Kelvin, (b) R1, and (c) westward-moving mixed Rossby gravity (WMRG) waves for Mal-22, Ind-22 and Phi-22 regions. (d) Kelvin, (e) R1, and (f) WMRG waves for Sumatra, Java, Borneo, and Sulawesi. (g) All wave types for Vietnam; (h) R1 and (i) WMRG waves for the Vietnamese subregions (Vie-S, Vie-C and Vie-N). [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/terms-and-conditions)]

- More than 60% of tropical cyclone genesis cases occur within pre-existing westward equatorial waves (Feng et al., 2023), although it is also the case that the strong off-equatorial vorticity of the cyclone will project onto R1, R2, and WMRG wave structures.

Vietnam

- Connection is to R1 and WMRG waves (Figure 7g).
- The wave connection is strongest for northern and central Vietnam for R1 waves (red and green areas in Figure 7h)—interestingly for northern Vietnam temporally mainly after the main peak of positive

vorticity—and for southern and central Vietnam for WMRG waves around the sector of peak positive vorticity (blue and red area in Figure 7i), although this connection with R1 and WMRG exists for all subregions for at least one season.

Ferrett et al. (2023) show similar results for the hybrid model with four sectors instead of eight sectors for the regions and wave types that have a potential in predicting HIW events, according to their results using BSS. They also highlighted the strong connection of HIW events and Kelvin waves for different parts of Indonesia for June–August (Ferrett et al., 2023, fig. 5a–d; Figure 7d) and a clear connection with R1 and WMRG waves for northern

and southern Philippines in December–February (Ferrett et al., 2023, fig. 5e–h; Figure 7b,c). This R1 connection for the Philippines further shows why it might be a good idea to use eight sectors instead of four sectors for the hybrid model, as there is a very sharp transition between the sector centred at 315° and 270° (Figure 2b), which does lead to more average values when using only four 90° sectors centred at 0° , 90° , 180° , and 270° . However, their further separation of the Philippines into the northern and southern parts shows a slightly different appearance within the wave-phase diagram, which might explain the broader range of sectors with increased values for our Phi-22 region, including both the northern and southern Philippines.

4 | FORECAST HIW EVENTS BY CP ENSEMBLES COMPARED WITH THE HYBRID METHOD

The general diagnostic used to evaluate the HIW forecasts of a specific model was presented in Figure 3, resulting in one value with uncertainty for a 1 to 5 day lead time ensemble forecast for a specific region, and in terms of the hybrid model associated with one specific wave type. In the following we will compare the performances of all models (hybrid model, MORGREPS-G, and CP ensembles) for different regions, equatorial waves, extreme definitions, and model resolutions. The hybrid model can be used to identify to what extent HIW forecast skill can be improved by incorporating equatorial wave information. The comparison between MORGREPS-G and CP can give additional insight into the importance of higher resolution for HIW forecasts. Further, owing to the short evaluation period of the CP model (6 months), the results of MORGREPS-G (applied to both the same short 6-month period and a longer 4-year period) can be used to check the robustness and generality of the CP results. Qualitative similar results of MORGREPS-G for the 6-month and 4-year periods can be used to justify the same general conclusions for the CP ensembles, even if the analysis is done only for one season. Because of this, in this section we will apply the analysis for both time ranges step by step to all regions (as defined in Figure 1). Using additionally a different definition of HIW events will further help to better understand where the strength of the different models are as well as include further variability in the analysis to improve our understanding about the extent of the robustness of the results.

An important result for the hybrid model can be seen if the area is modified from common “HIW” (70th percentile) to very rare HIW events (97th percentile). The general feature of the hybrid model seems to be that it

performs better for more extreme events (orange lines in Figure 8), indicating that rare HIW events are the events that can be best associated with specific sectors in the wave-phase diagram. For this figure we also included the performance of the CP model, and therefore restricted all models to the ONDJFM 2018–2019 period. The hybrid model and the associated HIW definition in this figure are now based on the high-resolution IMERG grid, same for the CP model. The HIW definition for the evaluation of the MORGREPS-G performance is still based on the low-resolution MORGREPS-G grid, meaning that in this case the IMERG precipitation is, as previously, interpolated onto the associated low-resolution MORGREPS-G grid.

The comparison of MORGREPS-G and CP (4.5 km resolution) suggests an overall better performance of the CP model (blue dashed and green dash-dotted line in Figure 8). Whereas the CP model does a very good job in predicting HIW events within Phi-22, both MORGREPS-G and CP are performing very poorly for rare extreme events within Mal-22 and Ind-22. However, the CP forecast is much better than both global and hybrid model for Kelvin waves over Indonesia up to the 90th percentile. One interpretation is that explicitly represented deep convection is much better for simulating convectively coupled Kelvin waves. As previously highlighted, the blended forecast (this time CP4.5 and hybrid model) does perform better or at least as good as the better of both individual models for all HIW percentiles (purple dash-dotted lines in all panels).

A crucial difficulty of interpreting the model performance of extreme events in this study is the limited time range of 6 months for the CP model forecast set. This difficulty can already be seen by the noisy signal within the Mal-22 and Ind-22 regions, with no identified very rare HIW event (97th percentile) in Ind-22 for the investigation period. One should also be aware that the Hybrid-RA analysis of such a short period might not be a good indicator of the potential of the hybrid model, as the smaller number of extreme events might not be well represented by the climatological relationship. For example, the somewhat surprising difference for Phi-22 in Figure 8a,b, with the Hybrid-RA showing clearly lower values than the Hybrid line, can be interpreted as the HIW events rather occurring in the wave-phase diagram sectors next to the sectors with the strongest climatological relationship. So it still contains useful information about the HIW events within a specific season; however, for the Hybrid-RA to be used as an indicator of the potential of the hybrid model, longer time periods should be used. Because of this, and to better understand the robustness of the signal and the possible conclusions that can be derived from this investigation, we repeated the same calculation for the MORGREPS-G

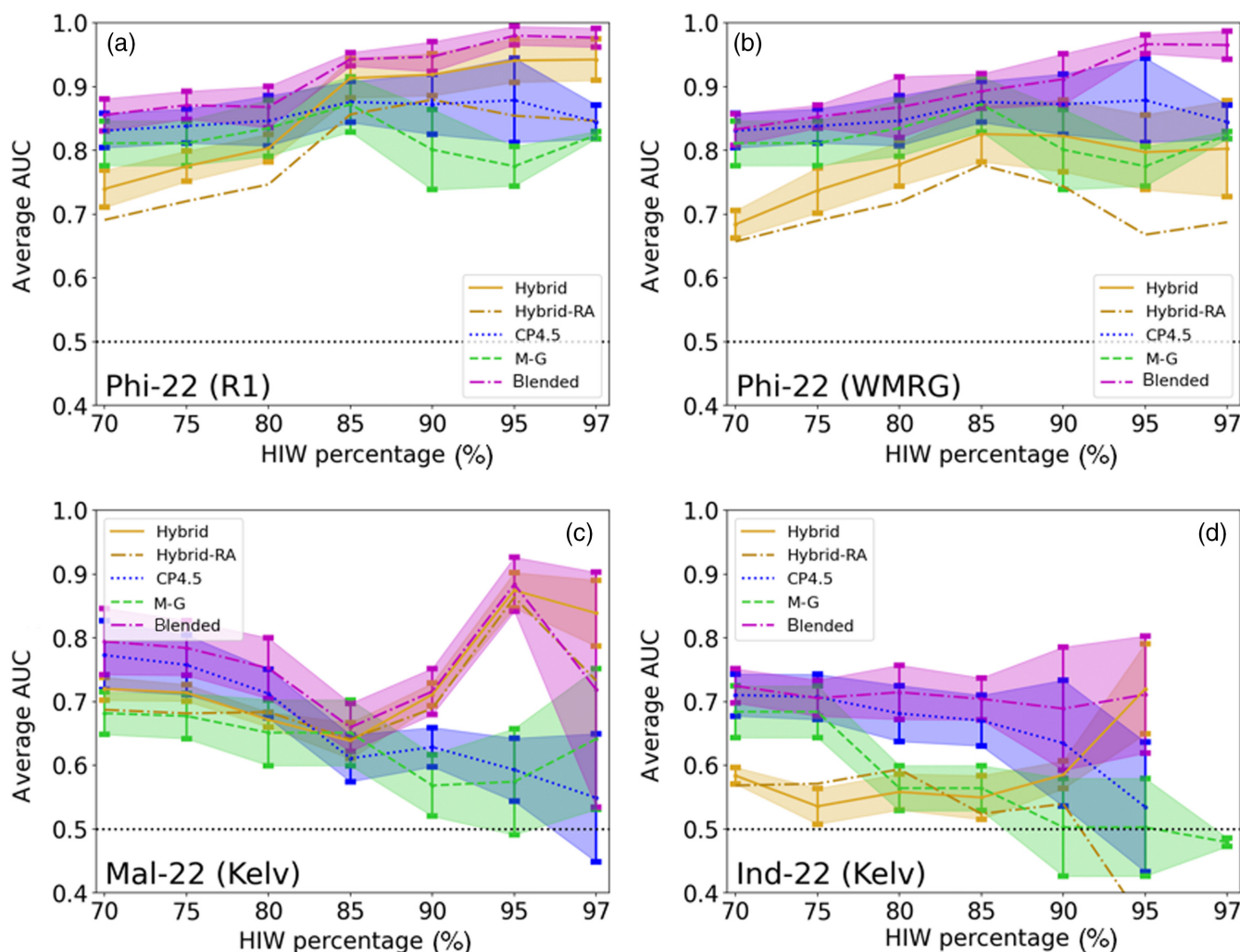


FIGURE 8 Impact of percentiles used for HIW_{area} on area under the receiver operating characteristic curve (AUC) diagnostic. The information about the underlying region (Phi-22, Mal-22, Ind-22) and hybrid-model-based wave state (Kelvin [Kelv], mode 1 Rossby [R1], westward-moving mixed Rossby gravity [WMRG]) are included in the lower left part of each panel. Blue line shows the values for the convection-permitting (CP) 4.5 km resolution model, dashed green line for Met Office Global and Regional Ensemble Prediction System, and orange lines for different hybrid models. Solid and dash-dotted orange lines represent hybrid model based on forecasts (solid line) and reanalysis data (dotted line), both using the high-resolution Integrated Multi-satellite Retrievals for Global Precipitation Measurement data. Purple dotted line represents the combined method blending the forecasts from the CP4.5 and the hybrid model. Evaluation period for all models is October–March 2018–2019. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/terms-and-conditions)]

grid-based curves (M-G and Hybrid) for the previously used 4-year period 2015–2019 (Figure 9). A comparison of the lines between Figures 8 and 9 can give some insight about the generality of the results. This additional analysis can therefore be used to confirm the better hybrid model performance for more extreme events. Although the qualitative behaviour of the hybrid model can be confirmed, its good performance for Mal-22 for the rare HIW events seems to be a special feature of this particular season (compare high percentile values for Figures 8c and 9c). An additional important result is the general underperformance of the hybrid model (compared with its values if based on reanalysis values Hybrid-RA) for Mal-22 and Ind-22,

which suggests that the Kelvin wave state is not adequately represented by the forecasts of MOGREPS-G (Figure 9c,d), confirming the analysis in Ferrett et al. (2023).

The underperformance of the hybrid model compared with its potential when based on the reanalysis wave state can be seen even more clearly when calculated for the Indonesian main islands separately (Figure 10). This identifies a crucial problem of the hybrid model in those regions but also highlights the general problem of the MOGREPS-G model in predicting the Kelvin wave state (Yang et al., 2021). Because of the model difficulty to predict the Kelvin wave state correctly, the hybrid model is performing below its potential that comes from

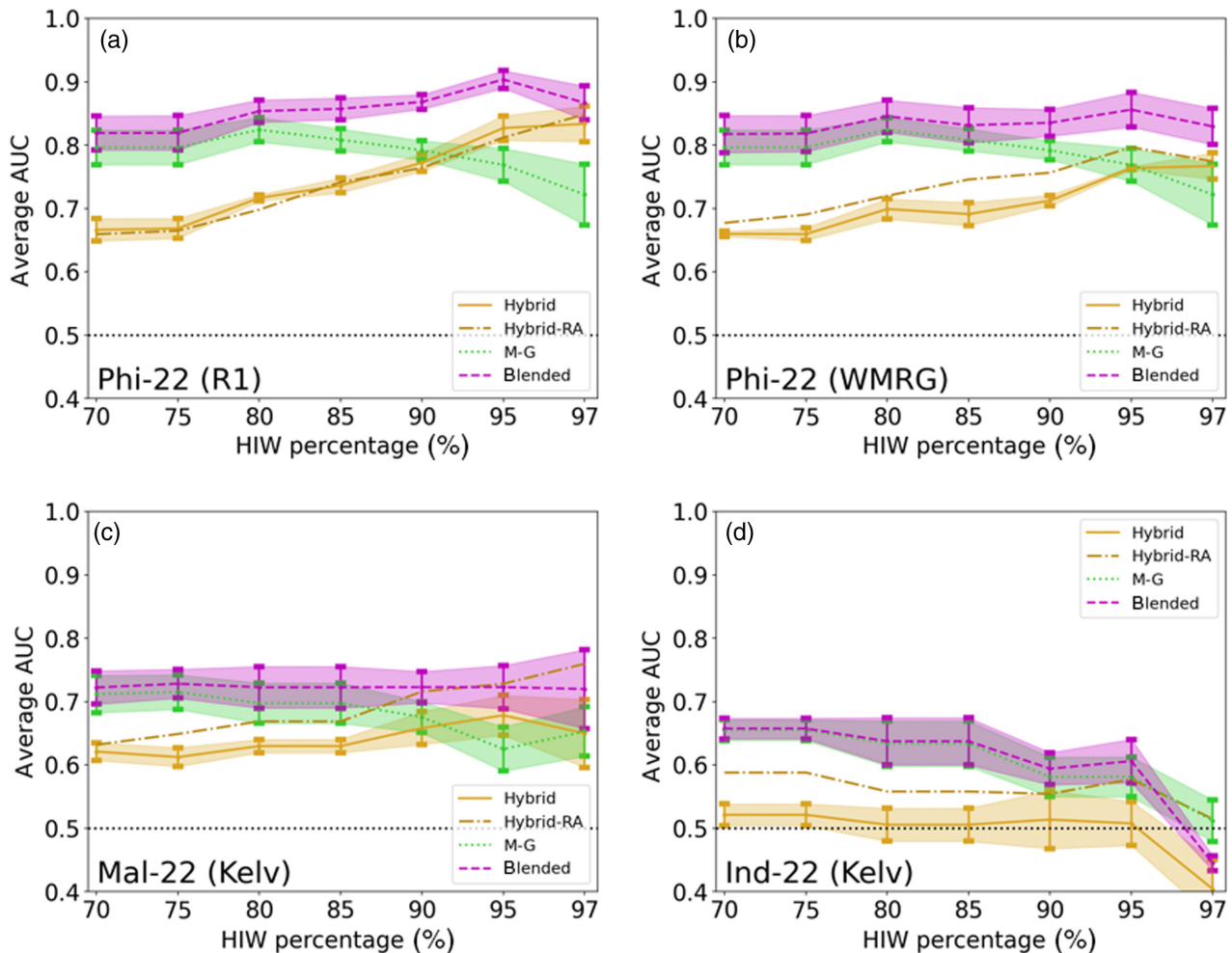


FIGURE 9 Same as Figure 8 but for the extended time range of October–March within 2015 to 2019 and without the convection-permitting model. All hybrid model lines are based on the Integrated Multi-satellite Retrievals for Global Precipitation Measurement data interpolated onto the M-G grid. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/terms-and-conditions)]

the identified dynamical connection between HIW events and the Kelvin wave state (as was indicated already in Figure 7a,d). Nevertheless, the hybrid model is still outperforming MOGREPS-G for the very rare HIW events; and as previously shown, the combined model (purple dashed line) leads to the best results for nearly all HIW percentages and all regions. The best performance of the hybrid model for Indonesia can be identified for Sumatra and Borneo (Figure 10a,c). The worst hybrid model performance can be seen for Java (Figure 10b), which explains the previous general statements about the bad hybrid model performance for Indonesia, which was based on Ind-22, the region with the largest overlap with Java (Figure 1). This worst hybrid model performance for Java is probably not that surprising, as this is the island further away from the Equator and therefore less influenced by the Kelvin wave mode.

For the Vietnamese regions, the direct precipitation forecasts of MOGREPS-G are, in general, a lot better than

the hybrid. One of the best hybrid model performances for Vietnam can be seen for Vie-C based on the R1 wave and Vie-S based on the WMRG wave (Figure 11). In general, the associated AUC values are quite low, except for central Vietnam for the more rare HIW events (Figure 11a). For southern Vietnam, the hybrid model would lead to better performances if based on the reanalysis wave state, indicating an underperformance of the hybrid model for the WMRG-wave-based predictions, as was the case for the Kelvin-wave-based prediction for all Indonesian sub-regions.

We also calculated the CP model performance based on the 8.8 km resolution for the restricted period of October 2018 to March 2019 for the Indonesian main islands, as shown in Figure 10 and the Vietnamese regions as shown in Figure 11. To be able to compare the results of the CP8.8 ensemble, we repeated the calculation for the MOGREPS-G ensemble for the same shorter 6-month time range. The results for Vietnam (Supporting Information

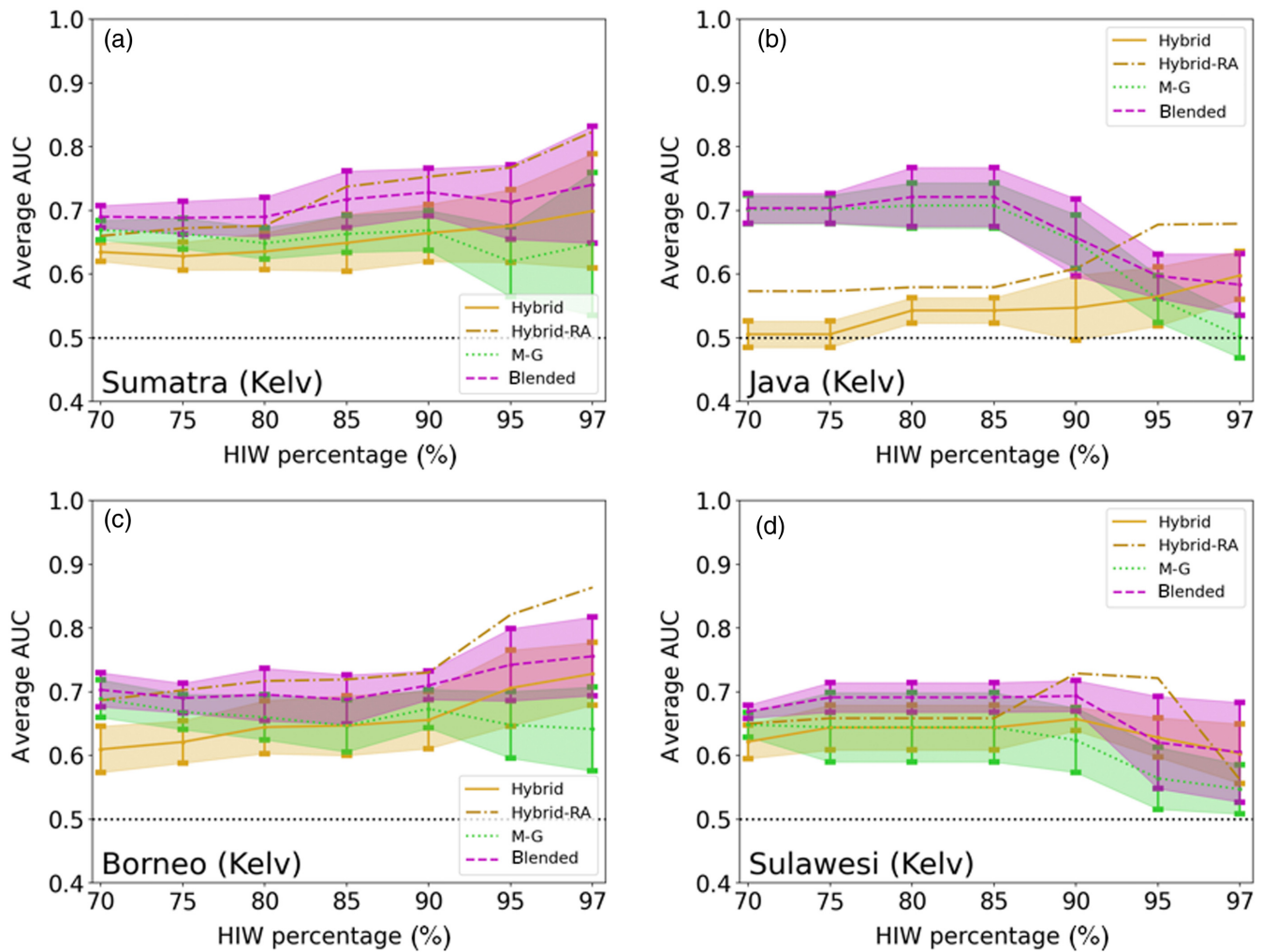


FIGURE 10 Same as Figure 9, but for different Indonesian parts with (a) Sumatra, (b) Java, (c) Borneo, and (d) Sulawesi. Additional results for the convection-permitting 8.8 km resolution simulations for the different Indonesian part are shown in Supporting Information Figure S5. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/terms-and-conditions)] [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/terms-and-conditions)]

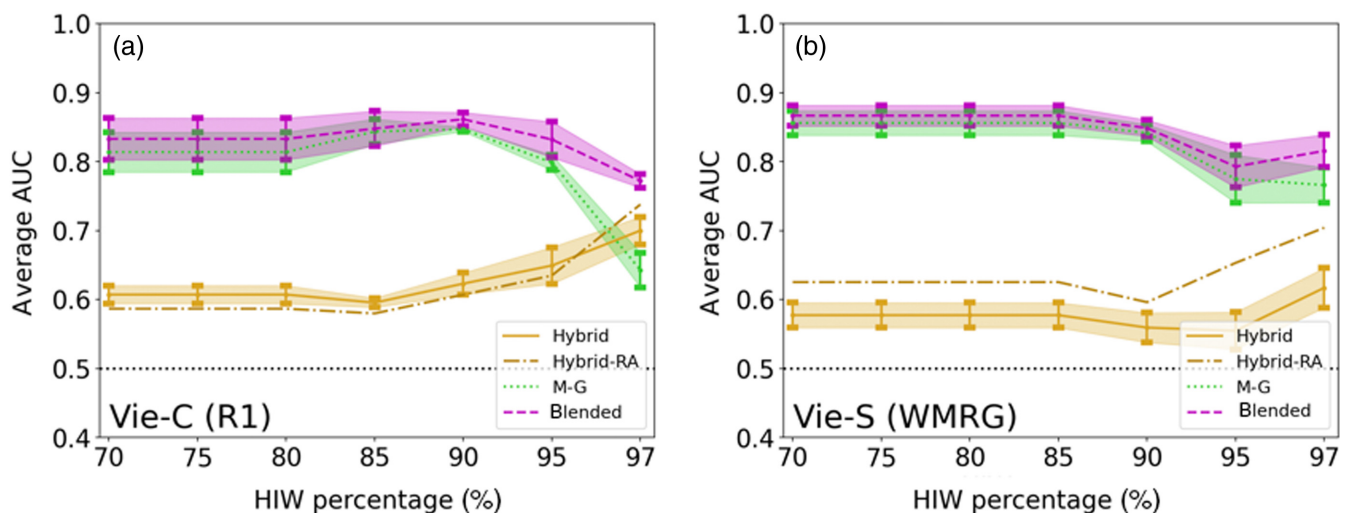


FIGURE 11 Same as Figure 10, but for two selected regions and wave states in Vietnam. (a) Results for Vie-C with hybrid model based on the R1 wave state and (b) for Vie-S with hybrid model based on the westward-moving mixed Rossby gravity (WMRG) wave state. Additional results for the convection-permitting 8.8 km resolution simulations for the different Indonesian part are shown in Supporting Information Figure S5. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/terms-and-conditions)]

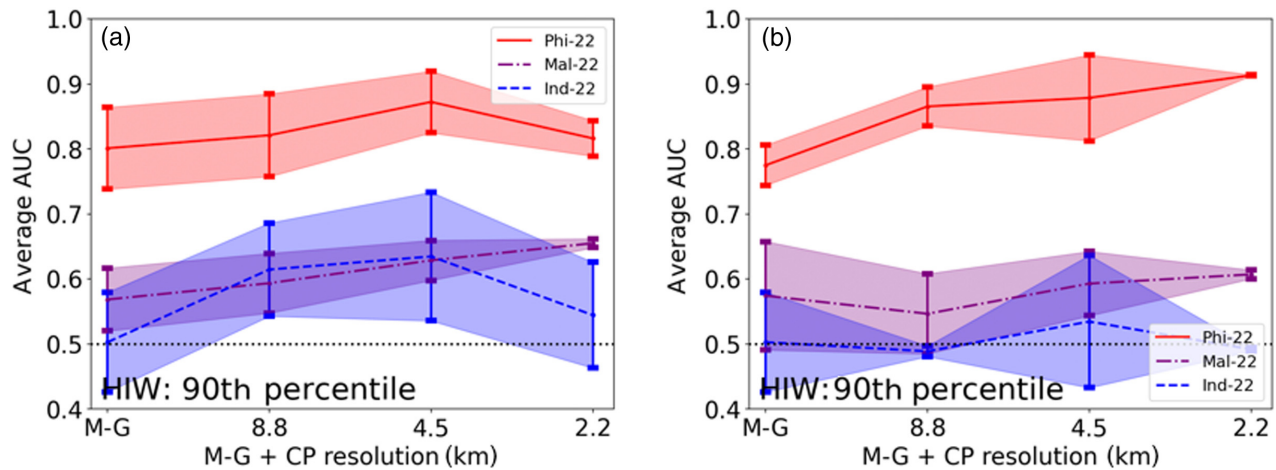


FIGURE 12 Impact of convection-permitting (CP) model resolution on area under the receiver operating characteristic curve (AUC) diagnostic for all 2.2 km model domains. High-impact weather (HIW) is defined by the (a) 90th and (b) 95th percentiles for HIW_{area} . For this analysis, only lead times up to 3 days (instead of 5 days) are used for averaging, as this is the available limit for the 2.2 km resolution CP ensemble. Different lines show the values for the different 22 regions, as given in the figure legends. M-G, Met Office Global and Regional Ensemble Prediction System (MOGREPS-G). [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/qj.4627)]

Figure S6) are in general similar to the extended time range (Figure 11), but with a noisier hybrid model performance and the CP8.8 ensemble showing slightly better performance than MOGREPS-G. The results of the MOGREPS-G ensemble for the shorter time range for Sumatra, Java, Borneo, and Sulawesi however show some larger differences compared to the extended time range (Supporting Information Figure S5). In general CP8.8 and MOGREPS-G show worse performances for rare HIW events, with the CP8.8 ensemble performing worse than MOGREPS-G and reaching AUC values of 0.5 for all regions for and above the 95th percentile, indicating that no such event was forecast. The hybrid model does show very good performances for such rare events for Sumatra, Borneo (as was the case for the extended time range), and Sulawesi, but unlike for the extended time range the hybrid model performance is very close to the reanalysis-based hybrid performance (Hybrid-RA). As these results deviate from the results of the extended time range, this seems to suggest that the HIW events for these regions and this specific season might be slightly unusual. Owing to the limited time range, it is difficult to draw conclusions from this beyond the ones already derived from Figure 10. However, the low performance of the CP8.8 model for Indonesia is rather surprising.

In terms of resolution for the CP model, higher resolution seems to lead to a further improvement (Figure 12) for the Phi-22 and Mal-22 regions. For Ind-22 this dependence is less clear with the 4.5 km resolution showing the best performance, but overall low values and huge uncertainties compared with the other regions. The improvement for Mal-22 is clearest with an increase in AUC of the order

of 0.1. These results suggest an improvement of the HIW forecast by increasing the model resolution; however, the time period and areas investigated are very limited (due to the expense of running CP ensemble forecast trials) and therefore conclusions about rare HIW events should be viewed with caution.

The applied HIW_{area} definition is a bit more sophisticated than a more commonly used simple precipitation threshold, as it does include a precipitation threshold at each grid point and an additional variable area criterion for the region. To understand the impact of this definition we also apply the alternative HIW definition HIW_{av} introduced in Section 2.2.1 that is based on average 24 hr accumulated precipitation over land without any further area criteria. The conclusion is very similar to the ones derived from varying HIW_{area} . The hybrid model is in general performing better for more extreme events, and the blended forecast of hybrid and MOGREPS-G is in general better than the better of both individual models (Supporting Information Figure S4). One can also identify again a stronger performance of the hybrid model if based on Kelvin waves for the reanalysis data for Mal-22 and Ind-22 (Supporting Information Figure S4c,d) and also for Phi-22, if based on WMRG waves (Supporting Information Figure S4b). An important difference, however, is that now MOGREPS-G is outperforming, or at least very similar to, the hybrid model for all HIW percentages and all regions. This means that there seems to be a relevant difference between localized HIW events and HIW events that occur over a larger area, with the hybrid model performing better than MOGREPS-G if the additional area criteria (HIW_{area} compared with HIW_{av}) is used. This difference

seems plausible, as the accompanying wave signal also occurs over a larger area.

5 | SUMMARY AND DISCUSSION

This work investigated the skill of probabilistic forecasts of high-impact rainfall events from the MOGREPS-G global ensemble, a CP limited-area version of the MOGREPS ensemble, and a hybrid statistical–dynamical forecast based on climatological rainfall conditioned on global ensemble forecasts of the phase and amplitude of tropical waves. The main findings are as follows,

The statistical–dynamical hybrid forecast model introduced is able to predict HIW events, characterised by widespread heavy precipitation linked with the passage of equatorial waves. Applied to the different regions within southeast Asia, this model illustrates the connections between relevant wave types and HIW. For Malaysia the main connection is with Kelvin waves, for the Philippines and Vietnam it is R1 and WMRG waves, and for Indonesia (Sumatra, Java, Borneo, and Sulawesi) it is Kelvin waves, with Java also showing a connection with WMRG waves. Applying the hybrid model to the different seasons further shows when those connections are strongest.

Owing to the strong connection between HIW and equatorial waves, the hybrid model can be used to generate dynamically conditioned probabilistic HIW forecasts, based on the global model prediction of the large-scale horizontal wind and geopotential height fields. Ferrett et al. (2023) have shown that their hybrid model, based on four angle sectors within the wave-phase diagram, does have higher skill for some regions within southeast Asia (mainly Kelvin waves in June–August for Indonesia, R1 and WMRG waves in December–February for the Philippines, Vietnam, and Thailand) compared with the direct simulation of precipitation in MOGREPS-G forecasts. In this study we investigated in detail the sensitivity of the hybrid model to specific parameter set-ups (training period, number of sectors in the wave-phase diagram) and introduced a reference hybrid model that is based on reanalysis wave data instead of forecasts. The introduction of this reanalysis-based hybrid model allows the identification of the extent to which the hybrid model reaches its expected theoretical potential or, if there is a problem in the global model, to correctly predict the underlying wave state. For the Philippines, the hybrid model based on the R1 wave state reaches its potential with very good performances. However, the reanalysis-based approach suggests a higher potential predictability for Malaysia and Indonesia that is not realised because the skill in forecasting Kelvin waves falls rapidly in global model forecasts (Yang et al., 2021). This result suggests that including a nested

high-resolution CP ensemble run for the Mal-22 or Ind-22 region will have a somewhat limited HIW forecast benefit as long as the large-scale conditions, which have been shown to be connected to HIW events, are not accurately represented.

We have now also investigated how the hybrid model (applied to MOGREPS-G equatorial wave data) performs in comparison with the higher resolution CP ensemble forecasts of the Met Office's Met-UM model. In general, the CP ensembles perform better than MOGREPS-G but still worse than the hybrid model, at least for predicting the rare HIW events. This means that a relevant part of the HIW events is controlled by the large-scale equatorial waves, explaining why such a simple statistical model is able to produce such good HIW forecasts. As a consequence, this means that a high-resolution nested CP model will not be able to predict those HIW events correctly as long as the global model does not represent the large-scale wave state correctly—for example, see the case study of Senior et al. (2023). Details about the evaluation of the CP ensemble forecasts and the aspects where they perform well relative to the global ensemble are presented in Ferrett et al. (2021). However, this does not explain why the hybrid model is in some instances able to outperform the CP model. A possible explanation might be that the sector approach of the hybrid model does allow for some temporal mismatch in the forecast of the wave state and the occurrence of the HIW event. The CP model shows systematic improvement in HIW forecasts as the resolution increases for the Philippines and Malaysia, but no clear tendency can be seen for Indonesia, with the best performance found for the 4.5 km model (Figure 12). Significantly, this model is much more skilful than the global or hybrid models over Indonesia (Figure 8d). This indicates that the convective coupling with Kelvin waves is much better represented in this CP model (but not as good in the 2.2 km model).

However, the limited investigation period and the rare occurrence of HIW events somewhat limit the possible conclusions. This is a general issue with the analysis presented, as the investigation of HIW events within a 6-month period will always be associated with large uncertainties and conclusions must be interpreted carefully. However, to further support our conclusions, we compared the results of the hybrid model and MOGREPS-G between the short 6-month period and an extended 4-year period. This comparison allows a better understanding of which identified model behaviour is robust or maybe rather a specific feature of the short investigation period. We further applied several sensitivity tests and different hybrid model set-ups or HIW definitions to allow the identification of the most robust results, which are all highlighted in this conclusion.

We are confident in the conclusion that the hybrid model is, in general, very capable of capturing HIW events, and for some regions within southeast Asia it will outperform (or at least lead to similar performances to) the MOGREPS-G and even the CP model. The hybrid model could therefore be used as a real benchmark test for any higher resolution precipitation forecasts in predicting HIW events. As the hybrid model forecasts are based on the larger scale horizontal flow, there seems to be the potential for the hybrid model to extend the lead time of HIW risk forecasts in situations where the global model is able to capture the large-scale dynamics of the most predictable components.

Interestingly, combining the probability forecasts of the hybrid model with MOGREPS-G or the CP model into one blended forecast actually led to an improved forecast (compared with both individual models) for nearly every region. This might be interpreted as evidence that the ensemble model forecasts are underdispersive, able to capture several HIW events but missing several other instances where the large-scale conditions suggest an increased climatologically based risk for an HIW event. In other words, the CP ensemble forecasts of HIW probability are sharper, but the hybrid forecasts have higher reliability: the blended forecast achieves an improved compromise between sharpness and reliability. This is in agreement with the investigated weighting for hybrid model and MOGREPS-G or CP model for the blended forecast, with the observed improvement resulting from small weighting for the hybrid model. In addition, the hybrid model gives useful insight into the role of equatorial waves in the occurrences of HIW events. The link between the risk of HIW and the large-scale precursor waves could be useful in the communication of forecast risk with stakeholders, making connections with past high-impact events (with similar large-scale conditions) and taking appropriate action to reduce impacts.

AUTHOR CONTRIBUTIONS

G. Wolf: formal analysis; investigation; software; visualization; writing – original draft. **S. Ferrett:** methodology; software; writing – review and editing. **J. Methven:** conceptualization; funding acquisition; supervision; writing – review and editing. **T.H.A. Frame:** conceptualization; funding acquisition; supervision; writing – review and editing. **C.E. Holloway:** conceptualization; funding acquisition; supervision; writing – review and editing. **O. Martinez-Alvarado:** funding acquisition; supervision; writing – review and editing. **S.J. Woolnough:** conceptualization; funding acquisition; supervision; writing – review and editing.

ACKNOWLEDGMENTS

The ERA5 equatorial wave dataset was produced by Dr Gui-Ying Yang of the National Centre for Atmospheric Science, University of Reading, supported by the Met Office Weather and Climate Science for Service Partnership (WCSSP) Southeast Asia as part of the Newton Fund. This work and its contributors were also supported by the Newton Fund WCSSP project Forecasting for Southeast Asia (FORSEA). O. Martinez-Alvarado and S. J. Woolnough were also supported by the National Centre for Atmospheric Science through the NERC National Capability International Programmes Award (NE/X006263/1).

ORCID

G. Wolf  <https://orcid.org/0000-0001-5704-6550>

S. Ferrett  <https://orcid.org/0000-0003-4726-847X>

C.E. Holloway  <https://orcid.org/0000-0001-9903-8989>

O. Martinez-Alvarado  <https://orcid.org/0000-0002-5285-0379>

REFERENCES

- Bengtsson, L., Dias, J., Gehne, M., Bechtold, P., Whitaker, J., Bao, J.-W. et al. (2019) Convectively coupled equatorial wave simulations using the ecmwf ifs and the noaa gfs cumulus convection schemes in the noaa gfs model. *Monthly Weather Review*, 147(11), 4005–4025. Available from: <https://doi.org/10.1175/MWR-D-19-0195.1> <https://journals.ametsoc.org/view/journals/mwre/147/11/mwr-d-19-0195.1.xml>
- Bowler, N.E., Arribas, A., Mylne, K.R., Robertson, K.B. & Beare, S.E. (2008) The mogreps short-range ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 134(632), 703–722. Available from: <https://doi.org/10.1002/qj.234>
- Cafaro, C., Frame, T.H.A., Methven, J., Roberts, N. & Bröcker, J. (2019) The added value of convection-permitting ensemble forecasts of sea breeze compared to a bayesian forecast driven by the global ensemble. *Quarterly Journal of the Royal Meteorological Society*, 145(721), 1780–1798. Available from: <https://doi.org/10.1002/qj.3531>
- Chang, C.-P., Harr, P.A. & Chen, H.-J. (2005) Ynoptic disturbances over the equatorial South China Sea and Western maritime continent during boreal winter. *Monthly Weather Review*, 133(3), 489–503. Available from: <https://doi.org/10.1175/MWR-2868.1>
- Dias, J., Gehne, M., Kiladis, G.N., Sakaeda, N., Bechtold, P. & Haiden, T. (2018) Equatorial waves and the skill of ncep and ecmwf numerical weather prediction systems. *Monthly Weather Review*, 146(6), 1763–1784. Available from: <https://doi.org/10.1175/MWR-D-17-0362.1>
- Feng, X., Yang, G.Y., Hodges, K. & Methven, J. (2023) Equatorial waves as useful precursors to tropical cyclone occurrence and intensification. *Nature Communications*, 14(1), 511. Available from: <https://doi.org/10.1038/s41467-023-36055-5>
- Ferrett, S., Frame, T.H.A., Methven, J., Holloway, C.E., Webster, S., Stein, T.H.M. et al. (2021) Evaluating convection-permitting ensemble forecasts of precipitation over southeast asia. *Weather and Forecasting*, 36(4), 1199–1217. Available from: <https://doi.org/10.1175/WAF-D-20-0216.1>

- Ferrett, S., Methven, J., Woolnough, S.J., Yang, G.-Y., Holloway, C.E. & Wolf, G. (2023) Hybrid dynamical-statistical forecasts of the risk of rainfall in southeast asia conditional on equatorial waves. *Monthly Weather Review*, 151(8), 2139–2152. Available from: <https://doi.org/10.1175/MWR-D-22-0300.1>
- Ferrett, S., Yang, G.-Y., Woolnough, S.J., Methven, J., Hodges, K. & Holloway, C.E. (2020) Linking extreme precipitation in southeast asia to equatorial waves. *Quarterly Journal of the Royal Meteorological Society*, 146(727), 665–684. Available from: <https://doi.org/10.1002/qj.3699>
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J. et al. (2020) The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. Available from: <https://doi.org/10.1002/qj.3803>
- Huffman, G.J., David, T.B., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C. et al. (2019) NASA Global Precipitation Measurement (GPM) Integrated Multi-satellite Retrievals for GPM (IMERG). Algorithm Theoretical Basis Doc., version 06, 38 pp., https://gpm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V06.pdf.
- Judt, F. (2020) Atmospheric predictability of the tropics, middle latitudes, and polar regions explored through global storm-resolving simulations. *Journal of the Atmospheric Sciences*, 77(1), 257–276. Available from: <https://doi.org/10.1175/JAS-D-19-0116.1>
- Juneng, L., Tangang, F.T. & Reason, C.J.C. (2007) Numerical case study of an extreme rainfall event during 9–11 December 2004 over the east coast of peninsular Malaysia. *Meteorology and Atmospheric Physics*, 98, 81–98. Available from: <https://doi.org/10.1007/s00703-006-0236-1>
- Kirschbaum, D., Stanley, T. & Zhou, Y. (2015) Spatial and temporal analysis of a global landslide catalog. *Geomorphology*, 249, 4–15. Available from: <https://doi.org/10.1016/j.geomorph.2015.03.016>
- Maier-Gerber, M., Fink, A.H., Riemer, M., Schoemer, E., Fischer, C. & Schulz, B. (2021) Statistical–dynamical forecasting of subseasonal North Atlantic tropical cyclone occurrence. *Weather and Forecasting*, 36(6), 2127–2142. Available from: <https://doi.org/10.1175/WAF-D-21-0020.1>
- Merz, B., Kuhlicke, C., Kunz, M., Pittore, M., Babeyko, A., Bresch, D.N. et al. (2020) Impact forecasting to support emergency management of natural hazards. *Reviews of Geophysics*, 58(4), e2020RG000704. Available from: <https://doi.org/10.1029/2020RG000704>
- Schlueter, A., Fink, A.H., Knippertz, P. & Vogel, P. (2019) A systematic comparison of tropical waves over northern Africa. Part i: influence on rainfall. *Journal of Climate*, 32(5), 1501–1523. Available from: <https://doi.org/10.1175/JCLI-D-18-0173.1>
- Senior, N.V., Matthews, A.J., Webber, B.G.M., Webster, S., Jones, R.W., Permana, D.S. et al. (2023) Extreme precipitation at Padang, Sumatra triggered by convectively coupled kelvin waves. *Quarterly Journal of the Royal Meteorological Society*, 149(755), 2281–2300. Available from: <https://doi.org/10.1002/qj.4506>
- Tangang, F.T., Juneng, L., Salimun, E., Vinayachandran, P.N., Seng, Y.K., Reason, C.J.C. et al. (2008) On the roles of the northeast cold surge, the Borneo vortex, the madden-Julian oscillation, and the Indian Ocean dipole during the extreme 2006/2007 flood in southern peninsular Malaysia. *Geophysical Research Letters*, 35(14). Available from: <https://doi.org/10.1029/2008GL033429>
- van der Linden, R., Fink, A.H., Pinto, J.G., Phan-Van, T. & Kiladis, G.N. (2016) Modulation of daily rainfall in southern Vietnam by the madden-Julian oscillation and convectively coupled equatorial waves. *Journal of Climate*, 29(16), 5801–5820. Available from: <https://doi.org/10.1175/JCLI-D-15-0911.1>
- Vogel, P., Knippertz, P., Fink, A.H., Schlueter, A. & Gneiting, T. (2020) Skill of global raw and postprocessed ensemble predictions of rainfall in the tropics. *Weather and Forecasting*, 35(6), 2367–2385. Available from: <https://doi.org/10.1175/WAF-D-20-0082.1>
- Walz, E.-M., Maranan, M., van der Linden, R., Fink, A.H. & Knippertz, P. (2021) An imerg-based optimal extended probabilistic climatology (epc) as a benchmark ensemble forecast for precipitation in the tropics and subtropics. *Weather and Forecasting*, 36(4), 1561–1573. Available from: <https://doi.org/10.1175/WAF-D-20-0233.1>
- Woodhams, B.J., Birch, C.E., Marsham, J.H., Bain, C.L., Roberts, N.M. & Boyd, D.F.A. (2018) What is the added value of a convection-permitting model for forecasting extreme rainfall over tropical east africa? *Monthly Weather Review*, 146(9), 2757–2780. Available from: <https://doi.org/10.1175/MWR-D-17-0396.1>
- Xavier, P., Rahmat, R., Cheong, W.K. & Wallace, E. (2014) Influence of madden-Julian oscillation on Southeast Asia rainfall extremes: observations and predictability. *Geophysical Research Letters*, 41(12), 4406–4412. Available from: <https://doi.org/10.1002/2014GL060241>
- Yang, G.-Y., Hoskins, B.J. & Slingo, J.M. (2003) Convectively coupled equatorial waves: a new methodology for identifying wave structures in observational data. *Journal of the Atmospheric Sciences*, 60(14), 1637–1654. Available from: [https://doi.org/10.1175/1520-0469\(2003\)060<1637:CCEWAN>2.0.CO;2](https://doi.org/10.1175/1520-0469(2003)060<1637:CCEWAN>2.0.CO;2)
- Yang, G.-Y., Ferrett, S., Woolnough, S., Methven, J. & Holloway, C. (2021) Real-time identification of equatorial waves and evaluation of waves in global forecasts. *Weather and Forecasting*, 36(1), 171–193. Available from: <https://doi.org/10.1175/WAF-D-20-0144.1>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Wolf, G., Ferrett, S., Methven, J., Frame, T., Holloway, C., Martinez-Alvarado, O. *et al.* (2023) Comparison of probabilistic forecasts of extreme precipitation for a global and convection-permitting ensemble and hybrid statistical–dynamical method based on equatorial wave information. *Quarterly Journal of the Royal Meteorological Society*, 1–20. Available from: <https://doi.org/10.1002/qj.4627>