Observation based temperature and freshwater noise over the Atlantic Ocean

Amber A. Boot[1] and Henk A. Dijkstra[1,2]

[1]Institute for Marine and Atmospheric research Utrecht, Department of Physics, Utrecht University, Utrecht, the Netherlands
[2]Center for Complex Systems Studies, Utrecht University, Utrecht, the Netherlands

Correspondence: Amber A. Boot (a.a.boot@uu.nl)

Abstract. The ocean is forced at the surface by a heat flux and freshwater flux field from the atmosphere. Short time-scale variability in these fluxes, i.e. noise, can influence ~~long term~~ long-term ocean variability and might even affect the Atlantic Meridional Overturning Circulation (AMOC). Often this noise is assumed to be Gaussian, but detailed analyses of its statistics ~~appears~~ appear to be lacking. Here we study the noise characteristics in reanalysis data for two fields which are commonly used to force ocean-only models: evaporation minus precipitation and 2 m air temperature. We construct several noise models for both fields, ~~of which a point wise~~ and a point-wise Normal Inverse Gaussian distribution ~~models shows~~ model gives the best performance. ~~A comparison with such noise in~~ An analysis of CMIP6 models shows that these models do a reasonable job in representing the standard deviation and skewness of the noise, but the excess kurtosis is more difficult to capture. The ~~point wise~~ point-wise noise model performs better than the CMIP6 models and can be used as forcing in ocean-only models to study, for example, noise-induced transitions of the AMOC.

1 Introduction

The ocean is forced at the surface by momentum, heat and freshwater fluxes from the atmosphere. Since the ocean responds relatively slowly to the atmospheric forcing, anomalies in this forcing can be modelled as a noise process (Hasselmann, 1976). This study is motivated by the role of such noise in causing ~~noise-induce~~ noise-induced transitions in the Atlantic Meridional Overturning Circulation (AMOC). The AMOC has a major influence on global, and specifically, Northern Hemispheric climate and has been identified as one of the potential major tipping points in the Earth System (Lenton et al., 2008; McKay et al., 2022). A collapse or strong weakening of the AMOC has major consequences for the climate system by changing e.g. global temperature patterns (van Westen et al., 2024b), atmospheric circulation (Orihuela-Pinto et al., 2022), Arctic sea ice cover (van Westen et al., 2024b), the global carbon cycle (Zickfeld et al., 2008; Boot et al., 2024b) and marine ecosystems (Schmittner, 2005; Boot et al., 2024a).

Simple box models have shown that the AMOC can show ~~noise induced~~ noise-induced transitions (Castellana et al., 2019; van Westen et al., 2024a) and probabilities of such transitions could be obtained using rare-event techniques. In these ~~type~~ types of studies, the noise is applied only in the freshwater flux and is often assumed to be white for simplicity. Recently, ~~also~~ noise induced transitions have also been studied in ~~a~~ an Intermediate Complexity Earth System Model (EMIC; Cini et al.,

2024) ~~usign~~ using rare event techniques. Ideally, one would want to study the transitions in full complexity, CMIP6-type, Earth System Models (ESMs). However, due to the complexity and cost of these models, it is not yet possible to systematically use these ESMs for these ~~type~~ types of studies. Recently, a study did look at AMOC tipping in a 10-member ensemble of the NASA-GISS ESM, showing that under the same forcing some ensemble members simulate an AMOC recovery under future emissions, and others show a consistent weakening (Romanou et al., 2023). However, the AMOC does not show a complete collapse in these ensemble members.

To determine the probability of noise-induced transitions using rare event techniques one is at the moment restricted to using ocean-only models and hence the specification of the atmospheric noise is crucial. However, to our knowledge, a detailed study on the properties of the noise in the actual fields relevant in the forcing of ocean models is lacking. Here we focus on noise in the freshwater flux (~~$E - P$~~ $E - P$) and in the 2 m air temperature (T_{2m}). Noise in the momentum flux related to surface winds might also be important for the AMOC. However, we do not consider this here for two main reasons: the statistical properties of the surface winds have been studied more thoroughly before (Sura, 2003; Monahan, 2004, 2018), and the noise in the momentum flux is less important for simulating ~~noise induced~~ noise-induced transitions of the AMOC.

Such a study is also ~~be~~ useful to determine whether EMICs and ESMs adequately ~~actually capture there~~ capture these noise fields. We know that these ~~type~~ types of models exhibit, sometimes very large, biases in their mean state, but also in variability on a whole range of time scales. For example, T_{2m} is biased too warm in the CMIP6 models over the Atlantic sector of the Southern Ocean and the Eastern South Atlantic, while there is a cold bias over much of the North Atlantic and Arctic Ocean (IPCC AR6 Chapter 3). The air temperature biases can also be seen in the sea surface temperatures (Zhang et al., 2023), thereby directly affecting the density structure of the ocean. For precipitation there is a consistent double Intertropical Convergence Zone (ITCZ) bias from CMIP3 to CMIP6 models (Tian and Dong, 2020). This means that in the Atlantic, the ITCZ, and therefore bands of high precipitation extend too much towards the south. Following the double ITCZ bias (Tian and Dong, 2020; Li et al., 2020), there is a strong positive freshwater flux bias north of the equator and a strong negative bias south of the equator in the CMIP6 multi model mean (MMM; Liu et al., 2022). Between 10° and 60°N, and the equator and 35°S the freshwater flux is typically positively biased in the CMIP6 MMM (Liu et al., 2022). These biases are among the reasons why the AMOC is thought to be too stable in CMIP6 type models (Weijer et al., 2019; van Westen and Dijkstra, 2024).

In this study we determine the statistical properties of the ~~$E - P$~~ $E - P$ and T_{2m} noise based on the ERA5 reanalysis data. We compare this ~~observed~~ observation-based noise with the noise simulated by coupled CMIP6 ESMs and identify relevant biases. Based on the ~~observed~~ ERA5 noise we construct a noise model that can be used to force ocean-only models. This product can be used to study the influence of short time scale atmospheric variability on long term ocean variability and eventually to study ~~noise induced~~ noise-induced transitions of the AMOC.

2 Methods

2.1 ERA5 reanalysis data

We analyse the noise in ~~$E - P$~~ $E - P$ and T_{2m} over the Atlantic Ocean between 60°S and 80°N. For this we use ERA5 reanalysis data (Hersbach et al., 2020), which is the most recent reanalysis product of the European Center for Medium-Range Weather Forecasts (ECMWF) and replaces the ERA-Interim reanalysis product. The ERA5 product is created by combining both satellite and ground observations with a numerical model used for weather forecasting. For the freshwater flux we determine the net freshwater forcing by taking the ~~negative of the summing of the~~ sum of the variables 'total precipitation' and 'evaporation', i.e. - (total precipitation + evaporation). Since evaporation is defined negative in ERA5 data, and total precipitation positive, this results in a dataset for ~~$E - P$~~ $E - P$ where net evaporation is positive, and net precipitation is negative. The datasets contain monthly data from 1940 to 2022 on a 0.25° rectilinear grid. To determine the noise in both fluxes, we first detrend each grid point by subtracting a ~~5 year~~ 5-year running mean. Next, we deseasonalise the data by subtracting a monthly climatology based on the detrended data. This results in a noisy dataset where each grid point has zero mean and no trend. We analyze the fields by looking at the standard deviation (σ), skewness and excess kurtosis of the noise, where Gaussian white noise would have zero skewness and zero excess kurtosis.

2.2 CMIP6 models

We compare the noise in the ERA5 data to that found in CMIP6 ESMs. In total we use 36 different models, where we note that we use two different realisations from the UKESM-1-0-LL model that is run by two different model groups (i.e. the Met Office Headly Centre (MOHC) and National Institute of Meteorological Sciences - Korea Meteorological Administration (NIMS-KMA)). For each model we determine the evaporation minus precipitation by using the variables 'evspsbl' and 'pr', and we use 'tas' for ~~T_{2m}~~ T_{2m}. We do this for the historical simulations between 1940 and 2014. We first regrid all models to a $1° \times 1°$ rectilinear grid. Next, we compute the noise in the models by following the same methodology as for the ERA5 data, i.e. we detrend and deseasonalise the data. We could also use simulations without forcing, i.e. the piControl simulations. However, for comparison with the ERA5 data it is better to use a similar methodology for both data sources, which includes the detrending in the ERA5 data.

When comparing the CMIP6 data to the ERA5 data we use the same time period in the ERA5 data as in the CMIP6 data, i.e. 1940 to 2014, and re-grid the ERA5 data to a 1° rectilinear grid. Due to its original higher resolution, the land mask in ERA5 captures small islands that are not captured by the CMIP6 land mask. To account for this, we mask out these small islands in both the ERA5 and the CMIP6 data. A full list of the models used, including which member and citations can be found in the Appendix (Table A1).

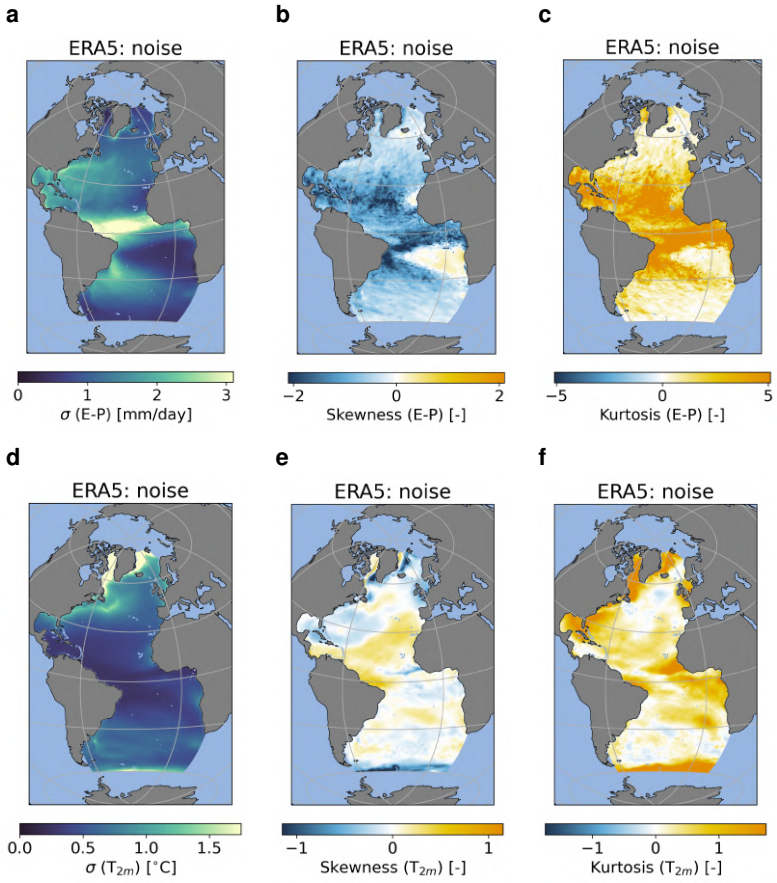

Figure 1. Standard deviation (σ), skewness and excess kurtosis over time for the ERA5 noise for the E – P flux (a-c) and T_{2m} (d-f).

3 Results

3.1 ERA5 reanalysis data

Fig. 1 (a-c) shows the standard deviation, skewness and excess kurtosis in the ~~$E - P$~~ $E - P$ noise. The highest standard deviation is found north of the equator in the ITCZ~~. Other~~ , and other regions with relatively high standard deviations are the western boundary currents. In the Northern Hemisphere, the strongest negative skewness is found between 10°N and 30°N (Fig. 1b). The negative skewness here indicates that the distribution is skewed towards extreme precipitation events, which is partially related to tropical storm activity in this region. In the Southern Hemisphere there is a strong negative skewness in the region of the South Equatorial Current. South of there, there is a small region of moderate ~~postive~~ positive skewness. The rest of the ocean generally shows a small negative or near zero skewness. The excess kurtosis shows a relatively similar pattern as the skewness except with the opposite sign (Fig. 1c). The ~~storngest positve~~ strongest positive excess kurtosis is found over

the entire latitudinal band 10°S to 30°N. Also this is an indication of high extremes, and because of the negative skewness it means indicates extreme precipitation events. The rest of the ocean has slightly positive or near zero excess kurtosis. Due to the non-zero skewness and excess kurtosis in the noise in most grid points, the noise cannot be classified as Gaussian white noise in these grid points.

For T_{2m} (Fig. 1d-f), the largest standard deviation in the noise is found in the (seasonally) sea ice covered regions in the high latitude North Atlantic (Fig. 1d). Also the Gulfstream region shows a relatively high standard deviation. Regions around the sea-ice edge, both in the Northern and Southern Hemispheres, show a relatively strong negative skewness (Fig. 1e), which means the distribution in these regions are skewed towards more cooling events. The pattern for the skewness in the South Atlantic is relatively patchy with both small negative and small positive values. In the North Atlantic, the regions around the trade winds show positive skewness, and the subtropical gyre shows negative skewness. The (seasonalyseasonally) sea ice covered regions show strong negative skewness. For the excess kurtosis (Fig. 1f) most of the Atlantic region shows (strong) positive values with the strongest signals over the sea ice covered regions and close to the seasonal sea-ice edge (also in the South Atlantic), and in the Gulf of Mexico. The combination of negative skewness and positive excess kurtosis in the sea ice covered regions suggests that in these regions strong cooling events can take place which is likely associated with strong increases in sea-ice cover. Just as for the freshwater flux, the excess kurtosis deviates from 0 zero in most regions in the ocean, which means that also the noise in T_{2m} is unlikely to be Gaussian white noise in most grid points.

To better understand the results, we look in the noise fields for regions with similar distributions. We do this by dividing both the $E-P$ and T_{2m} noise fields into 12 different clusters (Fig. 2). Thereto we use a k-means clustering algorithm where we use the standard deviation, skewness and kurtosis as input. The decision for 12 clusters is based on several methods (Fig. A1), i.e. the elbow method, the silhouette score, the gap statistic and visually inspecting the clusters while the number of clusters is varied. The probability density functions, standard deviation, skewness and excess kurtosis of the clusters are displayed in Fig. A2 - A5

For $E-P$ we find several relatively large clusters. The subpolar regions are divided into two clusters (clusters 1 and 4) where cluster 4 is more poleward. The main difference between the two clusters is the lower standard deviation in the higher latitude cluster. The high standard deviation region of the ITCZ also clearly stands out as a separate cluster (cluster 6). The subtropical region is divided into 9 different clusters. The 6 clusters that cover the North and South Equatorial Current stand out with strong positive excess kurtosis and strong negative skewness (clusters 2, 3, 5, 9, 10 and 11). These are regions that experience tropical storms and hurricanes, which are recorded as very strong extreme precipitation events in the $E-P$ noise fields. These extremes can be very local, explaining why 6 clusters are necessary for this region. The cluster closest to a Gaussian distribution is the cluster in the Southeastern subtropical region (cluster 12) with a skewness of 0.16 and an excess kurtosis of 0.90 (Fig. 2b). For all clusters, kurtosis is larger than 1.5 times the square of the skewness (Fig. 2b) which is consistent with multiplicative noise.

The 12 clusters for the T_{2m} noise do not show an overlap with the $E-P$ clusters and several of them appear to follow the general ocean circulation pattern. For example, cluster 12 is centered around the North Atlantic Current, and cluster 10 around the North Equatorial Current and the North Brazil Current. While for the $E-P$ noise several clusters are necessary

for the subtropics, for T_{2m} several clusters are necessary for regions covered by sea ice, or adjacent to these regions. Cluster 3 describes the regions in the Labrador and Greenland Seas that experience sea ice annually. Clusters 2, 5, and 6 all cover regions close to the sea-ice edge. The noise in these regions is likely affected by interannual variability in the sea-ice extent which leads to relatively strong positive excess kurtosis and relatively strong negative skewness. Two clusters (7 and 12) have near zero area weighted skewness and excess kurtosis and are therefore close to a Gaussian distribution (0.00 skewness for both, and 0.06 and 0.13 for excess kurtosis, respectively; Fig. 2d). Cluster 7 covers the South Atlantic between 30°S and 50°S, and parts of the Eastern North Atlantic between 30°N and 60°. Cluster 12 is the cluster around the North Atlantic Current, but does show some variability in both skewness and excess kurtosis in the cluster. Just as for the $E - P$ clusters, for all T_{2m} clusters, kurtosis is larger than 1.5 times the square of the skewness (Fig. 2d) consistent with multiplicative noise.

3.2 CMIP6 data

In this section we analyze the results for the ~~multi model~~ multi-model mean (MMM) of the CMIP6 models. We determine the MMM at the end of the analysis. This means that we for example first determine the skewness for each model, and then average over the 2D skewness fields of all the models to create the MMM. Each model has been given the same weight. Results for individual models can be found in the Appendix (Fig. A10 to Fig. A15).

The MMM for the noise in the ~~$E - P$~~ $E - P$ flux does not always represent the amplitude in the statistics of the ERA5 noise well (Fig. 3a-f), though the spatial patterns are relatively well resolved in the MMM. The standard deviation is underestimated over the entire ocean with the strongest underestimation in the ITCZ regions and over the western boundary currents (Fig. 3d). The multimodel mean shows a stronger negative skewness over the South Equatorial current that is also shifted more southward compared to ERA5 noise (Fig. 3e). Furthermore, the positive skewness over the eastern subtropical region is not captured by the CMIP6 MMM. The excess kurtosis is also positively biased in the CMIP6 MMM over the South Equatorial Current (Fig. 3f). In the region between 10°S and 25°N there is a patchy response where most regions see an underestimation of the excess kurtosis (red colors) and some regions an overestimation (blue colors) compared to the ERA5 noise.

The CMIP6 MMM does capture the spatial pattern and amplitude of the standard deviation of the noise in T_{2m} well compared to the ERA5 noise (Fig. 4a,b). The spatial pattern of the skewness is captured reasonably well in the Northern Hemisphere, but the amplitude is typically smaller than in the ERA5 noise (Fig. 4c, d). In the Southern Hemisphere the CMIP6 MMM shows mostly slightly positive skewness, whereas the ERA5 noise mostly shows small negative skewness. The absolute differences are not that large, but there is an important difference in sign. For excess kurtosis the spatial pattern is also relatively similar in the CMIP6 MMM compared to the ERA5 noise, however, the regions in the ERA5 noise with small negative kurtosis are not captured by the CMIP6 MMM (Fig. 4e,f). The amplitude of the excess kurtosis, however, is not as well resolved as the ~~special~~ spatial pattern. Most regions in the CMIP6 MMM show ~~a~~ an underestimation of the excess kurtosis compared to ERA5.

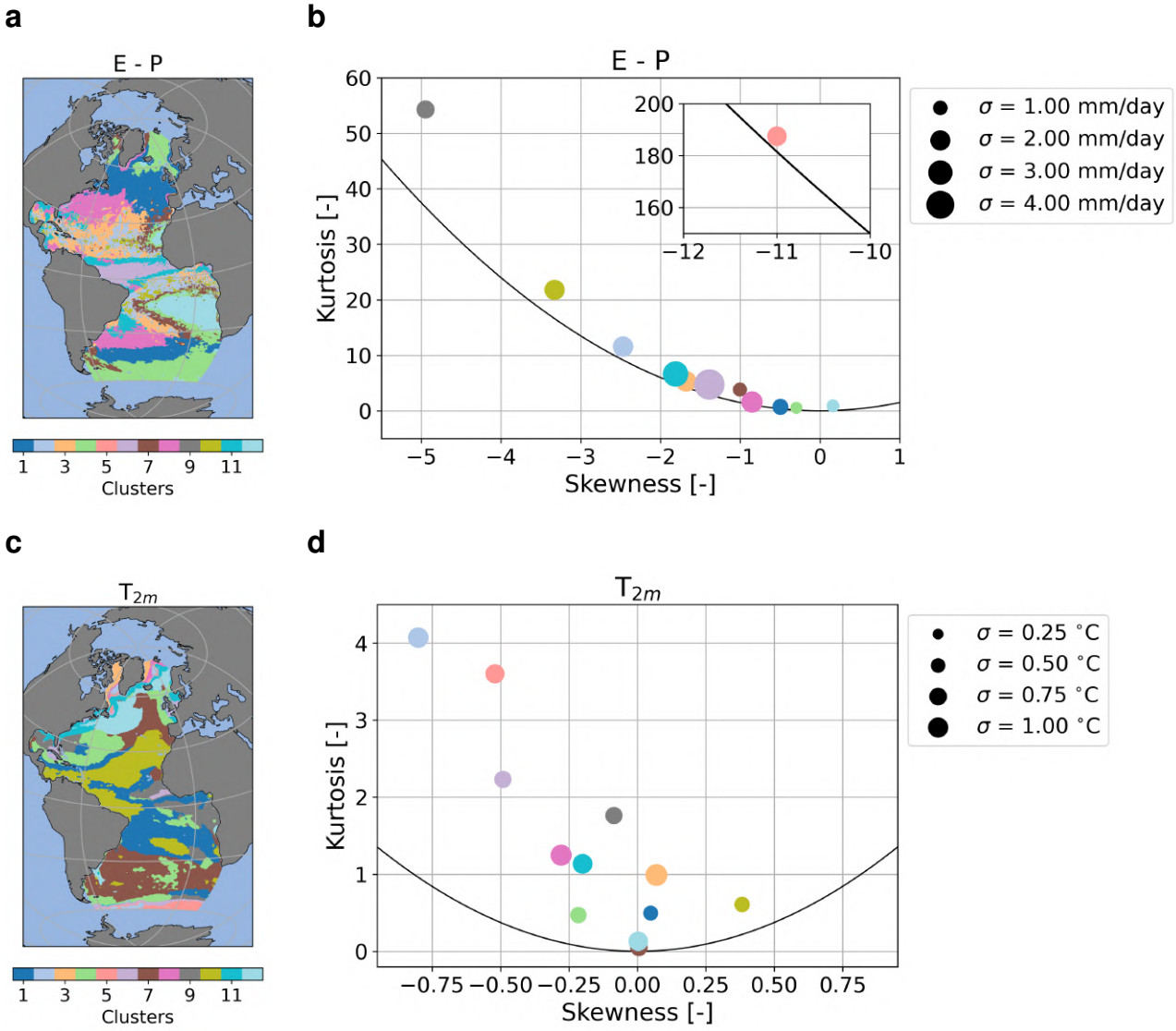

Figure 2. Overview of the clusters and corresponding statistics. (a) The clusters for the $E - P$ noise. (b) The skewness (S; x-axis) and excess kurtosis (K; y-axis) of the clusters in (a). The colors of the markers correspond to the color coding in (a). The size of ther markers represents the standard deviation in mm/day. (c) and (d) as in (a) and (b) but for the T_{2m} clusters. The unit of standard deviation in (d) is °C. The black line in (b) and (d) represents $K = 1.5S^2$.

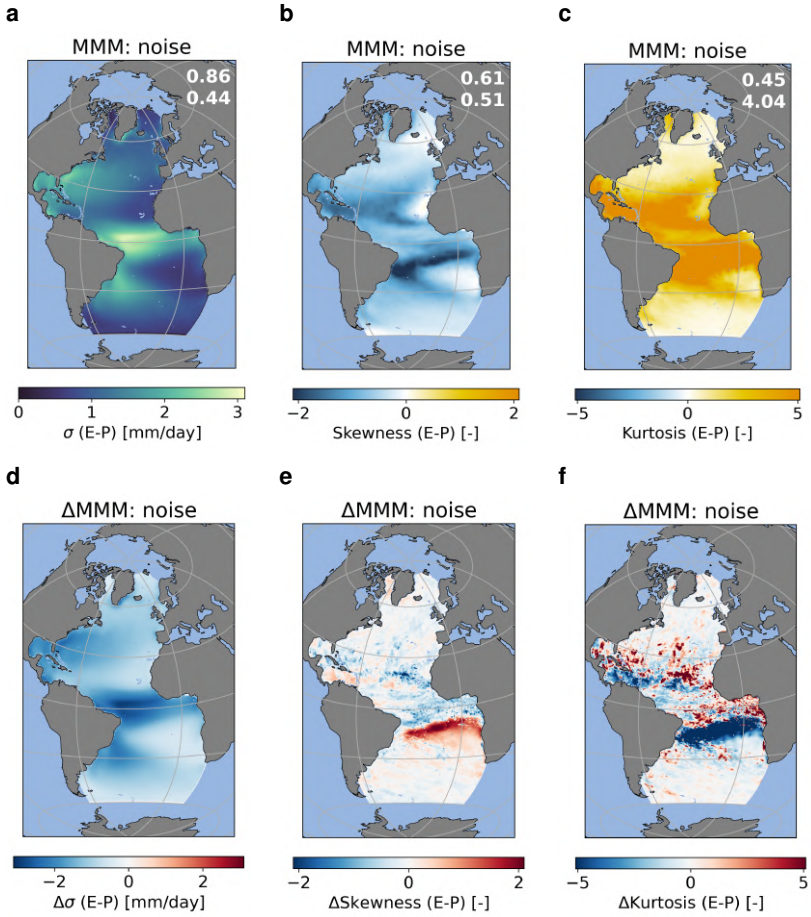

Figure 3. Standard deviation (σ), skewness and excess kurtosis of CMIP6 ~~multi model~~ multi-model mean (MMM) noise for the ~~$E-P$~~ $E-P$ flux (a) – (c). Differences with ERA5 data (i.e. ERA5 minus CMIP6 MMM) are shown in (d) – (f). The numbers in the top right corner of (a) – (c) reflect the spatial correlation and root mean square error. Units for (a) and (d) are mm/day.

4 Noise model

The CMIP6 MMM appears to do a decent job in capturing the ~~observed~~ observation-based noise field of both ~~$E-P$~~ $E-P$ and T_{2m}. However, there is still a large spread in the model ensemble, meaning not all models are able to capture these noise fields adequately. Our aim in this section is to develop a statistical model of the noise in both ~~$E-P$~~ $E-P$ and T_{2m} that can be used as forcing in ocean models. We have tried several methods to construct such a model and we will present four of those below. All these models are based on the ERA5 reanalysis data.

Three of the methods are based on a principal component analysis (PCA) in which we base the noise model on the Principal Components (PCs) and corresponding Empirical Orthogonal Functions (EOFs). The PCA is performed on the noise and is

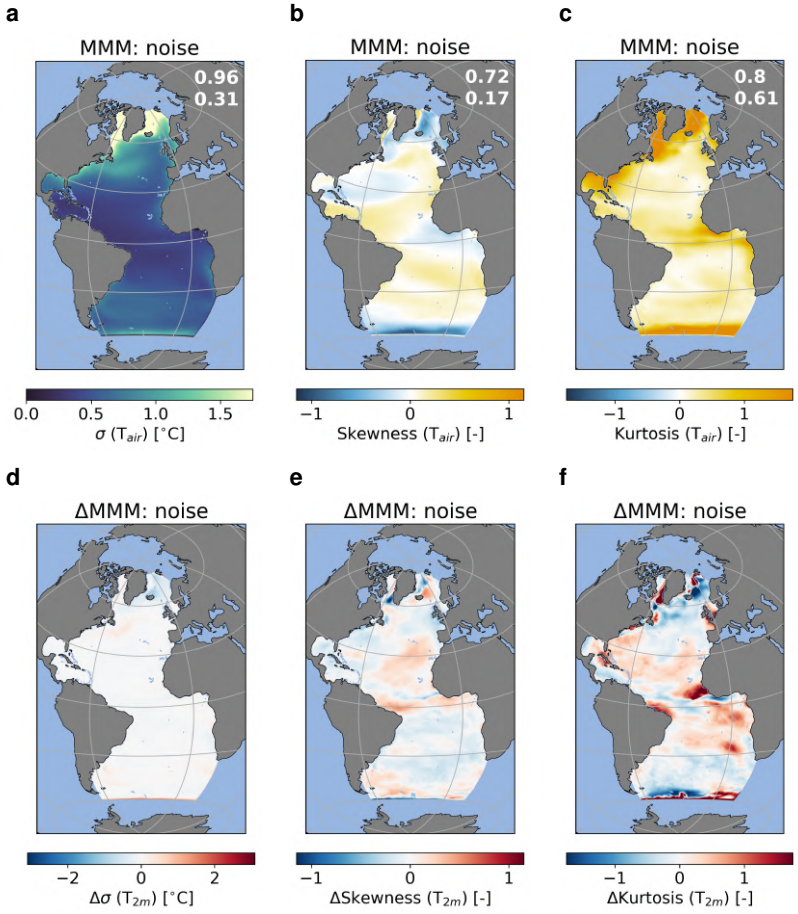

Figure 4. As Fig. 3 but for T_{2m} in °C.

weighted to account for the grid cell areas. For all three methods we use the number of EOFs necessary to explain 90% of the variance in the noise (i.e. 289 EOFs and PCs for ~~$E - P$~~$E - P$, and 53 for T_{2m}). For the first two methods we directly sample (using replacement) from the PCs. The first method we name PC (1), as we select one random time step (i.e. month) for all PCs. ~~This means we have 996 possible combinations for this method since~~ For the PC(1) method we uniformly sample one integer from 1 to the length of the PCs~~is~~, i.e. 996. We apply this integer for all PCs. For example, if our integer is 7, then we sample the 7th month of each PC to construct the noise model. Using this method we therefore have in total 996 ~~months~~different realizations to sample from, meaning this method is not strictly stochastic. The second method (PC (N)), we sample a random time step out of the PCs, but a different time step for each PC. For the third method (PC (NIG)), we fit a Normal Inverse Gaussian distribution to the individual PCs, and next sample randomly from these distributions in a similar fashion as the PC (N) method. The NIG distribution used in the PC (NIG) model has a probability density function determined

by

$$f(x,\alpha,\beta,\delta,\mu) = \frac{\alpha\delta K_1(\alpha\sqrt{\delta^2+(x-\mu)^2}}{\pi\sqrt{\delta^2+(x-\mu)^2}} e^{\delta\sqrt{\alpha^2-\beta^2}+\beta(x-\mu)}, \quad (1)$$

~~Where~~ Here α is a tail heaviness parameter, β an asymmetry parameter, μ regulates the shift of the distribution, and δ the scale of the distribution. K_1 represents a modified Bessel function of the second kind. Note that we have tried several other distributions as well, all of which performed worse than the NIG distribution. Among these distributions are the generalized hyperbolic distribution, the gamma distribution, the beta distribution and the skewed-t distribution.

We choose to use three different PCA-based models. The PC (1) model is used to test whether the PCAs can in fact capture the statistics of the noise well. However, since this method is not fully stochastic we also chose to use other models. The PC (N) model is in set-up very similar but more stochastic than the PC(1) method. As the PC (N) model also has a discrete number of values to sample from, we also used the PC (NIG) model, which does not have this problem. For all three methods, noise fields are constructed by multiplying the value sampled from the PCs with the spatial patterns captured by the EOFs and next summing over the number of PCs/EOFs. Results for the PC (1) and PC (N) models can be found in the Appendix (Fig. A6 to Fig. A9). The PC (NIG) ~~models~~ model shows a good agreement with the noise diagnosed from the ERA5 data for the spatial patterns of the standard deviation (Fig. 5a, d), but it is unable to capture the spatial patterns of the skewness (Fig. 5b, e) and excess kurtosis (Fig. 5c, f). The standard deviation in the noise is captured reasonably well (Fig. 5d). Looking at the skewness (Fig. 5e), and the excess kurtosis (Fig. 5f), we can see that this model is unable to represent these metrics correctly, since the PC (NIG) model simulates near zero skewness and excess kurtosis. Just as for the ~~$E-P$~~ $E-P$ flux, the PC (NIG) model represents the spatial pattern of the noise in the T_{2m} well in the standard deviation (Fig. 6a, d), but not in the skewness (Fig. 6b, e), and excess kurtosis (Fig. 6c, f). Again, the skewness and excess kurtosis are near zero in all regions, except for the excess kurtosis in the sea ice covered regions in the North Atlantic.

Since the models using the PCA show difficulty in representing the ERA5 noise we have, as the fourth method, also fitted several statistical distributions directly to the noise for each grid cell. For both the ~~$E-P$~~ $E-P$ and T_{2m}, the Normal Inverse Gaussian (NIG) distribution appeared to be the best fit. ~~Most~~

We have tested the goodness-of-fit with several measures (Fig. A16). Firstly, we have performed an Anderson-Darling test on normality. We find that for the $E-P$ noise, only 8% of the grid points pass ~~a~~ this test ($p < 0.05$) (Fig. A16b). For T_{2m} this is higher, i.e. 42% (Fig. A16f). Next we have tested whether the NIG provides a better fit than a normal distribution for each grid point. For this we use the following measure:

$$\chi_n = \frac{1}{N}\sum_{i=1}^{N}\frac{(f_i-m_i)^2}{m_i^2}, \quad (2)$$

where N is the number of bins used (i.e. 50), f_i the probability density function of the timeseries per grid point, and m_i the fitted probability density function which is either fitted to an NIG distribution or a normal distribution. We compute χ_n for both an NIG and Gaussian fit and compare the two. For 98% of the grid points the NIG fit performs better (i.e. χ_n is smaller for the

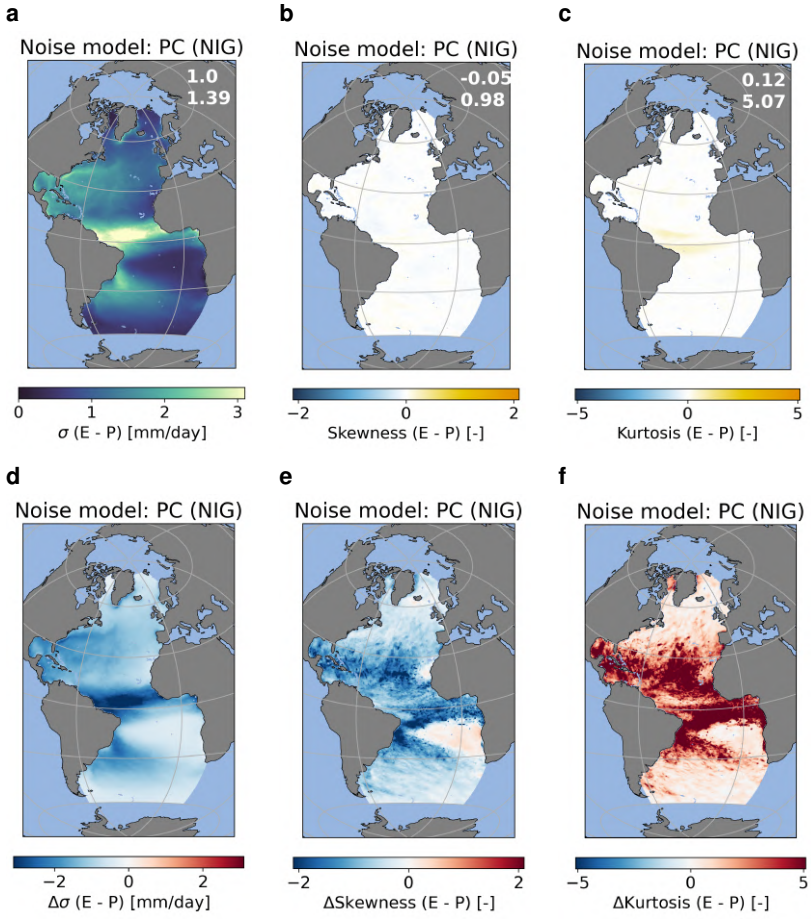

Figure 5. Standard deviation (σ), skewness and excess kurtosis of the noise from the PC (NIG) model for the ~~$E - P$~~ $E - P$ flux (a) – (c). Differences with ERA5 data (i.e. ERA5 minus PC (NIG)) are shown in (d) – (f). The statistics of the noise model are based on 10000 realisations (months). The numbers in the top right corner of (a) – (c) reflect the spatial correlation and root mean square error. Units for (a) and (d) are mm/day.

NIG fit) for the $E - P$ noise (Fig. A16a), and 94% for the T_{2m} noise (Fig. A16e). To test whether the NIG model is a good fit, we apply a Kolmogorov - Smirnov ~~goodness of fit~~ test. For the ~~$E - P$~~ $E - P$, only 27 grid points do not pass this test, and for T_{2m}, 8 grid points do not (out of 138,788 (ocean) grid points) ~~. With this method, we~~ ($p < 0.05$). However, the Kolmogorov - Smirnov test is not well suited for heavy tailed distributions as we find in our data. Ideally, we would like to perform an Anderson-Darling test, or a similar test, as a goodness-of-fit test to test whether the NIG fits are statistically significant, but this is computationally too expensive. For the Anderson - Darling test we need to compute critical values. For the Gaussian distribution these are known, and independent of the parameters describing the distribution (i.e. mean and standard deviation). For the NIG distribution, however, these critical values are dependent on the parameters of the distribution. This means that we

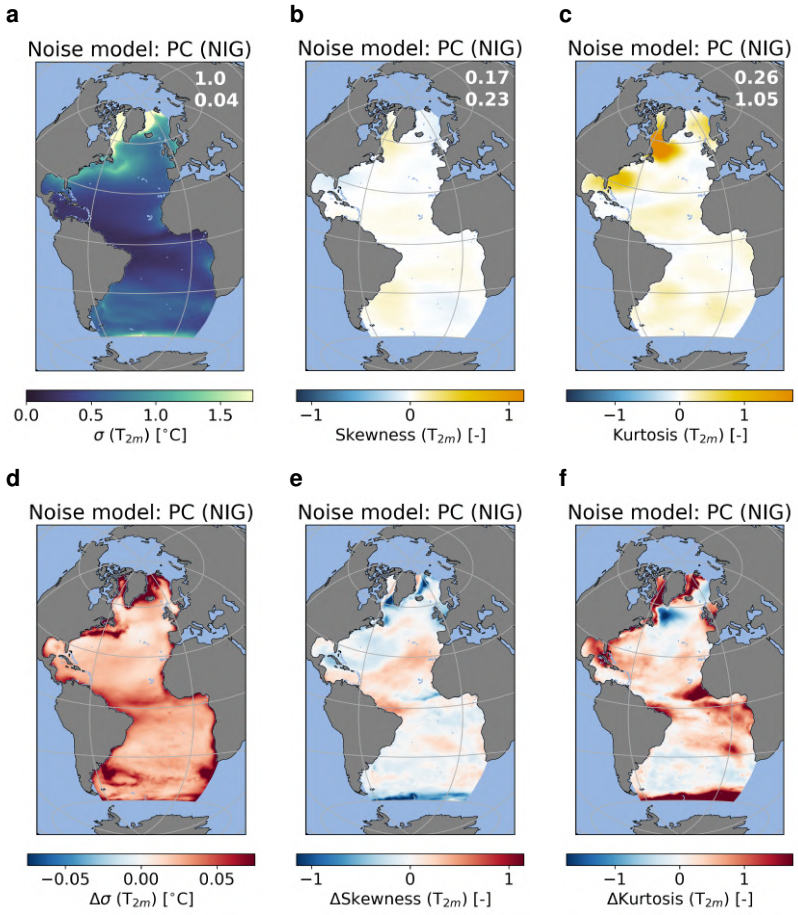

Figure 6. As Fig. 5 but for T_{2m} in °C

should compute the critical values for each grid point separately, which is computationally too expensive and therefore we do not use the Anderson - Darling test for the NIG distribution. Lastly, we test the significance of the skewness and kurtosis of the $E-P$ and T_{2m} noise. We do this by fitting an AR(1) model to the data, and subsequently generate sampling statistics from this model. The fitted (Gaussian) AR(1) model fails to provide good statistics ($p < 0.05$) for 93% and 85% of the grid points for the $E-P$ skewness and excess kurtosis (Fig. A16c, d), and 38 % and 53% for the T_{2m} skewness and excess kurtosis (Fig. A16g, h).

Based on this collection of tests, we think that for most of the grid points the NIG model provides a good fit to the data. Furthermore, following the Anderson-Darling test on normality and the fitted AR(1) model, most of the $E-P$ noise is non-Gaussian, and to a lesser degree this also applies to the T_{2m} noise. The grid points for the $E-P$ noise that are likely Gaussian are located in clusters 4 and 11, which are indeed clusters with skewness and excess kurtosis close to 0 (Fig. 2a, b). For the T_{2m} noise the grid points that show Gaussian behavior are mainly located in the sea-ice free subpolar Ocean. These

grid points mainly belong to clusters 1, 7 and 12 which are also the clusters with approximately zero skewness and near-zero excess kurtosis (Fig. 2c, d).

Using the fitted NIG distribution, we can generate a fully stochastic noise field for each month using the 4 parameters per grid cell.

~~The fitted NIG distribution~~ The model shows a good agreement with the noise diagnosed from the ERA5 data for the spatial patterns of the standard deviation (Fig. 7a, d), skewness (Fig. 7b, e) and excess kurtosis (Fig. 7c, f). Especially the standard deviation in the noise is captured well with only small deviations between 10°S and 25°N (Fig. 7d). The NIG distribution underestimates the regions with strong negative skewness over the latitude bands 0°N to 10°S and 10°N to 25°N (Fig. 7e). For excess kurtosis we see a similar underestimation in these regions, meaning that the excess kurtosis is higher in the ERA5 data (Fig. 7f). However, the region between these two latitude bands shows a much higher excess kurtosis in the NIG model compared to the ERA5 noise.

Just as for the ~~$E - P$~~ $E - P$ flux, the NIG model represents the spatial pattern of the noise in the T_{2m} well in the standard deviation (Fig. 8a, d), skewness (Fig. 8b, e), and excess kurtosis (Fig. 8c, f). Also here the standard deviation is captured very well by the NIG model with only very small differences in the sea ice covered regions (Fig. 8d). The same applies to the skewness, where we also see some deviations in these same regions (Fig. 8e). For most regions the NIG model captures the excess kurtosis quite well (Fig. 8f). However, for regions with a high excess kurtosis in the ERA5 noise, such as the sea ice covered regions and the Gulf of Mexico, the NIG model strongly overestimates the excess kurtosis.

5 Performance CMIP6 and NIG models

In this section, we compare the noise models and the CMIP6 models with the ERA5 noise using Taylor diagrams (Fig. 9) to provide a more ~~in depth~~ in-depth discussion on the performance of the individual models. We compare how well the different models represent the standard deviation (Fig. 9a, b), the skewness (Fig. 9c, d), and the excess kurtosis (Fig. 9e, f) found in the ERA5 noise. ~~In the Taylor diagram, the standard deviation of~~ Taylor diagrams are a good tool to better understand the performance of all the different models against the observation-based noise. In a Taylor diagram, three metrics are displayed: (1) the ~~specific field (e.g. skewness), the Root Mean Square Error (RMSE; circular contours) , and~~ spatial correlation coefficient~~are displayed. These metrics are determined using weights taking into account the area of each grid cell.~~

~~For E – P, the NIG model standard deviation correlates strongly (1.0) with the ERA5 noise, and also the RMSE (0.03) is very small. Also for the skewness the spatial correlation is very high (0.82), and the RMSE is also relatively low (0.36)~~, (2) the variation in the data as represented by the standard deviation, and (3) the root mean square error between the observation-based data and the model. The spatial correlation ~~between the NIG model and ERA5 noise for excess kurtosis (0.54) is lower compared to the skewness and standard deviation, and the RMSE is 3.60. For the standard deviation, the NIG model performs better than the PC (1) and PC (N) model, that both overestimate~~ coefficient is displayed on the outer circle in the figure and the straight dotted lines in Fig. 9, connecting the origin with the outer circle are lines of constant correlation. The standard deviation is displayed on both the x-axis and y-axis. Lines of equal standard deviation are circles with their center in the origin of the

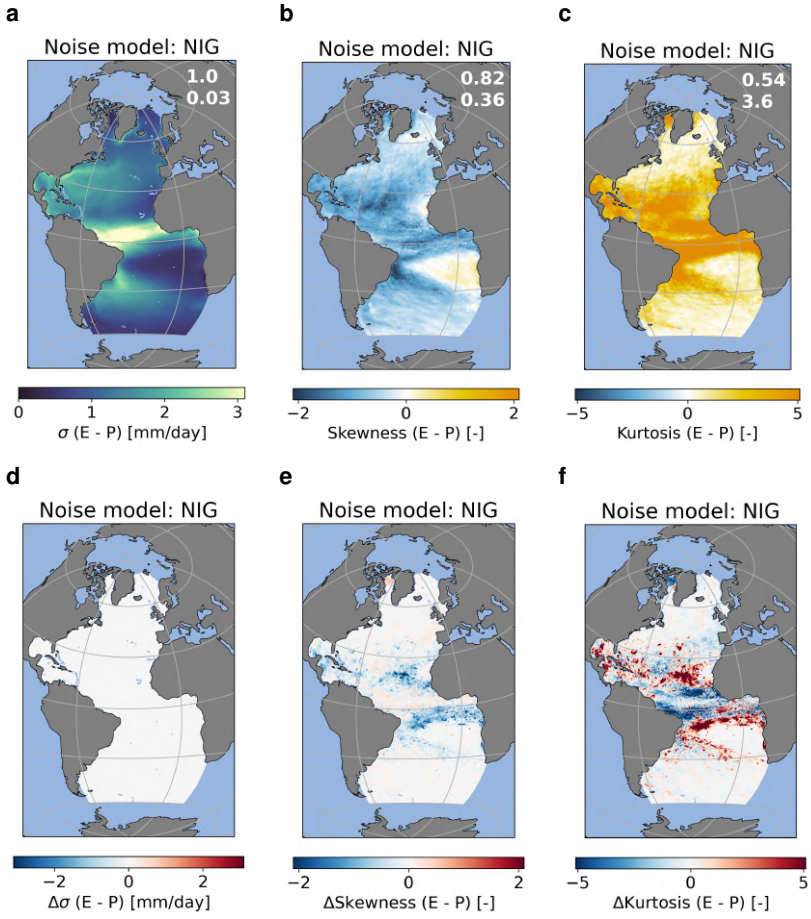

Figure 7. Standard deviation (σ), skewness and excess kurtosis of the noise from the NIG model for the ~~$E - P$~~ $E - P$ flux (a) – (c). Differences with ERA5 data (i.e. ERA5 minus NIG) are shown in (d) – (f). The statistics of the noise model are based on 5000 realisations (months). The numbers in the top right corner of (a) – (c) reflect the spatial correlation and root mean square error. Units for (a) and (d) are mm/day.

plot. The black dashed line in Fig. 9 displays the standard deviation ~~, and similar to the PC (NIG) model. For the skewness, the spatial correlation of the NIG model is slightly lower than those of the PC (1) and PC (N) models (i.e. 0.82 versus 0.93 and 0.91, respectively). The RMSE is similar for the NIG and PC (1) model but much larger for the PC (N) model (0.36 versus 0.34 and 0.74, respectively). As discussed in Section 4, the PC (NIG) model completely underestimates the skewness with a spatial correlation of -0.05 and an RMSE of 0.98. The spatial correlationis lower for the NIG model~~ in the observation-based noise. The RMSE is displayed with the black contour circles with their center in the observation-based noise marker. The location of each of the markers therefore provides information about three important metrics and therefore the performance of the individual models compared to the observation-based noise. Ideally, a model will be in the lower part of the graph,

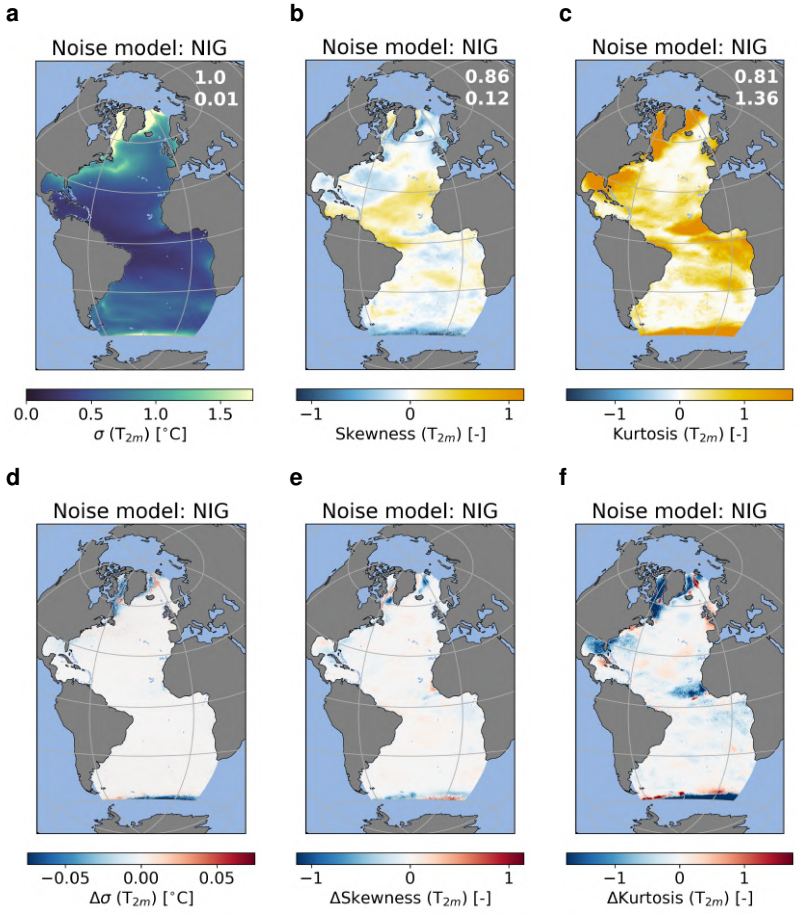

Figure 8. As Fig. 7 but for T_{2m} in °C.

since this indicates high spatial correlation, close to the black dashed lines, since this indicates similar variability compared to the ~~PC (1) and PC (N) models (0.54 versus 0.84 and 0.82, respectively). For the RMSE, the NIG model is outperformed by the~~ observation-based noise, and by combining these two the RMSE will consequently also be low. All three metrics are determined using weights considering the area of each grid cell.

For the $E - P$ noise models, the PC (1) model ~~, but outperforms~~ performs best for the standard deviation, skewness and excess kurtosis and the PC (~~N) model (3.60 versus 2.82 and 4.45, respectively), and again the PC (NIG) model fails to capture the excess kurtosis in the noise (spatial correlationof 0.12 and RMSE of 5.07)~~NIG) the worst (Fig. 9 a, c and e). The PC (N) model performs equally well for the spatial correlation, but strongly underestimates the variability in skewness and excess kurtosis. The NIG model has a lower spatial correlation, but is much better in capturing the variability in all three statistical moments. All models have trouble representing the excess kurtosis in the latitudinal bands 10°S to 25°N. The PC (1) and PC (N) models overestimate the excess kurtosis in almost the entire region, whereas the NIG model underestimates the excess

kurtosis over the ITCZ region ~~,~~ and overestimates it in the other regions. This is because this region can experience very extreme rainfall episodes with a very low number of occurrences which severely affects the excess kurtosis diagnosed from the ERA5 noise ~~.~~ as was also found with the clustering analysis (Fig. 2). Because these episodes only occur a few times in the time series, these are not represented well by the NIG model, and are also difficult to represent in the PC (1) model.

We can explain the failure of the PC (NIG) model to accurately resemble the ~~observed~~ observation-based skewness and excess kurtosis by the Central Limit Theorem. This theorem states that when summing over random variables, the distribution of this sum converges towards a Gaussian distribution, which, by definition, has zero skewness and excess kurtosis. What we do in these PC-models is that we sample values from the PCs, multiply those with the EOFs and sum these, which, following Central Limit Theorem, converges towards a Gaussian distribution. The same applies to the PC (N) model which performs well for spatial correlation skill, but (based on a timeseries of 10,000 realisations) underestimates the amplitude of the skewness and excess kurtosis. This underestimation increases when longer timeseries are used, and the model slowly converges to ~~one of Gaussian noise~~ a Gaussian one. Methods based on a PCA, except for the PC (1) model, will therefore be unable to represent the skewness and excess kurtosis in the ~~observed noise. Lastly, when a longer timeseries is used to determine the statistics, e.g. 20, 000 realizations instead of 10, 000 as used now,~~ observation-based noise. An alternative explanation as to why the PC-based models fail to capture the skewness and excess kurtosis is that the PCs might be (non-linearly) dependent on each other. To test this, we have calculated the distance correlation between the PCs, including whether the distance correlation is significant (p-value < 0.05) based on a permutation test of n = 1000 (Fig. A17). For both the $E-P$ and T_{2m} PCs, around 5% of the possible PC combinations experiences a significant dependence. However, the strongest distance correlation is only 0.14 for the PCs corresponding to the ~~statistics of the PC (1) model converge toward those of the observed noise~~ $E-P$ noise, and 0.11 for the PCs corresponding to the T_{2m} noise, meaning there is at best a very weak dependence between the PCs. We therefore do not expect that the weak non-linear dependence between some of the PCs is the reason why the PC-based models fail, but that the explanation mentioned before, i.e. the Central Limit Theorem, is the main reason.

For T_{2m}, Fig. 9 b, d and f show that the NIG and PC (1) models perform consistently best. All models capture the spatial pattern in the standard deviation of the noise as shown by the near unity spatial correlation coefficient, however, the ~~NIG model captures the standard deviation in the ERA5 very well with almost perfect metrics. The skewness is captured relatively well with a high spatial correlation (0.87) and low RMSE (0.12) . Compared to $E-P$, the excess kurtosis of the T_{2m} noise is better captured by the NIG model with a spatial correlation of 0.82, but a relatively high RMSE of 1.32. The NIG model performs better than the~~ PC (~~1~~N) and PC (~~N) model for the standard deviation and similar to the PC (~~NIG) model. For the skewness the spatial correlation is very similar to the~~ models both overestimate the variability in the standard deviation of the noise as shown by the high RMSE and larger standard deviation (Fig. 9b). The spatial pattern of the skewness is captured reasonably well by the NIG, PC (1) and PC (N) models~~(i.e. 0.87 versus 0.90, 0.86, respectively), and the RMSE is slightly lower (0.12 versus 0.13 and 0.17, respectively) . For the excess kurtosis the spatial correlation is also higher for the NIG model~~ , but not by the PC (NIG) model (Fig. 9d). The PC (N) shows a stronger underestimation of the variability compared to the PC (1) and ~~PC (N) models (0.82 versus 0.85 and 0.83, respectively). The RMSE is much higher though (1.32 versus 0.58 and 0.84, respectively~~NIG model. For the excess kurtosis a similar conclusion can be drawn, except that the NIG model strongly

overestimates the variability (Fig. 9f). The worse performance for excess kurtosis can be explained by the overestimation of the sea ice covered regions and the Gulf of Mexico by the NIG model. ~~Both the skewness and excess kurtosis are very poorly represented in the PC (NIG) model with a spatial correlation of -0.01 for the skewness and 0.28 for the excess kurtosis. The RMSEs are 0.23 for the skewness and 1.05 for the excess kurtosis and therefore relatively high compared to the other PC models. In this region~~ In these regions, the distribution of the ERA5 noise has a relatively broad, flat peak or sometimes a slightly bimodal peak. This is the reason the NIG fit does not perform very well in these regions. ~~In the Southern Ocean the estimates for excess kurtosis in the NIG model are off because of relatively extreme cold episodes in the ERA5 noise.~~ Similar as to the $E \breve{} P$ noise, the PC (N) and PC (NIG) models are unable to explain the variability in the skewness and excess kurtosis as explained ~~by Central Limit Theorem~~above.

~~For $E - P$~~As discussed in Section 3.2, the CMIP6 MMM captures the ~~standard deviation of the~~ ERA5 $E - P$ noise reasonably well, though the ~~MMM underestimates the standard deviation in large parts of the basin. The spatial correlation is high (0.86), and the RMSE is low (0.44). The performance for the skewness is less with a lower spatial correlation (0.61), and the RMSE is 0.51. Just as with the NIG model, the excess kurtosis is also the statistic that is not captured well. The spatial correlation is 0.45, and the RMSE 4.04. The strongest biases are found~~ performance decreases for the higher statistical moments. This is likely related to strong biases over the South Equatorial Current where the skewness is too negative in the CMIP6 MMM, and the excess kurtosis too positive compared to the ERA5 noise. This is potentially related to the double ITCZ bias present in most CMIP6 models (Tian and Dong, 2020). The latitudinal extent of the ITCZ is too southward in many models, which also causes a shift in the higher order statistical moments in this region resulting in relatively large biases. ~~The~~ From Fig. 9 we see that the individual models that consistently perform the best are CESM2-WACCM (30), CESM2 (31) and NorESM2-MM (32) (except for excess kurtosis where NorESM2-MM has quite a large RMSE). What these models have in common is that their atmospheric model is the Community ~~Atmopshere~~ Atmosphere Model 6 (CAM6), or in the case of CESM2-WACCM based on CAM6 and run on a nominal 1° horizontal resolution. This suggests that this atmospheric model is able to capture the ~~observed~~ observation-based noise reasonably well.

Liu et al. (2022) also found that these models are performing relatively well for precipitation biases which they suggest is due to the specific two-moment prognostic cloud microphysics scheme (Gettelman and Morrison, 2015) used in CAM6. TaiESM1, which uses CAM5 and an earlier version of the prognostic cloud microphysics scheme also performs relatively well. There are also two other CESM2 models that use a form of CAM6, i.e. CESM2 – WACCM – FV2 (27) and CESM2 – FV2 (29). These models perform less well as the other three, which might be explained by the fact that these models are run on a lower (i.e. 2°) resolution. The CMIP6 MMM has the same biases in the latitudinal band between 10°S and 25°N, though less strong in some regions. This is probably because the high rainfall episodes in the ERA5 data are smoothed when regridded to a 1° grid, which is done before comparing it to the CMIP6 models and MMM.

~~The~~ For T_{2m} the CMIP6 MMM ~~is able to capture the standard deviation in the~~ also performs reasonably well compared to the ERA5 ~~T_{2m} well with a high spatial correlation (0.96), and low RMSE (0.31). Also the skewness is captured relatively well with a spatial correlation of 0.72, and RMSE of 0.17, and the same applies to the excess kurtosis with a spatial correlation of 0.80, and RMSE of 0.61~~noise, and just as for the $E - P$ noise, performance is lower for higher statistical moments. The

strongest biases (both positive and negative) for the excess kurtosis are found over the sea ice covered regions. This might be related to biases in sea-ice cover in the CMIP6 models (Watts et al., 2021). For the individual models it is more difficult to point towards consistently well performing models. The UKESM1-0-LL (22) model simulations performed by the MOHC shows the most consistency. Other models that perform relatively well in 2 out of 3 statistical moments are CESM2-FV2 (29) and CAS-ESM2-0 (8). Interestingly, the UKESM1-0-LL (33) simulations performed by the NIMS-KMA are among the worst performing models. The only difference between the two models is the computer on which the model is run on, and the initial conditions. This suggests that there is also a ~~depedency~~ dependency on initial conditions in the performance of the CMIP6 models.

Except for the excess kurtosis in the T_{2m} noise, the NIG model outperforms the individual CMIP6 models and MMM which is due to the overestimation of the excess kurtosis over sea ice covered regions by the NIG model. The PC (NIG) model only outperforms the CMIP6 MMM for the standard deviation and is very poor for the skewness and excess kurtosis. The PC (1) model outperforms the CMIP6 models and MMM for the skewness and excess kurtosis. The PC (N) model ~~outerpeforms~~ outperforms the CMIP6 MMM for all moments with respect to the spatial correlation ~~,~~and is very similar to the CMIP6 MMM and the best CMIP6 models for RMSE. This means that we can capture 'realistic' noise better with the statistical model than the fully coupled Earth System Models. Among the PC-based models, the PC (1) model performs best and similar to the NIG model, but this model is not fully stochastic as the other noise models.

6 Summary and discussion

In this study we have analysed ERA5 evaporation minus precipitation (~~$E-P$~~ $E-P$) and 2 m air temperature (T_{2m}) fields to determine what ~~observed~~ observation-based noise is in these variables. We find that due to nonzero skewness and excess kurtosis, the noise in both variables typically cannot be classified as white and studies that assume white noise in either of the two variables might not resolve the response of the ocean to atmospheric noise realistically. We have analysed the noise in 36 different CMIP6 Earth System Models and the CMIP6 ~~multi model~~ multi-model mean (MMM) and compared those to the ERA5 noise. There is quite a spread in the performance of the CMIP6 models, but the MMM is performing relatively well compared to the individual models. Typically, the models perform best for the standard deviation and worst for the excess kurtosis. Furthermore, we have fitted a Normal Inverse Gaussian (NIG) distribution to the ERA5 noise of both variables. This results in a stochastic noise model that can be used as input in Ocean General Circulation Models (OGCMs). We have shown that the NIG model captures the standard deviation, skewness and excess kurtosis of the ERA5 noise reasonably well in both the ~~$E-P$~~ $E-P$ and T_{2m} except for the excess kurtosis in the T_{2m} noise where the NIG model strongly overestimates the positive excess kurtosis in sea ice covered regions. For most metrics and statistics, the NIG model performs better than the individual CMIP6 models and CMIP6 MMM.

Previous studies have looked into biases in CMIP6 models. However, these studies typically look into the biases in the mean state or the seasonality of the variables. Here, we have specifically looked at variability up to interannual timescales and specifically the distribution and related metrics (i.e. standard deviation, skewness and excess kurtosis). We found that biases in

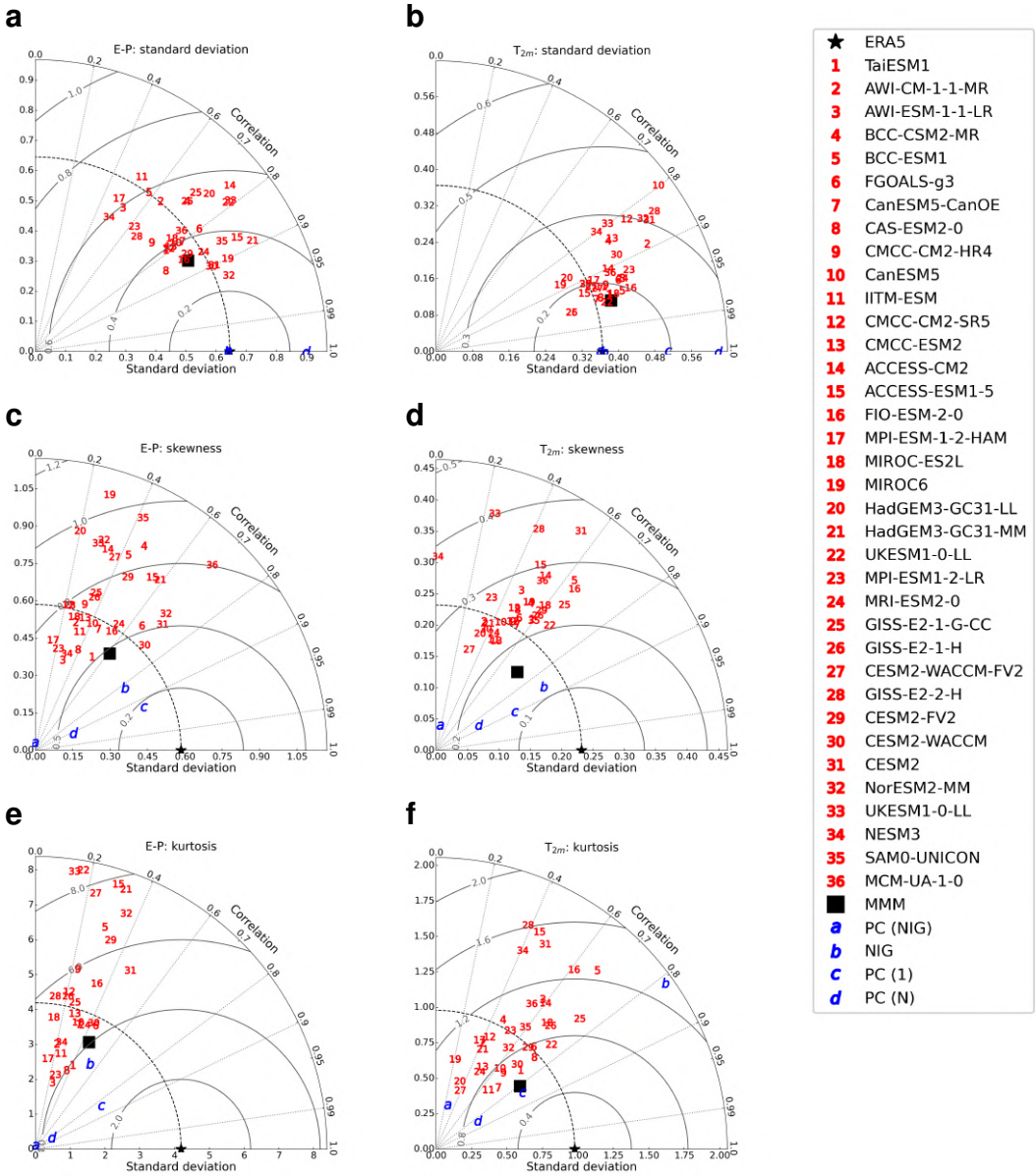

Figure 9. Taylor diagrams for statistics of the noise. (a) and (b) standard deviation. (c) and (d) skewness, and (e) and (f) excess kurtosis. (a), (c) and (e) are for the E – P noise, and (b), (d) and (f) for the noise in the T_{2m}. The star refers to the ERA5 data, the different red numbers refer to the different CMIP6 models, the blue letters to the noise models, and the square black marker represents the CMIP6 MMM. Note that UKESM1-0-LL number 22 is performed by the MOHC and number 33 by NIMS-KMA. Units for standard deviation in (a) are mm/day, and °C in (b).

these quantities are still to some extent connected to biases in the mean state. For example, the biases in skewness and excess kurtosis in the ~~$E - P$~~ $E - P$ noise in the South Atlantic are for example likely to be related to the double ITCZ bias described in earlier studies (Tian and Dong, 2020; Li et al., 2020). Differences in the excess kurtosis in sea ice covered regions can also be related to the biases in Arctic sea-ice thickness and cover (Watts et al., 2021).

In the development of a noise model the best variant turned out to be a point wise statistical fit of a Normal Inverse Gaussian (NIG) distribution. As shown in Section 3, the model performs relatively well in most grid points, but can still deviate quite a bit for especially the excess kurtosis. One major drawback of fitting a statistical distribution point wise to the data is that for the individual noise fields (i.e. one random realisation) we lose spatially coherent structures, and potentially auto-correlation in the noise. We have constructed alternative models based on a ~~principle~~ principal component analysis (PCA) where the corresponding Empirical Orthogonal Functions (EOFs) ~~captured~~ capture the spatial structures. However, these models underestimate the skewness and excess kurtosis in the noise fields because of the Central Limit Theorem, or (for the PC (1) model~~,~~) are not fully stochastic. Therefore, we eventually decided to fit a model to the data that can relatively accurately represent the standard deviation, skewness and excess kurtosis in the ERA5 noise. However, when the spatially coherent structures captured by the EOFs are deemed more important than an accurate representation of the skewness and kurtosis of the noise, PC-based models can be used. The loss of spatially coherent structures can be important when studying noise-induced transitions of the AMOC. Noise that is spatially coherent influence larger areas of ocean. This could, for example, mean that a freshening of the surface ocean could happen over a larger area of the ocean and therefore might be more efficient in inhibiting deep convection in the North Atlantic. Whether it is actually important should be tested in an ocean model but this is outside the scope of this study.

Similar studies that look into the characteristics of $E - P$and T_{2m} noise are sparse. In Sura and Sardeshmukh (2008), they investigate the non-Gaussianity of daily SST variability. The timescales assessed in Sura and Sardeshmukh (2008) are faster (i.e. daily versus monthly), and they look at SSTs, whereas we look at air temperatures. However, relatively similar results are achieved in our study compared to Sura and Sardeshmukh (2008). Skewness in daily SST variability is typically negative in the Atlantic Ocean, whereas the excess kurtosis is mostly positive, similar to what we find for the air temperature. They relate this to multiplicative noise in mixed layer dynamics. However, they make the assumption that daily fluctuations in air temperature are Gaussian. Our study shows, that at least on monthly timescales, this is not the case over most of the ocean. Whether the multiplicative noise signal we find in the T_{2m} noise originates from SST variability or atmospheric dynamics, or a combination of the two, is left for further study.

To conclude, we have provided an analysis of ~~observed~~ observation-based noise from ERA5 reanalysis data. Based on this realistic noise we have constructed a noise model based on a Normal Inverse Gaussian distribution fit to the ERA5 noise. This product is made publicly available in the repository related to this paper ~~(citation)~~(Boot and Dijkstra, 2024). The noise model can, for example, be used as a forcing on ocean models to ~~for example study noise induced~~ study noise-induced transitions of the AMOC under 'realistic' noise forcing.

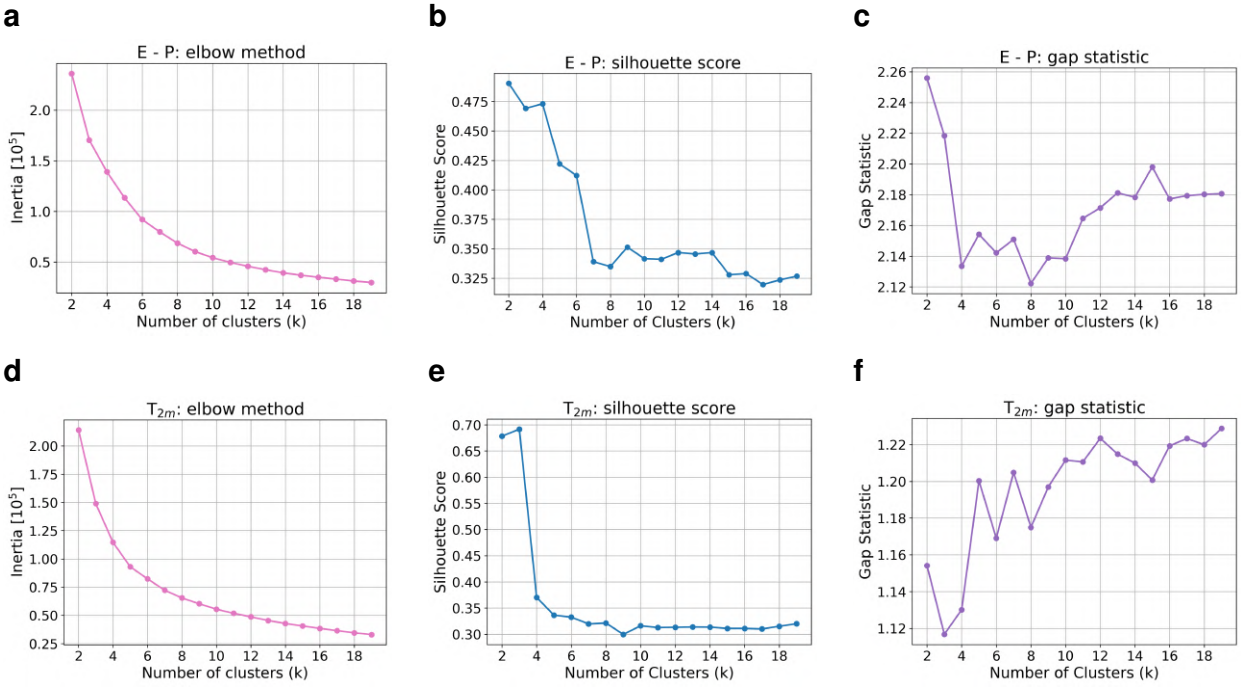

Figure A1. ~~Standard deviation (σ), skewness and excess kurtosis of~~ Metrics for the ~~noise from~~ k-means clustering method versus the ~~PC (1) model~~ number of clusters for the ~~$E - P$ flux~~ $E - P$ clusters (~~a~~a-c) ~~– (c)~~and the T_{2m} clusters. ~~Differences with ERA5 data~~ (~~i.e. ERA5 minus PC (1~~a) ~~) are shown in~~ and (d) ~~– (f). The statistics of~~ represent the ~~noise model are based on 10000 realisations~~ elbow method, (~~months~~b) ~~. The numbers in the top right corner of~~ and (~~a~~e) ~~– (c) reflect~~ represent the ~~spatial correlation~~ silhouette score, and ~~root mean square error. Units for~~ (~~a~~c) and (~~d~~f) ~~are mm/day~~the gap statistic.

Code and data availability. ERA5 data can be downloaded from the Copernicus Climate Data Store (CDS). CMIP6 data can be downloaded from the Earth System Grid Federation (ESGF) or using the scripts in the repository (Boot and Dijkstra, 2024). Directions on which exact data needs to be downloaded and all scripts used for analyses and making the figures can be found at Boot and Dijkstra (2024). Here also a script that contains the noise models can be found.

Author contributions. AAB and HAD conceptualized the study. AAB acquired the results. Both authors contributed to writing the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

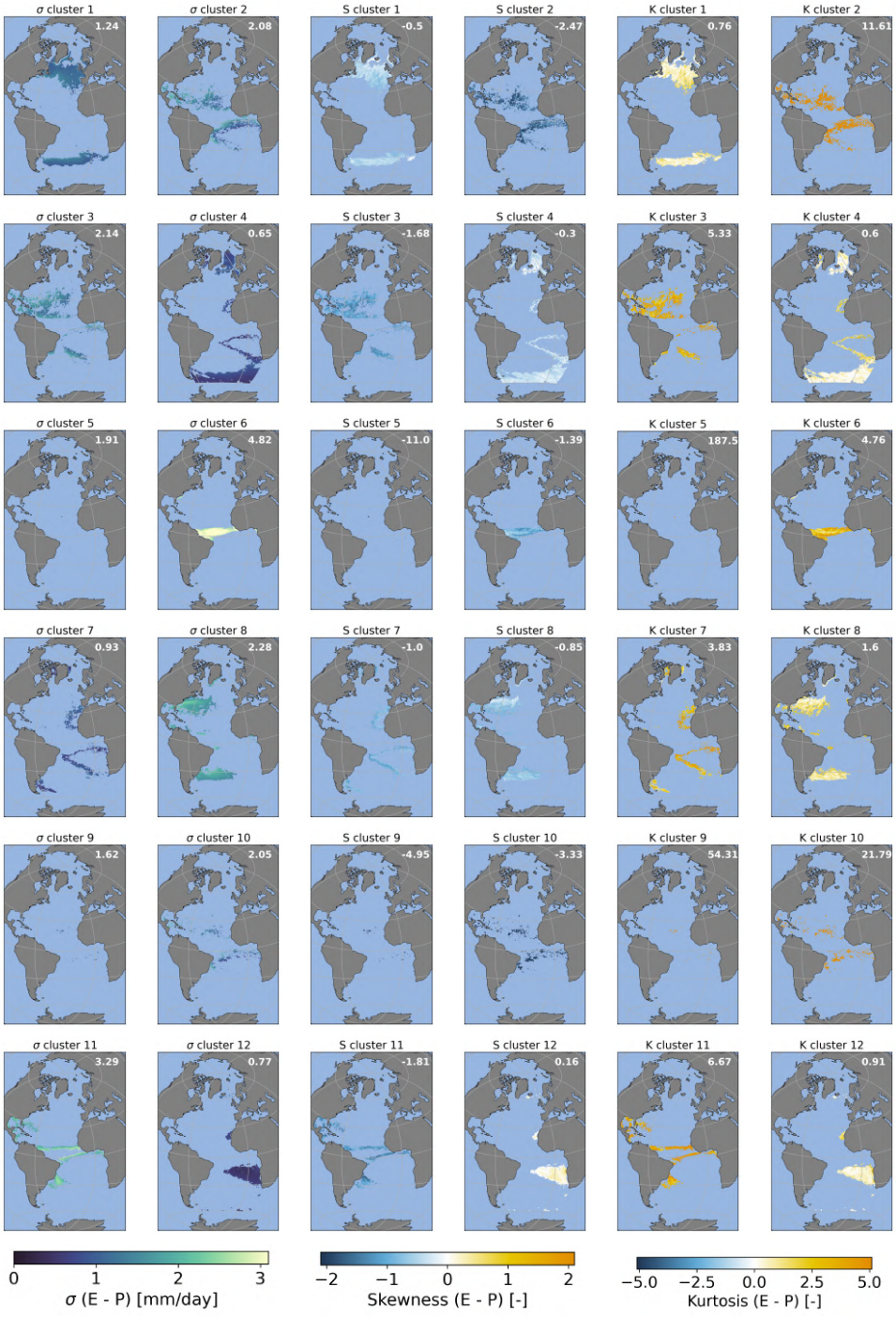

Figure A2. The 12 clusters for the $E - P$ noise fields. The columns 1 and 2 correspond to the standard deviation of the clusters, columns 3 and 4 represent the skewness, and columns 5 and 6 excess kurtosis. Numbers in the top right represent area weighted mean of the metric.

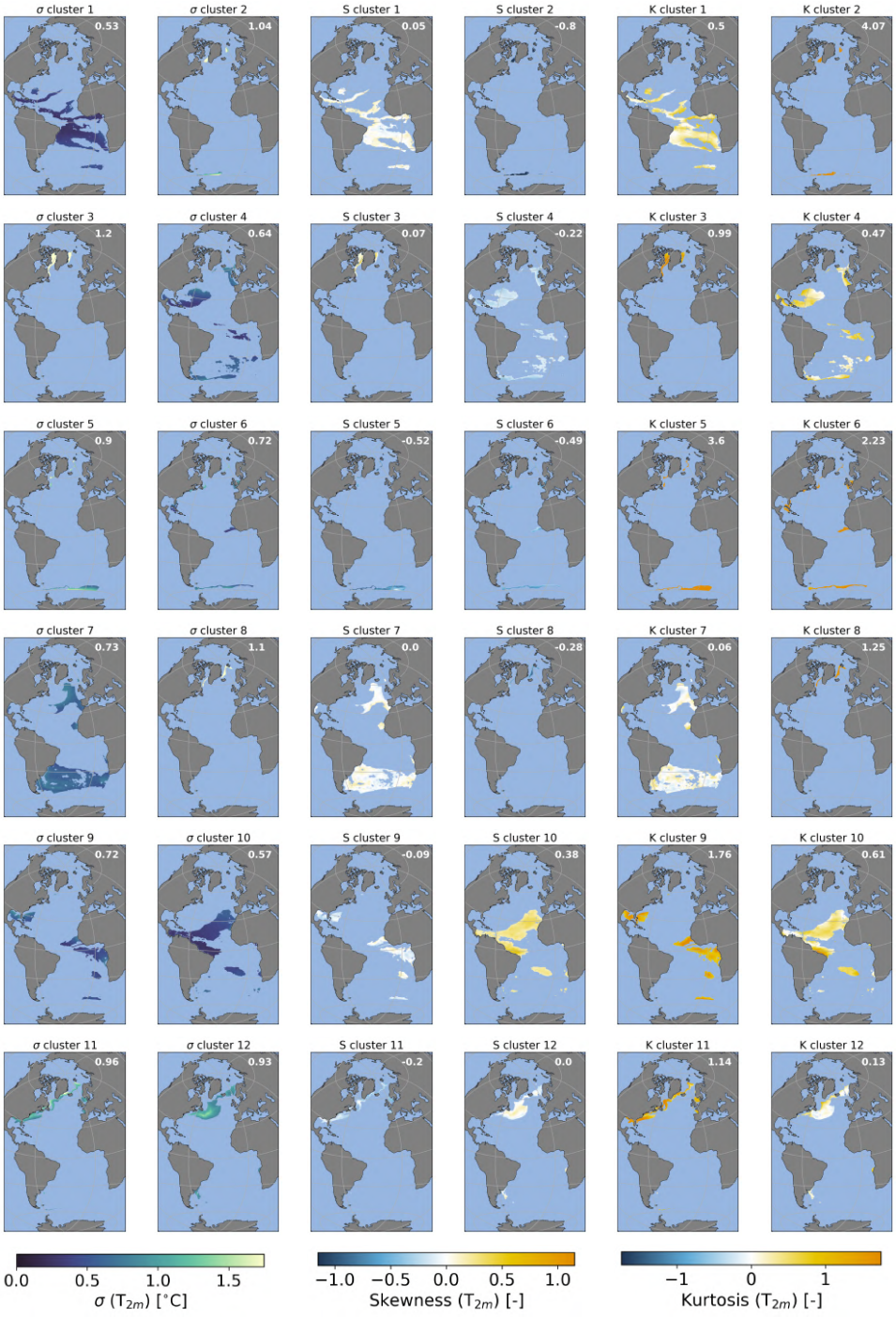

Figure A3. As Fig. A2 but for the T_{2m} clusters.

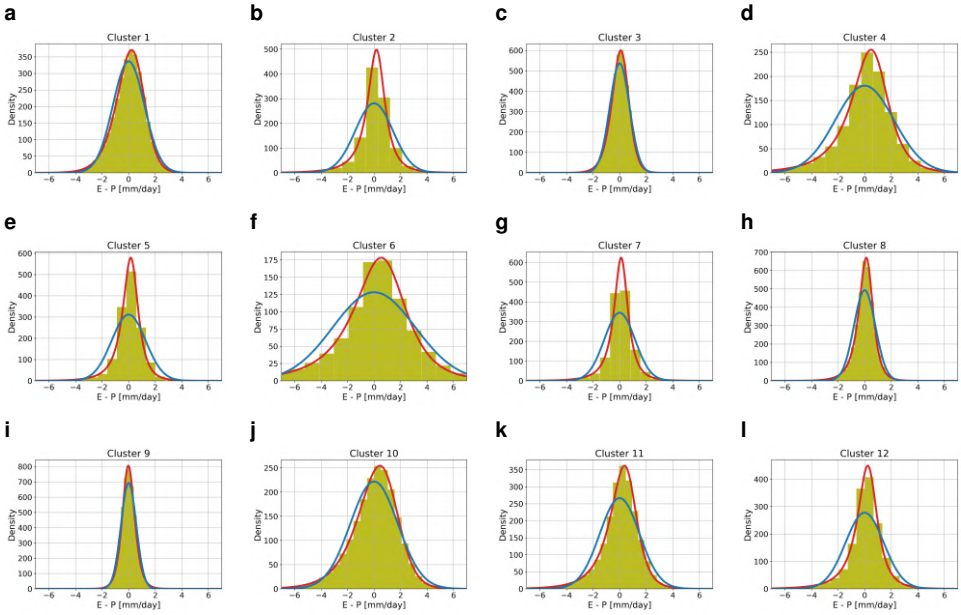

Figure A4. Probability density functions for the 12 clusters for the $E - P$ noise. Red lines represent a Normal Inverse Gaussian fit, blue lines a Gaussian fit, and the yellow histrogram the data (using 50 bins). The y-axis shows the density, and the x-axis the $E - P$ noise in mm/day.

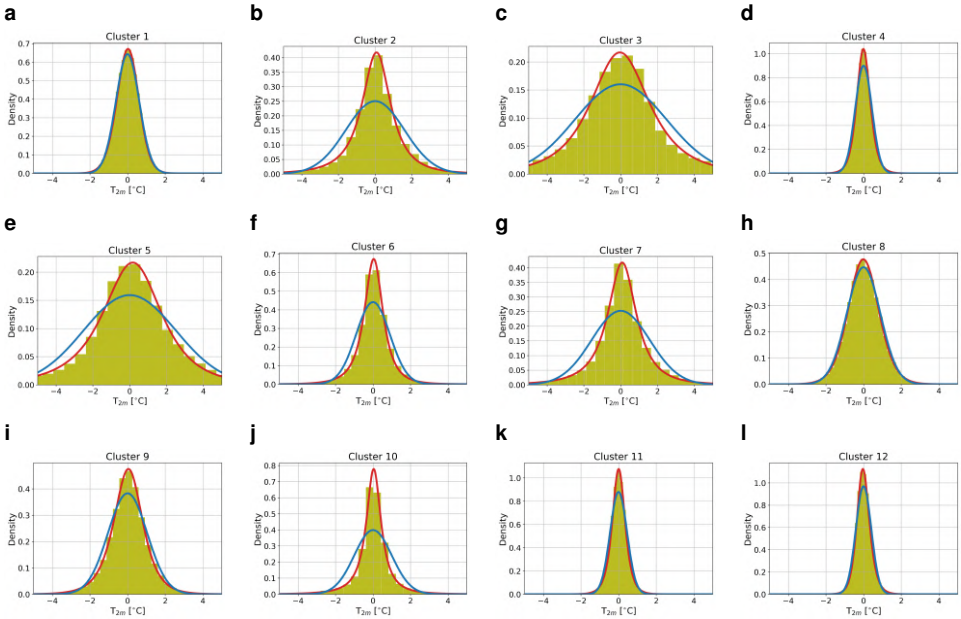

Figure A5. As Fig. A4 but for the T_{2m} clusters in °C instead of mm/day.

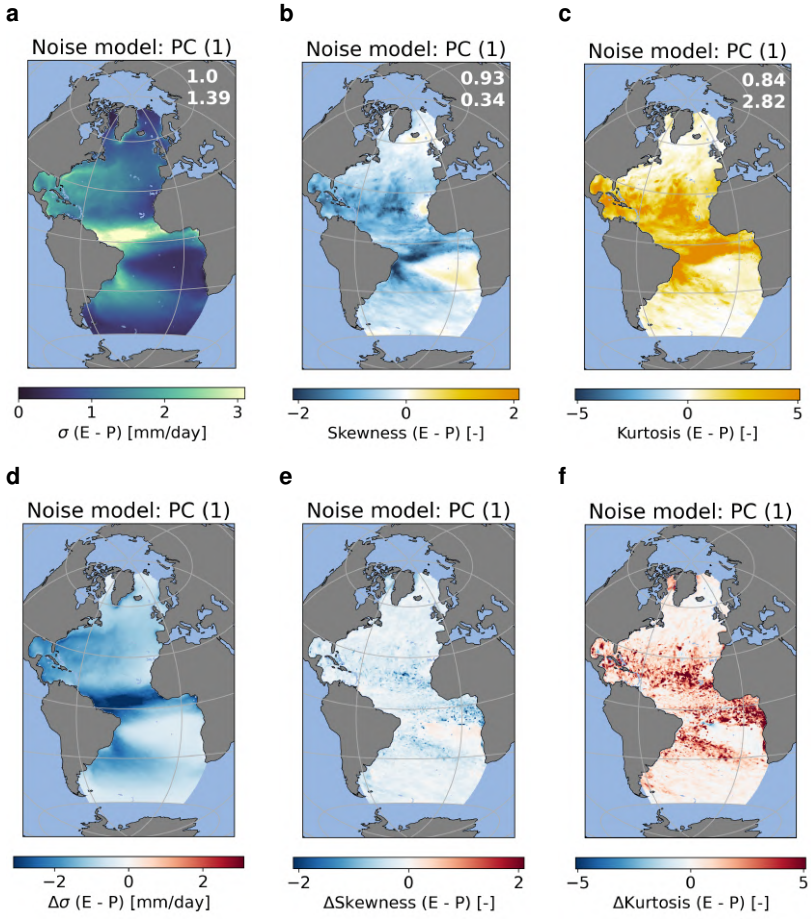

Figure A6. Standard deviation (σ), skewness and excess kurtosis of the noise from the PC (1) model for the $E - P$ flux (a) – (c). Differences with ERA5 data (i.e. ERA5 minus PC (1)) are shown in (d) – (f). The statistics of the noise model are based on 10000 realisations (months). The numbers in the top right corner of (a) – (c) reflect the spatial correlation and root mean square error. Units for (a) and (d) are mm/day.

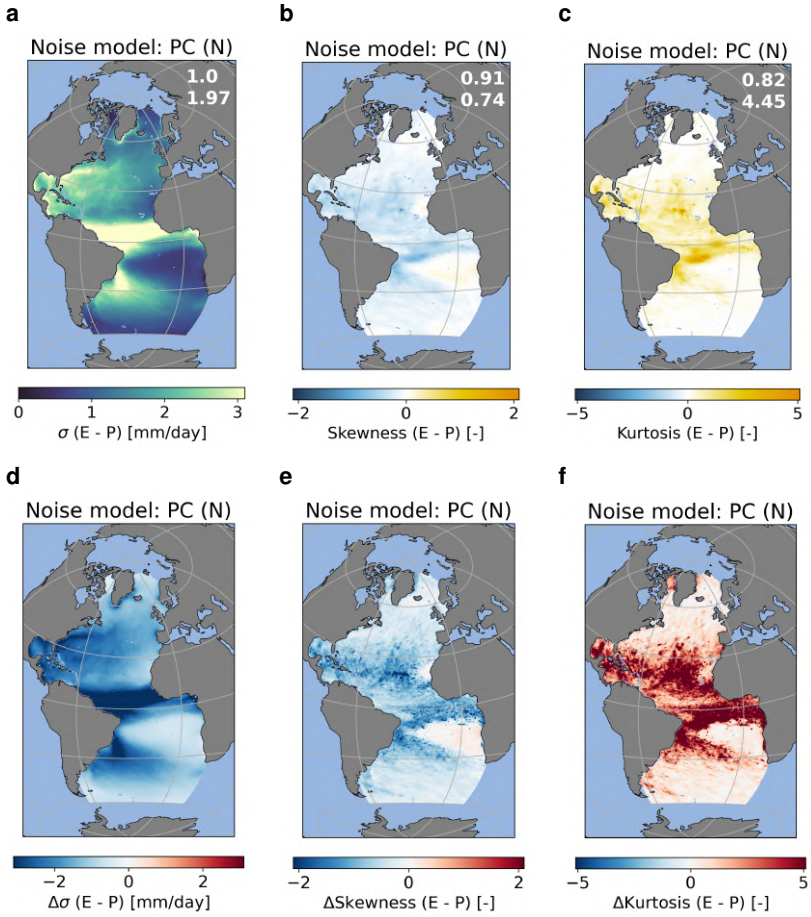

Figure A7. As Fig. A6 but for T_{2m}.

Financial support. This research has been supported by the European Research Council through the ERC-AdG project TAOC (PI: Dijkstra, project 101055096).

.

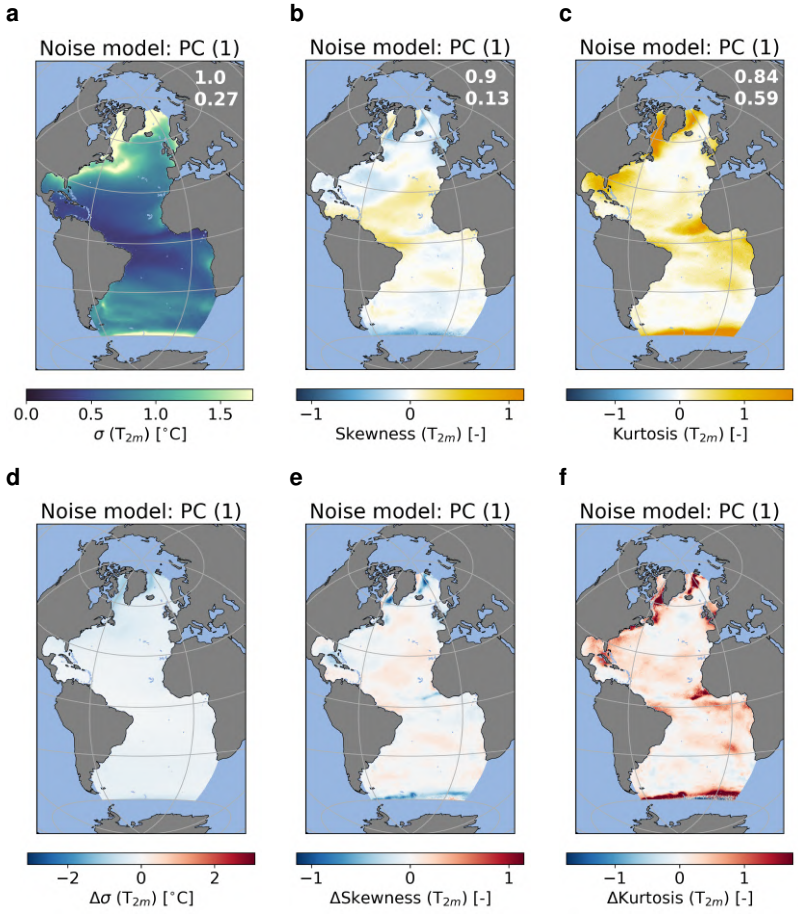

Figure A8. Standard deviation (σ), skewness and excess kurtosis of the noise from the PC (N) model for the ~~$E - P$~~ $E - P$ flux (a) – (c). Differences with ERA5 data (i.e. ERA5 minus PC (N)) are shown in (d) – (f). The statistics of the noise model are based on 10000 realisations (months). The numbers in the top right corner of (a) – (c) reflect the spatial correlation and root mean square error. Units for (a) and (d) are mm/day.

References

Bentsen, M., Oliviè, D. J. L., ?yvind Seland, Toniazzo, T., Gjermundsen, A., Graff, L. S., Debernard, J. B., Gupta, A. K., He, Y., Kirkevag, A., Schwinger, J., Tjiputra, J., Aas, K. S., Bethke, I., Fan, Y., Griesfeller, J., Grini, A., Guo, C., Ilicak, M., Karset, I. H. H., Landgren, O. A., Liakka, J., Moseid, K. O., Nummelin, A., Spensberger, C., Tang, H., Zhang, Z., Heinze, C., Iversen, T., and Schulz, M.: NCC NorESM2-MM model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.8040, 2019.

Boot, A. A. and Dijkstra, H. A.: ESD$_noise_2$024, https://doi.org/10.5281/$zenodo$.13148972, 2024.

Boot, A. A., Steenbeek, J. G., Coll, M., von der Heydt, A. S., and Dijkstra, H. A.: Global marine ecosystem response to a strong AMOC weakening under low and high future emission scenarios, https://doi.org/10.22541/essoar.171319366.64840276/v1, 2024a.

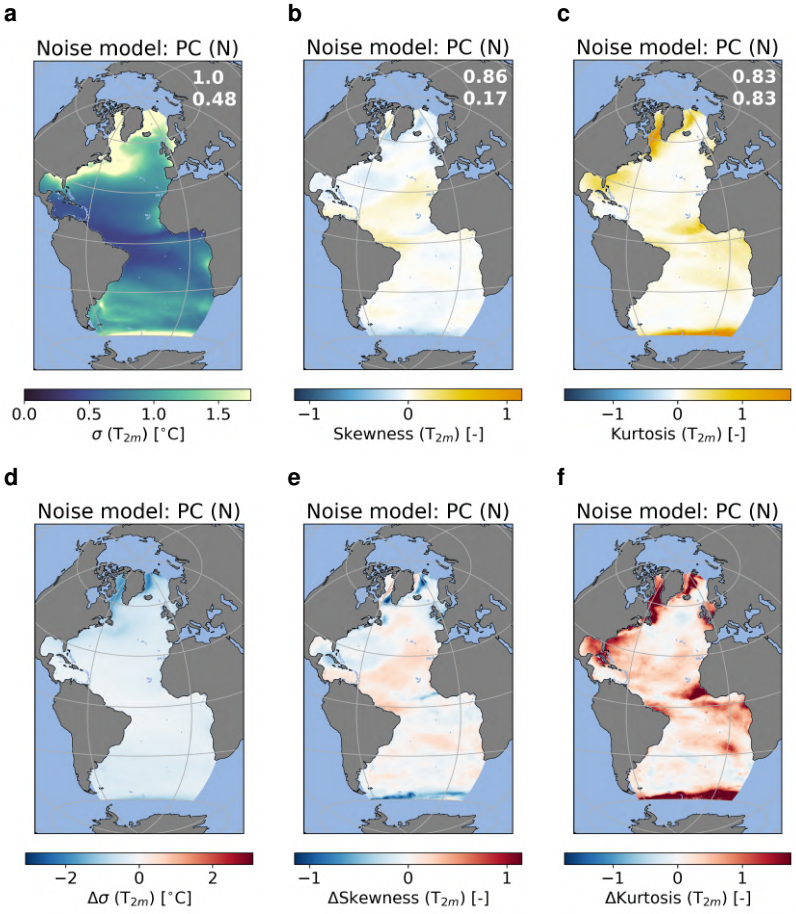

Figure A9. As Fig. A8 but for T_{2m}.

Boot, A. A., von der Heydt, A. S., and Dijkstra, H. A.: Response of atmospheric pCO_2 to a strong AMOC weakening under low and high emission scenarios, Climate Dynamics, https://doi.org/10.1007/s00382-024-07295-y, 2024b.

Byun, Y.-H.: NIMS-KMA UKESM1.0-LL model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.8379, 2020.

Cao, J. and Wang, B.: NUIST NESMv3 model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.8769, 2019.

Castellana, D., Baars, S., Wubs, F. W., and Dijkstra, H. A.: Transition Probabilities of Noise-induced Transitions of the Atlantic Ocean Circulation, Scientific Reports, 9, 20 284, https://doi.org/10.1038/s41598-019-56435-6, 2019.

Chai, Z.: CAS CAS-ESM1.0 model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.3353, 2020.

Choudhury, A. D., Raghavan, K., Gopinathan, P. A., Narayanasetti, S., Singh, M., Panickal, S., and Modi, A.: CCCR-IITM IITM-ESM model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.3708, 2019.

Cini, M., Zappa, G., Ragone, F., and Corti, S.: Simulating AMOC tipping driven by internal climate variability with a rare event algorithm, npj Climate and Atmospheric Science, 7, 31, https://doi.org/10.1038/s41612-024-00568-7, 2024.

Danabasoglu, G.: NCAR CESM2 model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.7627, 2019a.

Danabasoglu, G.: NCAR CESM2-WACCM-FV2 model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.11298, 2019b.

Danabasoglu, G.: NCAR CESM2-WACCM model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.10071, 2019c.

Danabasoglu, G.: NCAR CESM2-FV2 model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.11297, 2019d.

Danek, C., Shi, X., Stepanek, C., Yang, H., Barbi, D., Hegewald, J., and Lohmann, G.: AWI AWI-ESM1.1LR model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.9328, 2020.

Dix, M., Bi, D., Dobrohotoff, P., Fiedler, R., Harman, I., Law, R., Mackallah, C., Marsland, S., O'Farrell, S., Rashid, H., Srbinovsky, J., Sullivan, A., Trenham, C., Vohralik, P., Watterson, I., Williams, G., Woodhouse, M., Bodman, R., Dias, F. B., Domingues, C. M., Hannah, N., Heerdegen, A., Savita, A., Wales, S., Allen, C., Druken, K., Evans, B., Richards, C., Ridzwan, S. M., Roberts, D., Smillie, J., Snow, K., Ward, M., and Yang, R.: CSIRO-ARCCSS ACCESS-CM2 model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.4271, 2019.

Eyring, V., Gillett, N., Rao, K. A., Barimalala, R., Parrillo, M. B., Bellouin, N., Cassou, C., Durack, P., Kosaka, Y., McGregor, S., Min, S., Morgenstern, O., and Sun, Y.: Human Influence on the Climate System, pp. 423–552, Cambridge University Press, https://doi.org/10.1017/9781009157896.005, 2023.

Gettelman, A. and Morrison, H.: Advanced Two-Moment Bulk Microphysics for Global Models. Part I: Off-Line Tests and Comparison with Other Schemes, Journal of Climate, 28, 1268 – 1287, https://doi.org/10.1175/JCLI-D-14-00102.1, 2015.

Hajima, T., Abe, M., Arakawa, O., Suzuki, T., Komuro, Y., Ogura, T., Ogochi, K., Watanabe, M., Yamamoto, A., Tatebe, H., Noguchi, M. A., Ohgaito, R., Ito, A., Yamazaki, D., Ito, A., Takata, K., Watanabe, S., Kawamiya, M., and Tachiiri, K.: MIROC MIROC-ES2L model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.5602, 2019.

Hasselmann, K.: Stochastic climate models Part I. Theory, Tellus, 28, 473–485, https://doi.org/https://doi.org/10.1111/j.2153-3490.1976.tb00696.x, 1976.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G. D., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/https://doi.org/10.1002/qj.3803, 2020.

Lee, W.-L. and Liang, H.-C.: AS-RCEC TaiESM1.0 model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.9755, 2020.

Lenton, T. M., Held, H., Kriegler, E., Hall, J. W., Lucht, W., Rahmstorf, S., and Schellnhuber, H. J.: Tipping elements in the Earth's climate system, Proceedings of the National Academy of Sciences, 105, 1786–1793, https://doi.org/10.1073/pnas.0705414105, 2008.

Li, J.-L. F., Xu, K.-M., Richardson, M., Lee, W.-L., Jiang, J. H., Yu, J.-Y., Wang, Y.-H., Fetzer, E., Wang, L.-C., Stephens, G., and Liang, H.-C.: Annual and seasonal mean tropical and subtropical precipitation bias in CMIP5 and CMIP6 models, Environmental Research Letters, 15, 124 068, https://doi.org/10.1088/1748-9326/abc7dd, 2020.

Li, L.: CAS FGOALS-g3 model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.3356, 2019.

Liu, Y., Cheng, L., Pan, Y., Tan, Z., Abraham, J., Zhang, B., Zhu, J., and Song, J.: How Well Do CMIP6 and CMIP5 Models Simulate the Climatological Seasonal Variations in Ocean Salinity?, Advances in Atmospheric Sciences, 39, 1650–1672, https://doi.org/10.1007/s00376-022-1381-2, 2022.

Lovato, T. and Peano, D.: CMCC CMCC-CM2-SR5 model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.3825, 2020.

Lovato, T., Peano, D., and Butenschön, M.: CMCC CMCC-ESM2 model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.13195, 2021.

McKay, D. I. A., Staal, A., Abrams, J. F., Winkelmann, R., Sakschewski, B., Loriani, S., Fetzer, I., Cornell, S. E., Rockström, J., and Lenton, T. M.: Exceeding 1.5°C global warming could trigger multiple climate tipping points, Science, 377, eabn7950, https://doi.org/10.1126/science.abn7950, 2022.

Monahan, A. H.: A Simple Model for the Skewness of Global Sea Surface Winds, Journal of the Atmospheric Sciences, 61, 2037 – 2049, https://doi.org/10.1175/1520-0469(2004)061<2037:ASMFTS>2.0.CO;2, 2004.

Monahan, A. H.: Temporal Filtering Enhances the Skewness of Sea Surface Winds, Journal of Climate, 31, 5695 – 5706, https://doi.org/10.1175/JCLI-D-17-0814.1, 2018.

NASA/GISS): NASA-GISS GISS-E2-1-G-CC model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.11762, 2019.

NASA/GISS: NASA-GISS GISS-E2.1H model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.7128, 2019a.

NASA/GISS: NASA-GISS GISS-E2.2H model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.15871, 2019b.

Neubauer, D., Ferrachat, S., Drian, C. S.-L., Stoll, J., Folini, D. S., Tegen, I., Wieners, K.-H., Mauritsen, T., Stemmler, I., Barthel, S., Bey, I., Daskalakis, N., Heinold, B., Kokkola, H., Partridge, D., Rast, S., Schmidt, H., Schutgens, N., Stanelle, T., Stier, P., Watson-Parris, D., and Lohmann, U.: HAMMOZ-Consortium MPI-ESM1.2-HAM model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.5016, 2019.

Orihuela-Pinto, B., England, M. H., and Taschetto, A. S.: Interbasin and interhemispheric impacts of a collapsed Atlantic Overturning Circulation, Nature Climate Change, 12, 558–565, https://doi.org/10.1038/s41558-022-01380-y, 2022.

Park, S. and Shin, J.: SNU SAM0-UNICON model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.7789, 2019.

Ridley, J., Menary, M., Kuhlbrodt, T., Andrews, M., and Andrews, T.: MOHC HadGEM3-GC31-LL model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.6109, 2019a.

Ridley, J., Menary, M., Kuhlbrodt, T., Andrews, M., and Andrews, T.: MOHC HadGEM3-GC31-MM model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.6112, 2019b.

Romanou, A., Rind, D., Jonas, J., Miller, R., Kelley, M., Russell, G., Orbe, C., Nazarenko, L., Latto, R., and Schmidt, G. A.: Stochastic Bifurcation of the North Atlantic Circulation under a Midrange Future Climate Scenario with the NASA-GISS ModelE, Journal of Climate, 36, 6141 – 6161, https://doi.org/10.1175/JCLI-D-22-0536.1, 2023.

Schmittner, A.: Decline of the marine ecosystem caused by a reduction in the Atlantic overturning circulation, Nature, 434, 628–633, https://doi.org/10.1038/nature03476, 2005.

Scoccimarro, E., Bellucci, A., and Peano, D.: CMCC CMCC-CM2-HR4 model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.3823, 2020.

Semmler, T., Danilov, S., Rackow, T., Sidorenko, D., Barbi, D., Hegewald, J., Sein, D., Wang, Q., and Jung, T.: AWI AWI-CM1.1MR model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.2686, 2018.

Song, Z., Qiao, F., Bao, Y., Shu, Q., Song, Y., and Yang, X.: FIO-QLNM FIO-ESM2.0 model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.9199, 2019.

Stouffer, R.: UA MCM-UA-1-0 model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.8888, 2019.

Sura, P.: Stochastic Analysis of Southern and Pacific Ocean Sea Surface Winds, Journal of the Atmospheric Sciences, 60, 654 – 666, https://doi.org/10.1175/1520-0469(2003)060<0654:SAOSAP>2.0.CO;2, 2003.

Sura, P. and Sardeshmukh, P. D.: A Global View of Non-Gaussian SST Variability, Journal of Physical Oceanography, 38, 639 – 647, https://doi.org/10.1175/2007JPO3761.1, 2008.

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Solheim, L., von Salzen, K., Yang, D., Winter, B., and Sigmond, M.: CCCma CanESM5-CanOE model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.10260, 2019a.

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Solheim, L., von Salzen, K., Yang, D., Winter, B., and Sigmond, M.: CCCma CanESM5 model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.3610, 2019b.

Tang, Y., Rumbold, S., Ellis, R., Kelley, D., Mulcahy, J., Sellar, A., Walton, J., and Jones, C.: MOHC UKESM1.0-LL model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.6113, 2019.

Tatebe, H. and Watanabe, M.: MIROC MIROC6 model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.5603, 2018.

Tian, B. and Dong, X.: The Double-ITCZ Bias in CMIP3, CMIP5, and CMIP6 Models Based on Annual Mean Precipitation, Geophysical Research Letters, 47, e2020GL087 232, https://doi.org/https://doi.org/10.1029/2020GL087232, e2020GL087232 2020GL087232, 2020.

van Westen, R. M. and Dijkstra, H. A.: Asymmetry of AMOC Hysteresis in a State-Of-The-Art Global Climate Model, Geophysical Research Letters, 50, e2023GL106 088, https://doi.org/https://doi.org/10.1029/2023GL106088, e2023GL106088 2023GL106088, 2023.

van Westen, R. M. and Dijkstra, H. A.: Persistent climate model biases in the Atlantic Ocean's freshwater transport, Ocean Science, 20, 549–567, https://doi.org/10.5194/os-20-549-2024, 2024.

van Westen, R. M., Jacques-Dumas, V., Boot, A. A., and Dijkstra, H. A.: The Role of Sea-ice Processes on the Probability of AMOC Transitions, arXiv, 2024a.

van Westen, R. M., Kliphuis, M., and Dijkstra, H. A.: Physics-based early warning signal shows that AMOC is on tipping course, Science Advances, 10, eadk1189, https://doi.org/10.1126/sciadv.adk1189, 2024b.

Watts, M., Maslowski, W., Lee, Y. J., Kinney, J. C., and Osinski, R.: A Spatial Evaluation of Arctic Sea Ice and Regional Limitations in CMIP6 Historical Simulations, Journal of Climate, 34, 6399 – 6420, https://doi.org/10.1175/JCLI-D-20-0491.1, 2021.

Weijer, W., Cheng, W., Drijfhout, S. S., Fedorov, A. V., Hu, A., Jackson, L. C., Liu, W., McDonagh, E. L., Mecking, J. V., and Zhang, J.: Stability of the Atlantic Meridional Overturning Circulation: A Review and Synthesis, Journal of Geophysical Research: Oceans, 124, 5336–5375, https://doi.org/https://doi.org/10.1029/2019JC015083, 2019.

Wieners, K.-H., Giorgetta, M., Jungclaus, J., Reick, C., Esch, M., Bittner, M., Legutke, S., Schupfner, M., Wachsmann, F., Gayler, V., Haak, H., de Vrese, P., Raddatz, T., Mauritsen, T., von Storch, J.-S., Behrens, J., Brovkin, V., Claussen, M., Crueger, T., Fast, I., Fiedler, S., Hagemann, S., Hohenegger, C., Jahns, T., Kloster, S., Kinne, S., Lasslop, G., Kornblueh, L., Marotzke, J., Matei, D., Meraner, K., Mikolajewicz, U., Modali, K., Müller, W., Nabel, J., Notz, D., von Gehlen, K. P., Pincus, R., Pohlmann, H., Pongratz, J., Rast, S., Schmidt, H., Schnur, R., Schulzweida, U., Six, K., Stevens, B., Voigt, A., and Roeckner, E.: MPI-M MPI-ESM1.2-LR model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.6595, 2019.

Wu, T., Chu, M., Dong, M., Fang, Y., Jie, W., Li, J., Li, W., Liu, Q., Shi, X., Xin, X., Yan, J., Zhang, F., Zhang, J., Zhang, L., and Zhang, Y.: BCC BCC-CSM2MR model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.2948, 2018.

Wunderling, N., von der Heydt, A. S., Aksenov, Y., Barker, S., Bastiaansen, R., Brovkin, V., Brunetti, M., Couplet, V., Kleinen, T., Lear, C. H., Lohmann, J., Roman-Cuesta, R. M., Sinet, S., Swingedouw, D., Winkelmann, R., Anand, P., Barichivich, J., Bathiany, S., Baudena, M., Bruun, J. T., Chiessi, C. M., Coxall, H. K., Docquier, D., Donges, J. F., Falkena, S. K. J., Klose, A. K., Obura, D., Rocha, J., Rynders, S., Steinert, N. J., and Willeit, M.: Climate tipping point interactions and cascades: a review, Earth System Dynamics, 15, 41–74, https://doi.org/10.5194/esd-15-41-2024, 2024.

Yukimoto, S., Koshiro, T., Kawai, H., Oshima, N., Yoshida, K., Urakawa, S., Tsujino, H., Deushi, M., Tanaka, T., Hosaka, M., Yoshimura, H., Shindo, E., Mizuta, R., Ishii, M., Obata, A., and Adachi, Y.: MRI MRI-ESM2.0 model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.6842, 2019.

Zhang, J., Wu, T., Shi, X., Zhang, F., Li, J., Chu, M., Liu, Q., Yan, J., Ma, Q., and Wei, M.: BCC BCC-ESM1 model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.2949, 2018.

Zhang, Q., Liu, B., Li, S., and Zhou, T.: Understanding Models' Global Sea Surface Temperature Bias in Mean State: From CMIP5 to CMIP6, Geophysical Research Letters, 50, e2022GL100 888, https://doi.org/https://doi.org/10.1029/2022GL100888, e2022GL100888 2022GL100888, 2023.

Zickfeld, K., Eby, M., and Weaver, A. J.: Carbon-cycle feedbacks of changes in the Atlantic meridional overturning circulation under future atmospheric CO2, Global Biogeochemical Cycles, 22, https://doi.org/https://doi.org/10.1029/2007GB003118, 2008.

Ziehn, T., Chamberlain, M., Lenton, A., Law, R., Bodman, R., Dix, M., Wang, Y., Dobrohotoff, P., Srbinovsky, J., Stevens, L., Vohralik, P., Mackallah, C., Sullivan, A., O'Farrell, S., and Druken, K.: CSIRO ACCESS-ESM1.5 model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.4272, 2019.

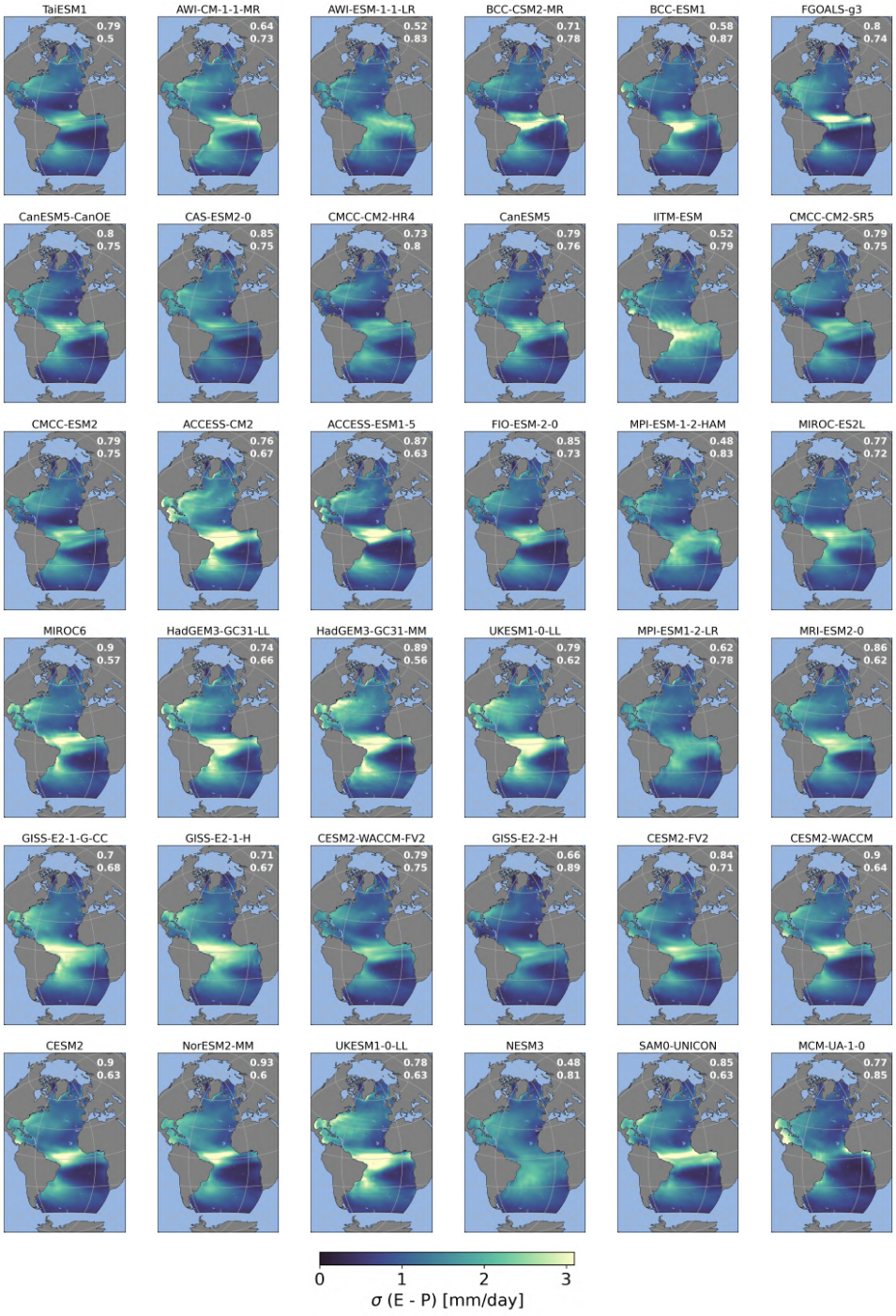

Figure A10. Standard deviation (σ) in the noise of the ~~$E - P$~~ $E - P$ for the analysed CMIP6 models. Numbers in the top right corner reflect the spatial correlation and root mean square error.

Figure A11. Skewness in the noise of the ~~$E - P$~~ $E - P$ for the analysed CMIP6 models. Numbers in the top right corner reflect the spatial correlation and root mean square error.

Figure A12. Excess kurtosis in the noise of the ~~$E - P$~~ $E - P$ for the analysed CMIP6 models. Numbers in the top right corner reflect the spatial correlation and root mean square error.

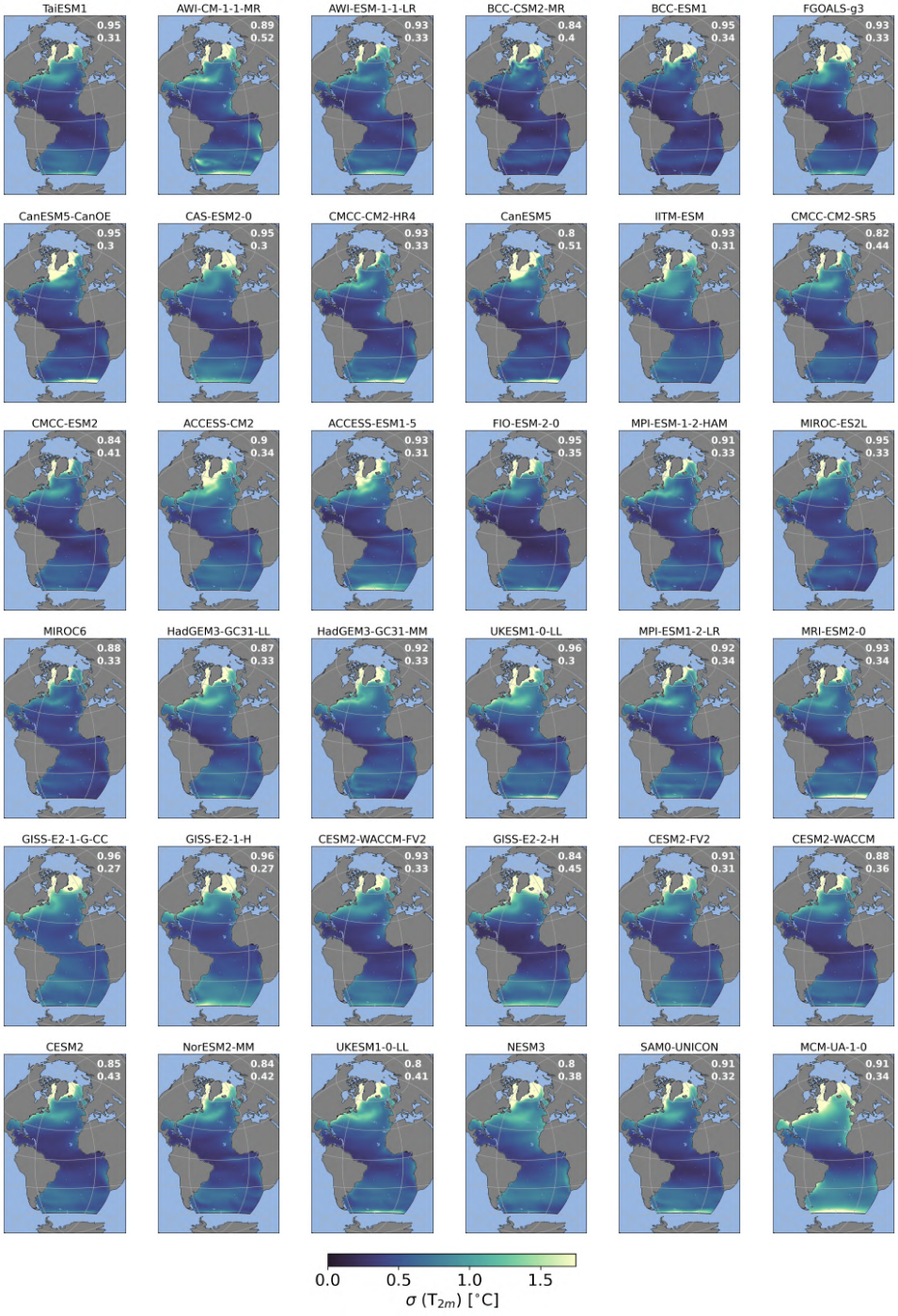

Figure A13. As Fig. A10 but for T_{2m}.

Figure A14. As Fig. A11 but for T_{2m}.

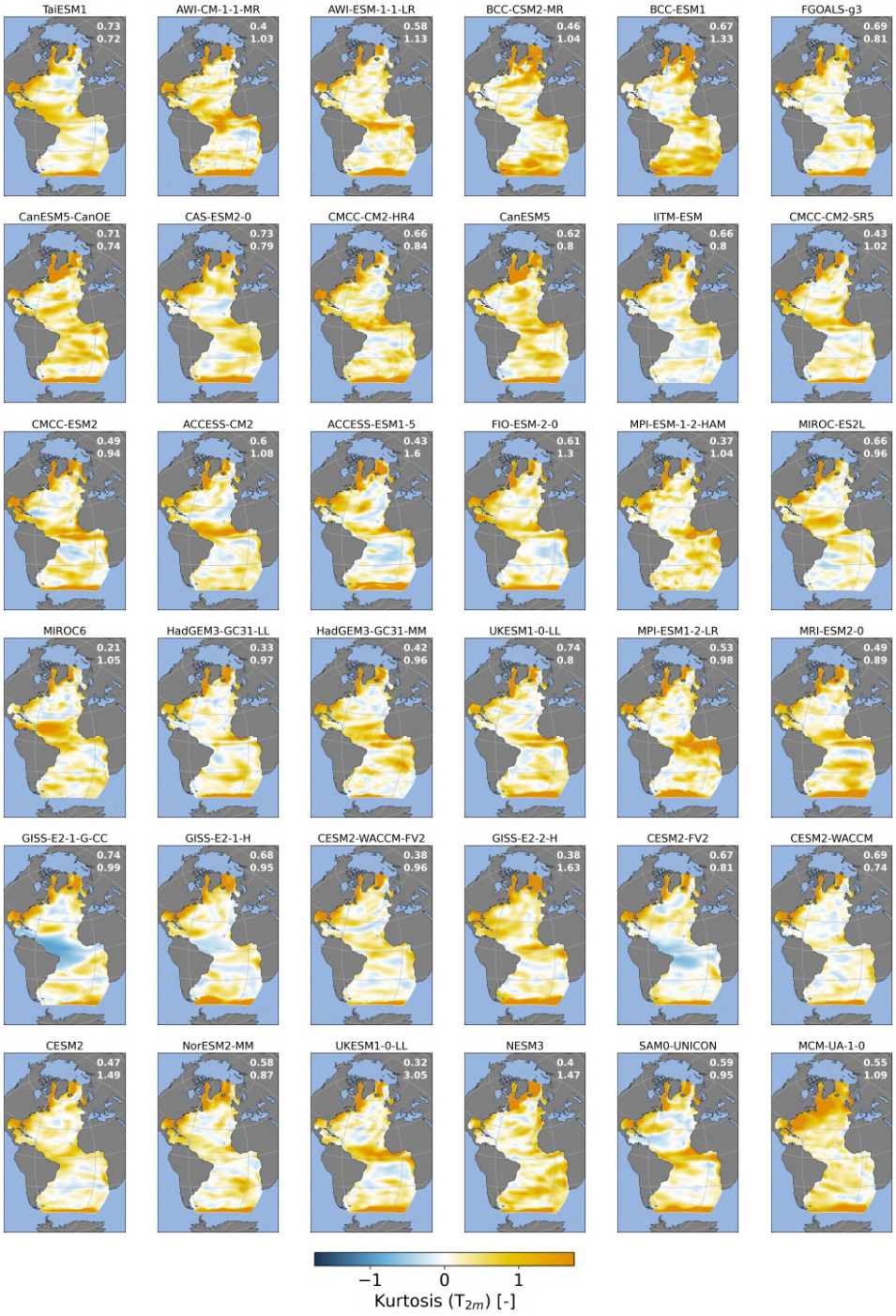

Figure A15. As Fig. A12 but for T_{2m}.

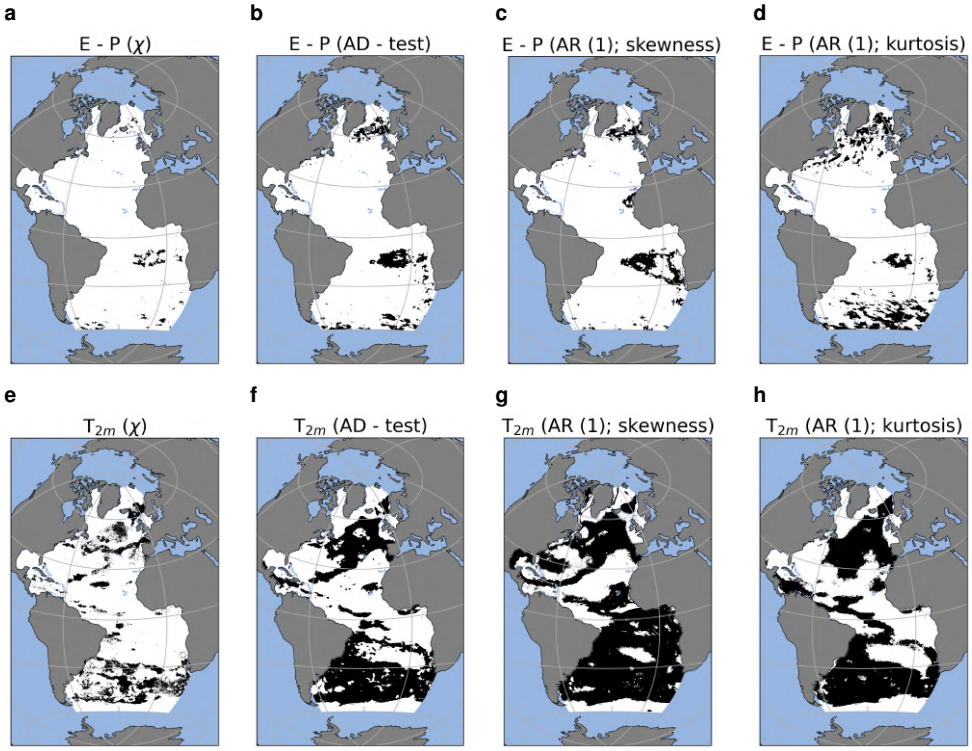

Figure A16. Collection of results for statistical tests. The top row is for the $E - P$ noise and the bottom row for the T_{2m} noise. (a, e) χ, where white regions represent grid points where the Normal Inverse Gaussian distribution provides a better fit than a Gaussian distribution. (b, f) Results from an Anderson-Darling test on normality where black regions represent grid points where the test is passed and the distribution is significantly Gaussian. (c, g) Results of the significance test of the skewness based on an AR(1) model where black regions represent that the AR(1) model provides a good fit. (d, h) as in (c, g) but for excess kurtosis.

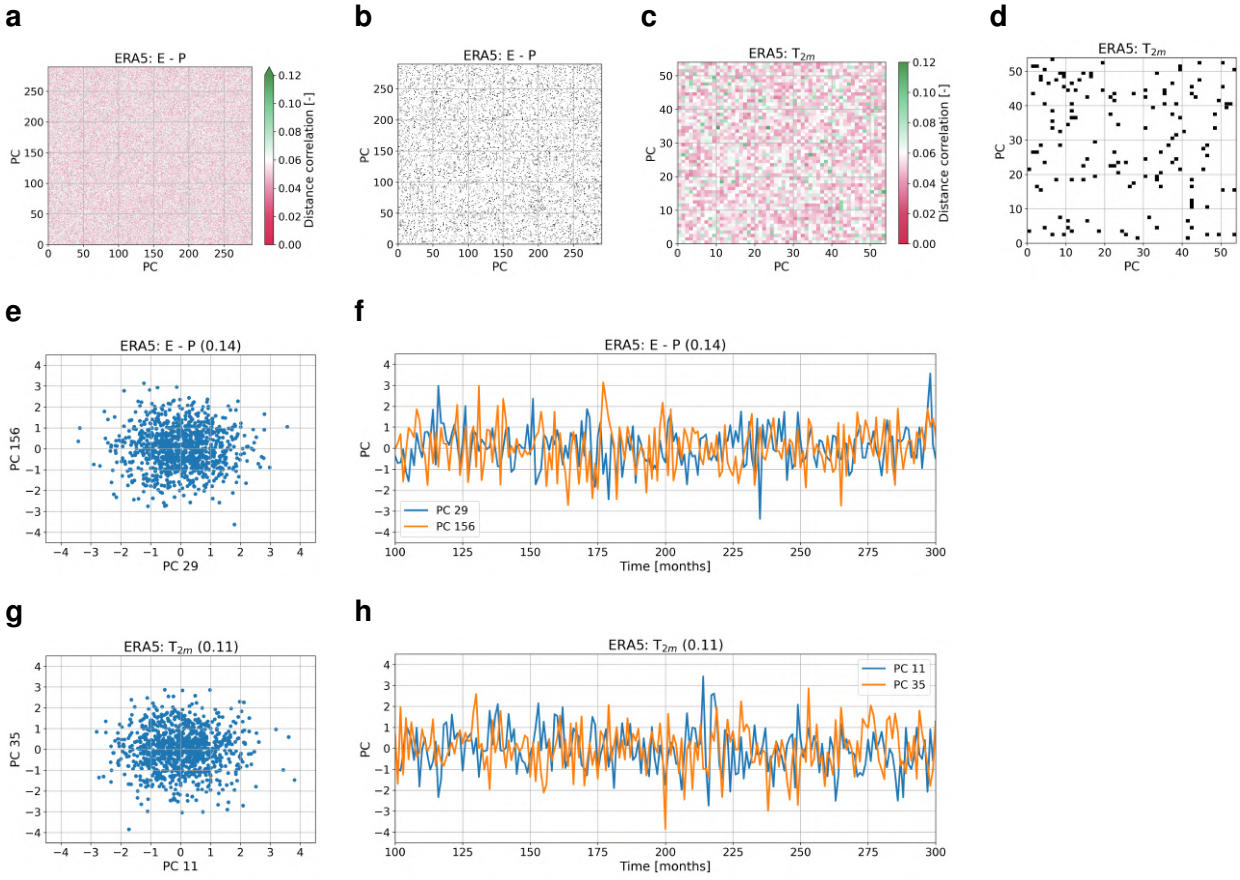

Figure A17. Results of the distance correlation analysis to test for dependencies between the PCs. (a) Distance correlation between the PCs for $E - P$. (b) Black squares represent a significant dependency ($p < 0.05$). (c) and (d) as (a) and (b) but for T_{2m}. (e) and (f) represent the two PCs that share the highest distance correlation (0.14) for $E - P$ with a scatter plot in (e) and a time series in (f). (g) and (h) are as (e) and (f) but for the PCs with the highest distance correlation (0.11) for T_{2m}. Note that a distance correlation has a range of 0 to 1, where 0 represents no relation, and 1 does represent a relation.

Table A1. CMIP6 model list.

Number	Name	~~Member~~ Reference
1.	TaiESM1	~~r1i1p1f1~~ Lee and Liang (2020)
2.	AWI-CM-1-1-MR	~~r1i1p1f1~~ Semmler et al. (2018)
3.	AWI-ESM-1-1-LR	~~r1i1p1f1~~ Danek et al. (2020)
4.	BCC-CSM2-MR	~~r1i1p1f1~~ Wu et al. (2018)
5.	BCC-ESM1	~~r1i1p1f1~~ Zhang et al. (2018)
6.	FGOALS-g3	~~r1i1p1f1~~ Li (2019)
7.	CanESM5-CanOE	~~r1i1p2f1~~ Swart et al. (2019a)
8.	CAS-ESM2-0	~~r1i1p1f1~~ Chai (2020)
9.	CMCC-CM2-HR4	~~r1i1p1f1~~ Scoccimarro et al. (2020)
10.	CanESM5	~~r1i1p1f1~~ Swart et al. (2019b)
11.	IITM-ESM	~~r1i1p1f1~~ Choudhury et al. (2019)
12.	CMCC-CM2-SR5	~~r1i1p1f1~~ Lovato and Peano (2020)
13.	CMCC-ESM2	~~r1i1p1f1~~ Lovato et al. (2021)
14.	ACCESS-CM2	~~r1i1p1f1~~ Dix et al. (2019)
15.	ACCESS-ESM1-5	~~r1i1p1f1~~ Ziehn et al. (2019)
16.	FIO-ESM-2-0	~~r1i1p1f1~~ Song et al. (2019)
17.	MPI-ESM-1-2-HAM	~~r1i1p1f1~~ Neubauer et al. (2019)
18.	MIROC-ES2L	~~r1i1p1f2~~ Hajima et al. (2019)
19.	MIROC6	~~r1i1p1f1~~ Tatebe and Watanabe (2018)
20.	HadGEM3-GC31-LL	~~r1i1p1f3~~ Ridley et al. (2019a)
21.	HadGEM3-GC31-MM	~~r1i1p1f3~~ Ridley et al. (2019b)
22.	UKESM1-0-LL (MOHC)	~~r1i1p1f2~~ Tang et al. (2019)
23.	MPI-ESM1-2-LR	~~r1i1p1f1~~ Wieners et al. (2019)
24.	MRI-ESM2-0	~~r1i1p1f1~~ Yukimoto et al. (2019)
25.	GISS-E2-1-G-CC	~~r1i1p1f1~~ NASA/GISS) (2019)
26.	GISS-E2-1-H	~~r1i1p1f1~~ NASA/GISS (2019a)
27.	CESM2-WACCM-FV2	~~r1i1p1f1~~ Danabasoglu (2019b)
28.	GISS-E2-2-H	~~r1i1p1f1~~ NASA/GISS (2019b)
29.	CESM2-FV2	~~r1i1p1f1~~ Danabasoglu (2019d)
30.	CESM2-WACCM	~~r1i1p1f1~~ Danabasoglu (2019c)
31.	CESM2	~~r1i1p1f1~~ Danabasoglu (2019a)
32.	NorESM2-MM	~~r1i1p1f1~~ Bentsen et al. (2019)
33.	UKESM1-0-LL (NIMS-KMA)	~~r13i1p1f2~~ Byun (2020)
34.	NESM3	~~r1i1p1f1~~ Cao and Wang (2019)
35.	SAMO0-UNICON	~~r1i1p1f1~~ Park and Shin (2019)
36.	MCM-UA-1-0 ~~r1i1p1f1~~	Stouffer (2019)