

## **Hadley Centre Technical Note 105**

# **Rainfall disaggregation via multiplicative random cascades in the JULES land surface model**

March 22, 2019

Ivan Paspaldzhiev, Karina Williams, Pete Falloon

## Contents

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introduction</b>   | <b>1</b>  |
| <b>2</b> | <b>Methodology</b>  | <b>2</b>  |
| 2.1      | Disaggregation in JULES via the IMOGEN approach . . . . .                 | 2         |
| 2.2      | Precipitation disaggregation via multiplicative random cascades . . . . . | 2         |
| 2.3      | Experimental setup . . . . .  | 3         |
| <b>3</b> | <b>Results</b>  | <b>5</b>  |
| 3.1      | Model water budget . . . . .  | 5         |
| 3.2      | Climate impacts . . . . .   | 15        |
| <b>4</b> | <b>Discussion</b>   | <b>17</b> |
| <b>5</b> | <b>Conclusion</b>   | <b>19</b> |
| <b>A</b> | <b>SPI and SRI calculated over different timescales</b>                   | <b>23</b> |

## 1 Introduction

The Joint UK Land Environment System (JULES) [Best et al., 2011, Clark et al., 2011] is a widely used land-surface model (LSM) formulation, designed to run at short timescales, so as to resolve the exchange of heat, moisture and momentum with the atmosphere. It can be run online as part of the Met Office Unified Model, or offline for impact studies. In the latter case, JULES meteorological forcings are provided as input files. As there is significant practical difficulty in storing very large volumes of forcing data, as is especially the case for global-scale studies, forcing datasets are often available at the 3-hourly scale (e.g. Weedon et al. [2014]). However, daily data is often the only available forcing for model intercomparison studies such as ISI-MIP [Warszawski et al., 2014] and EUPORIAS [Hewitt et al., 2013]. In order to make such data suitable for forcing JULES, which is known to be sensitive to forcings' temporal resolution [Compton and Best, 2011], data disaggregation is used internally in the model - namely the IMOGEN approach [Huntingford et al., 2010]. Multiple LSM studies have shown that inaccurate forcings, especially for precipitation, can introduce significant error in the model hydrological outputs [Robock et al., 2003, Pan and Wood, 2004, Berg et al., 2003, Fekete et al., 2004, Nijssen and Lettenmaier, 2004], which may have direct implications for using LSMs for studying hydrological climate impacts. Williams and Clark [2014] have shown that current method for disaggregating precipitation in JULES can yield good results, but is however strongly sensitive to its parameterisation, which itself is somewhat ad-hoc and not very physically defensible.

In previous work, Paspaldzhiev [2015] described an alternative precipitation disaggregator based on the multiplicative cascade procedure of Molnar and Burlando [2005], parameterising it using WFDEI-GPCC 3-hourly rainfall [Weedon et al., 2014] and assessing its ability to reproduce the statistics of

the parameterisation data. The procedure of Paspaldzhiev [2015] conserves the mean of the data by design, but was found to overestimate the variance and magnitude of extremes by a factor of 2 on average, as well as underestimate the lag-1 autocorrelation of the data, the length of longer rainfall events, as well as event magnitudes. Multiplicative cascades are based on the empirically observed scaling properties of rainfall [Schertzer and Lovejoy, 1987, Hubert et al., 1993], also observed for ERA-Interim data [Lovejoy et al., 2012](which WFDEI de-biases), which makes them attractive from the standpoint of physical plausibility. Further still, the adopted formulation of Paspaldzhiev [2015] has only two fitted parameters, offering improved parsimony over the IMOGEN approach.

With the aforementioned in mind, there is then utility in assessing the variability in JULES hydrological outputs in a disaggregation scenario if multiplicative cascades are to be adopted for disaggregating precipitation. As JULES is also often used in a climate impacts context, it would also be interesting to evaluate the influence of the introduced bias on the water budget on derived climate impact metrics. This report aims to assess this and build on the aforementioned work of Paspaldzhiev [2015].

## 2 Methodology

### 2.1 Disaggregation in JULES via the IMOGEN approach

Disaggregation of daily data exists internally in JULES from version 4.0 upward and can be switched on via `l_daily_disagg = TRUE`. In this analysis, we are only concerned with the disaggregation of precipitation, which is controlled by the `precip_disagg_method` in `JULES_DRIVE`. We use `precip_disagg_method = 3`, which restricts an entire day's worth of precipitation of each type  $x$  to a single event, which is deposited randomly at a time between the beginning of the day and  $\tau_x$  before the end of the day (in UTC),  $x$ =convective rain, large-scale rain, convective snow and large-scale snow respectively (in seconds). We use the JULES defaults  $\tau_{convrain} = 6 \times 3600s$ ,  $\tau_{lsrain} = \tau_{lssnow} = \tau_{convsnow} = 1 \times 3600s$ . Williams and Clark [2014] provide a full description of the disaggregator.

### 2.2 Precipitation disaggregation via multiplicative random cascades

Taking daily rainfall to be at scale  $n = 0$  with resolution  $\lambda_{n_0} = 24h$ , a discrete microcanonical multiplicative cascade procedure distributes rainfall to successively smaller subdivisions with a branching number  $b = 2$ . The  $i$ th subdivision at level  $n$  is denoted  $\Delta_{n,i}$ , there being  $i = 1, \dots, b^n$  subdivisions at level  $n$ . Distributed mass is then the product of a multiplicative process at all levels  $n$ , with the mass of  $\Delta_{n,i}$  given as:

$$\mu_n(\Delta_{n,i}) = R_0 \prod_{j=1}^n W_j(i) \quad \text{for } i = 1, 2, \dots, b^n; \quad n > 0 \quad (1)$$

$R_0$  is the initial depth of rainfall at  $n = 0$  and  $W$  is the cascade generator, where for a given  $\Delta_{n,i}$  timestep, there are two corresponding timesteps  $\Delta_{n+1,2i-1}$ ,  $\Delta_{n+1,2i}$ , for which  $W_j(2i-1) + W_j(2i) =$

1, which ensures that mass is conserved in the redistribution.  $W_j$  then denotes the *distribution* of weights for a transition between two given cascade levels, i.e.  $j$  in Equation 1 stands for  $n \rightarrow n + 1$ . The distribution  $W_j$  is identical to the proportions of  $R_n$  that get redistributed to its two corresponding timesteps at  $n + 1$ , calculated from non-overlapping adjacent pairs of rainfall measurements:

$$W_j(2i - 1) = \frac{R_n(i)}{R_{n+1}(2i - 1)}; \quad W_j(2i) = \frac{R_n(i)}{R_{n+1}(2i)}; \quad i = 1, \dots, N - 1 \text{ for } R_n(i) \neq 0 \quad (2)$$

$N$  being the length of the time series at scale  $n$ . The intermittency ( $p_{0,j}$ ) and variability ( $a_j$ ) of rain are parameterised independently for every scale transition  $j$ , so that  $p_{0,j} = P(W_j(2i - 1) = 0, 1 \text{ or } W_j(2i) = 0, 1)$ , i.e. the proportion of 0 and 1 to the total number of  $W_j$  instances, and  $a_j = \frac{1}{8\text{Var}(W_j)} - 0.5$  being the parameter of the symmetrical beta distribution  $f(w) = \frac{1}{B(a)} w^{a-1} (1 - w)^{a-1}$ . The formulation  $W_j(2i - 1) = x_1 / (x_1 + x_2)$  and  $W_j(2i) = x_2 / (x_1 + x_2)$  is utilised to generate two independent weights that sum to unity and are beta-distributed, where  $x_1$  and  $x_2$  are independent gamma-distributed random numbers with parameter  $a = a_j$  [Menabde and Sivapalan, 2000]. The full timeseries of rainfall is used for the parameterisation at a given grid cell.

See Paspaldzhiev [2015] for a more thorough description of the cascade formulation.

## 2.3 Experimental setup

The 0.5° spatial and 3-hourly temporal resolution, 1979-2010 WFDEI product Weedon et al. [2014] is randomly subsampled in space based on the regional classification scheme of Giorgi and Francisco [2000], including a new region "NMA" (Northern-most Asia) for completeness, as this is excluded in the original classification. 3% of the area of each bounding box in Figure 1 is sampled for a total of ~3000 points, found to be a good trade-off between global coverage and computational time. Please see Giorgi and Francisco [2000] for the meaning of the acronyms for each region.

Four model configurations are run for the full span of the WFDEI data (01/01/1970 00:00:00 - 31/12/2010 21:00:00), with the forcings interpolated (`interp`) to a 1h temporal resolution:

- 3h\_F - this run is forced with the original WFDEI 3-hourly forcings with interpolation flag `interp=nb`.
- 3h\_D - identical to 3h\_F, but with the rainfall forcing substituted with a timeseries generated by using JULES to disaggregate WFDEI daily means with `precip_disagg_method=3` and `interp=nb`, as outlined in Section 2.1<sup>1</sup>.
- 3h\_C - identical to 3h\_F, but with the rainfall forcing substituted with a timeseries generated by disaggregating WFDEI daily means with the cascade procedure outlined in Section 2.2<sup>2</sup>.
- 24h\_F - this run is forced with daily means of WFDEI data<sup>3</sup>, which are interpolated to the model timestep without disaggregation with `interp=nf`.

<sup>1</sup>This is necessary, as at the time of writing, JULES does not allow for disaggregating only a subset of the model forcings.

<sup>2</sup>This is necessary, as at the time of writing, the cascade disaggregator exists separately to JULES as standalone Python code (available on request).

<sup>3</sup>This makes for a somewhat unfair comparison as the interpolation of all forcings in addition to those for rainfall would add an additional source of bias. I'd like to remedy this in the future.

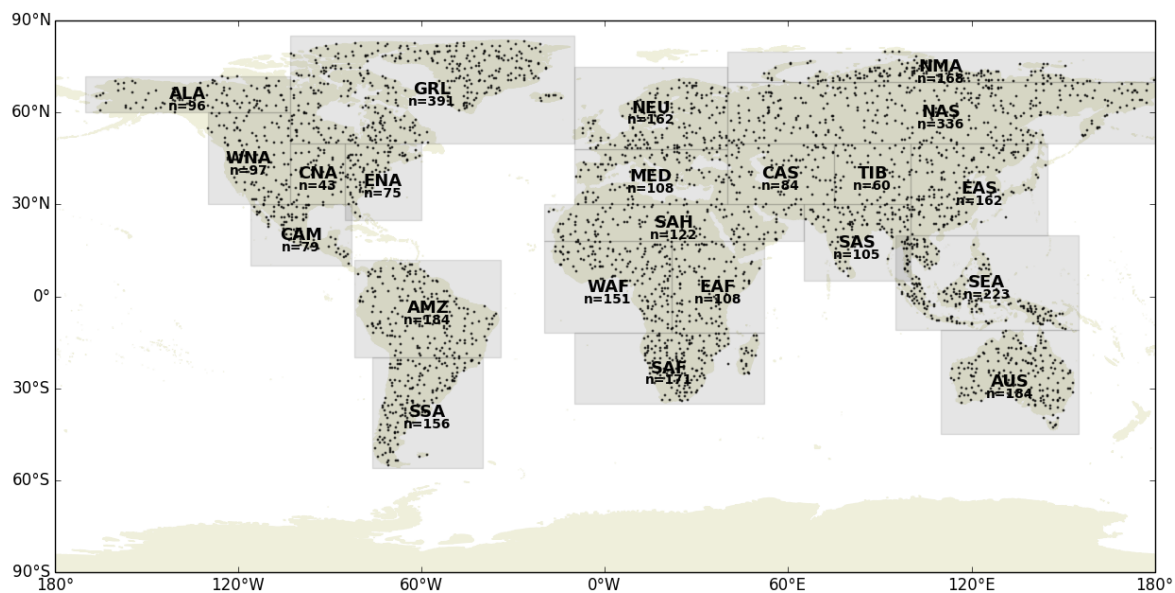


Figure 1: Map of sampled points for this analysis within each Giorgi region. For abbreviations, see Giorgi and Francisco [2000]. Note that under the regional definitions of Giorgi and Francisco [2000] there is a slight overlap between regions SAS and SEA. This does not mean that the points therein have been included in the analysis twice, it simply makes it ambiguous as to which region they belong to.

The chosen interpolation flags ensure that the output for each timestep correctly aligns with the time-series of the forcing. See the JULES manual for the definition of each flag. For all configurations, rainfall and snowfall totals are supplied, but only rainfall is disaggregated. Rainfall is redistributed to convective and large-scale components using the JULES default temperature threshold  $t_{\text{for\_con\_rain}} = 373.15$ . Convective snow is assumed to be zero (JULES default).

In order to avoid any spin-up effects (particularly to do with the model deep water store in TOP-MODEL - of direct relevance to the hydrological budget), all configurations are initialised from dump of a 60-year spin-up using the forcings of 3h\_F. The science settings for all runs are identical<sup>4</sup>.

The evaluation is performed by comparing the magnitude of the averages of the model hydrological fluxes (in  $\text{mm/day}$ ) calculated over the full integration period for each model run - evapotranspiration from the canopy/surface store ( $\text{ecan\_gb}$ ), evapotranspiration from soil ( $\text{esoil\_gb}$ ), total evapotranspiration ( $\text{etotal\_gb}$ ), surface runoff ( $\text{surf\_roff}$ ), sub-surface runoff ( $\text{sub\_surf\_roff}$ ) and total runoff ( $\text{total\_roff}$ ). We test whether the distribution of averages for each region is significantly different for all model runs compared to 3h\_F using the two-sample Kolmogorov-Smirnov (KS), as it does not make assumptions regarding the data's distribution. We adjust for multiple comparisons via the Holm-Bonferroni (HB) method at a significance level of  $p = 0.05$ . The HB method is regarded as a uniformly more powerful variant of the simple Bonferroni correction [Aickin and Gensler, 1996], offering a less conservative adjustment of the family-wise error rate, while still being valid under arbitrary assumptions.

<sup>4</sup>Rose suites are available for all model runs, please consult these for in-depth science settings. These are suites u-aa913, u-ab096, u-ab095 and u-aa916 for each run respectively.

[Holm, 1979]. Adjusted  $p$  values are reported.

Additionally, the degree of over or underestimation compared of the evaluation statistics compared to  $3h\_F$  is assessed by computing  $RUN\_ID/3h\_F$ , where  $RUN\_ID$  is each of the other model runs. Finally, in order to get a more comprehensive picture of the performance across the distribution for each variable, we compute quantiles 1 to 100 on the global aggregation of all sampled points for non-zero values of each variable, as well as the deviation of each model run for each quantile as  $RUN\_ID - 3h\_F$ . For the latter, the signed deviation is chosen instead of over/underestimation because at the lower end of the distribution, the model is able to produce very low non-zero values with unrealistically high precision (e.g. 3+ decimal places), where differences between the model runs would not have any practical significance, yet may be unrealistically exaggerated if ratios are used. To still present a picture at both ends of the distribution, we use a log scale for the actual magnitudes of the quantiles.

In addition to the assessment of the hydrological budget, in order to put our results in a climate impacts context, we compare each model for impact metrics to do with drought and high-end extremes. The Standardized Precipitation Index (SPI), Standardized Runoff Index (SRI) and Soil Moisture Anomaly (SMA) are used as in Burke and Brown [2008], Taylor et al. [2013] for assessing drought, and SPI and SRI are also used for high-end extremes. Drought at different severities is defined as a threshold at quantiles 1, 5, 10, 15 and 20 for each index at each grid square as per Burke and Brown [2008], Taylor et al. [2013], with high-end extremes defined as the inverse (80th, 85th, 90th, 95th, 99th quantile). SPI and SRI are assessed for monthly, seasonal, yearly and bi-yearly accumulations [Burke and Brown, 2008, Taylor et al., 2013], as it has been observed that the resultant magnitude of the impact metric is sensitive to the accumulation period on which it is calculated [Taylor et al., 2013]. Our analysis thus also allows us to assess if there is an additional variability introduced by different disaggregation options. Unlike Burke and Brown [2008], Taylor et al. [2013], we do not exclude cold regions, as these papers do so due to an index under comparison being sensitive to the presence or absence of frozen processes, which is not the case with the metrics we assess. We use the model rainfall forcing for calculation SPI and SRI, and the model soil moisture content at 1m depth (variable `smc_avail_top`) for SMA; SPI and SRI are calculated using 3-hourly outputs, while SMA is calculated on monthly averages.

## 3 Results

### 3.1 Model water budget

Figure 2 shows the results for the evapotranspiration components of the model hydrological budget. It can be seen that all model scenarios show very similar performance for `esoil_gb` (middle plot), though it is apparent that both  $3h\_C$  and  $3h\_D$  somewhat overestimate and  $24h\_F$  underestimates the  $3h\_F$  distribution, with this being least strong for  $3h\_C$ . Table 1 provides a more comprehensive picture and shows that the median value of  $3h\_C/3h\_F$  is closest to 1:1 for most regions, but the regional distributions of averages for all model scenarios are not significantly different from  $3h\_F$  in most regions, though this is

true for a higher number of regions for 3h\_C, plus also with higher  $p$  values for regions such as EAF, SAF and CAM. 24h\_F appears to exhibit higher  $p$  values for AMZ and SEA, also being singularly significant for the latter, plus appears visually closer to 3h\_F in Figure 2 (middle). However, 3h\_F in Table 1 shows a lower spread in the distribution, implying that more points are closer to 1:1. In the case of SEA at least, the significance of 24h\_F may be due to its overall negative bias, apparent for the high-end of the globally-aggregated distribution in Figure 3 (middle, right axis), pushing the median more toward 1:1. 3h\_C and 3h\_D perform similarly, but with lesser performance for 3h\_D. It can be seen in Figure 3 (middle), examining the globally-aggregated distribution, that both models show the same shape in the overestimation of higher quantiles but with 3h\_C having a lower amplitude (right axis). A very slight overestimation is also visible for quantiles below 80, though this would be somewhat exaggerated by the log scaling and is likely not physically significant. In linear terms, the apparent overestimation of 24h\_F in Figure 3 in the same range appears of a factor of 1.5 at its highest (left axis).

With regards to  $ecan\_gb$  (Figure 2, top), for 3h\_C the hypothesis of equality of the distribution to 3h\_F is not rejected for the largest amount of regions, though with quite tentative  $p$  values in at least SSA and WNA. From the lower bound of the distributions for 3h\_C and 3h\_D, we can infer that both models are negatively biased within each region, which appears to stem from underestimation of quantiles above the 40th in Figure 3 (top, right axis) for 3h\_C, and throughout the distribution for 3h\_D. If we are to judge from Table 1, these appear statistically significant in both models for multiple regions. As with  $esoil\_gb$ , both 3h\_C and 3h\_D show the same pattern of mismatch in the distribution, but the magnitude in 3h\_C is lesser (Figure 3). It is also the case that 3h\_C exhibits lower spread in over/underestimation for all regions, also evidenced from visually being a better match for 3h\_F in Figure 2. Inversely to the rest, 24h\_F shows a positive bias, in Figure 2; Table 1, also evident throughout the distribution up to the 95th+ quantiles in Figure 3, plus exhibits inflated spread in Figure 2. It is also interesting to note the performance for SAH in particular, where one sees a negative ratio of underestimation in Table 1 for 24h\_F. This implies that 24h\_F has negative values of evapotranspiration when computing  $24h\_F/3h\_F$ , physically implying deposition of moisture, which does not seem realistic for such a strongly arid region. We have not assessed this, but this may point to numerical issues within JULES. SAH also serves as a reminder that though differences are exhibited between the model formulations therein, due the very low values of evapotranspiration for all models, these may not actually be physically significant.

The results for  $etotal\_gb$  (Figure 2, bottom) show that both 3h\_D and 24h\_F statistically match the 3h\_F distribution for a large number of regions (Table 1). However, it can be seen that the results in over/underestimation for  $ecan\_gb$  and  $esoil\_gb$ , being opposite in sign (as more/less evaporated water from the canopy means less/more water available for evaporation from the soil), compensate for each other, leading to the observed closer 1:1 match in Table 1 and Figure 2, with 24h\_F evidently being the worse performing formulation in both cases. This also appears true for 3h\_C but less strongly, owing to the lower mismatches compared to 3h\_F for both components of total evapotranspiration. From Figure 3 (bottom), 24h\_F is evidenced to overestimate the 3h\_F distribution up to approximately the 95th quantile, from which strong underestimation occurs. As 24h\_F lacks a diurnal cycle, this is most strongly

apparent in regions such as AMZ and SEA. For these, 3h\_C also shows significant difference from 3h\_F, evidencing the importance of matching the diurnal cycle in the data in regions of predominantly convective rainfall origin. The general patterns of over/underestimation in 3h\_D and 24h\_F for the components of `etotal_gb` are broadly similar to those in Williams and Clark [2014], plus their compensatory effect, though with the caveat that therein the JULES disaggregator disaggregates all forcings in addition to rainfall, adding an extra source of bias.

Moving on to the runoff components of the hydrological budget Figure 4, we see a much stronger response in both runoff components in Table 1. With regards to `sub_surf_roff` (Figure 4, middle), 3h\_C appears to be the best visual match for most regions, though a mismatch with 3h\_C is still apparent, for example in underestimation for AUS, CAM and CNA, or overestimation for most of Africa (EAF, WAF, SAF) and WNA, plus less so AMZ and SSA. The results for 3h\_D are close to those of 3h\_C in Figure 4, though with more exaggerated patterns of mismatch. Turning to Figure 5 for a globally-aggregated picture, this is indeed seen to be the case, where both models show the same shape in the deviation (right axis) for quantiles above the median, with overestimations that for some regions in Figure 4 are of the same magnitude as the regional median (which gives the average for most points in the region). For the 90th+ quantiles however, there is a strong divergence in the strength of the response, with 3h\_D veering toward very strong underestimation. The substantially worse performance of 24h\_F can readily be seen in Figure 5, with the strong underestimation above the global median accounting for the underestimation of 3h\_F in Figure 4. However, if we are to evaluate statistical significance in the regional differences from 3h\_F (Table 1), then it seemingly appears that all model formulations are not statistically different from 3h\_F, and indeed so with very high  $p$  values for most regions. However, the models' mismatch is easily visible in Figure 4. Examining the 5th, 95th quantiles of over/underestimation of the 3h\_F distribution also makes this apparent - the lower bound for regions such as AMZ is zero for both 3h\_C, 3h\_D, implying that there is virtually no sub-surface runoff for some points compared to 3h\_F, and inversely so, for WNA for example, there appear to be very strong overestimations. The magnitude of mismatch is stronger in 3h\_D, reflecting Figure 4, Figure 5. From Figure 4, it can be seen that the regional distributions' medians are strongly shifted downward, implying a long-tailed distribution. It is well known that the KS test is weakly sensitive to variability in the tails [Mason and Schuenemeyer, 1983, Chicheportiche and Bouchaud, 2012], which would explain the non-significant results in Table 1, even though strong mismatch is observed. In TIB in particular, a very strong mismatch appears for two points in 24h\_F, however the values observed are of 5+ decimal point precision (Figure 4, which is unrealistic. Frozen soils are to be expected in TIB, and thus effectively zero sub-surface runoff. This may imply that at least some of the very strong underestimations in Table 1 may be for points where `sub_surf_roff` is close to zero but calculated to unreasonably high precision, where very small mismatches in magnitude would not be physically significant. The long sub-median tail in Figure 5 (middle, left axis) lends some credence to this.

The response surface runoff (Figure 4, top) shows 3h\_D and 3h\_C overestimating the 3h\_F distribution, the inverse true for 24h\_F. The pattern of mismatch in 3h\_C again mirrors that of 3h\_D, but is



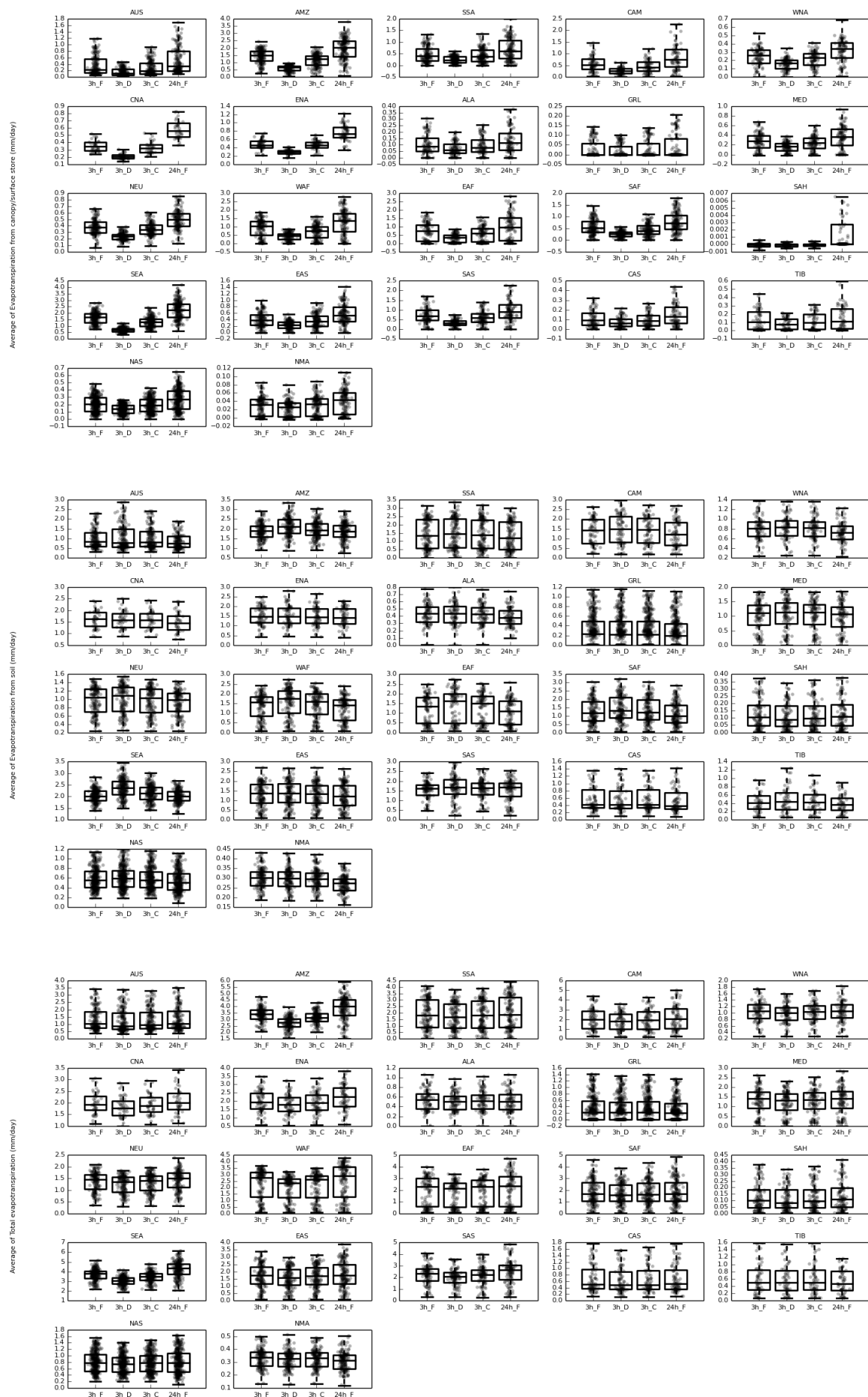


Figure 2: Regional distributions of averages for evapotranspiration variables over the integration period for every model scenario. Whiskers are the 5th, 95th confidence intervals and represent the spread of points within each region, with outlying points not shown. A higher spread simply indicates more variability within a region. Individual points are shown, with darker shading indicating a higher density.

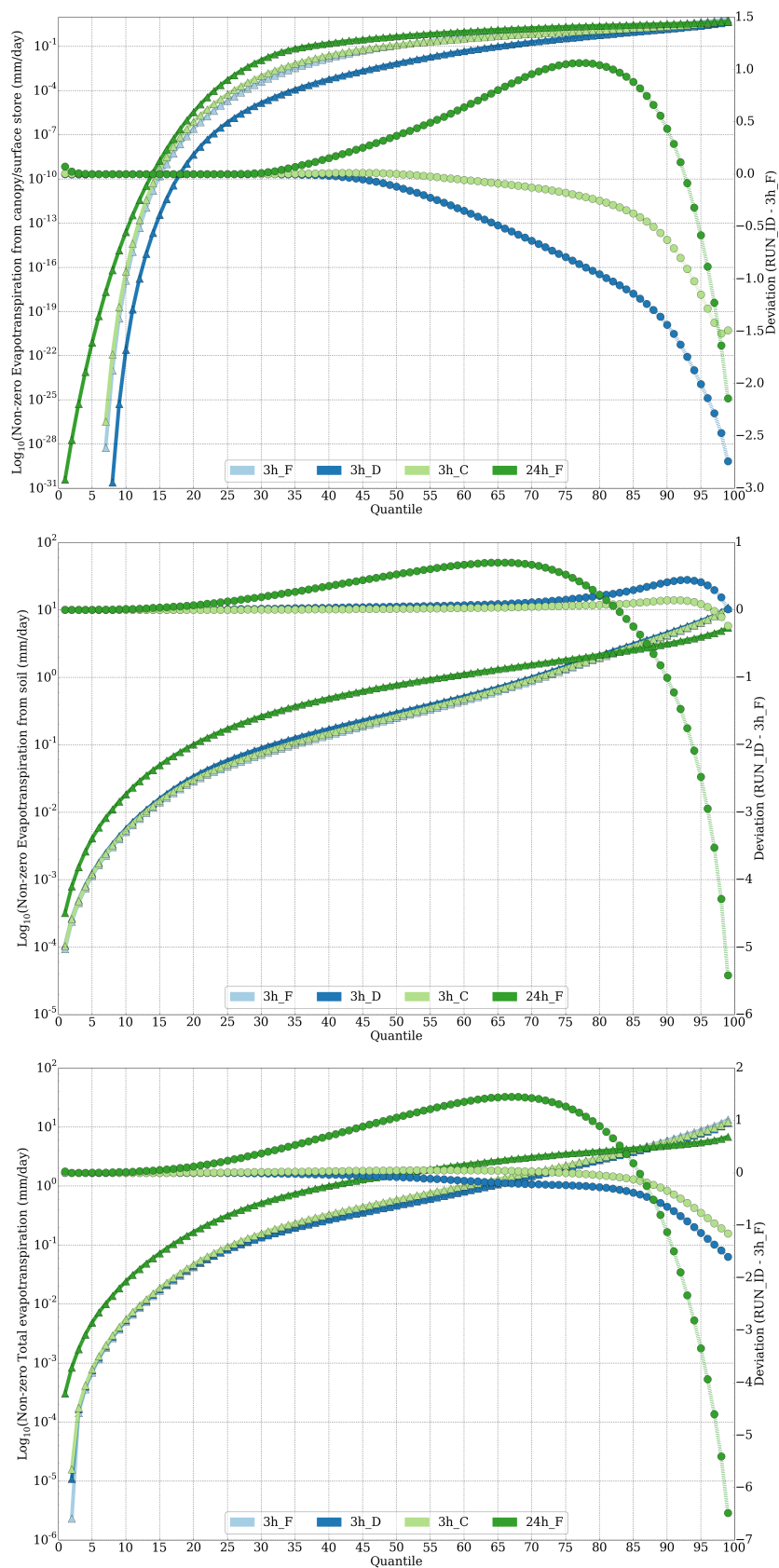


Figure 3: Quantiles 1 to 100 of the globally aggregated distribution of all sampled points for evapotran-  
spiration variables in every model scenario. Triangles (left axis): Actual magnitudes (log scale). Circles  
(right axis): Deviation with respect to 3h\_F.

smaller in magnitude. The magnitudes of over/underestimation are the largest of all water budget components (Table 1) with the consistent overestimation of  $3h\_D$  and  $3h\_C$  apparent from the bounds of over/underestimation. For TIB, CAM, WNA and SSA, the upper bounds of these models, especially for  $3h\_D$ , indicate that very strong overestimations are possible for some points, though the magnitudes for  $3h\_D$  are at least twice as large.  $24h\_F$  in Table 1 is roughly bounded upward by 1 for all regions bar NMA, indicating consistent underestimation. In terms of (non)significant differences from the regional  $3h\_F$  distribution,  $3h\_C$  shows this for the largest number of regions, as we can discount at least some of those for  $24h\_F$  both based on visual assessment, but also due to the aforementioned lack of sensitivity of the KS test for deviations in the tails (note that ratios scale logarithmically). For  $3h\_C$  regions with lowest  $p$  values (e.g. TIB, AMZ, CNA, ENA), these appear to be influenced by comparatively few overestimating points (Figure 4), though taking into account the overall results for the model, the general pattern of overestimation remains. Turning to Figure 5 (top), the globally-aggregated pattern of mismatch is visible for all models past the  $\sim 70$ th quantile, with the  $3h\_C$  deviation following the same pattern as  $3h\_D$ , but with a strongly dampened magnitude, with both models being positively biased. Conversely, the  $24h\_F$  negative bias is also visible. However, as it is roughly of the same magnitude as  $3h\_C$ , this does not explain the very strong lower bound in the ratio of  $24h\_F/3h\_F$  in Table 1. Closer visual inspection shows that  $24h\_F$  is negatively biased compared to  $3h\_C$  in the 0th to 10th quantile region (left axis). However, the magnitudes in this region are negligibly small (5+ decimal point precision), which may indicate that at least some of the strong mismatch observed is not physically significant. However, the case remains that  $24h\_F$  is upward-bounded at  $\sim 1:1$  in Table 1, thus the overall negative bias remains significant in at least part of the distribution.

The results for `total_roff` reflect the strong influence of `surf_roff`, seen both in the similar distributions in Figure 4 (bottom), plus also in the similar over/underestimation results in Table 1. The underestimation of `sub_surf_roff` seen by its below-one lower bound for  $3h\_D$  and  $3h\_C$  in Table 1 appears to be compensated by the values for `surf_roff`, though the upper bounds of overestimation remain of the same general magnitude as for `surf_roff`. From Figure 5, we can see that the bulk of the mismatch occurs for the 90th+ quantiles, the pattern for  $3h\_C$  being additive, while for  $3h\_D$  and  $24h\_F$  it is compensatory for the highest-most quantiles (right axis). Again, while mismatches may occur for quantiles below the  $\sim 30$ th (left axis), the very low actual amounts of runoff may not be physically significant. This again may point to the high upper bounds of overestimation partially stemming from mismatches in this unrealistically precise region of the distribution. With this in mind, we may conclude that both disaggregator options perform well in matching the average of  $3h\_F$  in Figure 4 in areas with predominantly large-scale rainfall (NEU, very cold areas such as ALA, GRL, NMA), though less so in regions with stronger hydrological extremes - drier/arid regions (AUS, SAH, MED), or regions with convective rainfall (tropics - AMZ, African regions). The overall pattern is one of  $3h\_C$  more closely matching both the median and the spread of the  $3h\_F$  distribution. The underestimation of sub-surface runoff for  $3h\_D$  is in accord with Williams and Clark [2014], though the overestimations suggested by the upper bounds in Table 1 are not visible in the anomaly plots they have presented, which may lend further

credence to the supposition that these strong overestimations are for the unduly precise regions of the distribution in Figure 5, though it may also point to a compensatory effect introduced from disaggregating forcings apart from rainfall in Williams and Clark [2014]. Again similarly to Williams and Clark [2014], the compensatory effect in `total_roff` e.g. for WAF, EAF can also be seen, plus that it is less strong than that for `etotal_gb`.

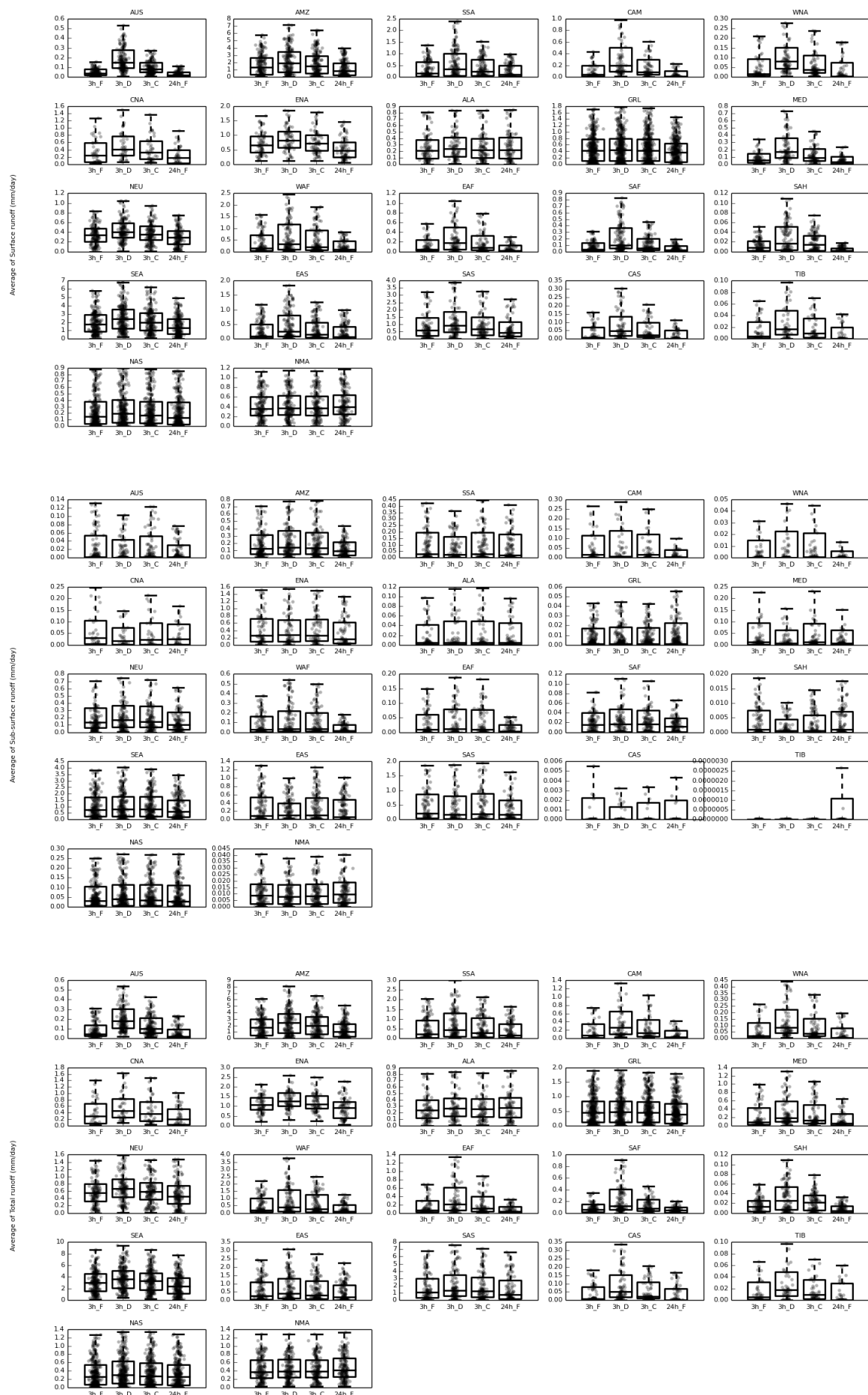


Figure 4: Regional distributions of averages for runoff variables over the integration period for every model scenario. Whiskers are the 5th, 95th confidence intervals and represent the spread of points within each region, with outlying points not shown. A higher spread simply indicates more variability within a region. Individual points are shown, with darker shading indicating a higher density.

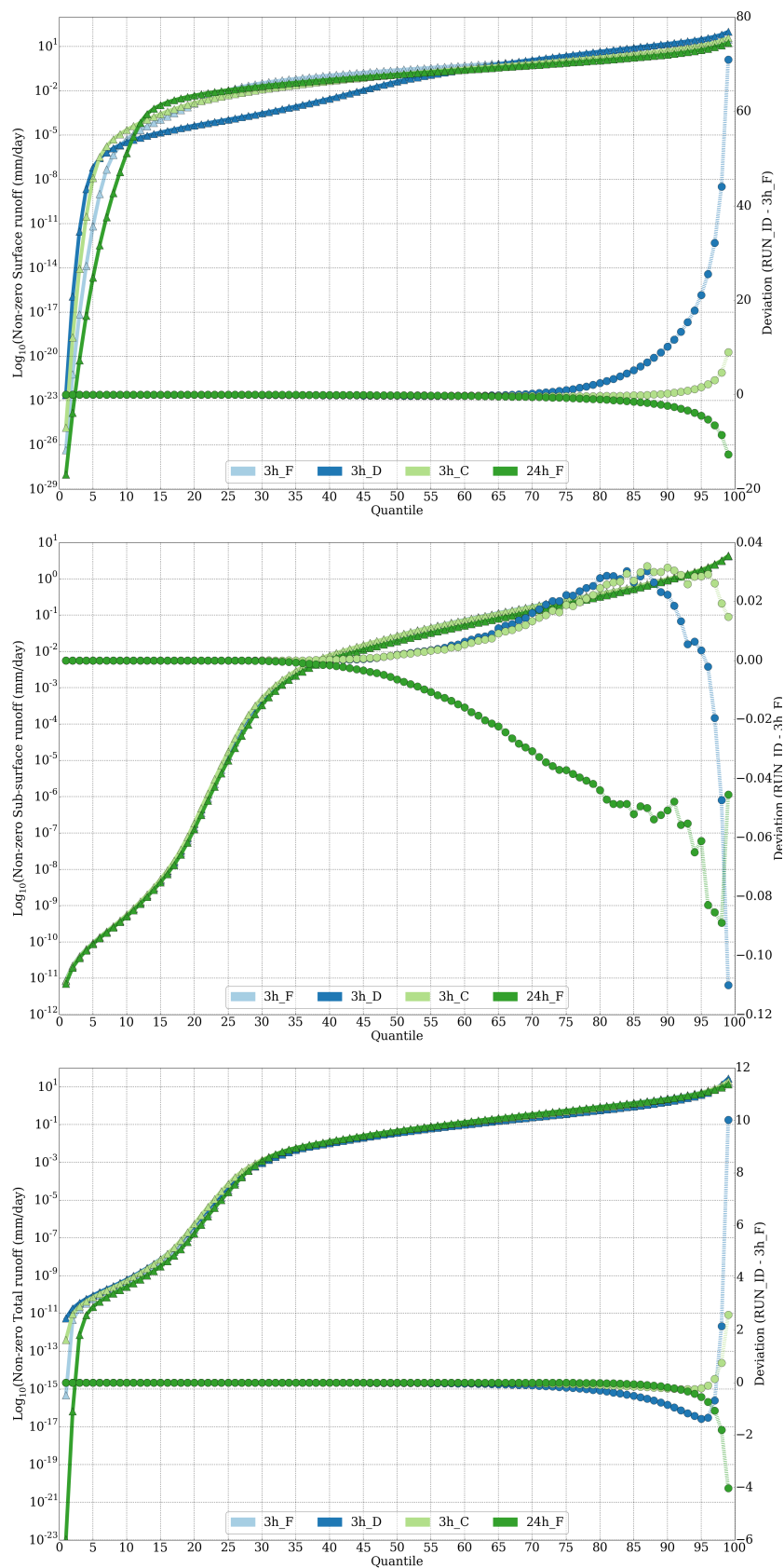


Figure 5: Quantiles 1 to 100 of the globally aggregated distribution of all sampled points for runoff variables in every model scenario. Triangles (left axis): Actual magnitudes (log scale). Circles (right axis): Deviation with respect to 3h\_F.

|     | ecan_gb                                  |   |   |   | esoil_gb                                |   |   |   | etotal_gb                               |   |   |       |
|-----|--|---|---|---|---|---|---|---|---|---|---|-------|
|     | 3h_D                                     | 3h_C                                    | 24h_F                                   | 3h_D                                    | 3h_C                                    | 24h_F                                   | 3h_D                                    | 3h_C                                    | 3h_D                                    | 3h_C                                    | 24h_F                                   | 24h_F |
| AUS | 0.50 (0.37,0.64)                         | 0.82 (0.71,0.91)                        | 1.49 (1.31,1.69)                        | <b>1.02</b> <sup>0.79</sup> (0.84,1.27) | <b>1.00</b> <sup>0.82</sup> (0.93,1.11) | 0.89 (0.81,0.97)                        | 0.89 (0.77,1.01)                        | <b>0.96</b> <sup>0.79</sup> (0.91,1.00) | <b>1.02</b> <sup>0.84</sup> (1.00,1.06) |   |   |       |
| AMZ | 0.44 (0.38,0.53)                         | 0.80 (0.70,0.99)                        | 1.40 (1.23,1.56)                        | 1.12 (0.97,1.24)                        | <b>1.03</b> <sup>0.53</sup> (0.93,1.12) | <b>0.97</b> <sup>0.73</sup> (0.80,1.14) | 0.83 (0.68,0.96)                        | 0.95 (0.87,0.99)                        | 1.14 (1.01,1.32)                        |   |   |       |
| SSA | 0.59 (0.41,0.79)                         | <b>0.93</b> <sup>0.06</sup> (0.75,1.05) | 1.45 (1.00,1.82)                        | <b>1.01</b> <sup>1.0</sup> (0.91,1.15)  | <b>0.99</b> <sup>1.0</sup> (0.95,1.06)  | <b>0.91</b> <sup>0.25</sup> (0.80,1.00) | 0.91 (0.76,1.00)                        | <b>0.97</b> <sup>0.73</sup> (0.95,1.00) | <b>1.03</b> <sup>0.73</sup> (0.93,1.12) |   |   |       |
| CAM | 0.50 (0.40,0.57)                         | <b>0.81</b> <sup>0.1</sup> (0.73,0.92)  | 1.44 (1.28,1.77)                        | <b>1.09</b> <sup>0.88</sup> (0.88,1.19) | <b>1.04</b> <sup>1.0</sup> (0.95,1.09)  | <b>0.88</b> <sup>0.88</sup> (0.80,1.02) | <b>0.91</b> <sup>1.0</sup> (0.79,0.96)  | <b>0.97</b> <sup>1.0</sup> (0.92,0.99)  | <b>1.03</b> <sup>1.0</sup> (1.00,1.16)  |   |   |       |
| WNA | 0.64 (0.52,0.70)                         | <b>0.88</b> <sup>0.06</sup> (0.78,1.00) | 1.32 (1.10,1.58)                        | <b>1.02</b> <sup>1.0</sup> (0.96,1.10)  | <b>1.00</b> <sup>1.0</sup> (0.96,1.04)  | 0.92 (0.82,1.07)                        | <b>0.94</b> <sup>0.73</sup> (0.87,0.98) | <b>0.98</b> <sup>1.0</sup> (0.96,0.99)  | <b>1.00</b> <sup>1.0</sup> (0.95,1.13)  |   |   |       |
| CNA | 0.61 (0.53,0.65)                         | <b>0.93</b> <sup>0.17</sup> (0.82,0.98) | 1.62 (1.48,1.79)                        | <b>0.99</b> <sup>1.0</sup> (0.90,1.05)  | <b>0.98</b> <sup>1.0</sup> (0.95,1.02)  | <b>0.90</b> <sup>0.81</sup> (0.82,1.03) | <b>0.93</b> <sup>0.51</sup> (0.83,0.95) | <b>0.97</b> <sup>1.0</sup> (0.94,0.98)  | <b>1.04</b> <sup>1.0</sup> (0.96,1.16)  |   |   |       |
| ENA | 0.61 (0.45,0.68)                         | <b>0.96</b> <sup>0.49</sup> (0.77,1.05) | 1.57 (1.38,1.76)                        | <b>0.99</b> <sup>1.0</sup> (0.96,1.10)  | <b>0.97</b> <sup>1.0</sup> (0.95,1.05)  | <b>0.97</b> <sup>1.0</sup> (0.90,1.05)  | <b>0.90</b> <sup>0.74</sup> (0.87,0.96) | <b>0.97</b> <sup>0.99</sup> (0.95,1.00) | <b>1.12</b> <sup>0.57</sup> (1.02,1.23) |   |   |       |
| ALA | <b>0.70</b> <sup>0.08</sup> (0.60,1.04)  | <b>0.92</b> <sup>0.65</sup> (0.87,1.05) | <b>1.24</b> <sup>0.48</sup> (0.14,1.40) | <b>1.01</b> <sup>1.0</sup> (0.98,1.06)  | <b>0.99</b> <sup>1.0</sup> (0.98,1.01)  | <b>0.90</b> <sup>0.38</sup> (0.77,0.99) | <b>0.96</b> <sup>1.0</sup> (0.88,0.98)  | <b>0.98</b> <sup>1.0</sup> (0.96,0.99)  | <b>0.97</b> <sup>1.0</sup> (0.85,1.05)  |   |   |       |
| GRL | <b>0.84</b> <sup>0.14</sup> (0.59,1.47)  | <b>1.00</b> <sup>1.0</sup> (0.85,1.58)  | 0.23 (-0.00,1.57)                       | <b>0.99</b> <sup>1.0</sup> (0.97,1.02)  | <b>0.99</b> <sup>1.0</sup> (0.98,1.01)  | <b>0.89</b> <sup>1.0</sup> (0.78,1.03)  | <b>0.93</b> <sup>1.0</sup> (0.88,1.09)  | <b>0.99</b> <sup>1.0</sup> (0.97,1.09)  | <b>0.92</b> <sup>1.0</sup> (0.80,1.17)  |   |   |       |
| MED | 0.59 (0.49,0.96)                         | <b>0.86</b> <sup>0.23</sup> (0.79,0.95) | 1.35 (-0.00,1.56)                       | <b>1.02</b> <sup>1.0</sup> (0.84,1.10)  | <b>1.00</b> <sup>1.0</sup> (0.93,1.04)  | <b>0.95</b> <sup>1.0</sup> (0.89,1.05)  | <b>0.93</b> <sup>1.0</sup> (0.83,1.00)  | <b>0.97</b> <sup>1.0</sup> (0.93,1.00)  | <b>1.02</b> <sup>1.0</sup> (0.99,1.10)  |   |   |       |
| NEU | 0.61 (0.53,0.68)                         | 0.91 (0.85,0.96)                        | 1.29 (1.18,1.40)                        | <b>1.03</b> <sup>0.08</sup> (1.00,1.07) | <b>0.99</b> <sup>1.0</sup> (0.97,1.02)  | 0.94 (0.87,1.04)                        | 0.91 (0.85,0.96)                        | <b>0.97</b> <sup>0.94</sup> (0.95,0.98) | <b>1.04</b> <sup>0.94</sup> (0.98,1.12) |   |   |       |
| WAF | 0.47 (0.40,0.60)                         | 0.75 (0.65,0.85)                        | 1.39 (1.13,1.67)                        | 1.15 (1.00,1.30)                        | <b>1.06</b> <sup>0.13</sup> (0.93,1.16) | <b>0.91</b> <sup>0.08</sup> (0.73,1.06) | 0.91 (0.74,1.00)                        | 0.97 (0.88,1.00)                        | 1.03 (1.00,1.25)                        |   |   |       |
| EAF | 0.47 (0.40,0.67)                         | 0.79 (0.70,0.90)                        | 1.37 (0.90,1.65)                        | <b>1.12</b> <sup>0.06</sup> (0.94,1.39) | <b>1.05</b> <sup>0.72</sup> (0.95,1.20) | <b>0.89</b> <sup>0.34</sup> (0.77,1.05) | 0.93 (0.76,1.00)                        | <b>0.98</b> <sup>0.3</sup> (0.89,1.00)  | <b>1.02</b> <sup>0.25</sup> (1.00,1.25) |   |   |       |
| SAF | 0.46 (0.39,0.63)                         | 0.76 (0.69,0.94)                        | 1.35 (1.24,1.59)                        | <b>1.13</b> <sup>0.14</sup> (0.97,1.39) | <b>1.06</b> <sup>0.35</sup> (0.98,1.21) | <b>0.87</b> <sup>0.17</sup> (0.79,0.97) | <b>0.94</b> <sup>0.85</sup> (0.85,1.00) | <b>0.98</b> <sup>1.0</sup> (0.95,1.00)  | <b>1.01</b> <sup>1.0</sup> (1.00,1.08)  |   |   |       |
| SAH | <b>0.87</b> <sup>1.0</sup> (0.16,1.35)   | <b>0.90</b> <sup>1.0</sup> (0.42,1.35)  | 0.00 (-0.53,2.89)                       | <b>0.91</b> <sup>1.0</sup> (0.80,1.00)  | <b>0.97</b> <sup>1.0</sup> (0.91,1.01)  | <b>1.01</b> <sup>1.0</sup> (0.92,1.14)  | <b>0.90</b> <sup>1.0</sup> (0.79,1.00)  | <b>0.97</b> <sup>1.0</sup> (0.91,1.01)  | <b>1.03</b> <sup>1.0</sup> (1.00,1.15)  |   |   |       |
| SEA | 0.42 (0.36,0.50)                         | 0.78 (0.71,0.92)                        | 1.32 (1.20,1.53)                        | 1.16 (1.04,1.30)                        | 1.06 (0.95,1.14)                        | <b>1.00</b> <sup>0.61</sup> (0.86,1.10) | 0.84 (0.70,0.94)                        | 0.94 (0.89,0.98)                        | 1.13 (1.04,1.29)                        |   |   |       |
| EAS | 0.60 (0.45,0.70)                         | <b>0.89</b> <sup>0.32</sup> (0.81,1.00) | 1.42 (1.18,1.64)                        | <b>1.02</b> <sup>0.95</sup> (0.95,1.11) | <b>0.99</b> <sup>1.0</sup> (0.96,1.04)  | <b>0.93</b> <sup>0.95</sup> (0.78,1.03) | <b>0.92</b> <sup>0.28</sup> (0.86,0.97) | <b>0.97</b> <sup>0.96</sup> (0.95,0.99) | <b>1.03</b> <sup>0.28</sup> (0.96,1.13) |   |   |       |
| SAS | 0.44 (0.34,0.68)                         | 0.80 (0.73,1.01)                        | 1.27 (1.08,1.50)                        | 1.09 (0.96,1.24)                        | <b>1.03</b> <sup>0.71</sup> (0.94,1.10) | <b>1.02</b> <sup>0.71</sup> (0.83,1.11) | 0.90 (0.77,0.98)                        | <b>0.97</b> <sup>0.55</sup> (0.91,1.00) | 1.10 (0.97,1.17)                        |   |   |       |
| CAS | <b>0.68</b> <sup>0.25</sup> (0.52,0.76)  | <b>0.89</b> <sup>0.82</sup> (0.82,0.99) | <b>1.42</b> <sup>0.67</sup> (1.02,1.81) | <b>0.99</b> <sup>1.0</sup> (0.87,1.13)  | <b>0.99</b> <sup>1.0</sup> (0.94,1.06)  | <b>0.92</b> <sup>0.73</sup> (0.85,1.05) | <b>0.95</b> <sup>1.0</sup> (0.84,0.99)  | <b>0.98</b> <sup>1.0</sup> (0.93,1.00)  | <b>0.99</b> <sup>1.0</sup> (0.95,1.07)  |   |   |       |
| TIB | 0.67 (0.57,0.89)                         | <b>0.89</b> <sup>1.0</sup> (0.81,0.99)  | <b>1.15</b> <sup>1.0</sup> (0.69,2.02)  | <b>1.04</b> <sup>1.0</sup> (0.89,1.15)  | <b>1.01</b> <sup>1.0</sup> (0.95,1.07)  | <b>0.89</b> <sup>1.0</sup> (0.75,1.03)  | <b>0.98</b> <sup>1.0</sup> (0.89,1.00)  | <b>0.99</b> <sup>1.0</sup> (0.94,1.00)  | <b>0.95</b> <sup>1.0</sup> (0.79,1.05)  |   |   |       |
| NAS | 0.67 (0.60,0.79)                         | <b>0.94</b> <sup>0.13</sup> (0.86,1.02) | 1.30 (1.14,1.44)                        | <b>1.01</b> <sup>1.0</sup> (0.98,1.12)  | <b>0.98</b> <sup>1.0</sup> (0.96,1.04)  | 0.90 (0.76,1.05)                        | <b>0.94</b> <sup>0.32</sup> (0.88,0.99) | <b>0.98</b> <sup>0.93</sup> (0.96,1.00) | <b>1.01</b> <sup>0.69</sup> (0.86,1.13) |   |   |       |
| NMA | 0.80 (0.71,1.04)                         | <b>1.02</b> <sup>0.97</sup> (0.83,1.29) | 1.32 (0.07,1.93)                        | <b>0.99</b> <sup>1.0</sup> (0.98,1.00)  | <b>0.98</b> <sup>1.0</sup> (0.97,0.99)  | 0.89 (0.84,0.95)                        | <b>0.98</b> <sup>0.99</sup> (0.95,1.00) | <b>0.99</b> <sup>0.99</sup> (0.98,1.00) | 0.94 (0.86,0.97)                        |   |   |       |
|     | surf_roff                                |   |   |   | sub_surf_roff                           |   |   |   | total_roff                              |   |   |       |
|     | 3h_D                                     | 3h_C                                    | 24h_F                                   | 3h_D                                    | 3h_C                                    | 24h_F                                   | 3h_D                                    | 3h_C                                    | 3h_D                                    | 3h_C                                    | 24h_F                                   | 24h_F |
| AUS | 4.36 (1.26,7.22)                         | 2.27 (1.08,3.47)                        | 0.42 (0.12,0.93)                        | <b>0.63</b> <sup>0.1</sup> (0.00,1.92)  | <b>0.90</b> <sup>1.0</sup> (0.00,1.42)  | <b>0.77</b> <sup>1.0</sup> (0.33,1.08)  | 4.00 (0.98,7.08)                        | 2.17 (0.97,3.46)                        | 4.00 (0.98,7.08)                        | 2.17 (0.97,3.46)                        | 0.44 (0.12,0.99)                        |       |
| AMZ | 1.42 (1.15,5.07)                         | <b>1.15</b> <sup>0.33</sup> (1.03,2.03) | 0.69 (0.31,0.94)                        | <b>1.09</b> <sup>1.0</sup> (0.66,2.13)  | <b>1.05</b> <sup>1.0</sup> (0.93,1.53)  | <b>0.69</b> <sup>0.31</sup> (0.30,0.98) | 1.34 (1.11,4.71)                        | <b>1.13</b> <sup>0.73</sup> (1.02,1.91) | 1.34 (1.11,4.71)                        | <b>1.13</b> <sup>0.73</sup> (1.02,1.91) | 0.69 (0.31,0.95)                        |       |
| SSA | <b>1.70</b> <sup>0.07</sup> (1.07,13.93) | <b>1.23</b> <sup>0.59</sup> (1.00,5.35) | <b>0.75</b> <sup>0.59</sup> (0.06,0.99) | <b>0.81</b> <sup>1.0</sup> (0.10,1.44)  | <b>0.97</b> <sup>1.0</sup> (0.49,1.21)  | <b>0.48</b> <sup>1.0</sup> (0.36,1.20)  | 1.51 (0.99,13.63)                       | <b>1.17</b> <sup>0.73</sup> (0.97,5.32) | <b>1.17</b> <sup>0.73</sup> (0.97,5.32) | <b>0.77</b> <sup>0.73</sup> (0.06,1.01) |   |       |
| CAM | 3.58 (1.22,15.27)                        | 1.87 (1.07,5.61)                        | 0.37 (0.17,0.87)                        | <b>0.90</b> <sup>1.0</sup> (0.00,1.83)  | <b>0.99</b> <sup>1.0</sup> (0.12,1.36)  | <b>0.48</b> <sup>1.0</sup> (0.13,0.95)  | 2.89 (1.17,15.27)                       | 1.57 (1.03,5.61)                        | 2.89 (1.17,15.27)                       | 1.57 (1.03,5.61)                        | 0.35 (0.15,0.88)                        |       |
| WNA | 4.73 (1.17,14.28)                        | 2.27 (1.04,5.70)                        | 0.30 (0.07,0.96)                        | <b>1.13</b> <sup>1.0</sup> (0.00,4.66)  | <b>1.10</b> <sup>1.0</sup> (0.36,3.33)  | <b>0.65</b> <sup>1.0</sup> (0.00,2.58)  | 4.11 (1.09,14.28)                       | 2.22 (1.01,5.70)                        | 4.11 (1.09,14.28)                       | 2.22 (1.01,5.70)                        | 0.31 (0.07,0.98)                        |       |
| CNA | <b>1.62</b> <sup>0.09</sup> (1.17,7.09)  | <b>1.22</b> <sup>0.54</sup> (1.06,3.47) | <b>0.64</b> <sup>0.54</sup> (0.22,0.84) | <b>0.74</b> <sup>1.0</sup> (0.05,1.16)  | <b>0.92</b> <sup>1.0</sup> (0.45,1.09)  | <b>0.72</b> <sup>1.0</sup> (0.32,0.98)  | <b>1.52</b> <sup>0.17</sup> (1.15,7.07) | <b>1.18</b> <sup>0.82</sup> (1.05,3.46) | <b>1.52</b> <sup>0.17</sup> (1.15,7.07) | <b>1.18</b> <sup>0.82</sup> (1.05,3.46) | <b>0.64</b> <sup>0.82</sup> (0.22,0.85) |       |
| ENA | <b>1.28</b> <sup>0.11</sup> (1.01,1.81)  | <b>1.07</b> <sup>0.49</sup> (0.98,1.28) | <b>0.74</b> <sup>0.11</sup> (0.42,0.98) | <b>1.06</b> <sup>1.0</sup> (0.55,1.60)  | <b>1.02</b> <sup>1.0</sup> (0.85,1.23)  | <b>0.70</b> <sup>0.57</sup> (0.17,1.12) | <b>1.16</b> <sup>0.07</sup> (1.04,1.51) | <b>1.05</b> <sup>0.76</sup> (1.00,1.19) | <b>1.16</b> <sup>0.07</sup> (1.04,1.51) | <b>1.05</b> <sup>0.76</sup> (1.00,1.19) | <b>0.79</b> <sup>0.07</sup> (0.39,0.98) |       |
| ALA | <b>1.09</b> <sup>1.0</sup> (1.02,2.23)   | <b>1.04</b> <sup>1.0</sup> (1.01,1.47)  | <b>1.08</b> <sup>0.11</sup> (0.54,1.36) | <b>0.99</b> <sup>1.0</sup> (0.51,1.93)  | <b>1.02</b> <sup>1.0</sup> (0.85,1.37)  | <b>1.15</b> <sup>1.0</sup> (0.53,2.24)  | <b>1.08</b> <sup>1.0</sup> (1.01,2.05)  | <b>1.04</b> <sup>1.0</sup> (1.00,1.43)  | <b>1.08</b> <sup>1.0</sup> (1.01,2.05)  | <b>1.04</b> <sup>1.0</sup> (1.00,1.43)  | <b>1.07</b> <sup>1.0</sup> (0.53,1.28)  |       |
| GRL | <b>1.01</b> <sup>1.0</sup> (0.96,1.19)   | <b>1.00</b> <sup>1.0</sup> (0.96,1.04)  | <b>0.98</b> <sup>0.2</sup> (0.44,1.11)  | <b>0.90</b> <sup>1.0</sup> (0.58,1.39)  | <b>0.99</b> <sup>1.0</sup> (0.85,1.17)  | <b>1.18</b> <sup>1.0</sup> (0.45,3.22)  | <b>1.01</b> <sup>1.0</sup> (0.96,1.19)  | <b>1.00</b> <sup>1.0</sup> (0.96,1.05)  | <b>1.01</b> <sup>1.0</sup> (0.96,1.19)  | <b>1.00</b> <sup>1.0</sup> (0.96,1.05)  | <b>1.00</b> <sup>0.89</sup> (0.45,1.12) |       |
| MED | 2.89 (1.06,12.72)                        | 1.59 (1.02,5.22)                        | 0.61 (0.11,0.99)                        | <b>0.77</b> <sup>1.0</sup> (0.23,1.51)  | <b>0.96</b> <sup>1.0</sup> (0.52,1.45)  | <b>0.74</b> <sup>1.0</sup> (0.29,1.34)  | 1.87 (1.03,11.54)                       | <b>1.29</b> <sup>0.06</sup> (1.01,4.70) | 1.87 (1.03,11.54)                       | <b>1.29</b> <sup>0.06</sup> (1.01,4.70) | <b>0.67</b> <sup>0.12</sup> (0.11,0.99) |       |
| NEU | 1.32 (1.05,2.17)                         | <b>1.10</b> <sup>0.47</sup> (1.00,1.45) | <b>0.85</b> <sup>0.31</sup> (0.50,0.99) | <b>1.15</b> <sup>0.95</sup> (0.77,1.50) | <b>1.09</b> <sup>0.96</sup> (0.95,1.27) | <b>0.84</b> <sup>0.95</sup> (0.41,1.05) | 1.22 (1.06,2.17)                        | <b>1.09</b> <sup>0.39</sup> (1.02,1.43) | 1.22 (1.06,2.17)                        | <b>1.09</b> <sup>0.39</sup> (1.02,1.43) | <b>0.85</b> <sup>0.1</sup> (0.50,1.01)  |       |
| WAF | 1.74 (1.07,4.67)                         | <b>1.30</b> <sup>0.22</sup> (1.02,2.88) | <b>0.59</b> <sup>0.2</sup> (0.15,0.97)  | <b>1.23</b> <sup>1.0</sup> (0.92,2.11)  | <b>1.18</b> <sup>1.0</sup> (0.98,1.59)  | <b>0.61</b> <sup>0.08</sup> (0.21,0.99) | <b>1.68</b> <sup>0.08</sup> (1.04,4.65) | <b>1.28</b> <sup>0.28</sup> (1.02,2.88) | <b>1.68</b> <sup>0.08</sup> (1.04,4.65) | <b>1.28</b> <sup>0.28</sup> (1.02,2.88) | <b>0.57</b> <sup>0.15</sup> (0.15,0.97) |       |
| EAF | 2.43 (1.11,8.80)                         | 1.47 (1.02,3.39)                        | 0.45 (0.04,0.96)                        | <b>1.20</b> <sup>1.0</sup> (0.38,2.78)  | <b>1.24</b> <sup>1.0</sup> (0.73,1.88)  | <b>0.51</b> <sup>0.19</sup> (0.08,0.96) | 2.35 (1.04,7.61)                        | <b>1.41</b> <sup>0.06</sup> (1.01,3.38) | 2.35 (1.04,7.61)                        | <b>1.41</b> <sup>0.06</sup> (1.01,3.38) | 0.47 (0.03,0.95)                        |       |
| SAF | 2.59 (1.27,15.47)                        | 1.47 (1.10,5.14)                        | 0.61 (0.12,0.90)                        | <b>0.94</b> <sup>1.0</sup> (0.25,1.81)  | <b>1.06</b> <sup>1.0</sup> (0.72,1.45)  | <b>0.79</b> <sup>0.68</sup> (0.39,1.04) | 2.20 (1.13,15.47)                       | 1.38 (1.03,5.07)                        | 2.20 (1.13,15.47)                       | 1.38 (1.03,5.07)                        | 0.66 (0.12,0.94)                        |       |
| SAH | 2.09 (0.99,6.00)                         | 1.49 (0.76,3.55)                        | 0.35 (0.04,1.07)                        | <b>0.74</b> <sup>0.38</sup> (0.02,1.01) | <b>0.85</b> <sup>0.94</sup> (0.17,1.08) | <b>1.01</b> <sup>0.99</sup> (0.53,6.16) | 1.77 (0.98,6.00)                        | 1.27 (0.93,3.55)                        | 1.77 (0.98,6.00)                        | 1.27 (0.93,3.55)                        | 0.45 (0.04,1.00)                        |       |
| SEA | 1.36 (1.13,2.11)                         | <b>1.11</b> <sup>0.61</sup> (1.04,1.34) | <b>0.79</b> <sup>0.15</sup> (0.41,0.94) | <b>1.01</b> <sup>1.0</sup> (0.74,1.37)  | <b>1.02</b> <sup>1.0</sup> (0.97,1.19)  | <b>0.84</b> <sup>0.28</sup> (0.38,1.00) | 1.21 (1.06,1.67)                        | <b>1.08</b> <sup>0.33</sup> (1.02,1.24) | 1.21 (1.06,1.67)                        | <b>1.08</b> <sup>0.33</sup> (1.02,1.24) | 0.81 (0.40,0.97)                        |       |
| EAS | 2.17 (1.23,12.02)                        | <b>1.33</b> <sup>0.24</sup> (1.07,4.89) | <b>0.72</b> <sup>0.24</sup> (0.09,0.96) | <b>0.75</b> <sup>1.0</sup> (0.22,2.02)  | <b>0.98</b> <sup>1.0</sup> (0.62,1.52)  | <b>0.80</b> <sup>1.0</sup> (0.18,1.17)  | 1.57 (1.08,11.66)                       | <b>1.20</b> <sup>0.32</sup> (1.02,4.89) | 1.57 (1.08,11.66)                       | <b>1.20</b> <sup>0.32</sup> (1.02,4.89) | <b>0.73</b> <sup>0.24</sup> (0.09,1.04) |       |
| SAS | 1.46 (1.06,4.34)                         | <b>1.13</b> <sup>0.55</sup> (1.02,2.42) | <b>0.75</b> <sup>0.23</sup> (0.46,1.01) | <b>0.85</b> <sup>1.0</sup> (0.12,1.26)  | <b>0.98</b> <sup>1.0</sup> (0.54,1.23)  | <b>0.77</b> <sup>1.0</sup> (0.41,1.17)  | <b>1.20</b> <sup>0.9</sup> (1.03,4.14)  | <b>1.07</b> <sup>0.98</sup> (1.00,2.11) | <b>1.20</b> <sup>0.9</sup> (1.03,4.14)  | <b>1.07</b> <sup>0.98</sup> (1.00,2.11) | <b>0.77</b> <sup>0.63</sup> (0.45,1.04) |       |
| CAS | 3.80 (1.05,10.68)                        | 1.95                                    |   |   |   |   |   |   |   |   |   |       |

### 3.2 Climate impacts

For using SPI to quantify drought thresholds (Figure A1), we see that there is virtually no difference between 3h\_D and 3h\_C compared to 3h\_F for every scale of accumulation. For 24h\_F, some negative bias occurs at the monthly scale (top-left) for drought-prone regions such as AUS, CAM, African regions (EAF, WAF, SAF, SAH), particularly for drought threshold at the 1st percentile of the distribution (i.e. the most severe form of drought under our definition). As all model formulations in our analysis preserve the mean of the rainfall forcing, this indicates that the variability is also important when calculating the SPI, where for 24h\_F this is entirely lacking as `interp=nf` keeps the diurnal cycle flat, though Paspaldzhiev [2015] found that 3h\_C overestimates the variance of the 3h\_F rainfall forcing, which seemingly does not introduce any differences in SPI, making this supposition uncertain. Note that though the 24h\_F run includes the interpolation of all model forcings and not just rainfall this would not be of any effect here, as calculation of the SPI uses the rainfall forcing output from the model only.

For the high-end extremes in SPI (Figure A2), there is mostly uniform lack of differences in the median of the regional distributions across timescales, with one exception at the 80th percentile of NMA at the seasonal scale for 3h\_C and 3h\_D, though this difference is not large. We can further see some discrepancies in the tails of the regional distributions e.g. for SAF at the yearly scale.

In summary for SPI, we do not have a concrete explanation as to why discrepancies are observed between the model formulations but if we adopt the widely used qualitative classification system for SPI magnitudes of McKee et al. [1993]<sup>5</sup>, then the distributions in the discrepant regions would fall within the same SPI category, thus the conclusion for severity of the drought/flood event would not change. This indicates that neither formulation is superior for SPI calculation, so results consistent with 3h\_F could be obtained even when forcing with daily means (24h\_F).

For SRI, there are obvious discrepancies, which we expect to reflect the discrepancies in the modelled `total_roff`, on which the SRI is calculated. With regards to drought Figure A3, this is true for e.g. AUS, CAM, and perhaps for TIB, NAS and NMA (at the monthly scale) if we are to attribute the strong positive bias in 24h\_F to the results for upper bounds in Table 1. However, for multiple regions, the pattern is seemingly shifted, with regions of over/underestimation of `total_roff` having negative/positive bias in SRI, more so in 3h\_D and 24h\_F, e.g. for SSA, CNA, SNA at the monthly scale. Examining Figure 5 shows discrepancies in the tail of the distribution (left axis) which would lead to such results but may not necessarily point to physical significance, as the runoff for these is very minute in actual magnitude, plus JULES cannot be expected to accurately model at such extreme levels of precision. Some shifts in the monthly pattern can be observed, e.g. a change in sign at low-percentile thresholds e.g. in AUS and CAM for 24h\_F at the yearly scale onward.

SRI results for high-end extremes show that the positive bias in Figure 5 for 3h\_D and 3h\_C broadly corresponds to positive bias in Figure A4 e.g. at the monthly scale for AUS, CAM, EAF across scales. The opposite can also be observed (e.g. NMA) - there would be regional differences not captured in Figure 5, but it may also be that the 70th to ~95th quantile region are contributing to this; the calculation

<sup>5</sup>i.e. extremely, very, moderately or near-normal wet/dry compared to the mean



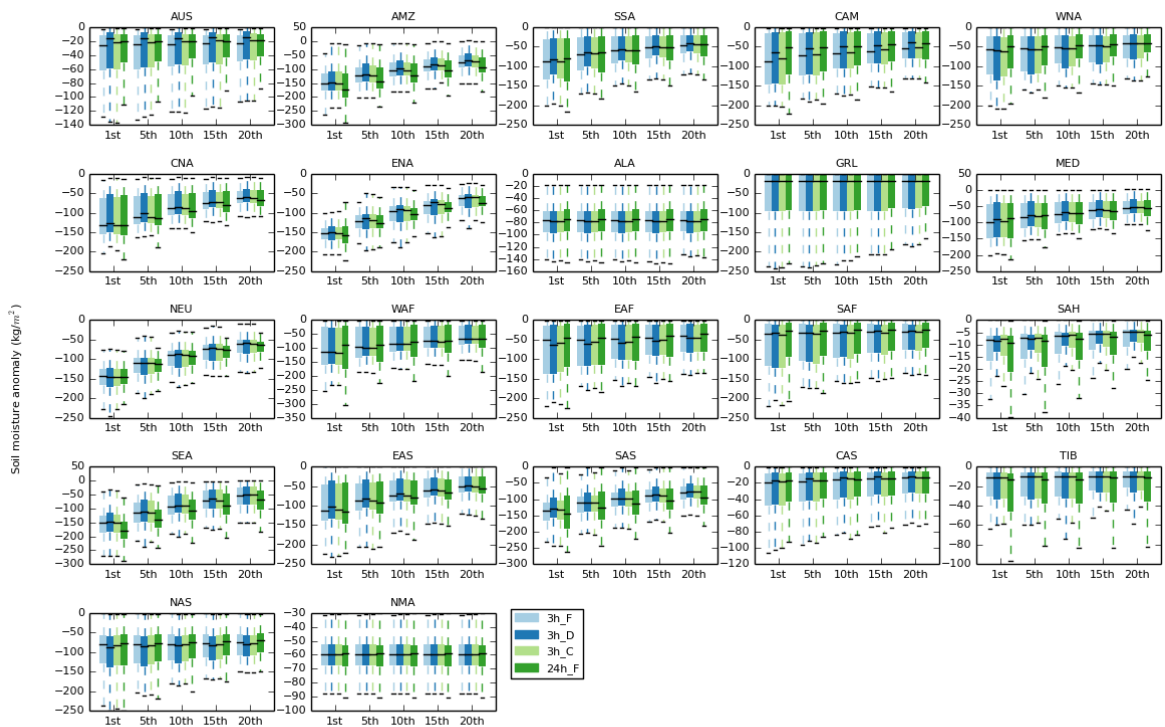


Figure 6: Regional distributions of SMA over the integration period for every model scenario. Whiskers are the 5th, 95th confidence intervals and represent the spread of points within each region, with outlying points not shown. A higher spread simply indicates more variability within a region.

of SPI/SRI involves fitting a theoretical distribution, so it is not necessarily the case that the quantiles of Figure 5 entirely correspond to those of the calculated impact statistic. For 24h\_F, there are no strong deviations from the regional median, apart from e.g. TIB and NMA at yearly and bi-yearly scales. For TIB in particular, 24h\_F also seems to inflate the variability of the region.

Summarising for SRI, wider discrepancies are observed compared to SPI, reflecting the added bias of disaggregation, plus that of lack of diurnal cycle in 24h\_F forcings, which also includes the influence of non-rainfall forcings. According to the classifications of McKee et al. [1993], these larger discrepancies may result in some mismatches, classifying 3h\_F and the rest of the model formulations in adjacent classes. The mismatch for 24h\_F appears greater for high-end percentiles, which we could expect, as daily means lack information about the extremes of the data. For 3h\_C and 3h\_D, the mismatch is roughly of the same magnitude.

Finally, with regards to SMA (Figure 6), we see generally consistent results between models, apart from e.g. AUS, where 3h\_D performs worse, with tied slightly positive bias in 3h\_C and 24h\_F, or negative bias for 24h\_F in AMZ, SAS and SEA. At their starkest, the mismatches are a factor of  $\sim 0.8$  underestimation for 3h\_D in AUS for the 1st percentile threshold, and a factor of  $\sim 1.2$  overestimation for 24h\_F at the 1st percentile in SEA, both for the medians, plus roughly the same magnitude for the tail-end of 24h\_F in SAH at the 1st percentile.

If we are to summarise for this brief assessment of climate impact metrics, we can tentatively conclude that some variability is introduced from employing different model formulations, but we cannot draw a robust conclusion as to whether this would be of physical significance. A more robust analysis of this would be comparing the proportion of time spent in drought as per Taylor et al. [2013], as it would both put all metrics on a common footing, plus would enable a more nuanced evaluation than the simple qualitative categorisation of McKee et al. [1993], though we do not attempt this herein. As we have seen some variation between the results for individual models at different percentile thresholds, it would be useful to employ the aforementioned analysis of Taylor et al. [2013], as it would allow us to distinguish whether this variability is of practical importance, the conclusion of Taylor et al. [2013] being that it is less important than the time period over which the metric is calculated, on which we would further be able to add whether the disaggregation approach (or lack thereof) is also of importance.

## 4 Discussion

Our evaluation of the model hydrological budget shows that both the current JULES disaggregator (3h\_D) and the cascade approach of Paspaldzhiev [2015](3h\_C) can add bias to the model hydrology, though generally outperform runs using just daily means (24h\_F), with the added caveat that the 24h\_F run includes the influence of disaggregating all forcings apart from rainfall. Even with the latter in mind, our results would be in line with Williams and Clark [2014] and would not be surprising, as using means does not carry information about the diurnal cycle, thus minima/maxima are dampened by virtue of the mean being a measure of central tendency. The drawback of the current JULES disaggregator, elucidated in Williams and Clark [2014] is its ad-hoc parameterisation, which has been shown to outperform forcings that are by design more informative of the true magnitude and pattern of precipitation, pointing to the possibility of overfitting and making up for biases elsewhere in JULES. As outlined by Paspaldzhiev [2015], the parameterisation of the cascade approach is more straightforward, plus it has good theoretical backing, making it superior to the current disaggregator in this regard. What is more, our evaluation has shown that the pattern of mismatch between both disaggregators is similar in shape (evident from examining the globally-aggregated analysis of the hydrological outputs' distributions), but that of the cascade approach is lesser in magnitude uniformly for all budget components. The combined benefit of a simpler and more physically-based approach, which also reduces the bias introduced, lends credence to the future application of multiplicative cascades for disaggregating JULES rainfall.

Focusing on the performance of the cascade approach in particular, Paspaldzhiev [2015] showed that it inflates the variance in the WFDEI rainfall forcing by a median factor of  $\sim 2$  across regions, which translates to positive bias in the high-end of the rainfall distribution and overly large annual maxima, plus negative bias at the low-end of the distribution (as the procedure conserves mass). We have seen that 3h\_C shows its strongest overestimation for `surf_roff`. We expect surface runoff to be sensitive to the extremes of the rainfall forcing, and an overestimation of higher forcing values, as shown for the cascade approach, is consistent with this. The then-observed negative bias in `ecan_gb` is not unex-

pected, as more runoff means that less water is available for transport to the atmosphere. Paspaldzhiev [2015] showed that the cascade approach also underestimates the magnitudes of rainfall events, an event defined as the number of consecutive non-zero timesteps. Smaller-magnitude events should lead to less throughfall and thus less water reaching the ground, as it would take longer for the canopy to saturate (Johannes Dolman and Gregory [1992] for the treatment of throughfall, cited in Best et al. [2011]), which may mean that the underestimation of rainfall event magnitudes may compensate for the inflation of timestep magnitudes for higher-quantile rainfall. In Paspaldzhiev [2015], it was shown that the lag-1 autocorrelation of the WFDEI rainfall forcing is underestimated, which would lead to non-preservation of the seriality in the data for longer events. We have not assessed what the exact nature of this relationship is in the current JULES disaggregator, but the larger bias it produces compared to the cascade procedure potentially points to either a larger overestimation of the variance in the rainfall forcing, and/or an underestimation of rainfall event lengths - thus larger volumes for individual "storms" (as mass is conserved), due to these being fixed throughout the integration period. A comparison of rainfall forcing statistics, as in Paspaldzhiev [2015], with the inclusion of the current JULES disaggregator, would allow for discerning these differences.

For all model formulations' regional bounds of over/underestimation in Table 1, though apparently very large in some instances, may not actually point to physically significant mismatch, as they may appear in regions of the distribution for a given hydrological component that is very near-zero in absolute magnitude, plus unreasonably precise for a land-surface model, as we have outlined for individual components in Section 3.1. The mismatch for the median of the regional distributions is however apparent from examining Figure 2 and Figure 4. Further, the 24h\_F scenario includes biases due to disaggregating forcings apart from rainfall, but the generation of a forcing dataset that excludes this did not fit in the timeframe of this work. The 24h\_F results should thus be taken as conservative estimates, but nevertheless should follow the same overall pattern as seen in Williams and Clark [2014]. With regards to testing statistical significance, the aforementioned lack of sensitivity of the KS test for deviation in the tails could be remedied for example via the utilisation of the k-sample Anderson Darling test (e.g. Stephens [1974], Engmann and Cousineau [2011]). In our analysis, we were limited due to practical considerations, as the Python implementation of the k-sample Anderson Darling test computes  $p$ -values by extrapolation, which are not necessarily accurate. Hence, the results of significance in Table 1 for soil evapotranspiration and sub-surface runoff should be taken with the caveats in Section 3 in mind, where the may well be significant differences in the regional distributions at least for sub-surface runoff.

Our analysis of climate impact metrics shows that overall, though there is variability introduced by the different model formulations in this analysis, roughly reflecting the biases in the modelled hydrological outputs, these may not necessarily be sufficiently large to be of practical importance. This supposition appears strongest for the Standardized Precipitation Index, whereas for the Standardized Runoff Index, the discrepancies may lead to some mismatching in classification if we are to employ the simple and commonly used system of McKee et al. [1993]. For the Soil Moisture Anomaly, as with the rest of the indices where discrepancies are observed, it appears that the cascade formulation 3h\_C is the

best performing, which should be expected if the results reflect those for the biases in the modelled hydrology. As our simple analysis does not permit drawing conclusions on the physical significance of the differences, a more thorough analysis as in Taylor et al. [2013] is a logical next step for evaluating the suitability of the disaggregators (or lack thereof) for use in impact studies.

Concerning disaggregation in a broader scope, our analysis purely highlights differences to do with the rainfall forcing in the current JULES disaggregator and the cascade approach. However, the impact of disaggregation on the rest of the JULES forcings is at present not entirely known, save for the brief assessment in Williams and Clark [2014], where it was shown that there is a temporal mismatch in the diurnal cycle of temperature, plus the amplitude of the diurnal profile for the long-wave radiation forcing is strongly dampened, and the diurnal cycle of specific humidity is not well-matched. What is more, a flat diurnal cycle is imposed on wind and pressure. An assessment of the added bias of the disaggregation for the combination of JULES forcings, plus their individual contributions, is needed for a robust picture of the (de)merits of applying the disaggregator in its entirety for hydrological and impact studies, both for the current JULES approach and for the cascade procedure. The treatment of the snow forcing in the model is also an open question, as the review of Paspalzhiev [2015] did not find any studies that explicitly consider the disaggregation of snowfall. In an operational case, JULES would either have to disaggregate total precipitation and subsequently redistribute between rainfall and snowfall, or treat snowfall with another of the current disaggregation options in the model. In the former case, this demands an assessment of the performance of the cascade model when used for total precipitation. Molnar and Burlando [2008] evaluated the difference in scaling behaviour between summer and winter precipitation in Switzerland, showing that winter precipitation is less variable, with longer events and close to monofractal in its behaviour, owing to the dominance of large-scale frontal events. The inflation of variance and underestimation of event lengths of the cascade approach [Paspalzhiev, 2015] may mean that similar mismatches in the hydrological budget may be introduced as shown in this report, though this may not necessarily be the case, as snowfall is qualitatively different in its behaviour than rainfall. Further still, if the model is parameterised on the full precipitation timeseries, in the same way in which this is currently performed for rainfall, it is unknown what weight each component of total precipitation would have on the model parameters, and if this would be of importance.

## 5 Conclusion

We have performed a comparison of the added bias to the model hydrological budget due to differing treatment of the rainfall forcing, between the current JULES disaggregator, a formulation based on multiplicative random cascades, as well as a non-disaggregated run forced with daily means, compared to forcing the model with 3-hourly WFDEI data. Our results show that the use of disaggregation adds utility to the modelling exercise, which we attribute to introducing a diurnal cycle which a non-disaggregated run forced with daily means would lack. Our analysis indicates that the response for the cascade procedure is of the same general shape and directionality as that of the current JULES disaggregator,

but with a dampened magnitude, which is consistent for all hydrological outputs, though the discrepancies may not be significant for all budget components. The cascade formulation is shown to strongly lessen the bias in the aboveground hydrological components (surface runoff and evapotranspiration from the canopy/surface store), with the response for sub-surface runoff and soil evapotranspiration being similar, though not necessarily significantly different at least for evapotranspiration based on visual assessment. Our analysis of climate impact metrics is not sufficient for a strong conclusion, but at least tentatively suggests that similar results are to be expected regardless of the model formulation utilised herein, most so for the Standardized Precipitation Index, though with potential discrepancies for the Standardized Runoff Index and Soil Moisture Anomaly. The exact magnitude of this mismatch remains to be robustly assessed. The cascade formulation generally yields the closest result to the 3-hourly WFDEI-forced run regardless.

In summary, we posit that the cascade approach presents an improvement for the disaggregation of rainfall in JULES and is worth considering for implementation in the model. Some further work is also suggested.

## Acknowledgements

We are thankful for the comments of Graham Weedon and Richard Betts whose inputs contributed to the direction of this work, plus for the kind support of fellow Climate Impacts Modelling colleagues Camilla Mathison and Ron Kahana. This work was supported by the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement no 603864 (HELIX: High-End cLimate Impacts and eXtremes; [www.helixclimate.eu](http://www.helixclimate.eu)) and the BEIS/Defra Met Office Hadley Centre Climate Programme (GA01101).

## References

- Mikel Aickin and Helen Gensler. Adjusting for multiple testing when reporting research results: the bonferroni vs holm methods. *American journal of public health*, 86(5):726–728, 1996.
- Aaron A. Berg, James S. Famiglietti, Jeffrey P. Walker, and Paul R. Houser. Impact of bias correction to reanalysis products on simulations of north american soil moisture and hydrological fluxes. *Journal of Geophysical Research: Atmospheres*, 108(D16):n/a–n/a, 2003. ISSN 2156-2202. doi: 10.1029/2002JD003334. URL <http://dx.doi.org/10.1029/2002JD003334>. 4490.
- M. J. Best, M. Pryor, D. B. Clark, G. G. Rooney, Essery, C. B. Ménard, J. M. Edwards, M. A. Hendry, A. Porson, N. Gedney, L. M. Mercado, S. Sitch, E. Blyth, O. Boucher, P. M. Cox, C. S. B. Grimmond, and R. J. Harding. The joint uk land environment simulator (jules), model description - part 1: Energy and water fluxes. *Geoscientific Model Development*, 4(3):677–699, September 2011. doi: 10.5194/gmd-4-677-2011. URL <http://dx.doi.org/10.5194/gmd-4-677-2011>.

- Eleanor J Burke and Simon J Brown. Evaluating uncertainties in the projection of future drought. *Journal of Hydrometeorology*, 9(2):292–299, 2008.
- Rémy Chicheportiche and Jean-Philippe Bouchaud. Weighted kolmogorov-smirnov test: Accounting for the tails. *Physical Review E*, 86(4):041115, 2012.
- D. B. Clark, L. M. Mercado, S. Sitch, C. D. Jones, N. Gedney, M. J. Best, M. Pryor, G. G. Rooney, R. L. H. Essery, E. Blyth, O. Boucher, R. J. Harding, C. Huntingford, and P. M. Cox. The joint uk land environment simulator (jules), model description - part 2: Carbon fluxes and vegetation dynamics. *Geoscientific Model Development*, 4(3):701–722, September 2011. doi: 10.5194/gmd-4-701-2011. URL <http://dx.doi.org/10.5194/gmd-4-701-2011>.
- E. Compton and M. Best. Impact of spatial and temporal resolution on modelled terrestrial hydrological cycle components. Technical Report Technical Report 44, Technical Report 44, WATCH report, July 2011.
- Sonja Engmann and Denis Cousineau. Comparing distributions: the two-sample anderson-darling test as an alternative to the kolmogorov-smirnov test. *Journal of Applied Quantitative Methods*, 6(3), 2011.
- Balázs M Fekete, Charles J Vörösmarty, John O Roads, and Cort J Willmott. Uncertainties in precipitation and their impacts on runoff estimates. *Journal of Climate*, 17(2):294–304, 2004.
- Filippo Giorgi and Raquel Francisco. Uncertainties in regional climate change prediction: a regional analysis of ensemble simulations with the hadcm2 coupled aogcm. *Climate Dynamics*, 16(2-3):169–182, 2000.
- Chris Hewitt, Carlo Buontempo, and Paula Newton. Using climate predictions to better serve society's needs. *Eos, Transactions American Geophysical Union*, 94(11):105–107, 2013. ISSN 2324-9250. doi: 10.1002/2013EO110002. URL <http://dx.doi.org/10.1002/2013EO110002>.
- Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pages 65–70, 1979.
- P. Hubert, Y. Tessier, S. Lovejoy, D. Schertzer, F. Schmitt, P. Ladoy, J. P. Carboneil, S. Violette, and I. Desrosne. Multifractals and extreme rainfall events. *Geophys. Res. Lett.*, 20(10):931–934, May 1993. doi: 10.1029/93gl01245. URL <http://dx.doi.org/10.1029/93gl01245>.
- C. Huntingford, B. B. Booth, S. Sitch, N. Gedney, J. A. Lowe, S. K. Liddicoat, L. M. Mercado, M. J. Best, G. P. Weedon, R. A. Fisher, M. R. Lomas, P. Good, P. Zelazowski, A. C. Everitt, A. C. Spessa, and C. D. Jones. Imogen: an intermediate complexity model to evaluate terrestrial impacts of a changing climate. *Geoscientific Model Development*, 3(2):679–687, 2010. doi: 10.5194/gmd-3-679-2010. URL <http://www.geosci-model-dev.net/3/679/2010/>.

- A Johannes Dolman and David Gregory. The parametrization of rainfall interception in gcms. *Quarterly Journal of the Royal Meteorological Society*, 118(505):455–467, 1992.
- S Lovejoy, J Pinel, and D Schertzer. The global space–time cascade structure of precipitation: Satellites, gridded gauges and reanalyses. *Advances in Water Resources*, 45:37–50, 2012.
- David M Mason and John H Schuenemeyer. A modified kolmogorov-smirnov test sensitive to tail alternatives. *The Annals of Statistics*, pages 933–946, 1983.
- Thomas B McKee, Nolan J Doesken, John Kleist, et al. The relationship of drought frequency and duration to time scales. In *Proceedings of the 8th Conference on Applied Climatology*, volume 17, pages 179–183. American Meteorological Society Boston, MA, USA, 1993.
- Merab Menabde and Murugesu Sivapalan. Modeling of rainfall time series and extremes using bounded random cascades and levy-stable distributions. *Water Resour. Res.*, 36(11):3293–3300, November 2000. doi: 10.1029/2000wr900197. URL <http://dx.doi.org/10.1029/2000wr900197>.
- Peter Molnar and Paolo Burlando. Preservation of rainfall properties in stochastic disaggregation by a simple random cascade model. *Atmospheric Research*, 77(1-4):137–151, September 2005. ISSN 01698095. doi: 10.1016/j.atmosres.2004.10.024. URL <http://dx.doi.org/10.1016/j.atmosres.2004.10.024>.
- Peter Molnar and Paolo Burlando. Variability in the scale properties of high-resolution precipitation data in the alpine climate of switzerland. *Water Resources Research*, 44(10), 2008.
- Bart Nijssen and Dennis P. Lettenmaier. Effect of precipitation sampling error on simulated hydrological fluxes and states: Anticipating the global precipitation measurement satellites. *Journal of Geophysical Research: Atmospheres*, 109(D2):n/a–n/a, 2004. ISSN 2156-2202. doi: 10.1029/2003JD003497. URL <http://dx.doi.org/10.1029/2003JD003497>. D02103.
- M Pan and EF Wood. Water budget estimation by assimilating multiple observations and hydrological modeling using constrained ensemble kalman filtering. In *AGU Spring Meeting Abstracts*, volume 1, page 04, 2004.
- I. Paspaldzhiev. Disaggregating daily accumulations of 3-hourly wfdei-gpcc rainfall via a multiplicative cascade procedure. Met Office Hadley Centre internal report, November 2015. Contact [karina.williams@metoffice.gov.uk](mailto:karina.williams@metoffice.gov.uk).
- Alan Robock, Lifeng Luo, Eric F. Wood, Fenghua Wen, Kenneth E. Mitchell, Paul R. Houser, John C. Schaake, Dag Lohmann, Brian Cosgrove, Justin Sheffield, Qingyun Duan, R. Wayne Higgins, Rachel T. Pinker, J. Dan Tarpley, Jeffery B. Basara, and Kenneth C. Crawford. Evaluation of the north american land data assimilation system over the southern great plains during the warm season. *Journal of Geophysical Research: Atmospheres*, 108(D22):n/a–n/a, 2003. ISSN 2156-2202. doi: 10.1029/2002JD003245. URL <http://dx.doi.org/10.1029/2002JD003245>. 8846.

Daniel Schertzer and Shaun Lovejoy. Physical modeling and analysis of rain and clouds by anisotropic scaling multiplicative processes. *J. Geophys. Res.*, 92(D8):9693–9714, 1987.

Michael A Stephens. Edf statistics for goodness of fit and some comparisons. *Journal of the American statistical Association*, 69(347):730–737, 1974.

IH Taylor, E Burke, L McColl, PD Falloon, GR Harris, and D McNeall. The impact of climate mitigation on projections of future drought. *Hydrology and Earth System Sciences*, 17(6):2339, 2013.

Lila Warszawski, Katja Frieler, Veronika Huber, Franziska Piontek, Olivia Serdeczny, and Jacob Schewe. The inter-sectoral impact model intercomparison project (isi-mip): Project framework. *Proceedings of the National Academy of Sciences*, 111(9):3228–3232, 2014. doi: 10.1073/pnas.1312330110. URL <http://www.pnas.org/content/111/9/3228.abstract>.

Graham P Weedon, Gianpaolo Balsamo, Nicolas Bellouin, Sandra Gomes, Martin J Best, and Pedro Viterbo. The wfdei meteorological forcing data set: Watch forcing data methodology applied to era-interim reanalysis data. *Water Resources Research*, 50(9):7505–7514, 2014.

K. Williams and D. Clark. Disaggregation of daily data in jules. Met Office Hadley Centre Technical Note 96, July 2014.

## **A SPI and SRI calculated over different timescales**



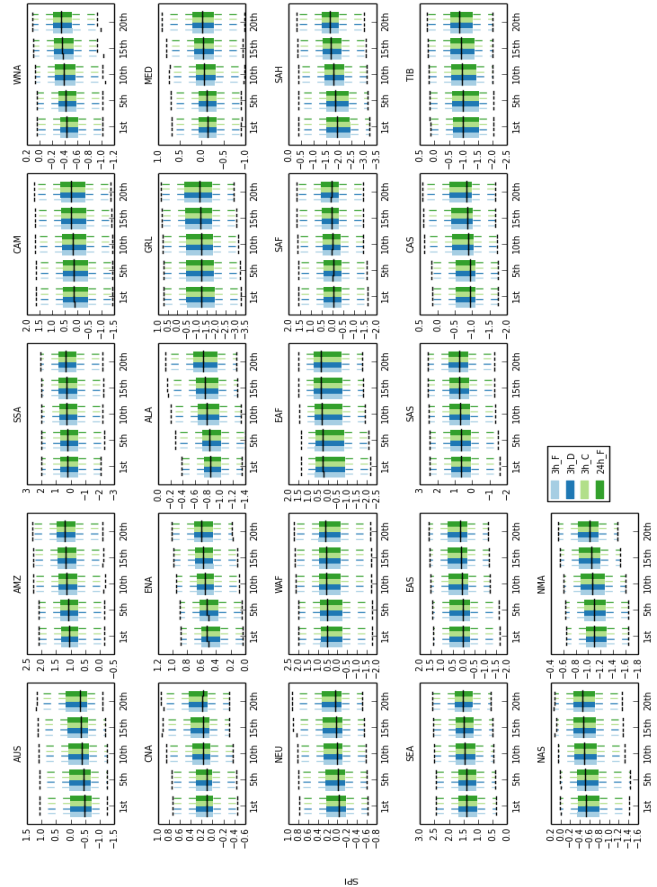
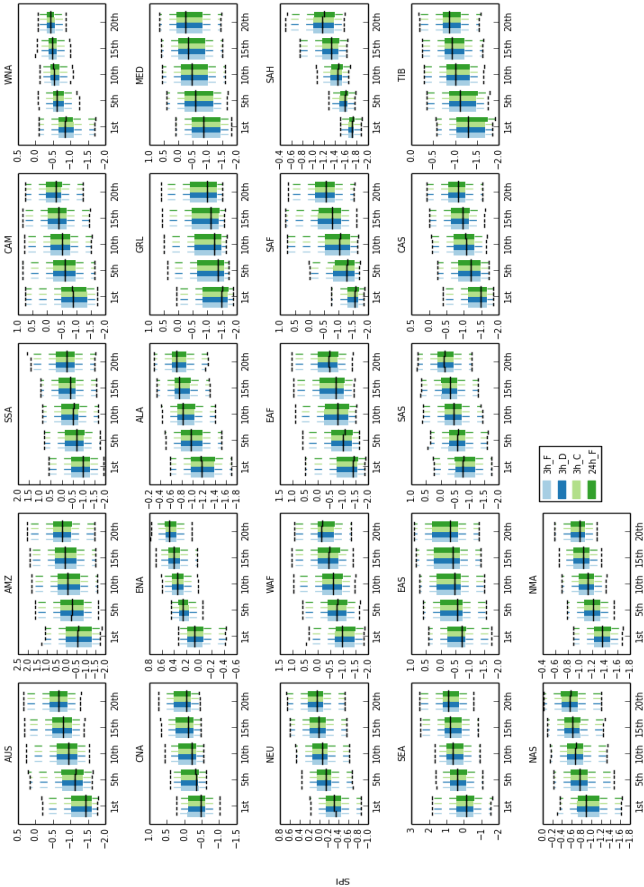
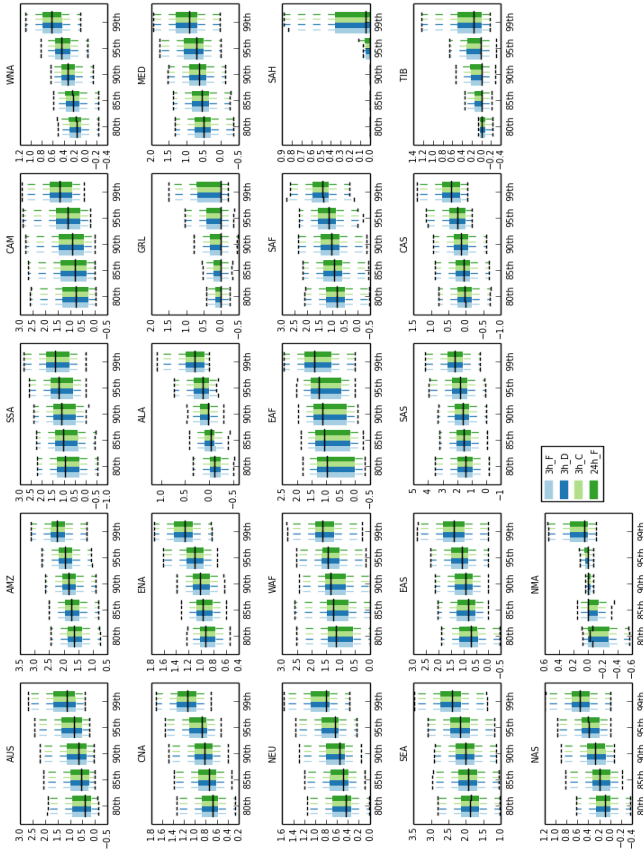
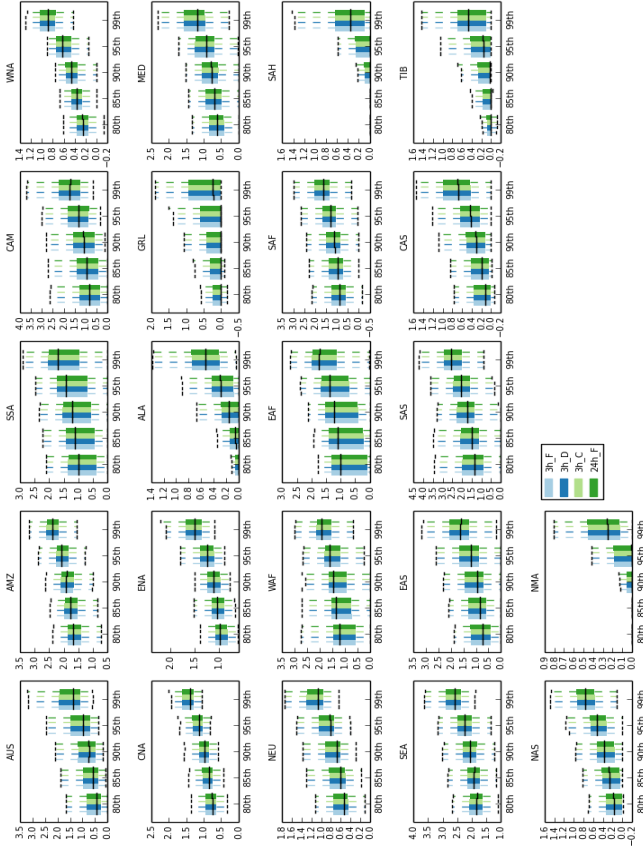


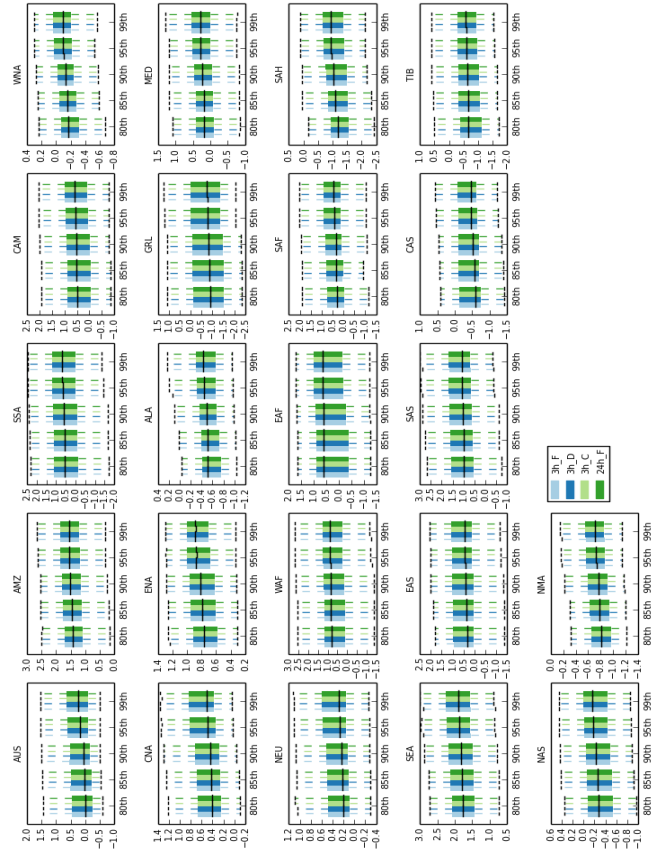
Figure A1: Regional distributions of SPI for drought integration period for every model scenario. Whiskers are the 5th, 95th confidence intervals and represent the spread of points within each region, with outlying points not shown. A higher spread simply indicates more variability within a region.



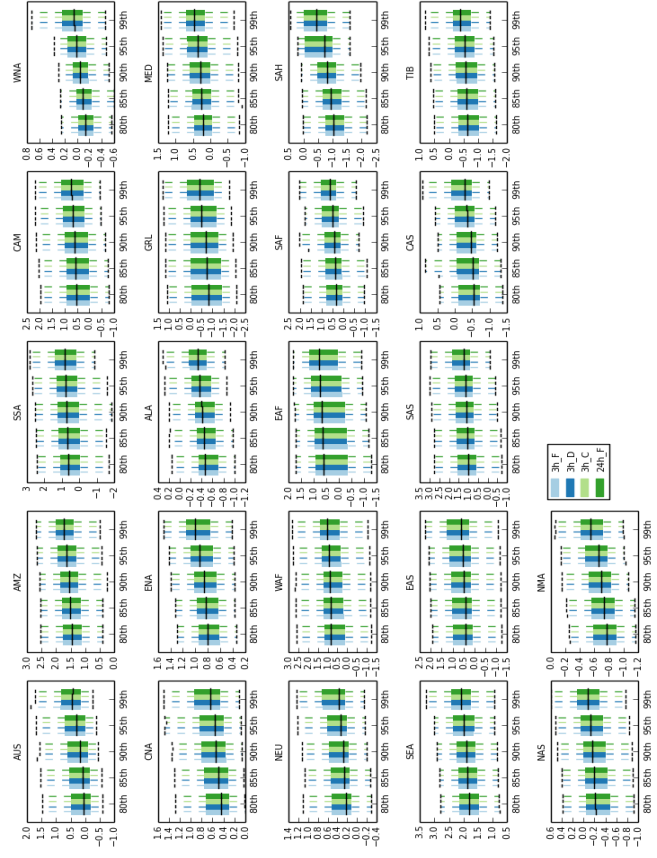
ids



ids

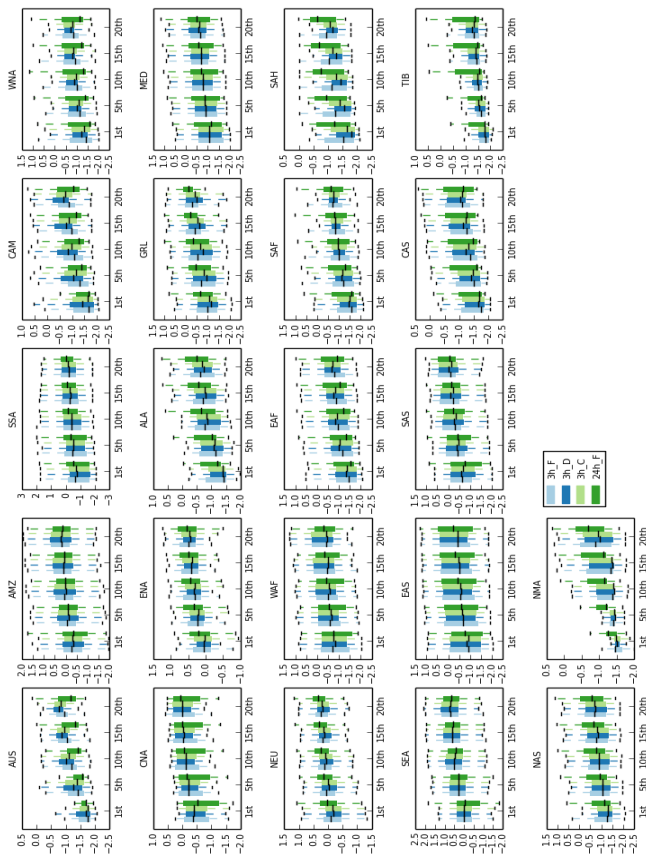


ids

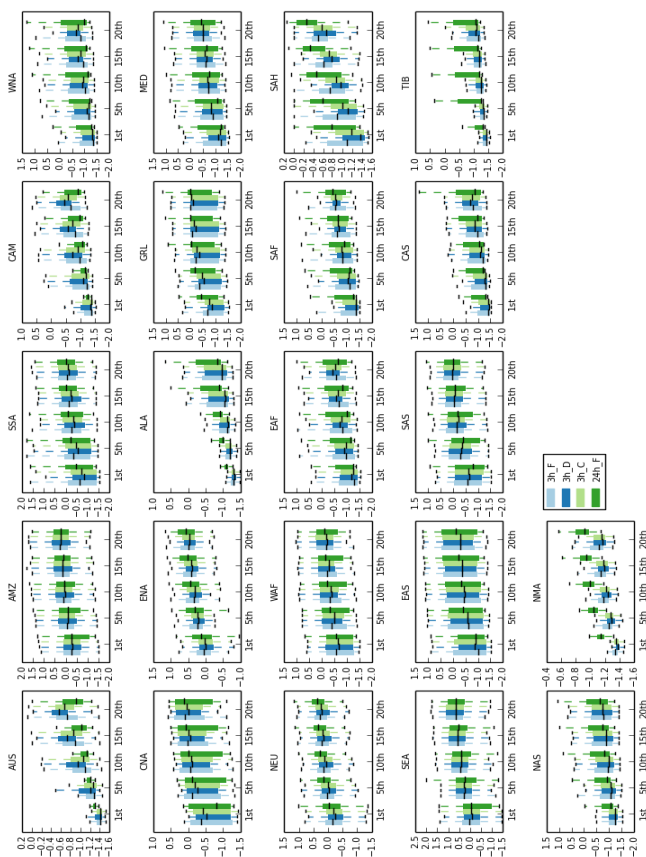


ids

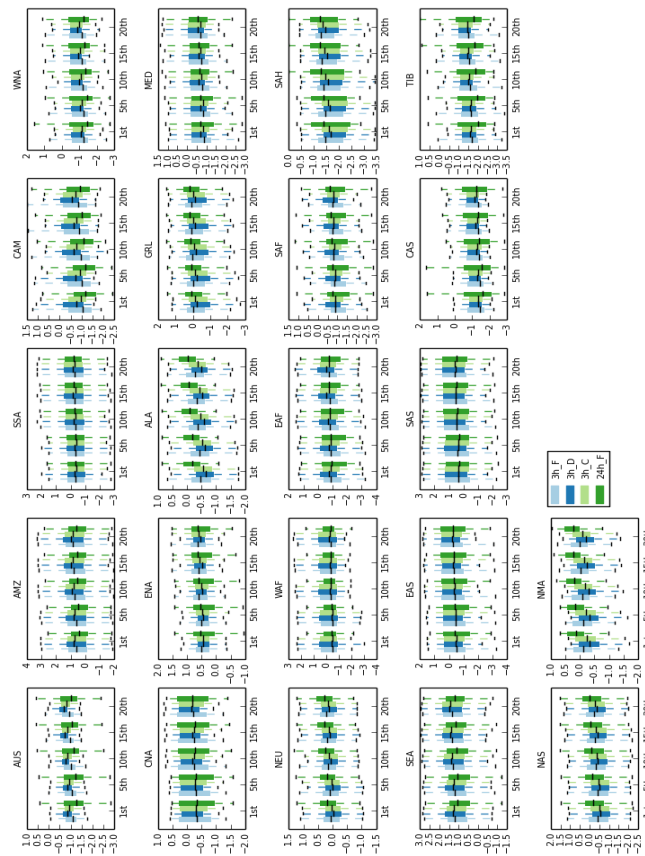
Figure A2: Regional distributions of SPI for high-end extremes percentiles over the integration period for every model scenario. Whiskers are the 5th, 95th confidence intervals and represent the spread of points within each region, with outlying points not shown. A higher spread simply indicates more variability within a region. Calculated over monthly, seasonal, yearly and bi-yearly timescales (top-left, top-right, bottom-left, bottom-right)



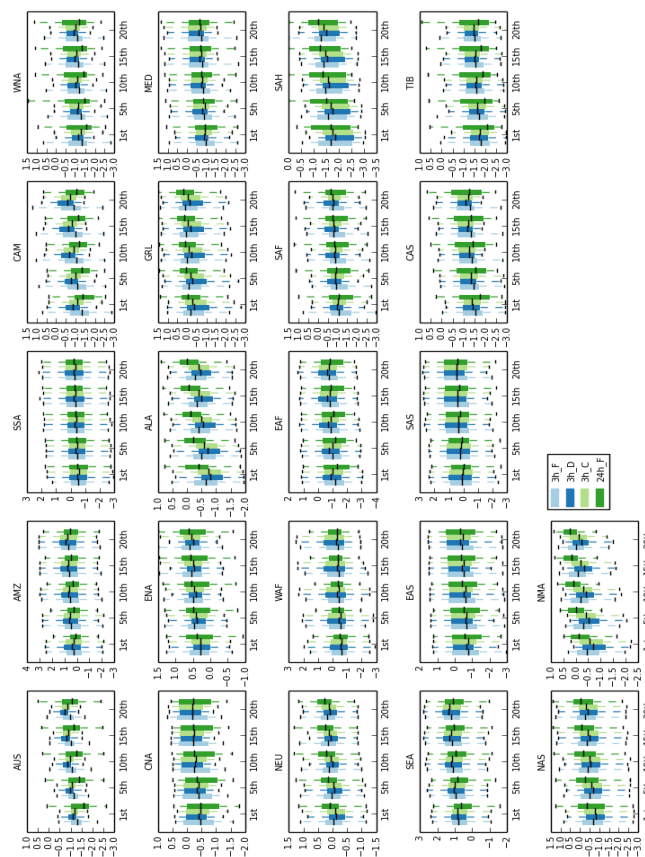
SRI



SRI



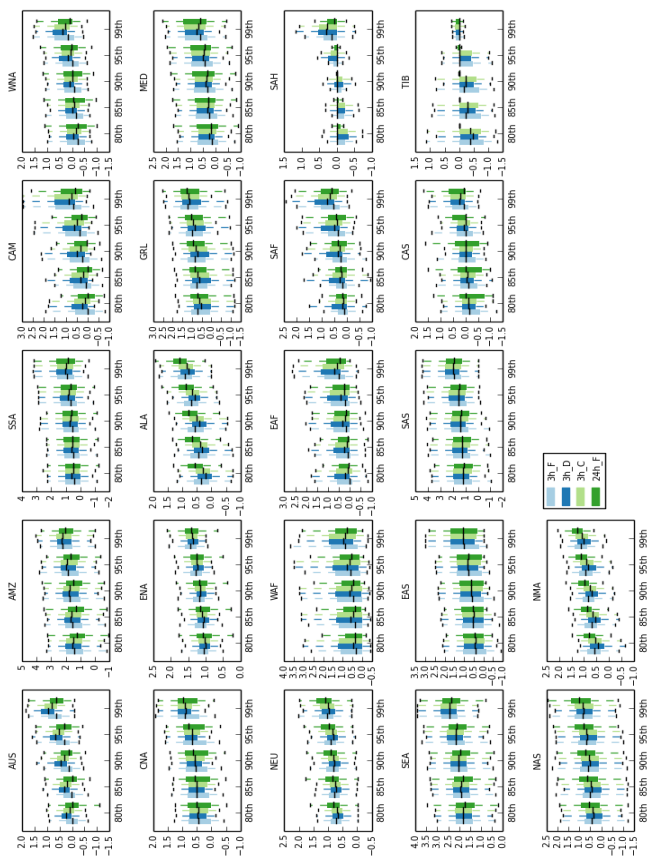
SRI



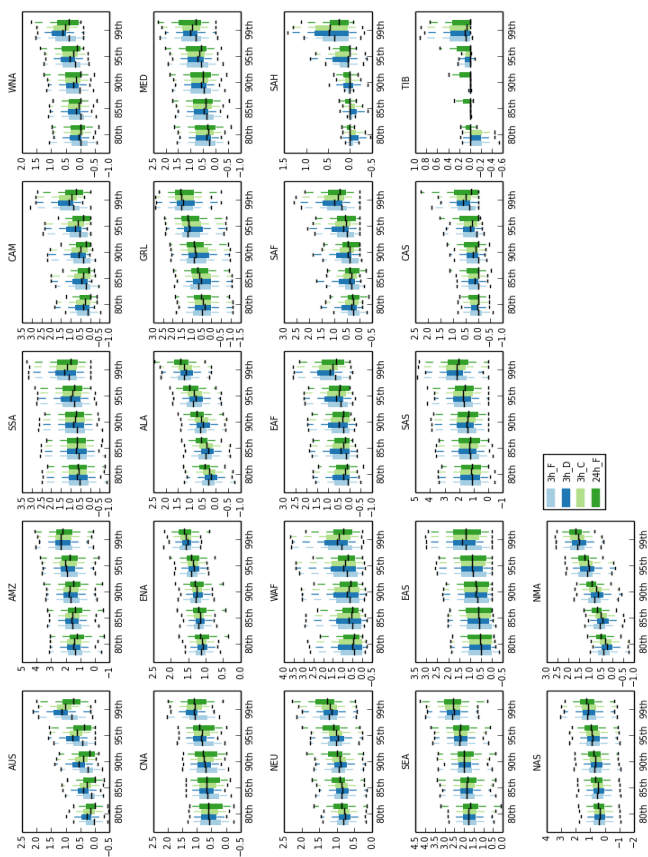
SRI

Figure A3: Regional distributions of SRI for drought percentiles over the integration period for every model scenario. Whiskers are the 5th, 95th confidence intervals and represent the spread of points within each region, with outlying points not shown. A higher spread simply indicates more variability within a region. Calculated over monthly, seasonal, yearly and bi-yearly timescales (top-left, top-right, bottom-left, bottom-right)

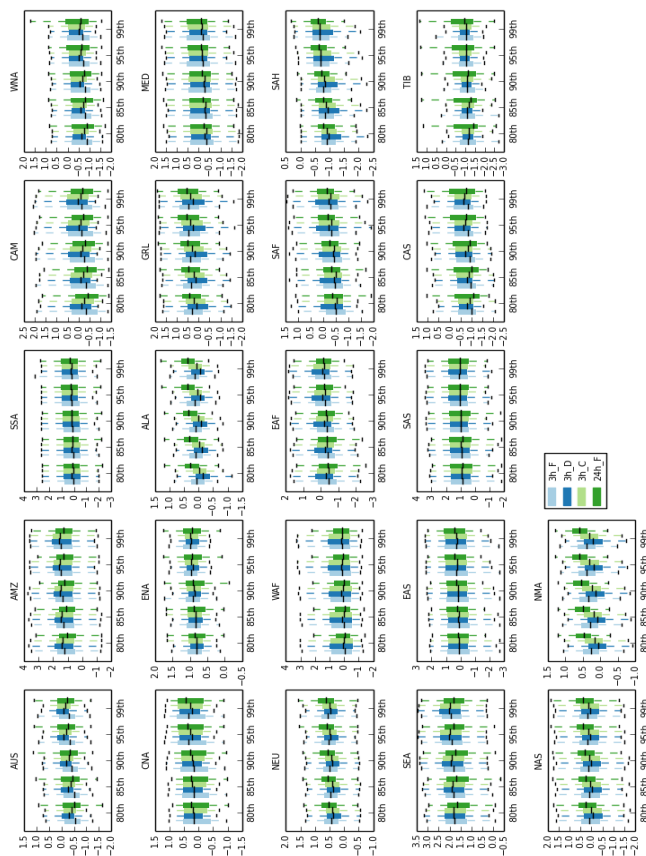




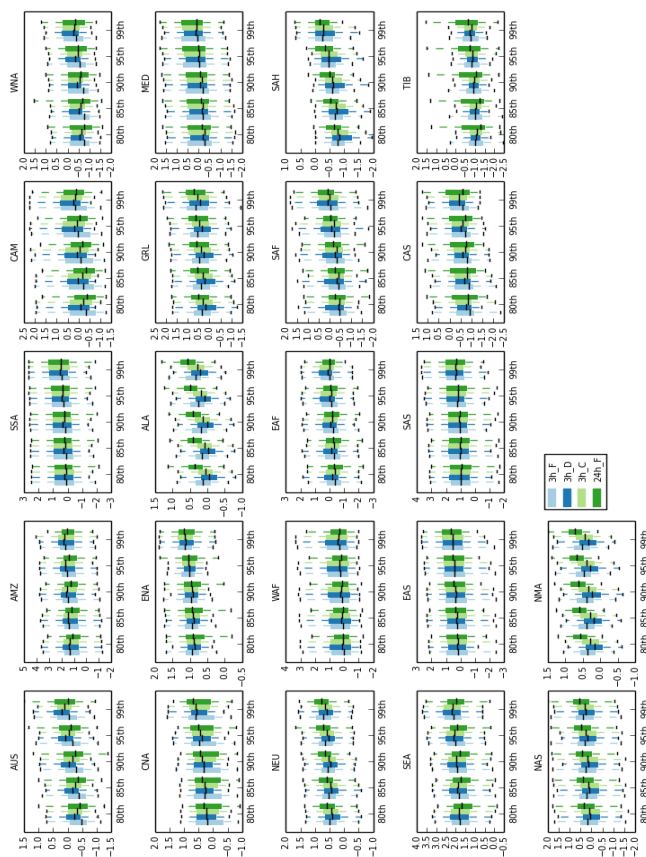
INS



INS



INS



INS

Figure A4: Regional distributions of SRI for high-end extremes percentiles over the integration period for every model scenario. Whiskers are the 5th, 95th confidence intervals and represent the spread of points within each region, with outlying points not shown. A higher spread simply indicates more variability within a region. Calculated over monthly, seasonal, yearly and bi-yearly timescales (top-left, top-right, bottom-left, bottom-right).

Met Office  
FitzRoy Road  
Exeter  
Devon  
EX1 3PB  
United Kingdom