Evaluation of Asian summer precipitation in different configurations of a high-resolution GCM at a range of decision-relevant spatial scales

Mark R. Muetzelfeldt[1], Reinhard Schiemann[1], Andrew G. Turner[1], Nicholas P. Klingaman[1], Pier Luigi Vidale[1], and Malcolm J. Roberts[2]

[1]National Centre for Atmospheric Science, Department of Meteorology, University of Reading, Reading, UK
[2]Met Office Hadley Centre, Exeter, UK

Correspondence: Mark R. Muetzelfeldt (mark.muetzelfeldt@reading.ac.uk)

Abstract. High-resolution general circulation models (GCMs) can provide new insights into the simulated distribution of global precipitation. We evaluate how summer precipitation is represented over Asia in global simulations with a grid length of $14\,\mathrm{km}$. Three simulations were performed: one with a convection parametrization, one with convection represented explicitly by the model's dynamics, and a hybrid simulation with only shallow and mid-level convection parametrized. We evaluate

5   the mean simulated precipitation and the diurnal cycle of the amount, frequency and intensity of the precipitation against satellite observations of precipitation from the Climate Prediction Center morphing method (CMORPH). We also compare the high-resolution simulations with coarser simulations that use parametrized convection.

The simulated and observed precipitation is averaged over spatial scales defined by the hydrological catchment basins; these provide a natural spatial scale for performing decision-relevant analysis that is tied to the underlying regional physical

10   geography. By selecting basins of different sizes, we evaluate the simulations as a function of the spatial scale. A new BAsin-Scale Model Assessment ToolkIt (BASMATI) is described, which facilitates this analysis.

We find that there are strong wet biases (locally up to $72\,\mathrm{mm\,day^{-1}}$ at small spatial scales) in the mean precipitation over mountainous regions such as the Himalayas. The explicit convection simulation worsens existing wet and dry biases compared to the parametrized convection simulation. When the analysis is performed at different basin scales, the precipitation bias

15   decreases as the spatial scales increase for all simulations; the lowest-resolution simulation has the smallest root mean squared error compared to CMORPH.

In the simulations, a positive mean precipitation bias over China is primarily found to be due to too frequent precipitation for the parametrized convection simulation, and too intense precipitation for the explicit convection simulation. The simulated diurnal cycle of precipitation is strongly affected by the representation of convection: parametrized convection produces a

20   peak in precipitation too close to midday over land, whereas explicit convection produces a peak that is closer to the late afternoon peak seen in observations. At increasing spatial scale, the representation of the diurnal cycle in the explicit and hybrid convection simulations improves when compared to CMORPH; this is not true for any of the parametrized simulations.

Some of the strengths and weaknesses of simulated precipitation in a high-resolution GCM are found: the diurnal cycle is improved at all spatial scales with convection parametrization disabled; the interaction of the flow with orography exacerbates

25 existing biases for mean precipitation in the high-resolution simulations; and parametrized simulations produce similar diurnal cycles regardless of their resolution. The need for tuning the high-resolution simulations is made clear. Our approach for evaluating simulated precipitation across a range of scales is widely applicable to other GCMs.

1 Introduction

Simulating the summer precipitation over Asia in atmospheric models poses many challenges. Atmospheric phenomena span-

35 ning a range of spatial and temporal scales are important. The large-scale circulation is dominated by the Indian summer monsoon (ISM) and East Asian summer monsoon (EASM) flows. At subseasonal and interannual timescales, both of these flows are affected by the boreal summer intraseasonal oscillation (Ajayamohan et al., 2008; Hsu et al., 2016), El Niño–Southern Oscillation (Wu and Wang, 2002; Xavier et al., 2007), and the Indian Ocean Dipole (Ajayamohan et al., 2008; Ding et al., 2010). At shorter time and space scales, the presence of depressions, tropical cyclones and mesoscale convective systems can

40 influence both monsoons (Zhong and Hu, 2007; Yang et al., 2015; Virts and Houze Jr, 2016; Zhang et al., 2018); the EASM can be affected by Tibetan Plateau vortices (Curio et al., 2018). The presence of high mountains affects the circulation and leads to extremes in orographically induced precipitation. Traditional General Circulation Models (GCMs) are designed to represent the larger scales, but at low resolution cannot capture the smaller scales, the effects of orography or the interactions between scales. Regional climate models may simulate the small-scale features, but cannot represent the upscale interactions and inherit

45 biases from their driving GCMs. To address both of these issues, we analyse the precipitation produced by a new generation of high-resolution global models, which can represent more of the scales of monsoon flow in a physically consistent way.

High-resolution GCM simulations (here, 14 km grid length at 30° N) are now able to simulate multiple years of global atmospheric conditions. This resolution is comparable to high-resolution regional climate models from approximately 2010 (Hurkmans et al., 2010). Due to the computational cost of running these simulations, it was only feasible to run them for four

50 years, and so we have chosen to analyse the seasonal mean and the diurnal cycle of precipitation. Both regionally and globally, simulated precipitation has been found to improve with increased resolution in some GCMs (e.g., Haarsma et al., 2016; Zhang et al., 2016; Schiemann et al., 2018; Vannière et al., 2019), but not in others run globally (e.g., Bador et al., 2019); we compare the high-resolution simulations with coarser ones to address this point. The high resolution and global nature of these simulations makes them well-suited, in principle, to address the challenges of simulating precipitation over Asia. For example,

55 orography is more accurately resolved, which can improve simulated precipitation (Schiemann et al., 2018; Vannière et al.,

2019). Scale interactions between large-scale atmospheric modes of variability can be modelled without artificial boundaries introduced when a large-scale GCM provides boundary conditions for a regional climate model.

The high-resolution simulations use grid lengths that can be considered "convection permitting". That is, they can simulate convection using the model's internal dynamics of atmospheric motions and grid cell cloud parametrization scheme, by disabling the model's convection parametrization scheme. These convection parametrization schemes are responsible for modelling the sub-grid moist convection; they represent the statistical effects of cumulus clouds at the grid scale, and generally rely upon the quasi-equilibrium assumption stated in Arakawa and Schubert (1974). Although a grid length of 14 km is quite coarse to use without a convection parametrization scheme, studies have been carried out at comparable and coarser resolutions with the convection parametrization scheme disabled to understand the effects that this has on regional models and GCMs (Holloway et al., 2013; Satoh et al., 2014; Webb et al., 2015; Becker et al., 2017; Schär et al., 2020; Vergara-Temprado et al., 2020). Simulations with 12 km grid lengths and explicit convection have produced more realistic West African and Indian monsoons than equivalent parametrized simulations (Marsham et al., 2013; Willetts et al., 2017), with the diurnal cycle of precipitation in particular being improved.

The representation of precipitation in GCMs is usually evaluated on a grid-point basis, as defined by the grid columns of the GCM, possibly with upscaling to the coarsest resolution dataset being used (Li et al., 2018). Klingaman et al. (2017) and Martin et al. (2017) recommend upscaling to coarser than the coarsest dataset, so that all datasets are subject to some upscaling. Here, we take a different approach. In the same manner as Schiemann et al. (2018), we choose to use hydrological catchment basins as the spatial units for our analysis. Basins determine how the river network experiences precipitation, and thus represent the relevant interface between the atmosphere and the land surface. They aggregate the incoming precipitation, and so represent decision-relevant spatial scales over which to analyse the precipitation, with possible implications for flooding, droughts, hydroelectricity and agriculture. They are closely tied to the underlying physical geography of the region, such as the orography and the land surface. To this end, we have developed the BAsin-Scale Model Assessment ToolKit (BASMATI, Sect. 2.4.2), which is freely available. BASMATI can select catchment basins of different sizes, which allows us to perform scale-selective analysis.

Many studies have shown that the diurnal cycle of precipitation is not well represented in climate models (e.g., Yang and Slingo, 2001; Dai and Trenberth, 2004; Marsham et al., 2013; Covey et al., 2016). The convection parametrization scheme has been identified as one of the main reasons for this deficiency (e.g. Stirling and Stratton, 2012; Marsham et al., 2013; Willetts et al., 2017; Li et al., 2018). Furthermore, the representation of the diurnal cycle of precipitation varies with model resolution, with higher-resolution models producing more realistic diurnal cycles (Khairoutdinov et al., 2005; Sato et al., 2009; Ploshay and Lau, 2010). Since the diurnal cycle of precipitation captures the sub-daily precipitation, its representation is important to provide suitable input to land surface models (Sheffield et al., 2006; Reichle et al., 2011). For these reasons, we evaluate the simulated diurnal cycle of the precipitation against the observations, and perform scale-selective analysis of the diurnal cycle over different sizes of catchment basin using BASMATI. By doing this, we can ascertain how the diurnal cycle is affected by the presence of a convection parametrization scheme and model resolution, and we can determine how it responds over larger decision-relevant spatial scales in a similar manner to Covey et al. (2016).

The aim of this study is to evaluate the precipitation produced by different configurations of a high-resolution GCM over Asia, by comparing simulations against satellite observations. We do this by varying the resolution and representation of convection in order to learn about how the precipitation is affected by these, and how the 14 km simulations perform. We assess both the mean seasonal precipitation and diurnal cycle. Using BASMATI, we carry out the analysis over a hierarchy

95 of spatial scales defined by catchment basins. Our results evaluate the new capabilities offered by high-resolution models in representing precipitation and its diurnal cycle over Asia at different spatial scales. Additionally, our results should inform land surface and hydrological modelling, as the spatial distribution of precipitation and its diurnal cycle are key drivers of these, although evaluating such coupled setups is beyond the scope of this study.

The remainder of this study is structured as follows. In Sect. 2, we describe the observations (Sect. 2.1.1) and simulations

100 (Sects. 2.2.1 and 2.2.2) which we analyse and compare. In Sects. 2.3 and 2.4, we describe respectively the amount, frequency and intensity analysis and the basin-scale analysis methods which we employ. The results are set out in Sects. 3 and 4, focusing on Asia and south-eastern China respectively. In Sects. 3.1 and 3.2, we compare the mean June–July–August (JJA) precipitation over Asia, as represented by the observations and the simulations, presenting both grid-point and basin-scale analyses. In Sect. 3.3 we compare the diurnal cycle of the amount, frequency and intensity of precipitation over Asia. In Sect. 3.4, we

105 present analysis comparing the diurnal cycle of the amount of precipitation over different basin scales between the observations and simulations. In Sects. 4.1 and 4.2, we present results showing the amount, frequency and intensity of precipitation and their diurnal cycle over south-eastern China. We discuss our work in context of the wider literature and give some suggestions for future work in Sect. 5. Finally, in Sect. 6, we finish with our conclusions.

2 Data and methods

2.1 Observations

2.1.1 CMORPH

To evaluate the representation of precipitation in the simulations, we use the Climate Prediction Center morphing method (CMORPH; Joyce et al., 2004) observational dataset. This dataset uses high-quality precipitation estimates from microwave satellite data, which are modified (morphed) by the information from geostationary infrared satellites by using a time-weighted linear interpolation to provide a higher temporal resolution. The morphing also allows the construction of a spatially and temporally complete precipitation dataset.

Due to the high spatial and temporal resolution (8 km and 30 minute respectively), CMORPH is well suited to analysis of the diurnal cycle of precipitation. For example, Dai et al. (2007) found that the use of infrared data to improve sampling does not significantly affect the mean precipitation amount, frequency or intensity. Furthermore, satellite products such as CMORPH that include infrared data have smaller biases in the phase of the diurnal cycle. They note that CMORPH has a wet bias over warm season land areas compared with the Global Precipitation Climatology Project (GPCP; Adler et al., 2003). They summarize by saying that the satellite products they evaluate: "capture much of the sub-daily variations in precipitation amount, frequency, and intensity, although quantitative differences in the diurnal phase and amplitude exist among the different products and with surface observations". They also note that the diurnal cycle in these products is biased towards convective precipitation instead of detecting total precipitation, which is due to the microwave frequency detecting larger hydrometeors and the infrared frequency picking up cold cloud tops. Our simulations do not distinguish between convective and total precipitation, so this bias must be borne in mind when we compare to CMORPH.

We use CMORPH to produce a 21 y climatology of JJA precipitation over Asia from 1998–2018. We produce climatologies of the mean precipitation and of the diurnal cycle of precipitation, against which we evaluate the representation of precipitation in the simulations. To ease comparison with the high-resolution simulations, we upscale the spatial resolution of CMORPH to match their resolution, using area-weighted interpolation. We note that, although we use the full available timeseries for CMORPH (1998–2018), a shorter duration of 4 y (2006–2009) produces very similar values of amount, frequency and intensity (Sect. 2.3) as the full timeseries (Figs. S5 and S6 in the Supplement). We therefore infer that our results are robust with respect this choice of analysis period, because there is not much difference between the mean precipitation in these two durations and we are interested in the average precipitation or its diurnal cycle over many seasons.

2.1.2 APHRODITE

Some results shown in the Supplement compare the simulations with the observational product Asian Precipitation - Highly-Resolved Observational Data Integration Towards Evaluation (APHRODITE; Yatagai et al., 2012), which is a gridded gauge-based precipitation product available at a daily temporal resolution. We used the latest version (V1901) and the years 1988–

Table 1. List of simulations. All simulations were performed with the UM. The JJA seasons from 2005–2008 are used. The resolution is denoted by the equivalent longitudinal length of the grid cells at 30° N. Our focus is on the two high-resolution simulations N1280-PC and N1280-EC, although we compare with N1280-HC and the coarser simulations. All the simulation settings are described in the main text.

Simulation	Resolution (km)	Convection	References
N1280-PC	14	parametrized	
N1280-HC	14	hybrid	
N1280-EC	14	explicit	Field et al. (2018)
N512-PC	34	parametrized	Roberts (2017a); Roberts et al. (2019)
N216-PC	81	parametrized	Roberts (2017c); Roberts et al. (2019)
N96-PC	180	parametrized	Roberts (2017b); Roberts et al. (2019)

140   2015. We describe this dataset here for completeness. As with CMORPH, we do not expect that the results are sensitive to the exact choice of the multi-year analysis period.

2.2   Simulations

With increasing computing power, it is possible to run global simulations that have grid lengths of $O(10 \text{ km})$, comparable to that of regional weather simulations run two decades ago (Golding, 2000) or regional climate models from one decade ago

145   (Hurkmans et al., 2010). These high-resolution simulations are listed in Table 1, and described below in Sect. 2.2.1. Coarser-resolution simulations are described in Sect. 2.2.2.

2.2.1   High-resolution simulations

We have performed simulations with the UK Met Office Unified Model (UM), run in its global climate configuration – HadGEM3-GC3.1 (Hadley Centre Global Environment Model 3 – Global Climate 3.1; Williams et al., 2018). At these resolu-

150   tions, it is insightful to run simulations both with and without parametrized convection (N1280-PC and N1280-EC respectively, see Table 1), as previous studies have shown that the convection scheme can have a drastic effect on the diurnal cycle (Li et al., 2018), and the spatial and temporal variability of the precipitation (Klingaman et al., 2017; Martin et al., 2017). Additionally, a simulation with a hybrid representation of convection (N1280-HC) was run. These global simulations are performed for 4 y, allowing us to sample some interannual variability and to assess their representation of the mean precipitation and the diurnal

155   cycle of precipitation over Asia. As shown in Figs. S5 and S6 in the Supplement, a 4 y timeseries of CMORPH observations is very similar to the full 21 y timeseries. From this, we infer that our 4 y simulations will be long enough to provide a representative duration which we can compare against observations.

The N1280-PC simulation with parametrized convection is based on the standard Global Atmosphere (GA) 7.1 settings (Walters et al., 2019), which is the atmospheric component of HadGEM3-GC3.1 with some modifications to allow it to be

160   used at a higher resolution. It broadly follows the HighResMIP protocol (Haarsma et al., 2016), and in particular uses the

HadISST 2.2 sea ice and sea surface temperature dataset for the oceanic boundary conditions (Titchner and Rayner, 2014; Kennedy et al., 2017), which has a 0.25° horizontal grid length and a daily temporal frequency. The simulation uses the coupled model intercomparison project 6 (CMIP6) greenhouse gas, ozone and solar values (Haarsma et al., 2016). However, unlike in the HighResMIP protocol that recommends the use of MACv2-SP aerosols (Stevens et al., 2017), N1280-PC uses

165 the GLOMAP-mode aerosol scheme (Mulcahy et al., 2018). The N1280 designation means it has 2×1280 longitudinal grid cells, and 1.5×1280 latitudinal grid cells, which corresponds to a grid length of $0.14° \times 0.09°$ respectively ($16 \, \mathrm{km} \times 10 \, \mathrm{km}$ at the equator). It uses 85 vertical levels. The simulation is run with a 4 minute timestep. Apart from these changes that were required to run this simulation at high-resolution and the different aerosol scheme, this simulation uses settings that are the same as the ones described in Sect. 2.2.2. This simulation was not tuned to produce a realistic climate, as is done for the original

170 HadGEM3-GC3.1 model. For this simulation, three ensemble members are available. We mainly focus on one of the ensemble members below – referring to this as "the N1280-PC simulation". Figures 5 and 8 below show the spread of ensemble members (light blue shading), and this is typically smaller than the difference between different simulations, which justifies this choice.

The N1280-HC simulation uses hybrid convection, meaning that the deep convection scheme is disabled, but the shallow and mid-level convection schemes are enabled. The convective available potential energy closure timescale is also increased

175 from 3600 to 5800 s for the enabled convection schemes, as this was found to improve the representation of African easterly waves (Tomassini, 2018). Apart from these, it uses an identical configuration to the N1280-PC simulation. The results of this simulation are typically quite close to the N1280-EC simulation, as can be seen in Figs. 5 and 8. In most figures below we omit this simulation. We describe and show the differences where they are substantial.

The N1280-EC simulation with explicit convection, which is described in Field et al. (2018), uses the same configuration as

180 the N1280-PC simulation apart from the following differences. The convection parametrization scheme is completely disabled. Hence all convection is a product of the model's dynamics; the large-scale microphysics scheme is responsible for producing precipitation. Some further modifications are made: a 3 minute timestep is used for numerical stability; both stochastic perturbation schemes are disabled (stochastic kinetic energy backscatter 2 and stochastic perturbed tendencies), as these depend on the convection parametrization scheme to function; two-dimensional Smagorinsky horizontal diffusion (Smagorinsky, 1964)

185 is used; and the model uses a prognostic representation of graupel, as in Field et al. (2018).

2.2.2 Coarser-resolution simulations

Resolution has been shown to have a large effect on the simulated precipitation (e.g., Vannière et al., 2019). To investigate this, we compare with previous UM simulations, whose grid lengths at 30° N are shown in Table 1. These were performed as part of the European Commission Horizon H2020 PRIMAVERA project (Roberts et al., 2018), and are fully described in Roberts et al.

190 (2019); here we note some of the details. They are run with similar settings as the high-resolution simulations, using the same GA7.1 atmospheric science configuration (Walters et al., 2019), where the main differences between the simulations are those needed for numerical stability of the simulations at higher resolutions and that they use the MACv2-SP aerosol scheme (Stevens et al., 2017). They all have parametrized convection; the scheme is the same as the one used in the N1280-PC simulation. As with HighResMIP (Haarsma et al., 2016), and the high-resolution simulations described above, little tuning was done between

195 the different resolutions. Note, even though they are coarser than the high-resolution simulations described above, the N512-PC

simulation is still a similar resolution to what is currently considered "high resolution" for climate simulations (Haarsma et al.,

2016).

2.3   Amount, frequency and intensity analysis

Following many other studies, we partition the precipitation into three measures: amount (A), frequency (F), and intensity (I).

200 This requires the use of a threshold, for which we use $0.1 \ \mathrm{mm \, h^{-1}}$ in line with previous studies (e.g., Li et al., 2018). A is the

thresholded precipitation over a particular period. It is very close to the mean over a particular period, although precipitation

rates lower than the threshold do not contribute to it. F is the percentage of time for which the threshold is exceeded. I is a

measure of precipitation intensity for events above the threshold. The three measures are related by: $A = F \times I$.

Each measure of precipitation can be calculated over a sub-daily window, e.g. for each hour of all days during JJA as is

205 done here for the simulations. Thus, the diurnal cycle of each measure can be computed at every grid cell, where the diurnal

cycle is calculated by using the mean value of the measure over a particular time period. For example, for the simulations,

the diurnal cycle for A is calculated as the mean amount over each 1 h window across the 24 h day. Furthermore, the diurnal

cycle can be summarized by two quantities: the phase of the peak and the amplitude of the cycle. We use harmonic analysis

to compute this information, as in e.g. Dai and Wang (1999), where these are given by the phase and amplitude of the first

210 harmonic respectively. The phase information is converted to Local Solar Time (LST) using the longitude of each grid cell.

In the diurnal cycle figures below (Figs. 6, 7 and 10), we partition the amplitude of the diurnal cycle into three: strong

(top third), medium (middle third) and weak (bottom third). Each of these is defined relative to the amplitude of the diurnal

cycle separately for each dataset, over the complete Asia analysis domain shown in Fig. 2. As this is calculated separately for

each dataset, the visual representation of the amplitude is a relative measure of the strength of the diurnal cycle within each

215 dataset. However, when comparing the amplitude of the diurnal cycle using the RMSE (Sect. 2.4.4) in Fig. 8, the amplitudes

are compared on an absolute basis between the datasets.

2.4 Basin-scale analysis

2.4.1 HydroBASINS dataset

We use the HydroBASINS catchment basin dataset (Lehner, 2014), which is a subset of the Hydrological data and maps based
on SHuttle Elevation Derivatives (HydroSHEDS) dataset (Lehner and Grill, 2013). HydroBASINS is based on a high-resolution
(15 arc-second) digital elevation model (DEM), which is stored in a raster (gridded) data format. The basins are generated from
the DEM using the scheme of Verdin and Verdin (1999), which works by calculating the steepest-descent direction from the
DEM, and then calculating how many tributaries drain into a given grid cell. An area threshold of 100 km^2 is applied to
delineate the basin boundaries. The resulting raster basins are then vectorized and stored in a vector format.

The basins are represented by 12 different levels, with the lowest levels representing the largest-scale features (Lehner and
Grill, 2013), and with levels 1, 2 and 3 assigned manually. Level 1 distinguishes continents: there are 9 of these. We only use
the Asian basins, which are denoted by a 4 at the top level using the Pfafstetter coding system (Verdin and Verdin, 1999). Level
2 splits each continent into nine large sub-units; level 3 splits these into the largest river basins. Beyond that, it follows the
traditional Pfafstetter coding system, with minor modifications for islands, endorheic basins, coastal basins and sub-basin size
consistency. There is, however, no guarantee that basins at the same level will have similar sizes.

2.4.2 BASMATI

To facilitate the basin-scale analysis, and to make similar analysis easier for other studies in the future, we developed the
BAsin-Scale Model Assessment ToolkIt (BASMATI – available from https://github.com/markmuetz/basmati). BASMATI is
written in Python 3 and uses some key libraries to interact with the underlying data: `pandas`, `geopandas` and `rasterio`.
BASMATI simplifies downloading and interacting with the HydroBASINS dataset (Sect. 2.4.1), which provides the underlying
data about catchment basins.

BASMATI adds some key capabilities to the HydroBASINS dataset. As already stated (Sect. 2.4.1), the dataset is split into
different levels, where the largest basins are at the lowest levels (e.g. the top level is level 1). However, basins at a given level
are not all the same size. For example, basins at level 4 range in size from 4 to 1000000 km^2. As we want to compare basins
that are of similar sizes, it is necessary to select basins from different levels that fall within a given size range. To that end, we
implemented a simple area selection algorithm. This selects basins within a given size range, e.g. 2000–20000 km^2. It works
by starting at the top level of basin size, and if the basin is larger than the upper size limit, splitting the basin into its sub-basins.
It does this iteratively until all the basins are below the upper size limit. It then removes basins if they are lower than the lower
limit, meaning that the total area covered by the top-level HydroBASINS region is not completely covered (see e.g. Fig. 4).
However, most of the area is covered (at least 92 % of the total area); the variation in area between the different basin scales is
small (Table 2).

Table 2 shows the principal basin scales used in this study, as well as some information about each scale. For some of the
analysis, extra basin scales are used in between the principal basin scales (e.g. Fig. 5); these are chosen on a sliding scale of
logarithmically equally spaced basin areas.

Table 2. List of principal basin scales used, showing information about the properties of the basins, based on all basins which are part of the HydroBASINS Asian domain (shown in Fig. 2).

Basin scale	Area range (km^2)	Number of basins	Median basin area (km^2)	Total area ($\times 10^6$ km^2)
small	2000–20000	2990	5040	19.5
medium	20000–200000	287	54600	19.2
large	200000–2000000	27	553000	19.3
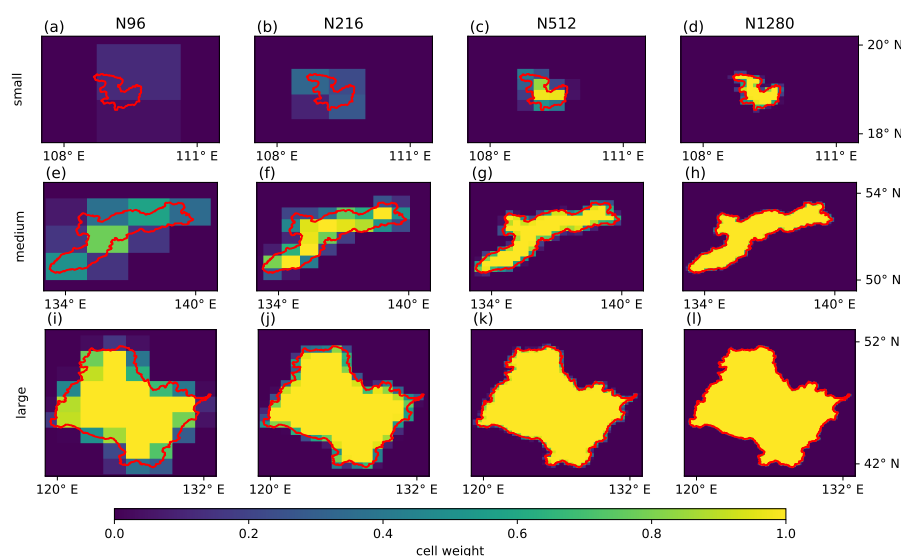
Figure 1. Weights used in the gridded representation of example basins at different resolutions. The basin with the median area from each of the basin scales in Table 2 is shown in each row, and each column shows the different simulation resolutions used in this study. The small, medium and large basins shown have areas of 5040, 54600 and 553000 km^2 respectively.

250   The HydroBASINS catchment basins are stored in the Environmental Systems Research Institute (ESRI) shapefiles (ESRI, 1998) vector format. The output from a GCM or other atmospheric model is typically on a latitude/longitude grid. To convert from a vector format to a gridded format, the basins must be *rasterized*. This produces a gridded field of weights for each grid cell that can be used to produce e.g. the mean precipitation over a given basin at any resolution. We do this by rasterizing the vector data onto a grid that is 10 times finer in both the latitudinal and longitudinal direction than the resolution for which we

255   want to produce a weighted raster, and using this to produce weights (accurate to 1 %) for each resolution. The weights for a given basin are zero outside the basin and one inside, with a fractional value on the boundary. These weights are shown in Fig. 1 for all resolutions for the median-sized basin at each of the basin scales in Table 2.

It is clear that, at lower resolutions, the smallest basins cover substantially less than one grid cell (e.g. Fig. 1a). We choose to perform the analyses at these scales for the coarse resolutions because, even though each basin is poorly represented by any

260   one grid cell, the statistical picture that is created by aggregating over many such basins should still be accurate. This choice is somewhat validated by the evaluations between CMORPH and the simulations shown in Figs. 5 and 8, which show that the different resolution simulations exhibit plausible relationships, even at the finest basin scale.

2.4.3 Spatial averages over basins

We complete the analysis by computing spatial means over each of the basins. For JJA mean precipitation, we use an area-
265 weighted mean over each basin, also using the basin weights:

$$\overline{P}^i = \frac{\sum_{j=1}^{N} W_{\text{basin},j}^i \, W_{\text{area},j} \, P_j}{\sum_{j=1}^{N} W_{\text{basin},j}^i \, W_{\text{area},j}}. \tag{1}$$

Here, P_j is the precipitation in the jth cell, and $\overline{P}^i$ is the mean precipitation in the ith basin. The summation is over all N
grid cells, and the area weights are the same for all cells at a given latitude. The area weights for each grid cell, $W_{\text{area},j}$, are
calculated to ensure that grid cells further north, which will be smaller on a latitude-longitude grid, contribute less to the mean.
270 The basin weights for each basin i and each grid cell j, $W_{\text{basin},j}^i$, are as described above (Sect. 2.4.2).

Producing the basin-weighted diurnal cycle requires producing a composite diurnal cycle over the weighted grid cells that
comprise each basin. We do this by taking the spatial mean of the diurnal cycle in each grid cell, weighted by the basin weights
and an area weighting as above. For example, for the basin-mean diurnal cycle of amount, $\overline{A}^i(t)$, at each time t over the day,

$$\overline{A}^i(t) = \frac{\sum_{j=1}^{N} W_{\text{basin},j}^i \, W_{\text{area},j} \, A_j(t)}{\sum_{j=1}^{N} W_{\text{basin},j}^i \, W_{\text{area},j}}, \tag{2}$$

275 where $A_j(t)$ is the amount at time t in grid cell j. The diurnal cycle of frequency and intensity have the same equation,
replacing A with F or I. From the diurnal cycle of A over a basin, the phase of the peak and the amplitude of the diurnal cycle
are calculated using harmonic analysis as above (Sect. 2.3).

2.4.4 Basin-scale error statistics

To compare the simulations with the CMORPH observations, we use basin-level error metrics. For example, we use the root
280 mean squared error (RMSE) to compare the mean JJA precipitation between CMORPH and each of the simulations, as defined
by:

$$\text{RMSE} = \sqrt{\sum_{i=1}^{N_{\text{basin}}} \frac{(\overline{P}_{\text{obs}}^i - \overline{P}_{\text{sim}}^i)^2}{N_{\text{basin}}}}. \tag{3}$$

Here, there are N_{basin} basins, and so the summation is over all basins. $\overline{P}_{\text{obs}}^i$ is the mean precipitation in basin i in the observa-
tions, and $\overline{P}_{\text{sim}}^i$ is the mean precipitation in basin i for the simulation.
285 When comparing the phase ϕ of the peak of A, F or I between simulations and observations, it is necessary to take into
account the fact that this is a circular quantity. That is, a phase in LST of 2300 should be 2 h away from 0100, not 22 h away.
This is achieved by applying a base-24 circular difference ($\text{circular } \text{diff}_{24}(x)$) operation to the difference between observations
and simulations to calculate a circular RMSE:

$$\text{circular RMSE} = \sqrt{\sum_{i=1}^{N_{\text{basin}}} \frac{\text{circular diff}_{24}(\phi_{\text{obs}}^i - \phi_{\text{sim}}^i)^2}{N_{\text{basin}}}}, \qquad (4)$$

290 $$\text{circular diff}_{24}(x) = \begin{cases} x \ (\mathrm{mod}\ 24) & \text{for } x \ (\mathrm{mod}\ 24) \leqslant 12 \\ 24 - (x \ (\mathrm{mod}\ 24)) & \text{for } x \ (\mathrm{mod}\ 24) > 12 \end{cases}. \qquad (5)$$
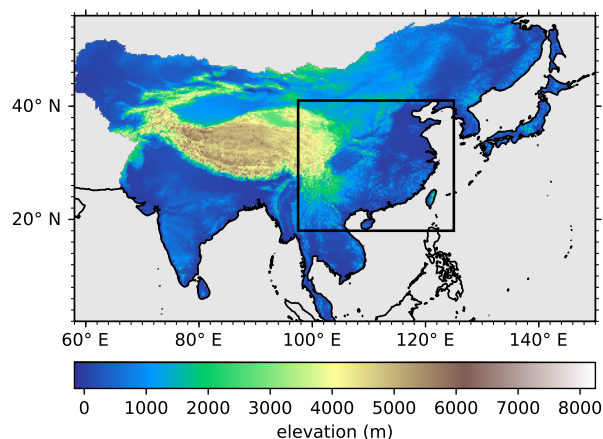
Figure 2. Asia, as represented by the HydroBASINS level 1 dataset (coloured), showing the terrain elevation. The black rectangle is the region of interest over south-eastern China shown in Figs. 9 and 10. Prominent features which are discussed are the Tibetan Plateau, particularly the Himalayas on its southern flank. Also, the diurnal cycle over the Sichuan Basin (centred on 30° N, 106° E) will be discussed. The 30 arc-second elevation data from the HydroSHEDS dataset (Sect. 2.4.1) is shown.

3   Precipitation over Asia

In Fig. 2, the regions of interest in this study are shown. The full domain encompasses the top-level HydroBASINS Asia region. Results from this region are discussed in Sect. 3. Because the orography is important for determining where and when precipitation occurs, the terrain over the top-level HydroBASINS Asia region is shown. The most prominent feature is the Tibetan Plateau; we are interested in the effect that it has on precipitation, particularly that of its southern flank. The Sichuan Basin can also be distinguished, where the diurnal cycle of precipitation exhibits a clear phase propagation (Li et al., 2020). In Sect. 4, we will focus on the south-east of China, shown by the black rectangle, in Figs. 9 and 10 below, so as to directly compare with previous results from a regional UM configuration (Li et al., 2018).
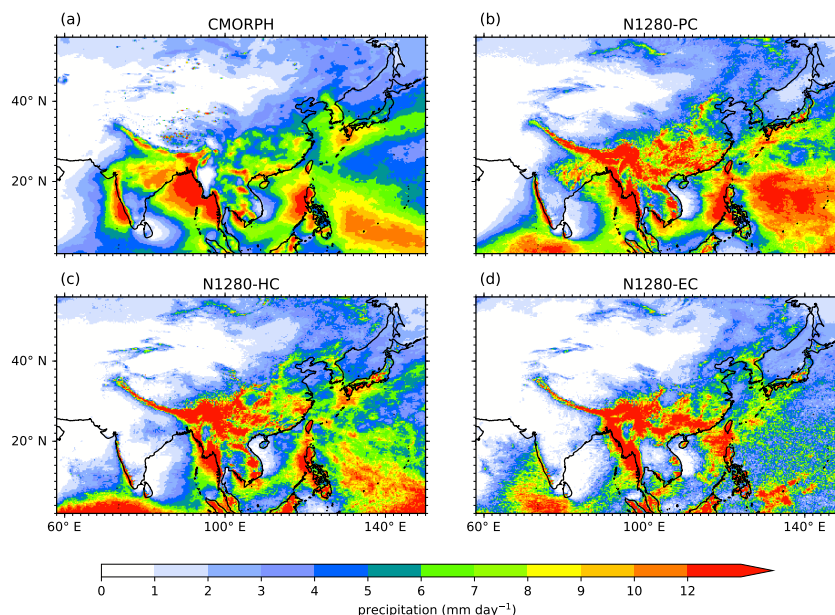
Figure 3. Mean JJA precipitation over Asia for CMORPH observations (a), and the parametrized (b), hybrid (c) and explicit (d) convection simulations. The observations are taken from 1988–2018, and the simulations are all from 2005–2008.

3.1 Mean precipitation

300   In Fig. 3, the mean JJA precipitation over Asia from, observations and simulations, is shown. In the following, we mainly discuss the features over land. From the observations, several prominent features are visible. There is a strong band of precipitation off the western coast of India, which is related to the ISM flow. Precipitation falls at a rate of 2–4 $\mathrm{mm\,day^{-1}}$ over southern India east of the Western Ghats mountains, with substantially higher rates over northern India. Both of these are consistent with other studies of the ISM (e.g., Mitra et al., 2013). Along the Himalayas, there are very high local maxima of precipitation of

305   12 $\mathrm{mm\,day^{-1}}$. This is due to the ISM flow interacting with the high orography and associated steep gradients of the Tibetan Plateau and the Himalayas (Fig. 2). Some scattered points with high precipitation rates over the body of the Tibetan Plateau are an artefact of the satellite retrieval method and due to the snow cover in this region (Joyce et al., 2004).

   Over south-eastern China, high precipitation rates of 6–12 $\mathrm{mm\,day^{-1}}$ are seen. This is associated with the EASM, and the propagation of the Meiyu front, both of which are active during JJA. Further inland, there is a small secondary maximum at

310   around 28° N, 102° E, which is related to the orography of the Tibetan Plateau on the north-western boundary of the Sichuan Basin.

   A large area of intermediate summer rainfall (2–5 $\mathrm{mm\,day^{-1}}$) is seen over north-eastern China at higher latitudes. A maximum in seasonal JJA precipitation is seen over and near to the Korean Peninsula. Over the Gobi and Taklamakan deserts, less than 1 $\mathrm{mm\,day^{-1}}$ occurs in JJA.

315   Over the ocean, the western tongue of the intertropical convergence zone (ITCZ) is clearly visible. Precipitation rates are high to the west of the Philippines, off the southern coast of Japan, and over a large area of the Bay of Bengal.

The simulations broadly reproduce the observed Asian summer precipitation distribution. The low precipitation rate over the deserts is well matched. The magnitude and extent of precipitation in north-eastern China is accurately reproduced, although N1280-EC produces some localized areas of higher precipitation and N1280-PC fails to produce the maxima over the Korean

320   Peninsula. However, at lower latitudes there are substantial differences. Over China, the simulations produce too much precipitation, with maximum rain rates of over $12 \ \mathrm{mm \, day^{-1}}$ over much larger areas than in the observations. Precipitation rates over India are much lower in the simulations than the observations – a known bias of the UM (Bush et al., 2015). N1280-EC in particular has very low rain rates over India (particularly north-eastern India), with no signal of the higher rates seen at 20° N in observations. An intriguing possibility is that this is due to a lack of Bay of Bengal depressions forming in N1280-EC,

325   although we have not investigated this here. Clearly the lack of a convection parametrization affects the simulation to a large degree over this region. There are signs of a large, spurious maximum of precipitation over the Indian Ocean in all simulations, which has been linked to the dry bias over India in the UM (Bush et al., 2015) and also in multiple other GCMs (e.g., Bollasina and Ming, 2013; Levine et al., 2013). Over land, N1280-EC and N1280-HC closely resemble each other.

All simulations produce high rates of precipitation over the Himalayas, although N1280-PC produces a band of precipitation which is too wide.

330   Both N1280-EC and N1280-HC produce maximum precipitation rates which are too high, particularly close to 100° E. This is consistent with Willetts et al. (2017), who found that shorter-duration explicit simulations with $12 \ \mathrm{km}$ grid lengths produced excessive precipitation over the Himalayas.

Over the western Pacific, the N1280-PC simulation produces far too heavy precipitation over far too large an area, whereas the opposite is true for N1280-EC. N1280-HC has smaller precipitation biases against observations. All simulations produce

335   heavier precipitation west of the Philippines, consistent with observations. All simulations produce less precipitation than observed over the Bay of Bengal.
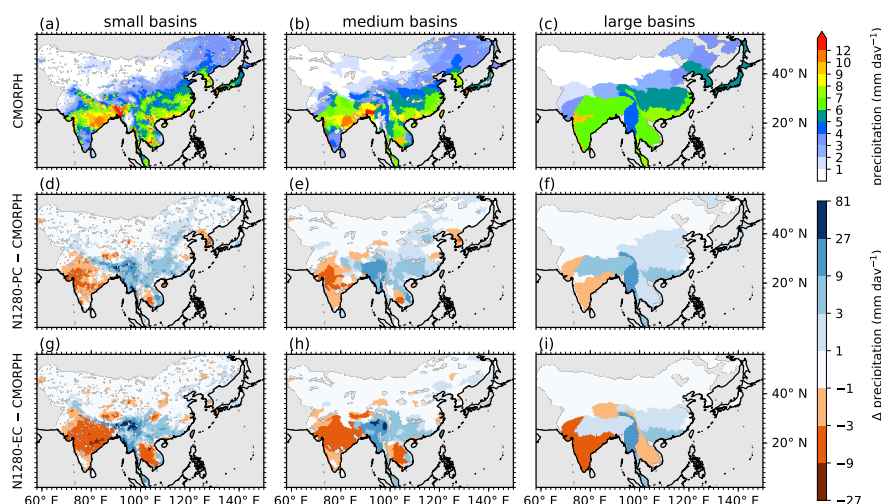
Figure 4. Mean JJA precipitation over Asia averaged over basins of different scales showing CMORPH (top), and difference between simulations and CMORPH for N1280-PC (middle) and N1280-EC (bottom). One basin scale is shown in each column, and one dataset is shown in each row. Areas shown in grey are not covered by selected basins (see Sects. 2.4.1 and 2.4.2).

3.2 Mean precipitation over catchment basins

In Fig. 4, the precipitation rates in Fig. 3 are averaged over basins of different scales. Thus, Fig. 4a, which shows the CMORPH observations averaged over the small basin scale (see Table 2), resembles Fig. 3a. As noted in Sect. 2.4.2, the basin-selection
340   algorithm cannot pick basins in a given scale range that completely cover the Asian land, hence there are gaps at each basin scale. As is clear from Figs. 4a–c, averaging over larger basin scales reduces the maximum precipitation rates at that scale, since high precipitation rates are averaged with lower rates. Note, as N1280-EC and N1280-HC produce similar JJA precipitation patterns over land (Figs. 3 and S1 in the Supplement), this simulation is not shown here, but the minor differences are discussed below.

345   The small basin scale column provides some more details of what was seen in Fig. 3. Both simulations produce too little precipitation over India; this is particularly the case for N1280-EC at around $20°$ N. Both simulations produce too much precipitation on the Himalayas, with N1280-EC producing up to $72 \, \mathrm{mm \, day^{-1}}$ more precipitation at $95°$ E than CMORPH. The basins with the maximum precipitation rates in N1280-PC and N1280-EC have rates that are respectively 7 and 11 times higher than CMORPH. These biases dominate the differences between the simulations and observations; the increased precipitation
350   over south-eastern China in the simulations is present but is smaller in magnitude. Similarly, the differences in precipitation rates between the simulations and CMORPH in north-eastern China are barely visible, as the magnitude of the precipitation in this region is typically low to begin with in both observations and simulations (Figs. 3 and 4a, d and g). There are few differences between N1280-EC and N1280-HC (see Fig. S1 in the Supplement). Both simulations produce a dry bias over the

Indochina Peninsula, although the spatial extent of this is larger for N1280-EC. Likewise, the dry bias over the Sichuan Basin
355   in N1280-EC is larger than that in N1280-HC.

At the small basin scale, the HadGEM3-GC3.1 simulations at coarser resolutions (N512-PC, N216-PC and N96-PC) pro-
duce mean JJA precipitation over Asia that most closely resembles N1280-PC (not shown). The highest-resolution of these
simulations, N512-PC, is most similar to N1280-PC, although it has slightly enhanced precipitation ($1\,\mathrm{mm\,day^{-1}}$) over India,
and slightly reduced precipitation ($1\text{–}4\,\mathrm{mm\,day^{-1}}$) over China. N216-PC is less similar to N1280-PC than N512-PC. N216-PC
360   produces less precipitation than N512-PC and N1280-PC over the Western Ghats on the west coast of India, presumably due to
under-resolved orography, and produces more precipitation than N512-PC and N1280-PC over north-eastern China. N96-PC
is least similar to N1280-PC. Again, there is less precipitation compared to N1280-PC over the Western Ghats, presumably
because of under-resolved orography. A lack of precipitation over India is evident, as in the simulations of Bush et al. (2015).
These broad comparisons are true at all basin scales.
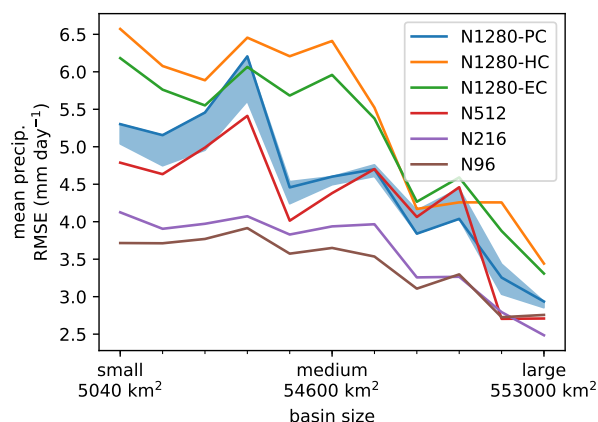
Figure 5. Basin RMSE (Sect. 2.4.4) of each simulation compared to CMORPH shown as a function of basin scale. Light blue shading shows the maximum spread for the three N1280-PC ensemble members. The basin name and median basin size are shown on the x-axis.

365   For medium and large basin scales, there are smaller discrepancies between the simulations and observations as precipitation is averaged over larger basin scales. This is partly because averaging over a larger area smooths out the signal of localized maxima in precipitation, as mentioned above. We also expect to extract useful information about how the model represents precipitation at different spatial scales by calculating statistics about the agreement between the simulations and observations. In Fig. 5, basin-scale RMSE values (Sect. 2.4.4), calculated between the JJA mean precipitation in CMORPH and the simulations, are shown as a function of scale. As noted in Sect. 2.4.2, we perform the analysis at 11 different basin scales, ranging from small to large, spaced equally on a logarithmic scale. We also include the three lower resolution simulations: N512-PC, N216-PC and N96-PC (Table 1). For all simulations, averaging over a larger basin scale improves the error statistics by reducing the RMSE as the basin scale increases.

We note that the N96-PC simulation performs best by these metrics. This is because N96-PC produces much lower precipitation maxima, and therefore is penalized less than the other five simulations for producing too much precipitation over e.g. the Himalayas. Indeed, the error increases with resolution for all spatial scales. The spread between the three N1280-PC ensemble members is typically smaller than the difference between that simulation and the other simulations. When the three high-resolution simulations are compared, N1280-EC and N1280-HC perform worse than N1280-PC, with N1280-HC performing slightly worse than N1280-EC. Again, this is due to excessive precipitation in the simulations with explicit deep convection. This shows that disabling the convection parametrization has an important bearing on performance for climatological JJA precipitation.

We have performed identical analysis comparing the simulations to APHRODITE (Sect. 2.1.2). The equivalent figure to Fig. 5 (Fig. S2 in the Supplement) is very similar, and qualitatively the same conclusions would be drawn. This shows that the difference for this metric between the simulations and observations is larger than the differences between the observational products.

3.3 Diurnal cycle

Figure 6 shows the diurnal cycle over Asia in the CMORPH observations. As also seen below in Sect. 4.2 for south-eastern China, the amount and intensity of precipitation are highly similar. For both of these, there is a marked difference between land and ocean, both in the phase and amplitude of the diurnal cycle: over land the amplitude is larger and the phase is later.

390   The phase over the ocean tends to be either early morning (0300–0700 LST) over the ITCZ and off the coast of Japan, or close to midday off the coast of China. There is an interesting phase delay off the east coast of India over the Bay of Bengal, as the phase of the diurnal cycle goes from 0800 to 1700 LST. This feature has been noted in previous studies (e.g., Yang and Slingo, 2001); it is thought to be related to the coupling between gravity waves and convection, leading to long-lived mesoscale convective systems (Houze Jr., 2004).

395   Over land, the diurnal cycles of the amount and frequency of precipitation show some clear features. North of 35° N, the amplitude of the diurnal cycle is typically weak and the peak has a phase of 1500–1800 LST. Closer to the equator, the amplitude of the diurnal cycle is generally strong, consistent with the stronger diurnal solar forcing in the tropics. India can be divided into two main regions. Central India shows a peak of precipitation in the late evening. Coastal western India, north-eastern India and north of about 20° N show an early evening peak. The southern flank of the Tibetan Plateau shows an

400   interesting phase delay from north to south, with a late-night peak in precipitation close to the Plateau top progressing to an early morning peak (0800 LST) further south.

The diurnal cycle field for intensity of precipitation is noisier. In general, late-night peaks in amount, frequency and intensity are co-located. We speculate that this could be due to the activity of mesoscale convective systems, which are associated with both convective and stratiform precipitation. The convective precipitation would likely affect the intensity of precipitation,

405   whereas the combined convective and stratiform precipitation would affect the amount, frequency and intensity of precipitation. For intensity, some areas show broadly similar signals as for amount and frequency, such as the phase delay over the Bay of Bengal, the pattern over the ocean and the phase delay on the southern flank of the Tibetan Plateau. However, some regions show substantial differences. At higher latitudes over land, the peak of intensity is 4–6 h later than the peaks of amount and frequency. This potentially indicates that convective precipitation is dominant later during the day in these regions.

410   We can compare the simulations to the observations (Fig. 6), noting again that the observations span 21 years, and the simulations span four years. This leads to the simulations being noisier than the observations. Before presenting a detailed analysis, we note that N1280-EC produces a more-realistic diurnal cycle in most regards than either N1280-PC or N1280-HC. In N1280-EC, over land the phase of the diurnal cycle better matches the observations for both amount and frequency, and is marginally better for intensity. The land–sea contrast is better represented, for example off the coast of south-eastern China.

415   The similarity of the diurnal cycles of amount and frequency in N1280-EC is closer to that of the observations. Thus, in terms of producing a realistic diurnal cycle, the lack of parametrized convection is clearly an advantage.

Perhaps the most striking differences between the simulations and observations are for N1280-PC frequency and amount. For frequency, this simulation shows a peak over almost all of Asia that is far too uniform, and too early, being close to local midday. It is well known that convection parametrization schemes respond to the peak in insolation forcing by producing
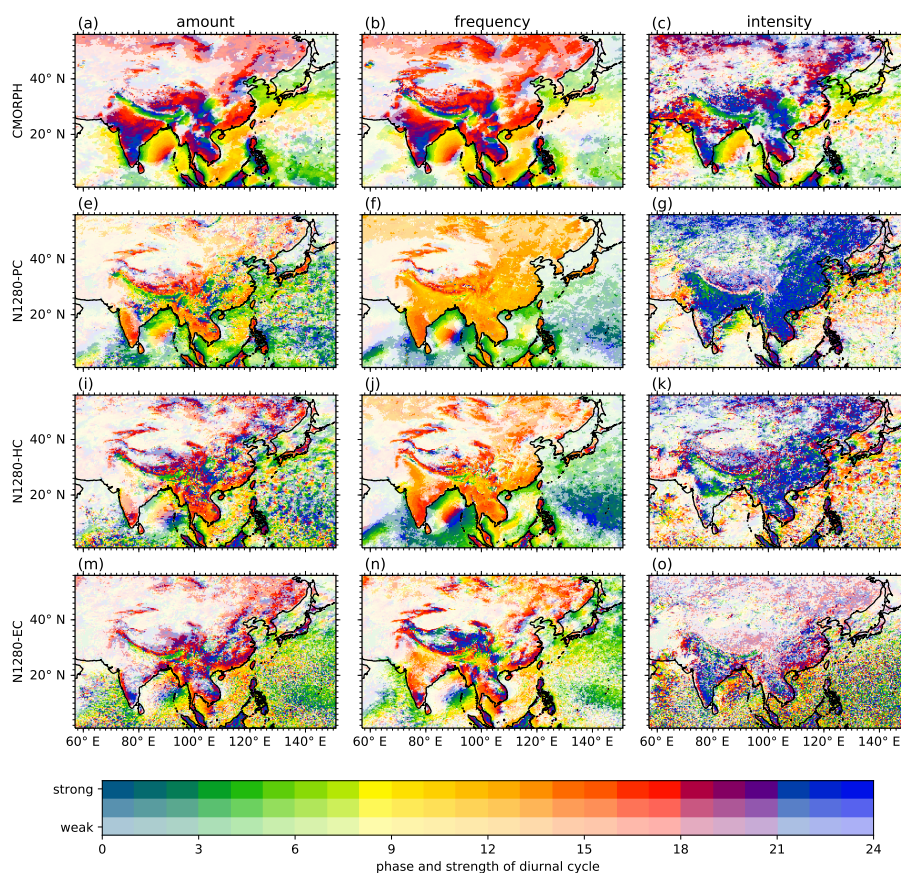
Figure 6. The diurnal cycle of the amount, frequency and intensity of precipitation over Asia for CMORPH and the three high-resolution simulations, showing the phase in LST in colour, and the amplitude by the opacity of the colour. The diurnal cycle is considered strong if its amplitude is in the top third of the Asia-wide amplitudes over the whole domain (including oceans), weak if it is in the bottom third and medium otherwise (Sect. 2.3). Note, the strength of the diurnal cycle is calculated separately for each dataset. Thus, it is possible to say that all datasets have a weak diurnal cycle for the amount of precipitation over north-western region of the domain, but not that the magnitude of the amount is similar where the diurnal cycle is considered strong. An absolute comparison is done in Fig. 8.

420   convective precipitation (e.g., Yang and Slingo, 2001; Stirling and Stratton, 2012; Bechtold et al., 2014), which is certainly one of the reasons for this signal. For intensity, the peak over almost all of Asia is close to local midnight. For N1280-PC, the peak in precipitation amount is also clearly biased to be too early, although it does show greater spatial variation than the frequency or intensity over this region. Additionally, there are differences between amount and frequency, even though this is not seen in the observations. There are few differences between the different N1280-PC ensemble members (not shown). The

425   largest difference is for the phase of precipitation amount over eastern India, although in general the phase and amplitude are very similar among ensemble members.

For N1280-EC over land, the phase of the peaks for amount and frequency broadly agree, although this simulation appears to produce too little late night precipitation. Additionally, the phases for all three precipitation measures over the ocean match the observed phases quite closely, although they are noisier for the simulation. Likewise, the amount and frequency are fairly

430   similar for this simulation, although the similarity is not as strong as it is for the observations. N1280-EC does capture some aspects of the phase of the intensity field in the observations, such as the late-night peaks over India and the Indochina Peninsula. However, there are pronounced differences between N1280-EC and the observations. From the peak in amount of precipitation over India, the amplitude of the diurnal cycle is too weak. This is probably due to the dry bias in this region (Fig. 3c). N1280-EC produces a peak in intensity of precipitation that is too noisy, which is probably related to the shorter duration of

435   the simulation. The peak in intensity is also weak, particularly over the Tibetan Plateau. However, this is a region with known biases in CMORPH (Joyce et al., 2004), so any comparisons to CMORPH in this region should be made with caution.

N1280-HC shares some similarities with both N1280-PC and N1280-EC, which makes sense given its hybrid representation of convection. Its diurnal cycle of the amount of precipitation resembles N1280-EC. This could indicate that the convective precipitation that is caused by deep convection is primarily responsible for the diurnal cycle of the amount of precipitation,

440   as both simulations represent deep convection explicitly. For the diurnal cycle of the frequency of convection, N1280-HC more closely resembles N1280-PC, although it shows more spatial variation over land. The resemblance potentially indicates that the shallow and mid-level parametrizations of convection are responsible for this diurnal cycle. The fact that the N1280-HC amount and frequency diurnal cycles do not closely resemble each other means N1280-HC is less like the observations (Fig. ??). The diurnal cycle for the intensity is stronger over land in N1280-HC than N1280-EC, meaning N1280-HC matches

445   the observations more closely and indicating that there may be issues with the representation of this diurnal cycle in N1280-EC.

Over the ocean, the N1280-PC simulation does not match the observations particularly closely, whereas the N1280-EC simulation performs better, with N1280-HC falling in between. All simulations show hints of a phase delay of the peak in amount and frequency of precipitation over the Bay of Bengal, indicating that the simulations might capture the important aspects of the coupling between convection and gravity waves. However, all show significant biases: N1280-PC is too early

450   close to the coast of India, and N1280-EC and N1280-HC are too weak and too late. N1280-PC produces diurnal cycles which have substantial biases in phase and amplitude over all other areas of the ocean, for all three of amount, frequency and intensity. N1280-EC in general produces more realistic phases, and produces slightly more realistic peaks in amount and frequency over the western Pacific, and between the Philippines and the Indochina Peninsula. However, it is noisier than the observations, which is probably due to its shorter duration. The phase of the intensity of precipitation in N1280-EC does

455   not match observations particularly closely, being stronger, too uniform, and in general occurring later in N1280-EC than the observations. N1280-HC is again closer to N1280-EC in amount, and N1280-PC in frequency and intensity. It has a particularly strong diurnal cycle near midnight for frequency over the western Pacific that is not evident in the observations, which is perhaps related to its wet bias at that location (Fig. 3).
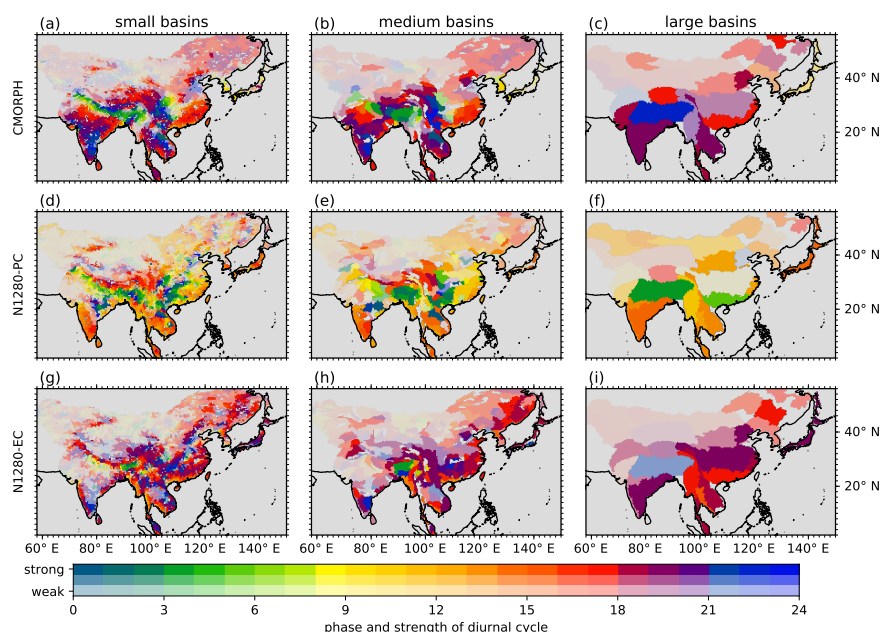
Figure 7. Amount of precipitation over Asia averaged over different basin scales (columns) for CMORPH, N1280-PC and N1280-EC (rows). As described in Sect. 2.3 and as in Fig. 10, a visual representation of the amplitude is given by its strength, dependent on whether it is strong, medium or weak. This is calculated separately for each dataset and scale (i.e. for each panel).

3.4 Diurnal cycle over catchment basins

460 As with the mean precipitation (Sect. 3.1), the phase and amplitude of the diurnal cycle can be averaged over catchment basins of different spatial scales (see Sect. 2.4.2 for details). Figure 7 shows this over Asia for CMORPH, N1280-PC and N1280-EC over small, medium and large basin scales for the amount of precipitation. Figure 7a, the CMORPH observations, is similar to Fig. 6a, as the basins are small, and likewise for the simulations (Figs. 7d and g). Note, as can be seen from Fig. 6g and j for amount over land N1280-HC resembles N1280-EC and is not shown here.

465 For the observations, averaging the diurnal cycle over larger basin scales yields useful information about the phase and amplitude of the diurnal cycle at different scales. This information is similar in spirit to that presented in Covey et al. (2016), although the method they used relied on vector averaging, whereas the method in this study uses direct averaging of the diurnal cycle at each grid point (Sect. 2.4.2). They recommend using their method to compare CMIP5 simulations with observations, to give a sense of how well the diurnal cycle is represented. However, whereas they average over all land and ocean grid points,

470 we use a much finer-grained approach of averaging over catchment basins over land. This allows us to distinguish between the phase and amplitude of the diurnal cycle over specific regions, and allows us to compare simulations and observations as a function of spatial scale.

For the observations, as the diurnal cycle is averaged over larger scales, certain information comes to the fore and at the same time location-specific detail is lost. For example, over northern India and Nepal, from Fig. 7a there is fine-scale detail in terms

475   of the phase delay on the southern flank of the Tibetan Plateau, whereas this is no longer detectable over medium-sized basins – where it averages to around 0600 LST. Furthermore, comparing Figs. 7b and c, the detail of the late-night phase over southern India is clearly lost at the large scales. However, at the large scale, a continent-wide pattern emerges that is hard to discern at the finer scales – that over southern India on average the diurnal cycle is strong and occurs at 1900 LST. Thus, the large basin scale yields information about the overall behaviour of the diurnal cycle, which can be compared against its behaviour

480   in simulations – similar to Covey et al. (2016). Additionally, averaging over larger spatial scales will reduce the noise in the shorter-duration simulations, and so should produce a fairer comparison between the simulations and the observations.

At small basin scales N1280-PC performs poorly for the diurnal cycle in the amount of precipitation (it performs worse for frequency and intensity of precipitation, not shown). It does not improve when the analysis is performed at the large basin scale. From the large scale, N1280-PC produces a diurnal cycle that peaks too early in the day, e.g. over India where it occurs

485   at 1400 LST in the simulation and 1900 LST in the observations. In comparison, at the large scale, N1280-EC broadly agrees with the observations, producing phases over different regions such as southern India and the south coast of China which are much closer to the observations. There are regions that do not match so well, such as northern India, where the diurnal cycle in the simulation is far too weak, although the phase matches more closely. This weak diurnal cycle can be attributed to the dry bias in this simulation (Sect. 2.4.2). N1280-HC produces a diurnal cycle of amount of precipitation that is quite similar to that

490   of N1280-EC (not shown). Some differences are evident at the smallest basin scales. The phase of the diurnal cycle is earlier over coastal south-eastern China, which means that N1280-HC matches CMORPH more closely. However, the phase is also earlier over the Indochina Peninsula, which means that it matches CMORPH less closely.

The N96-PC, N216-PC and N512-PC simulations all strongly resemble N1280-PC (not shown), suggesting that the convection parametrization scheme, and not model resolution, is responsible for producing the diurnal cycle in the simulations.
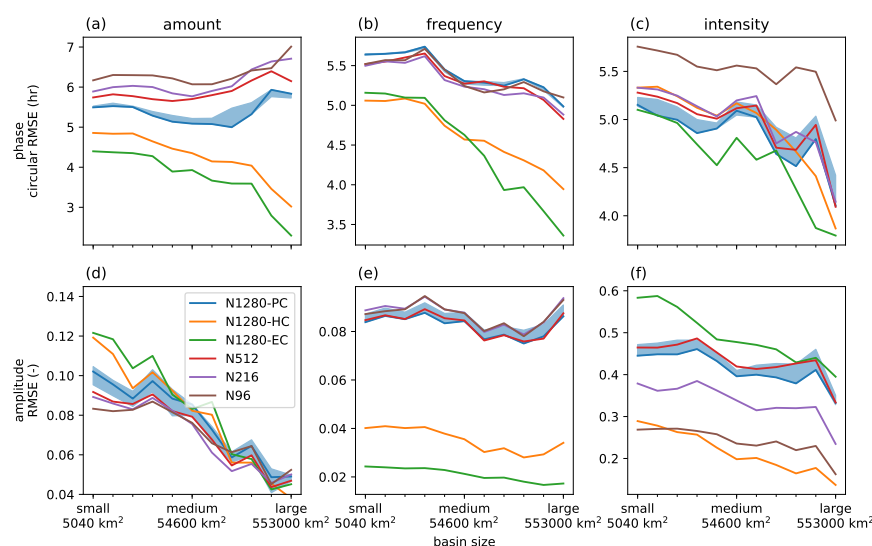
Figure 8. Basin error statistics (Sect. 2.4.4) of the diurnal cycle phase and amplitude of each simulation compared to CMORPH shown as a function of basin scale. Amount, frequency and intensity are shown in each column, and the error statistic appropriate for phase, amplitude and both combined are shown in each row. Light blue shading shows the maximum spread for the three N1280-PC ensemble members. The basin name and median basin size are shown on the x-axis.

495   In Fig. 8, we compare the simulations to observations at different spatial scales. Error statistics appropriate for the phase and amplitude are shown for amount, frequency and intensity for all simulations against CMORPH as a function of basin scale (the error statistics are described in Sect. 2.4.4).

For the phase, the clear signal is that for amount, frequency and intensity N1280-EC performs best, with N1280-HC performing almost as well. Indeed, for amount of precipitation, all the simulations with parametrized convection perform worse as

500   spatial scale increases to the large basin scale; only N1280-EC and N1280-HC improve as spatial scale increases. For frequency and intensity, all simulations improve as spatial scale increases, although again N1280-EC shows the most improvement (particularly for frequency). It is remarkable that the parametrized convection simulations span a range of resolutions from 14 to 180 km, yet their error statistics are very similar across all three measures of precipitation (particularly so for frequency). Where there is some resolution sensitivity, the performance improves as resolution increases – e.g. amount at all spatial scales.

505   The main difference between N1280-EC and N1280-HC is that the latter performs better for the amplitude of the intensity. For the amount of precipitation, all simulations perform better at larger spatial scales. N1280-EC performs worst at the finer scales. The same is broadly true for intensity, although the improvement at larger scales is less pronounced. For frequency, all the parametrized simulations perform poorly.

4 Precipitation over south-eastern China

510 In this section, we focus on south-eastern China (Fig. 2, black rectangle) for straightforward comparison with Li et al. (2018). In Sect. 4.1, we analyse the amount, frequency and intensity of precipitation, and in Sect. 4.2 we analyse their diurnal cycles.

4.1 Amount, frequency and intensity of precipitation

The amount, frequency and intensity of precipitation are shown in Fig. 9 for JJA in south-eastern China. The threshold for the analysis is $0.1\,\mathrm{mm\,h^{-1}}$. The amount, averaged over one day, is very similar to the mean precipitation, although the thresholding

515 means that the values are not identical. Thus, CMORPH amount, Fig. 9a, is effectively the same as Fig. 3a but for China instead of Asia, and likewise for the N1280-PC and the N1280-EC amount. The results shown here are directly comparable to those in Li et al. (2018), who analyse precipitation in two regional UM simulations against gauge-based observations. Their simulations are run at 4.4 and 13 km grid lengths, with explicit and parametrized convection respectively, for the warm season of 2009, encompassing JJA. Both their simulations use boundary conditions provided by a global UM run with a grid length of 0.2°. All

520 their simulations use the GA6.1 science settings (Walters et al., 2017). To facilitate the comparison, we only show and discuss results from N1280-PC and N1280-EC here, omitting N1280-HC (Figs. S3 and S4 in the Supplement show N1280-HC and are equivalent to Figs. 9 and 10).

For the CMORPH amount, there are localized maxima near south-facing coasts, indicating that the moist EASM flow in JJA produces precipitation when it passes over land. These are linked to higher intensity precipitation near the coast (Fig. 9c).

525 There is a maximum near 23° N, 104° E, which appears to be related to particularly frequent precipitation. For amount, there is generally a decreasing gradient in precipitation going further inland, with local inland maxima and minima typically related to the orography.

Amount, frequency and intensity in CMORPH over south-eastern China are similar to Zhou et al. (2008), Figs. 2c, f and i, which show the Tropical Rainfall Measuring Mission (TRMM) satellite product. This is despite the fact that they use a different

530 satellite product, a threshold of $0.2\,\mathrm{mm\,h^{-1}}$ (which will affect frequency and intensity), and a shorter time period of 2000–2004. The similarity indicates three things: that amount, frequency and intensity in CMORPH are broadly similar to those in TRMM; that a shorter time period is able to represent the mean of these precipitation measures; and that the quantitative values are affected by the threshold, but the qualitative conclusions are broadly the same for different thresholds.

From N1280-PC and N1280-EC amount, both simulations produce too much precipitation over the majority of south-eastern

535 China. In the Sichuan Basin, both simulations produce amounts of precipitation that are too low, and this is pronounced in N1280-EC. This is due to less frequent precipitation in this region (Figs. 9e and h). The overestimation of precipitation in N1280-PC is due to too frequent precipitation, as can be seen by comparing Figs. 9b and e. However, in N1280-EC the frequency matches CMORPH more closely east of 104° E (notwithstanding the bias in the Sichuan Basin), and the bias in precipitation rates is due more to the intensity bias (Figs. 9c and i). West of 104° E, the wet bias in N1280-EC appears to be

540 related to both frequency and intensity being too high. N1280-EC produces very intense precipitation over the sea, and there is a marked land–sea contrast that is not present in either CMORPH or N1280-PC.
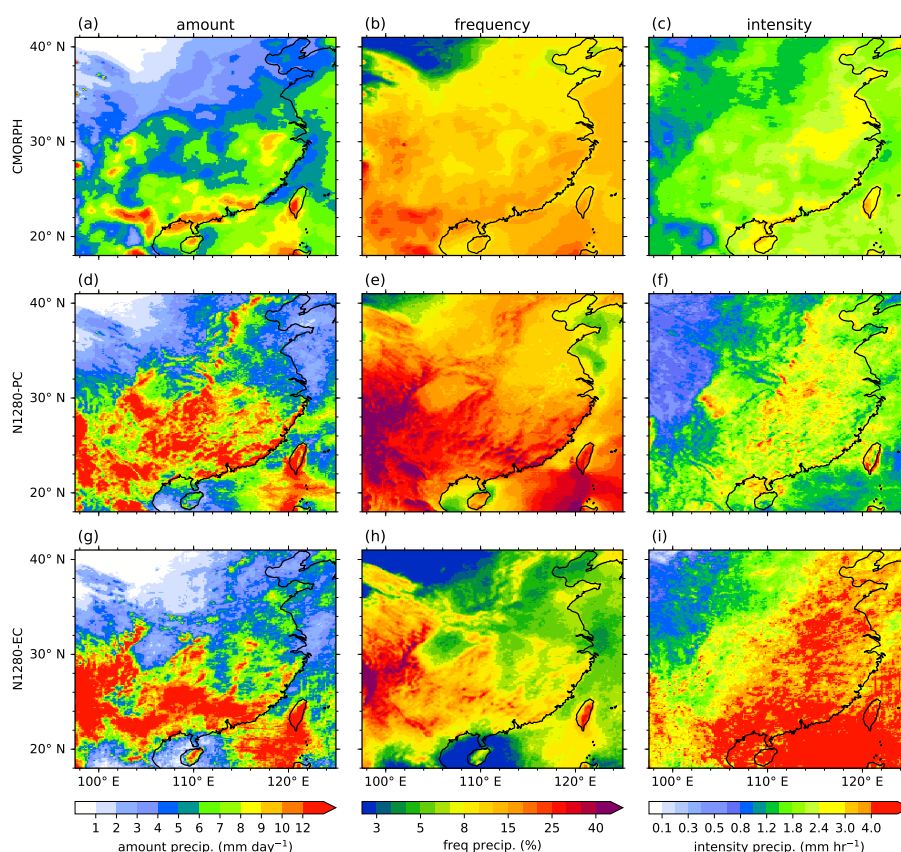
Figure 9. Amount, frequency and intensity of precipitation (columns) over China for CMORPH (top), N1280-PC (middle) and N1280-EC (bottom).

Comparing with Fig. 2 from Li et al. (2018) is instructive. Comparing the observations (CMORPH here and gauge stations in their study) reveals that they both produce similar patterns of amount of precipitation. In Li et al. (2018) they only sample from one year, whereas this study uses a 21 y duration, and hence their Fig. 2a is noisier. We note that the frequency and intensity

545 fields in both studies are qualitatively similar. However, in Fig 9b the frequency is generally higher than that in Li et al. (2018), and correspondingly the intensity is less in this study. This could be due to the method they used to turn the point gauge observations into a continuous field, as both studies use the same precipitation threshold. The similarity between the explicit and parametrized simulations in both studies is striking, despite the difference in resolution between the explicit simulation in this study and that in Li et al. (2018). In both studies, both the parametrized and explicit simulations clearly overestimate

550 precipitation amount: the parametrized simulation because precipitation is too frequent, and the explicit simulation because precipitation is too intense. Indeed, Figs. 9e and f here match Figs. 2f and i in Li et al. (2018) very closely, and Figs. 9h and

i here match Figs. 2e and h in Li et al. (2018) very closely. This is despite the differences in simulation design and duration, which demonstrates that this is a robust bias of the UM.
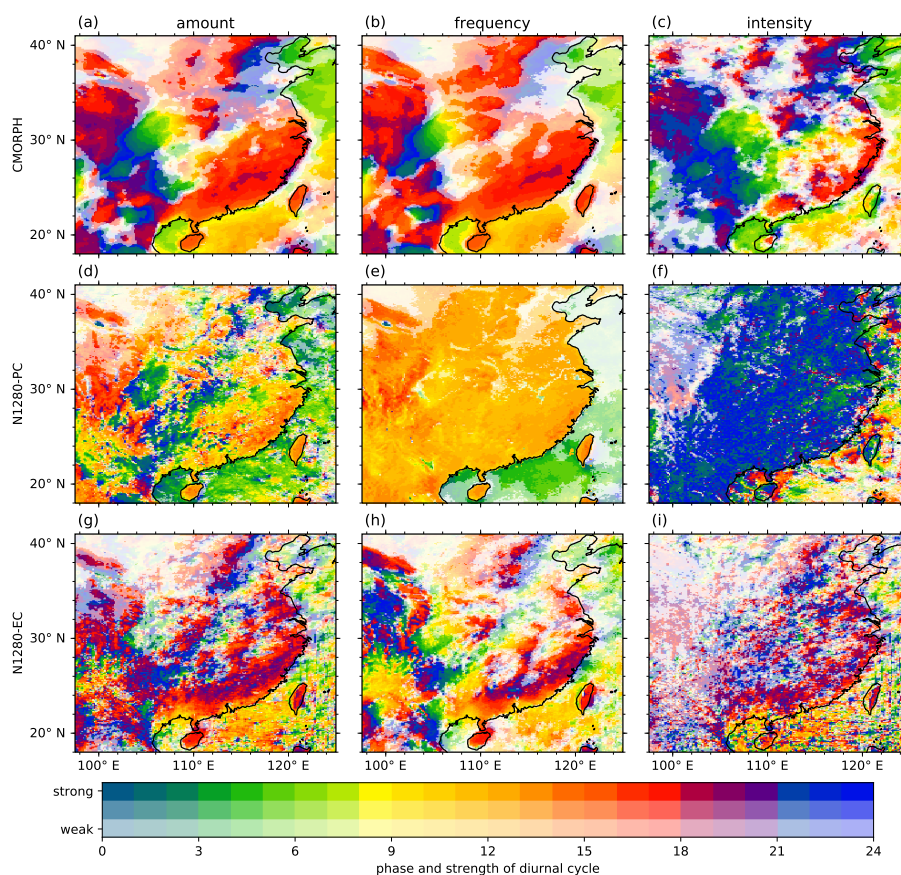
Figure 10. Diurnal cycle of amount, frequency and precipitation over south-eastern China. Layout as in Fig. 9. A visual representation of the amplitude of the diurnal cycle is given by the opacity. This is calculated separately for each dataset over the entire Asian domain (Sect. 2.3), not for the subregion shown here.

4.2 Diurnal cycle

555 As in Sect. 4.1, we focus on south-eastern China (Fig. 10) to investigate in detail the diurnal cycle of precipitation. In the observations, the phases of peak amount and frequency are again very similar, as in e.g. Zhou et al. (2008) and Li et al. (2018). The coastal region shows a peak in precipitation in the early evening for amount, frequency and intensity. The region of early-evening peak precipitation covers a larger area for amount and frequency. A phase delay is evident going south-west to north-east across the Sichuan Basin, with a peak in amount and frequency at 2300 LST in the south-west of the basin shifting to

560 1000 LST in the north-east. This has been observed in other studies (e.g., Li et al., 2018, 2020). Potential mechanisms include the interaction between the mean wind and the orography, and the steering-level winds at 700 hPa affecting the propagation of mesoscale convective systems.

As in Yu et al. (2007), their Fig. 3, and Chen et al. (2010), their Fig. 2a, a phase delay is also evident along 28° N, from 100–120° E for the diurnal cycle of the amount of precipitation. At the western end, the phase is in the late evening to midnight

565   and the diurnal cycle is strong. Further east, between 107–113° E, the amplitude weakens and the phase is around 0600–0900 LST. At the eastern end, the phase is around 1800 LST and the diurnal cycle is strong. There is a clear divide between the diurnal cycle over land and over ocean for all of amount, frequency and intensity, with oceanic precipitation peaking much closer to midday.

As in Sect. 3.3, N1280-PC produces a diurnal cycle of precipitation that is poorly matched with the observations across all

570   three precipitation fields over both land and ocean. The frequency and intensity are both too uniform, and generally have the wrong phase. For amount, there is more spatial variation but the phase rarely matches that observed, being too close to midday. A phase delay at 28° N is difficult to discern.

N1280-EC bears a stronger resemblance to the observations for amount and frequency, although it produces too much late-night precipitation near the coast. The contrast between land and ocean is closer to the observed contrast than it is for

575   N1280-PC. There are some signs of a phase delay going south-west to north-east across the Sichuan Basin, particularly in the frequency field, but it is not as clear as in the observations. This could be because N1280-EC produces too little precipitation in the Sichuan Basin (Fig. 9). Again, there is little clear sign of a phase delay at 28° N.

We can again compare directly with Li et al. (2018), their Fig. 3. We note that they do not show the strength of the diurnal cycles. Comparing the observations, their diurnal cycles of amount and frequency of precipitation are similar, and so are those

580   in this study. However, the phase of the peak is later over coastal China using CMORPH. This is consistent with Dai et al. (2007), who found that CMORPH had a delayed peak compared to gauge observations. Furthermore, the general patterns of amount, frequency and intensity are similar for both studies, although their fields are noisier (particularly frequency), which is to be expected given the shorter duration of their analysis.

Comparing the simulations to those of Li et al. (2018), the parametrized simulations produce diurnal cycles of amount,

585   frequency and intensity that are very similar to each other (Figs. 10d, e and f here; their Figs. 3c, f and i). They see a slightly later peak in intensity in coastal China; however, this could be due to the particular year they have analysed or the boundary conditions provided by their coarser driving model. For the explicit simulations, more differences between this study and Li et al. (2018) are evident. Although both studies show a strong similarity between amount and frequency, the peak of the diurnal cycle of these occurs often later in this study over coastal regions than in Li et al. (2018). A similar comparison holds for

590   intensity: there is a later peak in this study than in Li et al. (2018). Thus, even though both explicit simulations produce a similar results for amount, frequency and intensity of precipitation (Sect. 4.1), this does not hold as strongly for their diurnal cycles. This could be due to the different resolutions, the shorter simulations or the boundary conditions they imposed on their regional model.

5 Discussion

595 The value of aggregating precipitation over catchment basins was demonstrated in Schiemann et al. (2018). Here, we have extended that approach to allow for scale-selective analysis, which can be used to estimate the scales over which model simulations agree with observations and is independent of the resolutions and grids of the datasets that are evaluated. This is similar in spirit to the Fractional Skill Score metric (FSS), which can be used to determine the scale over which simulated precipitation is skilful (Roberts, 2008). However, there are important differences. For precipitation, the use of catchment basins

600 provides a direct route into integrating the results from simulations into hydrological impact assessments, for example to predict large-scale flooding (e.g., Grams et al., 2014) or to provide useful information for the construction and management of dam networks (e.g., Zabalza-Martínez et al., 2018); these are not possible using the FSS. Furthermore, our approach could be used to provide a quantification of how the risks of widespread flooding or dams exceeding their limits will change due to a changing climate, in a way which takes into account the physical geography of a region.

605 We found that different configurations of a high-resolution GCM over Asia have some similar biases to coarser models. The biases are very similar when the same convection parametrization scheme is used, reinforcing the finding in e.g. Klingaman et al. (2017) and Martin et al. (2017) that this scheme is a key candidate for improving GCM performance. Disabling the scheme clearly produces a mean state that has larger biases than when it is enabled, however the diurnal cycle of precipitation that is produced bears a stronger resemblance to the observational diurnal cycle. The high-resolution GCM configurations were

610 not tuned, as the purpose is to compare the simulations on an equal footing following the CMIP6 HighResMIP philosophy (Haarsma et al., 2016). Given the mean biases identified in this study, some tuning of the N1280-resolution models will be necessary before using them in climatological and climate service applications.

The representation of simulated orographic precipitation seems to be particularly important (Sects. 3.1, 4.1 and 4.2); this has been noted in previous studies (Schiemann et al., 2018; Vannière et al., 2019) and is something we would like to inves-

615 tigate further. The nature of the precipitation is also important when considering the potential impacts: extreme convective precipitation increases the risk of land slides and localized flooding (He et al., 2018), whereas extreme large-scale precipitation increases the risk of widespread flooding and stresses dam networks (Hunt and Menon, 2020). It is likely that simulations with different configurations will have an effect on the distribution of precipitation; investigating this over catchment basins would be an interesting way to follow up our study.

620 Extending the simulation length would permit further analysis. Extremes of precipitation are difficult to analyse over only four summer seasons – with a longer duration more robust statistics on these could be generated, as in Schiemann et al. (2018). A longer simulation would also allow a better characterization of the climatology by sampling more interannual variability. However, previous works suggests that a longer simulation might not change the biases that we have identified, as the systematic errors which affect climate simulations develop after only a few days (Martin et al., 2010). The computational cost of such

625 simulations would be high, however the benefits would include an improved understanding of how various processes improve with increasing resolution, and could also lay the foundations of the next generation of climate models.

6 Conclusions

We have compared the precipitation produced by new high-resolution GCM simulations against the observed precipitation from the CMORPH satellite dataset. The simulations were performed using the HadGEM3-GC3.1 Met Office Unified Model (UM) with longitudinal grid length of 14 km at 30° N for 2005–2008. Three simulations were run: one with the standard UM convection parametrization scheme (N1280-PC), one with the deep convection scheme disabled but shallow and mid-level convection still parametrized (N1280-HC), and one where the convection scheme was disabled so convection was simulated explicitly by the model (N1280-EC). To determine the effect of resolution, we included results from other coarser-resolution HadGEM3-GC3.1 simulations, which also use the convection parametrization scheme.

6.1 Mean summer precipitation over Asia

We compared how the mean JJA precipitation over Asia was represented in the observations and simulations. We found that the simulations broadly reproduce observed Asian summer precipitation distribution. N1280-PC exhibited substantial biases compared with CMORPH, producing too much precipitation over the Indian Ocean, too little precipitation over India, and too much precipitation over south-eastern China. This is similar to biases seen in the UM at coarser resolutions, for example Bush et al. (2015) found similar biases over India at N96 resolution. The N1280-PC simulation produced a band of precipitation on the Himalayas, indicating that it was representing some aspects of the interaction between the monsoon flow and orography, although the band was too wide and the precipitation rates were too high. The N1280-EC simulation worsened the existing biases in the UM, producing very little precipitation over India, and maximum precipitation rates over the east end of the Himalayas that were far in excess of observations. The N1280-HC simulation performed similarly to N1280-EC over land.

Using the newly developed BAsin-Scale Model Assessment ToolkIt (BASMATI), we averaged the precipitation field over hydrological catchment basins. The basins were chosen so that they were within a given size range, ranging from 2000–20000 km^2 to 200000–2000000 km^2 for the small and large basin scales respectively. This allowed for the mean summer precipitation in the simulations to be compared against the observed precipitation from CMORPH as a function of spatial scale. We found that all simulations improved as the spatial scale of analysis was increased, and that the lowest resolution simulation (N96-PC) produced the smallest error statistics, due mainly to its lack of very high precipitation rates.

6.2 Diurnal cycle of summer precipitation over Asia

Diurnal cycles of amount, frequency and intensity of precipitation were produced. We found that, for the summer diurnal cycle of precipitation over Asia, there were substantial differences between the observations and the simulations, with N1280-EC and N1280-HC generally performing far better than N1280-PC. For N1280-PC, the representation of the diurnal cycle is poor for all three precipitation measures. Over land across Asia, the peak in amount is too early at close to local midday, while the peak in frequency is too early at close to local midday and too uniform. The intensity is too late at local midnight and too uniform. For N1280-EC, the phase of the peak is more realistic for all three precipitation measures. However, the amplitude of the diurnal cycle is too weak over India, which we attributed to dry biases in this region. N1280-HC produced diurnal cycles

that resemble both of the other simulations, yielding useful information about which aspect of the representation of convection

660 is responsible for each of the diurnal cycles of amount, frequency and intensity. For the diurnal cycle of amount, N1280-HC closely resembles N1280-EC, indicating that the explicit deep convection is primarily responsible for producing this signal. For frequency, N1280-HC resembles N1280-PC more closely, indicating that the shallow and mid-level parametrizations of convection are responsible for this signal. However, in N1280-HC there is a clear difference between the diurnal cycles of amount and frequency, whereas in observations these fields are very similar.

665 Using BASMATI, diurnal cycles of amount, frequency and intensity were compared against CMORPH as a function of scale. For phase, N1280-EC and N1280-HC perform best for all three precipitation measures. These are the only simulations that perform better at larger spatial scale for the amount of precipitation, which we attributed to their lack of convection parametrization scheme. For the other simulations, higher resolution slightly improves the phase of the amount of precipitation. For the amplitude of amount, N1280-EC performs worst at small scales, but improves more rapidly than the others as scale

670 is increased, so that at the largest scales all simulations perform similarly well. N1280-EC performs worst for intensity at all scales. The simulations which use a convection parametrization scheme at resolutions between N96 and N1280 (180 and 14 km grid length at 30° N respectively) perform almost identically and similarly poorly compared to CMORPH for frequency. This is consistent with the overestimation of precipitation frequency in parametrized simulations, as shown in Fig. 9 and e.g. Martin et al. (2017).

675 **6.3 Summer precipitation over south-eastern China**

Focusing on south-eastern China, the three precipitation measures in CMORPH matched similar analysis using gauge data in Li et al. (2018). The similarity was greatest for amount, which is least sensitive to the choice of threshold, whereas the frequency and intensity were generally lower and higher than that in Li et al. (2018), respectively. Li et al. (2018) used a regional version of the UM with both parametrized and explicit convection, and the explicit convection simulation had a finer

680 resolution than the resolution used here. Similar to Li et al. (2018), N1280-PC and N1280-EC overestimated the amount of precipitation, which we attributed to the overestimations of frequency for N1280-PC, and to the overestimation of intensity for N1280-EC, consistent with Li et al. (2018). The similarities are apparent despite the differences in setup between this study and Li et al. (2018), which indicates that these are robust biases of the UM.

In N1280-PC, the land–sea contrast of the phase in maximum precipitation is unrealistic, whereas it is more realistic in

685 N1280-EC. N1280-EC produces phase and amplitude for diurnal cycles of amount and frequency that are closer to the observed values, although the intensity is too weak.

Code and data availability. The UM is available for use under licence. A number of research organisations and national meteorological services use the UM in collaboration with the Met Office to undertake basic atmospheric process research, produce forecasts, develop the UM code and build and evaluate Earth system models. For further information on how to apply for a licence see https://www.metoffice.gov.

690 uk/research/approach/collaboration/unified-model/partnership.

The N1280 UM simulation output data are available upon request, as each simulation is over 80 TB in size and stored on the Met Office Managed Archive Storage System (MASS). The other simulations are part of CMIP6 HighResMIP, and are available on Earth System Grid Federation (ESGF) with DOIs as shown in https://www.primavera-h2020.eu/modelling/ and cited in Table 1, where N512-PC in this study refers to HadGEM3-GC31-HM, N216-PC refers to HadGEM3-GC31-MM, and N96-PC refers to HadGEM3-GC31-LM. All of these used

695   the highresSST-present configuration.

CMORPH data are available from ftp://ftp.cpc.ncep.noaa.gov/precip/CMORPH_V1.0/CRT/8km-30min/.

The Basin-Scale Model Analysis ToolkIt (BASMATI) is freely available: https://github.com/markmuetz/basmati (v0.4.0 – https://github.com/markmuetz/basmati/archive/v0.4.0.zip).

The analysis packages that made use of BASMATI and produced all of the figures are also freely available: https://github.com/markmuetz/

700   cosmic (v0.4.0 – https://github.com/markmuetz/cosmic/archive/v0.4.0.zip); and https://github.com/markmuetz/remake (v0.4.1 – https://github.com/markmuetz/remake/archive/v0.4.1.zip).

Author contributions.   Contributions by the different authors include conceiving the study (MRM and RS), development of BASMATI, all data analysis and plotting of figures, and writing of the manuscript (MRM), performing the simulations (MJR and PLV), and feedback on the analysis as it progressed and on the manuscript (RS, NPK, AGT, PLV, MJR).

705   *Competing interests.*   The authors declare that there are no competing interests.

Acknowledgements.   MM, RS, AGT, NPK and PLV were supported by the COSMIC project through the Met Office Climate Science for Service Partnership (CSSP) China as part of the Newton Fund, contract number P106301. NPK was supported by a Natural Environmental Research Council (NERC) Independent Research Fellowship (NE/L010976/1) and by the ACREW programme of the National Centre for Atmospheric Science. MJR acknowledges PRIMAVERA funding received from the European Commission under grant agreement 641727

710   of the Horizon 2020 research programme.

The simulations were run using Met Office supercomputer (https://www.metoffice.gov.uk/about-us/what/technology/supercomputer), the ARCHER UK National Supercomputing Service (http://www.archer.ac.uk), and MONSooN (https://www.metoffice.gov.uk/research/approach/collaboration/jwcrp/monsoon-hpc). All analysis was done on JASMIN (https://www.jasmin.ac.uk/).

References

715 Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P.-P., Janowiak, J. E., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P., and Nelkin, E.: The Version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979–Present), Journal of Hydrometeorology, 4, 1147–1167, https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2, 2003.

Ajayamohan, R. S., Rao, S. A., and Yamagata, T.: Influence of Indian Ocean dipole on poleward propagation of boreal summer intraseasonal
720 oscillations, Journal of Climate, 21, 5437–5454, https://doi.org/10.1175/2008JCLI1758.1, 2008.

Arakawa, A. and Schubert, W. H.: Interaction of a cumulus cloud ensemble with the large-scale environment, Part I, Journal of the Atmospheric Sciences, 31, 674–701, https://doi.org/10.1175/1520-0469(1974)031<0674:IOACCE>2.0.CO;2, 1974.

Bador, M., Boé, J., Terray, L., Alexander, L. V., Baker, A., Bellucci, A., Haarsma, R., Koenigk, T., Moine, M.-P., Lohmann, K., et al.: Impact of higher spatial atmospheric resolution on precipitation extremes over land in global climate models, Journal of Geophysical Research:
725 Atmospheres, p. e2019JD032184, https://doi.org/10.1029/2019JD032184, 2019.

Bechtold, P., Semane, N., Lopez, P., Chaboureau, J.-P., Beljaars, A., and Bormann, N.: Representing equilibrium and nonequilibrium convection in large-scale models, Journal of the Atmospheric Sciences, 71, 734–753, https://doi.org/10.1175/JAS-D-13-0163.1, 2014.

Becker, T., Stevens, B., and Hohenegger, C.: Imprint of the convective parameterization and sea-surface temperature on large-scale convective self-aggregation, Journal of Advances in Modeling Earth Systems, 9, 1488–1505, https://doi.org/10.1002/2016MS000865, 2017.

730 Bollasina, M. A. and Ming, Y.: The general circulation model precipitation bias over the southwestern equatorial Indian Ocean and its implications for simulating the South Asian monsoon, Climate Dynamics, 40, 823–838, https://doi.org/10.1007/s00382-012-1347-7, 2013.

Bush, S. J., Turner, A. G., Woolnough, S. J., Martin, G. M., and Klingaman, N. P.: The effect of increased convective entrainment on Asian monsoon biases in the MetUM general circulation model, Quarterly Journal of the Royal Meteorological Society, 141, 311–326, https://doi.org/10.1002/qj.2371, 2015.

735 Chen, H., Yu, R., Li, J., Yuan, W., and Zhou, T.: Why nocturnal long-duration rainfall presents an eastward-delayed diurnal phase of rainfall down the Yangtze River valley, Journal of Climate, 23, 905–917, https://doi.org/10.1175/2009JCLI3187.1, 2010.

Covey, C., Gleckler, P. J., Doutriaux, C., Williams, D. N., Dai, A., Fasullo, J., Trenberth, K. E., and Berg, A.: Metrics for the diurnal cycle of precipitation: Toward routine benchmarks for climate models, Journal of Climate, 29, 4461–4471, https://doi.org/10.1175/JCLI-D-15-0664.1, 2016.

740 Curio, J., Chen, Y., Schiemann, R., Turner, A. G., Wong, K. C., Hodges, K., and Li, Y.: Comparison of a manual and an automated tracking method for Tibetan Plateau vortices, Advances in Atmospheric Sciences, 35, 965–980, https://doi.org/10.1007/s00376-018-7278-4, 2018.

Dai, A. and Trenberth, K. E.: The diurnal cycle and its depiction in the Community Climate System Model, Journal of Climate, 17, 930–951, https://doi.org/10.1175/1520-0442(2004)017<0930:TDCAID>2.0.CO;2, 2004.

Dai, A. and Wang, J.: Diurnal and semidiurnal tides in global surface pressure fields, Journal of the Atmospheric Sciences, 56, 3874–3891,
745 https://doi.org/10.1175/1520-0469(1999)056<3874:DASTIG>2.0.CO;2, 1999.

Dai, A., Lin, X., and Hsu, K.-L.: The frequency, intensity, and diurnal cycle of precipitation in surface and satellite observations over low-and mid-latitudes, Climate Dynamics, 29, 727–744, https://doi.org/10.1007/s00382-007-0260-y, 2007.

Ding, R., Ha, K.-J., and Li, J.: Interdecadal shift in the relationship between the East Asian summer monsoon and the tropical Indian Ocean, Climate Dynamics, 34, 1059–1071, https://doi.org/10.1007/s00382-009-0555-2, 2010.

750 ESRI, E.: Shapefile technical description, An ESRI white paper, 39, https://support.esri.com/en/white-paper/279, 1998.

Field, P. R., Roberts, M. J., and Wilkinson, J. M.: Simulated lightning in a convection permitting global model, Journal of Geophysical Research: Atmospheres, 123, 9370–9377, https://doi.org/10.1029/2018JD029295, 2018.

Golding, B. W.: Quantitative precipitation forecasting in the UK, Journal of Hydrology, 239, 286–305, https://doi.org/10.1016/S0022-1694(00)00354-1, 2000.

755 Grams, C. M., Binder, H., Pfahl, S., Piaget, N., and Wernli, H.: Atmospheric processes triggering the central European floods in June 2013, Natural Hazards and Earth System Sciences, 14, 1691–1702, https://doi.org/10.5194/nhess-14-1691-2014, 2014.

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N. S., Guemas, V., et al.: High resolution model intercomparison project (HighResMIP v1. 0) for CMIP6, Geoscientific Model Development, 9, 4185–4208, https://doi.org/10.5194/gmd-9-4185-2016, 2016.

760 He, B., Huang, X., Ma, M., Chang, Q., Tu, Y., Li, Q., Zhang, K., and Hong, Y.: Analysis of flash flood disaster characteristics in China from 2011 to 2015, Natural Hazards, 90, 407–420, https://doi.org/10.1007/s11069-017-3052-7, 2018.

Holloway, C. E., Woolnough, S. J., and Lister, G. M. S.: The effects of explicit versus parameterized convection on the MJO in a large-domain high-resolution tropical case study. Part I: Characterization of large-scale organization and propagation, Journal of the Atmospheric Sciences, 70, 1342–1369, https://doi.org/10.1175/JAS-D-12-0227.1, 2013.

765 Houze Jr., R. A.: Mesoscale convective systems, Reviews of Geophysics, 42, RG4003, https://doi.org/10.1029/2004RG000150, 2004.

Hsu, P.-C., Lee, J.-Y., and Ha, K.-J.: Influence of boreal summer intraseasonal oscillation on rainfall extremes in southern China, International Journal of Climatology, 36, 1403–1412, https://doi.org/10.1002/joc.4433, 2016.

Hunt, K. M. and Menon, A.: The 2018 Kerala floods: a climate change perspective, Climate Dynamics, 54, 2433–2446, https://doi.org/10.1007/s00382-020-05123-7, 2020.

770 Hurkmans, R., Terink, W., Uijlenhoet, R., Torfs, P., Jacob, D., and Troch, P. A.: Changes in streamflow dynamics in the Rhine basin under three high-resolution regional climate scenarios, Journal of Climate, 23, 679–699, https://doi.org/10.1175/2009JCLI3066.1, 2010.

Joyce, R. J., Janowiak, J. E., Arkin, P. A., and Xie, P.: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution, Journal of Hydrometeorology, 5, 487–503, https://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2, 2004.

775 Kennedy, J., Titchner, H., Rayner, N., and Roberts, M. J.: input4MIPs.MOHC.SSTsAndSeaIce.HighResMIP.MOHC-HadISST-2-2-0-0-0, https://doi.org/10.22033/ESGF/input4MIPs.1221, 2017.

Khairoutdinov, M., Randall, D., and DeMott, C.: Simulations of the atmospheric general circulation using a cloud-resolving model as a superparameterization of physical processes, Journal of the Atmospheric Sciences, 62, 2136–2154, https://doi.org/10.1175/JAS3453.1, 2005.

780 Klingaman, N. P., Martin, G. M., and Moise, A. F.: ASoP (v1. 0): A set of methods for analyzing scales of precipitation in general circulation models, Geoscientific Model Development, 10, 57–83, https://doi.org/10.5194/gmd-10-57-2017, 2017.

Lehner, B.: HydroBASINS: Global watershed boundaries and sub-basin delineations derived from HydroSHEDS data at 15 second resolution. Technical Documentation Version 1. c (with and without inserted lakes), https://hydrosheds.org/images/inpages/HydroBASINS_TechDoc_v1c.pdf, 2014.

785 Lehner, B. and Grill, G.: Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems, Hydrological Processes, 27, 2171–2186, https://doi.org/10.1002/hyp.9740, 2013.

Levine, R. C., Turner, A. G., Marathayil, D., and Martin, G. M.: The role of northern Arabian Sea surface temperature biases in CMIP5 model simulations and future projections of Indian summer monsoon rainfall, Climate Dynamics, 41, 155–172, https://doi.org/10.1007/s00382-012-1656-x, 2013.

790 Li, J., Li, Y., Zhao, T., Schiemann, R., Jiang, X., and Muetzelfeldt, M.: Northeastward propagation of nocturnal precipitation over the Sichuan Basin, International Journal of Climatology, https://doi.org/https://doi.org/10.1002/joc.6886, 2020.

Li, P., Furtado, K., Zhou, T., Chen, H., Li, J., Guo, Z., and Xiao, C.: The diurnal cycle of East Asian summer monsoon precipitation simulated by the Met Office Unified Model at convection-permitting scales, Climate Dynamics, pp. 1–21, https://doi.org/10.1007/s00382-018-4368-z, 2018.

795 Marsham, J. H., Dixon, N. S., Garcia-Carreras, L., Lister, G. M. S., Parker, D. J., Knippertz, P., and Birch, C. E.: The role of moist convection in the West African monsoon system: Insights from continental-scale convection-permitting simulations, Geophysical Research Letters, 40, 1843–1849, https://doi.org/10.1002/grl.50347, 2013.

Martin, G. M., Milton, S. F., Senior, C. A., Brooks, M. E., Ineson, S., Reichler, T., and Kim, J.: Analysis and reduction of systematic errors through a seamless approach to modeling weather and climate, Journal of Climate, 23, 5933–5957, 800 https://doi.org/10.1175/2010JCLI3541.1, 2010.

Martin, G. M., Klingaman, N. P., and Moise, A. F.: Connecting spatial and temporal scales of tropical precipitation in observations and the MetUM-GA6, Geoscientific Model Development, 10, 105–126, https://doi.org/10.5194/gmd-10-105-2017, 2017.

Mitra, A. K., Momin, I. M., Rajagopal, E. N., Basu, S., Rajeevan, M. N., and Krishnamurti, T. N.: Gridded daily Indian monsoon rainfall for 14 seasons: Merged TRMM and IMD gauge analyzed values, Journal of Earth System Science, 122, 1173–1182, 805 https://doi.org/10.1007/s12040-013-0338-3, 2013.

Mulcahy, J. P., Jones, C., Sellar, A., Johnson, B., Boutle, I. A., Jones, A., Andrews, T., Rumbold, S. T., Mollard, J., Bellouin, N., et al.: Improved aerosol processes and effective radiative forcing in HadGEM3 and UKESM1, Journal of Advances in Modeling Earth Systems, 10, 2786–2805, https://doi.org/10.1029/2018MS001464, 2018.

Ploshay, J. J. and Lau, N.-C.: Simulation of the diurnal cycle in tropical rainfall and circulation during boreal summer with a high-resolution 810 GCM, Monthly Weather Review, 138, 3434–3453, https://doi.org/10.1175/2010MWR3291.1, 2010.

Reichle, R. H., Koster, R. D., De Lannoy, G. J. M., Forman, B. A., Liu, Q., Mahanama, S. P. P., and Touré, A.: Assessment and enhancement of MERRA land surface hydrology estimates, Journal of Climate, 24, 6322–6338, https://doi.org/10.1175/JCLI-D-10-05033.1, 2011.

Roberts, M. J.: MOHC HadGEM3-GC31-HM model output prepared for CMIP6 HighResMIP, https://doi.org/10.22033/ESGF/CMIP6.446, 2017a.

815 Roberts, M. J.: MOHC HadGEM3-GC31-LM model output prepared for CMIP6 HighResMIP, https://doi.org/10.22033/ESGF/CMIP6.1321, 2017b.

Roberts, M. J.: MOHC HadGEM3-GC31-MM model output prepared for CMIP6 HighResMIP, https://doi.org/10.22033/ESGF/CMIP6.1902, 2017c.

Roberts, M. J., Vidale, P. L., Senior, C., Hewitt, H. T., Bates, C., Berthou, S., Chang, P., Christensen, H. M., Danilov, S., Demory, M.-E., 820 et al.: The benefits of global high resolution for climate simulation: process understanding and the enabling of stakeholder decisions at the regional scale, Bulletin of the American Meteorological Society, 99, 2341–2359, https://doi.org/10.1175/BAMS-D-15-00320.1, 2018.

Roberts, M. J., Baker, A., Blockley, E. W., Calvert, D., Coward, A., Hewitt, H. T., Jackson, L. C., Kuhlbrodt, T., Mathiot, P., Roberts, C. D., et al.: Description of the resolution hierarchy of the global coupled HadGEM3-GC3. 1 model as used in CMIP6 HighResMIP experiments, Geoscientific Model Development, 12, 4999–5028, https://doi.org/10.5194/gmd-12-4999-2019, 2019.

825 Roberts, N.: Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model, Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling, 15, 163–169, https://doi.org/10.1002/met.57, 2008.

Sato, T., Miura, H., Satoh, M., Takayabu, Y. N., and Wang, Y.: Diurnal cycle of precipitation in the tropics simulated in a global cloud-resolving model, Journal of Climate, 22, 4809–4826, https://doi.org/10.1175/2009JCLI2890.1, 2009.

830 Satoh, M., Tomita, H., Yashiro, H., Miura, H., Kodama, C., Seiki, T., Noda, A. T., Yamada, Y., Goto, D., Sawada, M., et al.: The non-hydrostatic icosahedral atmospheric model: Description and development, Progress in Earth and Planetary Science, 1, 18, https://doi.org/10.1186/s40645-014-0018-1, 2014.

Schär, C., Fuhrer, O., Arteaga, A., Ban, N., Charpilloz, C., Di Girolamo, S., Hentgen, L., Hoefler, T., Lapillonne, X., Leutwyler, D., et al.: Kilometer-scale climate models: Prospects and challenges, Bulletin of the American Meteorological Society, 101, E567–E587,
835 https://doi.org/10.1175/BAMS-D-18-0167.1, 2020.

Schiemann, R., Vidale, P. L., Shaffrey, L. C., Johnson, S. J., Roberts, M. J., Demory, M.-E., Mizielinski, M. S., and Strachan, J.: Mean and extreme precipitation over European river basins better simulated in a 25km AGCM, Hydrology and Earth System Sciences, 22, 3933–3950, https://doi.org/10.5194/hess-22-3933-2018, 2018.

Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-year high-resolution global dataset of meteorological forcings for land surface
840 modeling, Journal of Climate, 19, 3088–3111, https://doi.org/10.1175/JCLI3790.1, 2006.

Smagorinsky, J.: Some aspects of the general circulation, Quarterly Journal of the Royal Meteorological Society, 90, 1–14, https://doi.org/10.1002/qj.49709038302, 1964.

Stevens, B., Fiedler, S., Kinne, S., Peters, K., Rast, S., Müsse, J., Smith, S. J., and Mauritsen, T.: MACv2-SP: A parameterization of anthropogenic aerosol optical properties and an associated Twomey effect for use in CMIP6, Geoscientific Model Development, 10, 433–452,
845 https://doi.org/10.5194/gmd-10-433-2017, 2017.

Stirling, A. J. and Stratton, R. A.: Entrainment processes in the diurnal cycle of deep convection over land, Quarterly Journal of the Royal Meteorological Society, 138, 1135–1149, https://doi.org/10.1002/qj.1868, 2012.

Titchner, H. A. and Rayner, N. A.: The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations, Journal of Geophysical Research: Atmospheres, 119, 2864–2889, https://doi.org/10.1002/2013JD020316, 2014.

850 Tomassini, L.: Mesoscale circulations and organized convection in African easterly waves, Journal of the Atmospheric Sciences, 75, 4357–4381, https://doi.org/10.1175/JAS-D-18-0183.1, 2018.

Vannière, B., Demory, M.-E., Vidale, P. L., Schiemann, R., Roberts, M. J., Roberts, C. D., Matsueda, M., Terray, L., Koenigk, T., and Senan, R.: Multi-model evaluation of the sensitivity of the global energy budget and hydrological cycle to resolution, Climate Dynamics, 52, 6817–6846, https://doi.org/10.1007/s00382-018-4547-y, 2019.

855 Verdin, K. L. and Verdin, J. P.: A topological system for delineation and codification of the Earth's river basins, Journal of Hydrology, 218, 1–12, https://doi.org/10.1016/S0022-1694(99)00011-6, 1999.

Vergara-Temprado, J., Ban, N., Panosetti, D., Schlemmer, L., and Schär, C.: Climate Models Permit Convection at Much Coarser Resolutions Than Previously Considered, Journal of Climate, 33, 1915–1933, https://doi.org/10.1175/JCLI-D-19-0286.1, 2020.

Virts, K. S. and Houze Jr, R. A.: Seasonal and intraseasonal variability of mesoscale convective systems over the South Asian monsoon
860 region, Journal of the Atmospheric Sciences, 73, 4753–4774, https://doi.org/10.1175/JAS-D-16-0022.1, 2016.

Walters, D., Boutle, I., Brooks, M., Thomas, M., Stratton, R. A., Vosper, S., Wells, H., Williams, K., Wood, N., Allen, T., et al.: The Met Office Unified Model Global Atmosphere 6.0/6.1 and JULES Global Land 6.0/6.1 configurations, Geoscientific Model Development, 10, 1487, https://doi.org/10.5194/gmd-10-1487-2017, 2017.

Walters, D., Baran, A. J., Boutle, I., Brooks, M., Earnshaw, P., Edwards, J., Furtado, K., Hill, P., Lock, A. P., Manners, J., Morcrette, C.,
865   Mulcahy, J., Sanchez, C., Smith, C., Stratton, R. A., Tennant, W. J., Tomassini, L., Van Weverberg, K., Vosper, S., Willett, M. R., Browse, J., Bushell, A., Carslaw, K., Dalvi, M., Essery, R., Gedney, N., Hardiman, S., Johnson, B., Johnson, C., Jones, A., Jones, C., Mann, G., Milton, S., Rumbold, H., Sellar, A., Ujiie, M., Whitall, M. A., Williams, K., and Zerroukat, M.: The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations, Geoscientific Model Development, 12, 1909–1963, https://doi.org/10.5194/gmd-12-1909-2019, https://www.geosci-model-dev.net/12/1909/2019/, 2019.

870   Webb, M. J., Lock, A. P., Bretherton, C. S., Bony, S., Cole, J. N. S., Idelkadi, A., Kang, S. M., Koshiro, T., Kawai, H., Ogura, T., et al.: The impact of parametrized convection on cloud feedback, Philosophical Transactions of the Royal Society A, 373, 20140414, https://doi.org/10.1098/rsta.2014.0414, 2015.

Willetts, P. D., Marsham, J. H., Birch, C. E., Parker, D. J., Webster, S., and Petch, J.: Moist convection and its upscale effects in simulations of the Indian monsoon with explicit and parametrized convection, Quarterly Journal of the Royal Meteorological Society, 143, 1073–1085,
875   https://doi.org/10.1002/qj.2991, 2017.

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., Davis, P., Graham, T., Hewitt, H. T., Hill, R., et al.: The Met Office global coupled model 3.0 and 3.1 (GC3.0 and GC3.1) configurations, Journal of Advances in Modeling Earth Systems, 10, 357–380, https://doi.org/10.1002/2017MS001115, 2018.

Wu, R. and Wang, B.: A contrast of the East Asian summer monsoon–ENSO relationship between 1962–77 and 1978–93, Journal of Climate,
880   15, 3266–3279, https://doi.org/10.1175/1520-0442(2002)015<3266:ACOTEA>2.0.CO;2, 2002.

Xavier, P. K., Marzin, C., and Goswami, B. N.: An objective definition of the Indian summer monsoon season and a new perspective on the ENSO–monsoon relationship, Quarterly Journal of the Royal Meteorological Society, 133, 749–764, https://doi.org/10.1002/qj.45, 2007.

Yang, G.-Y. and Slingo, J.: The diurnal cycle in the tropics, Monthly Weather Review, 129, 784–801, https://doi.org/10.1175/1520-0493(2001)129<0784:TDCITT>2.0.CO;2, 2001.

885   Yang, X., Fei, J., Huang, X., Cheng, X., Carvalho, L. M. V., and He, H.: Characteristics of mesoscale convective systems over China and its vicinity using geostationary satellite FY2, Journal of Climate, 28, 4890–4907, https://doi.org/10.1175/JCLI-D-14-00491.1, 2015.

Yatagai, A., Kamiguchi, K., Arakawa, O., Hamada, A., Yasutomi, N., and Kitoh, A.: APHRODITE: Constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges, Bulletin of the American Meteorological Society, 93, 1401–1415, https://doi.org/10.1175/BAMS-D-11-00122.1, 2012.

890   Yu, R., Zhou, T., Xiong, A., Zhu, Y., and Li, J.: Diurnal variations of summer precipitation over contiguous China, Geophysical Research Letters, 34, https://doi.org/10.1029/2006GL028129, 2007.

Zabalza-Martínez, J., Vicente-Serrano, S. M., López-Moreno, J. I., Borràs Calvo, G., Savé, R., Pascual, D., Plá, E., Morán-Tejeda, E., Domínguez-Castro, F., and Tague, C.: The Influence of Climate and Land-Cover Scenarios on Dam Management Strategies in a High Water Pressure Catchment in Northeast Spain, Water, 10, 1668, https://doi.org/10.3390/w10111668, 2018.

895   Zhang, L., Wu, P., Zhou, T., Roberts, M. J., and Schiemann, R.: Added value of high resolution models in simulating global precipitation characteristics, Atmospheric Science Letters, 17, 646–657, https://doi.org/10.1002/asl.715, 2016.

Zhang, Q., Gu, X., Li, J., Shi, P., and Singh, V. P.: The impact of tropical cyclones on extreme precipitation over coastal and inland areas of China and its association to ENSO, Journal of Climate, 31, 1865–1880, https://doi.org/10.1175/JCLI-D-17-0474.1, 2018.

Zhong, Z. and Hu, Y.: Impacts of tropical cyclones on the regional climate: An East Asian summer monsoon case, Atmospheric Science
900   Letters, 8, 93–99, https://doi.org/10.1002/asl.158, 2007.

Zhou, T., Yu, R., Chen, H., Dai, A., and Pan, Y.: Summer precipitation frequency, intensity, and diurnal cycle over China: A comparison of
       satellite data with rain gauge observations, Journal of Climate, 21, 3997–4010, https://doi.org/10.1175/2008JCLI2028.1, 2008.