![Barcelona Supercomputing Center - Centro Nacional de Supercomputación]

CHARACTERIZATION OF NEAR SURFACE WIND SPEED AND TEMPERATURE STATISTICAL DISTRIBUTIONS

BSC-ESS-2016

Raül Marcos, Veronica Torralba, Nicola Cortesi, Nube González, Albert Soret and Francisco J. Doblas-Reyes

Earth Sciences Department
Barcelona Supercomputing Center - Centro Nacional de Supercomputación (BSC-CNS)

04 April 2017

TECHNICAL REPORT

Series: Earth Sciences (ES) Technical Report

A full list of ES Publications can be found on our website under:

https://earth.bsc.es/wiki/doku.php?id=library:external:technical_memoranda

Summary

In this work we have studied the properties of the statistical distributions of 10m wind speed and 2m temperature from the ERA-Interim reanalysis and S4 seasonal forecast system. Additionally, we have compared both datasets by assessing the different distributions of the two variables at seasonal time scales, considering their inter-annual and intra-seasonal variability. This technical note focuses on 10m wind speed and two seasons, JJA and DJF, because this information could be the most relevant for wind energy applications. That said, the figures for the other seasons and for 2m temperature can be found in the web-catalog (www.bsc.es/ESS/catalogue). The 10m wind speed distribution has been characterized in terms of the four main moments of the probability distribution (mean, standard deviation, skewness and kurtosis). We have also computed the coefficient of variation and goodness of fit Shapiro-Wilks test to identify the regions with the higher wind variability and Gaussian behaviour. These parameters are important to provide useful climate information in wind energy decision-making processes which use simple assumptions of the wind speed frequency distribution to properly estimate the wind energy potential. Besides, this study also illustrates where the discrepancies of the distributions of the seasonal predictions and the reference dataset are higher and, thus, which might need special attention from a bias correction perspective.

Contents

Index of figures

1. Introduction

In the framework of seasonal forecast verification, knowing whether the climatology represented in the models is similar to the observed world is essential to decide about the application of the optimal bias correction methods (Ruffault et al. 2013) . However, without an absolute reference dataset, the selection of an appropriate reliable database becomes a critical decision (Trenberth et al. 2008). In this regard, it is valuable to firstly determine the way our potential candidate represents its own world and, afterwards, check whether the climatology of the forecasts is similar to the chosen reference or not.

Considering that among the objectives of RESILIENCE are the assessment and forecasting of 10m wind speed and 2m temperature variables, we have characterized the statistical distributions of global 10m wind speed and 2m temperature from the ERA-Interim reanalysis (Dee et al. 2011) and the ECMWF System4 (S4) seasonal prediction system (Molteni, Stockdale, M. A. Balmaseda, et al. 2011). This choice has been motivated by the fact that 10m wind speed and 2m temperature are chief variability drivers of wind power production and energy demand (Archer 2005; Giannakopoulos & Psiloglou 2006; Hekkenberg et al. 2009; Lu et al. 2009) and, hence, they are the key variables to determine, forecast and manage the wind energy resource.

Concerning the chosen datasets, the ERA-Interim reanalysis is our observational reference database. This is the reanalysis product from the European Centre for Medium Range Weather Forecasts (ECMW**F)** and it covers the period 1979-nowadays. It has been used in a wide range of research fields: climatology (e.g. Škerlak et al. 2014), climate change (e.g. Andres et al. 2014), characterization of extremes (e.g. Cornes & Jones 2013), etc. Concerning wind-power production, it has also been thoroughly applied with positive results (e.g. Kiss et al. 2009; Rose & Apt 2015; Lorenz & Barstad 2016).

On the other hand, S4 is the most recent instalment of ECMWF in the seasonal forecasting field and it covers the period 1981-nowadays (Molteni, Stockdale, M. A. Balmaseda, et al. 2011). It is the evolution of the well-considered ECMWF System 3 (Stockdale et al. 2011) and its full potential is still being unfolded (e.g. Torralba et al. 2015; Tompkins & di Giuseppe 2015; Marcos et al. 2016; Ogutu et al. 2016).

In this work we have portrayed the main properties of the 10m wind speed and 2m temperature from both datasets in an inter-annual and intra-seasonal basis to distinguish the contribution of each variability source (e.g. Achuthavarier & Krishnamurthy 2010; Luo et al. 2011). This has been achieved through a set of statistical parameters that are commonly used to characterize probability density functions, such as the standard deviation, the Shapiro-Wilks test or the excess kurtosis analysis (Wilks 2006). Besides, we have also computed the differences between each pair of dataset parameters to highlight the hot-spot discordant regions.

This technical report only discusses the most relevant results obtained for 10m wind speed because there is a limited number of studies that characterize the probability density function of wind speed at global scales. The distribution of 2m temperature has also been explored and the results can be found in the web-catalog (www.bsc.es/ESS/catalogue). In fact, this evaluation provides uncertainty estimates in the different parameters that should be considered in the seasonal verification processes, which is an aspect non readily available to the climate research and climate services communities and, thus, that can be relevant for many scientists and wind energy users who attempt to evaluate the wind speed variability at seasonal time scales.

This technical note is organized as follows. After this first introductory section there is the 'Data and Methodology' section that is focused on the description of the different datasets and methods used. Afterwards, we present and discuss the results obtained and, finally, there is the 'Conclusions' section, which contains a summary of the main results and reflections on the task performed.

2. Data and methodology

We have characterized the probability density function of the global surface 10m wind speed at inter-annual and intra-seasonal time scales for the ERA-Interim reanalysis (Dee et al. 2011) and S4 seasonal prediction system(Molteni et al. 2011). This analysis has been performed through six statistical parameters: mean value, standard deviation, kurtosis, skewness, coefficient of variation, and Shapiro-Wilks goodness of fit test.

2.1. Data

2.1.1. ERA-Interim

The ERA-Interim reanalysis dataset (Dee et al. 2011) is the latest global atmospheric reanalysis issued by the European Centre for Medium-Range Weather Forecasts (ECMW**F)**. It spans from 1979 to nowadays, being updated on a real-time basis. In comparison to the previous system, ERA-40 (Uppala et al. 2005), it shows multiple improvements such as the incorporation of the four-dimension variational data assimilation approach (4D-Var), the increase of model resolution (~80km) or the enhancement of the forecast model physics.

2.1.2. ECMWF System-4

S4 seasonal prediction system (Molteni, Stockdale, M. Balmaseda, et al. 2011) is a fully coupled general circulation model that provides operational multi-variable seasonal forecasts in a real-time basis. In this study we focus on period 1981-2015. Last 35 years of forecasts proceed from the combination of the 30-years hindcasts with the 5-years regular contemporary pool of predictions. All forecasts have a minimum of 15-member ensemble (51 members for the start months of February, May, August and November, and at every month since May 2011) and 7 months forecast horizon. The predictions used for the discussion are those initialized the 1st of November and the 1st of May (1 month lead tim**e)** and have been interpolated to the ERA-Interim resolution.

2.2. Methodology

The characterization of the statistical distributions of the 10m wind speed and 2m temperature at inter-annual and intra-seasonal time scales has been achieved through the computation of six statistical parameters: mean, standard deviation, coefficient of variation, skewness, excess kurtosis and normal goodness-of-fit through the Shapiro-Wilks test (Wilks 2006).

Inter-annual statistics reflect the properties of the inter-annual distribution, i.e., how values vary from year to year for a particular season, whereas intra-seasonal statistics reflect the average properties of the distribution inside a season, i.e., how values vary among the months of the season. The difference between two time scales is that at the intra-seasonal scale, we do not average the wind speed of the three months in a

season. We concatenate these three values in a single time series from 1981 to 2015 to compute each statistical parameter.

The mean is the first moment of the distribution and a useful measure in order to characterize which are the usual conditions in one particular location. The mean of a variable x for N values ($x_1,\dots x_N$) can be defined by the equation 1:

$$\bar{x} = \frac{1}{N} \sum_{j=1}^{N} x_j \tag{1}$$

The evaluation of the discrepancies between the climatologies (mean) for the seasonal predictions and the reference dataset can be useful to illustrate the systematic errors of the seasonal prediction systems.

Having computed the distribution's central value we have characterized its variability around the mean based on the standard deviation (st**d)**, which is the second central moment of the distribution and it can be defined:

$$std = \frac{1}{N-1} \sum_{j=1}^{N} (x_j - \bar{x})^2 \tag{2}$$

Although two different distributions have the same mean, it is helpful to understand if the data points spread far from the mean or if they are clustered around the mean, because this information could be of crucial importance in order to assess the variability of a variable in a particular location.

Skewness is a measure of degree of symmetry related to the normal distribution, which indicates whether a sample has more values to the left of its mean (i.e., it is left-skewed), to the right (right-skewed) or it is symmetrically distributed around the mean (zero skewness) and it is defined:

$$skw = \frac{1}{N} \sum_{j=1}^{N} \left[\frac{x_j - \bar{x}}{\sigma} \right]^3 \tag{3}$$

Kurtosis is the four moment of the distribution and it measures the weight of the tails of a sample distribution It can be defined as:

$$kurt = \left\{ \frac{1}{N} \sum_{j=1}^{N} \left[\frac{x_j - \bar{x}}{\sigma} \right]^4 \right\} - 3 \tag{4}$$

A low kurtosis reflects a flat distribution, whereas a high kurtosis reflects a

distribution with a sharp peak. Kurtosis is often compared to that of a Standard Normal, which equals 3, by subtracting 3 from the kurtosis value, yielding what is known as 'excess kurtosis'.

In addition to the main moments of the distributions we have used two more measures that can provide extra information about the features of the probability distributions: the coefficient of variation and the goodness of fit estimations.

The coefficient of variation (cv) is a standardized measure of dispersion. It is often expressed as a percentage, and is defined as the ratio of the standard deviation to the mean:

$$cv = \frac{\sigma}{\bar{x}} \qquad (5)$$

The smaller coefficient of variation the less dispersion is in the probability density function. The coefficient of variation expresses similar information to the standard deviation but it has the main advantage that it is independent of the variable units, therefore coefficient variations allows the comparison of the dispersion of different locations and variables.

The goodness of fit assesses how closely a sample resembles a specific distribution. In our case we have assessed how close is our sample data to a normal distribution. Although instantaneous values or daily accumulations of many climatic variables do not have Gaussian distributions, at inter-annual or intra-seasonal time scales the distributions tend to be 'near-normal'. It is for that reason a normal distribution is tested.

The evaluation of the goodness of fit is based on the Shapiro-Wilks test (Shapiro & Wilk 1965), which is an specific test to evaluate how similar is the distribution to a normal one. This test is a statistical hypothesis testing process which null hypothesis is whether the distribution of a x variable for N values ($x_1,\dots x_N$) comes from a normally distributed population.

The statistic that is employed in the hypothesis contrast is calculated as follows:

$$W = \frac{\left(\sum_{j=1}^{N} a_j x_{(j)}\right)^2}{\sum_{j=1}^{N}\left(x_j - \bar{x}\right)^2} \qquad (6)$$

Where $x_{(j)}$ are the ordered sample values ($x_{(1)}$ is the smallest) and the a_j are constants that depend on the mean, variance and covariance matrix. The null hypothesis is that the population is normally distributed. Therefore, if the p-value computed with the statistic (equation 6) is lower than the significance level, then the null hypothesis is rejected and the distribution is not normal.

Figure 1 shows how these calculations are performed at seasonal time scale, for the four meteorological seasons separately (DJF, MAM, JJA and SON), considering that forecasts for DJF are assigned to the year of its start date.

The same six statistics are likewise obtained for the 10m wind speed and 2m temperature from ERA-Interim reanalysis and S4 seasonal prediction system for each grid-point and lead time. Both of them are calculated in the same way, except for the presence of ensemble members in the seasonal prediction dataset. In this case, they are sampled with the temporal dimension to form a single distribution. Finally, the differences between each parameter from the ERA-Interim and S4 have been obtained to identify the regions with greater discrepancies.
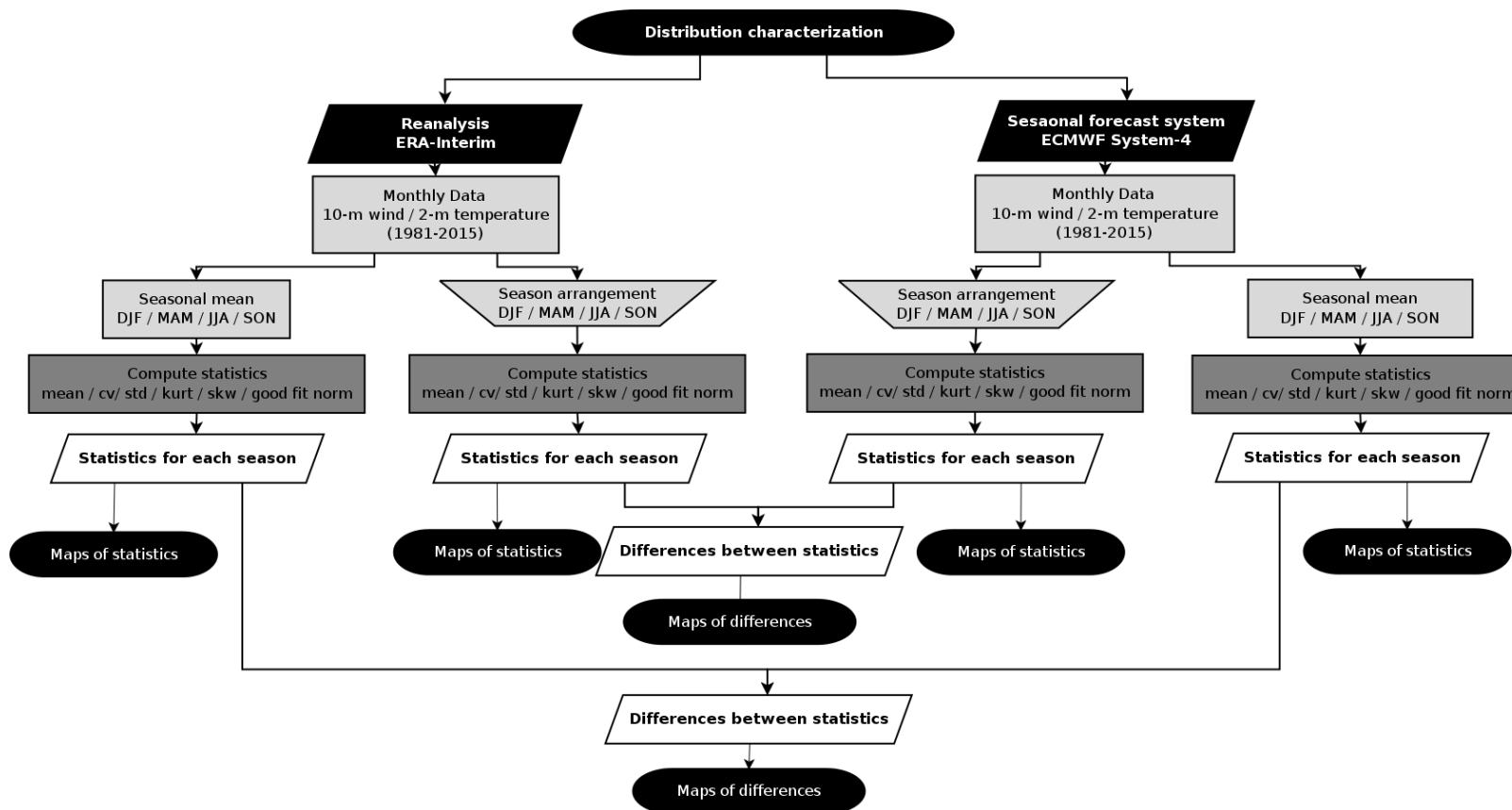
Figure 1. *Work-flow diagram.*

Work-flow diagram for the characterization of the statistical distributions of the 10m wind and 2m temperature at inter-annual and intra-seasonal time scales for the ERA-Interim reanalysis and ECMWF System-4 forecast system. Six statistical parameters have been computed: mean value (mean), standard deviation (std), coefficient of variation (cv), excess kurtosis (kurt), skewness (skw) and Shapiro-Wilks goodness of fit test. The differences between the parameters computed parameters for each dataset have also been obtained. In the 'season arrangement' we do not average the variables of the three months belonging to the same season.

3. Results

In this section we present the results for DJF and JJA for 10m wind speed since these two seasons allow the illustration of the seasonal changes in wind speed variability The corresponding discussion focuses on each statistical parameter at inter-annual and intra-seasonal time scales. The six statistics are explored for both datasets: ERA-Interim reanalysis and the seasonal predictions from S4. Finally the differences between the distribution of 10m wind speeds for ERA-Interim and S4 are also computed.

3.1. Climatology (mean)

In this case there is no difference between inter-annual and intra-seasonal computation and, hence, only one column is presented (named inter-annual).

3.1.1. DJF

ERA-Interim and S4 show very similar patterns (**Figure 2**a and **Figure 2**b). There is a widespread area of higher wind speeds in the northern and southern extra-tropical oceans, with secondary maximums around 20 degrees north and south. In all the cases the winds are stronger in the Northern Hemisphere. The maximum wind speeds are in the oceans while the minimum values lie over land. This behavior is due to continental roughness, which is bigger than the oceanic counterpart. In fact, the speed over land is generally less than half the observed over the oceans. As for the differences between the reanalysis and the forecasting model, 95% of the grid points lie between 1 and -1 m/s (**Figure 2**c). It seems the S4 shows a systematic negative bias because it tends to overestimate the speed values globally. The greatest discrepancies are in the inter-tropical oceans and eastern Siberia and Canada. Regarding the areas where the S4 underestimates wind speeds, they are notorious in the Himalayan region and the eastern Equatorial Atlantic and Pacific, in front of African and South American coasts.

3.1.2. JJA

The wind speed patterns for this season are also very similar in ERA-Interim and S4 (**Figure 3**a and **Figure 3**b). In this case, however, the maximum wind speeds are restricted to the Southern Hemisphere, around the extra-tropics and in the Indian Ocean (where the Monsoon structure is clear). It is worth noting that the maximum wind speed difference between both hemispheres is more noticeable in JJA than in DJF. Nevertheless, in both seasons the maximum wind speeds are still observed over the oceans, with the minimum values settled over land. As for the differences between the reanalysis and the prediction system, 95% of the grid points lie between 1 and -1 m/s (**Figure 3**c). It seems the S4 overestimates the values over the oceans, especially in the inter-tropics and the Australia southern oceanic region. Regarding the areas where the S4 underestimates wind speeds, we find them in the eastern Pacific (near the Colombian coast), in the northern tropical Atlantic, in a strip of land between 20 and 40 degrees north covering African and Asian territories and, also, in the Arctic. It seems this behavior arises in the areas with most prominent orography.

Figure 2. *10m wind speed climatology for DJF.*

*This figure depicts 10m wind speed climatology for DJF spanning the period 1981-2015 **a)** ERA-Interim **b)** S4 November start date c) Result of comparing a) minus **b)** maps. Note these figures have different colorbars.*
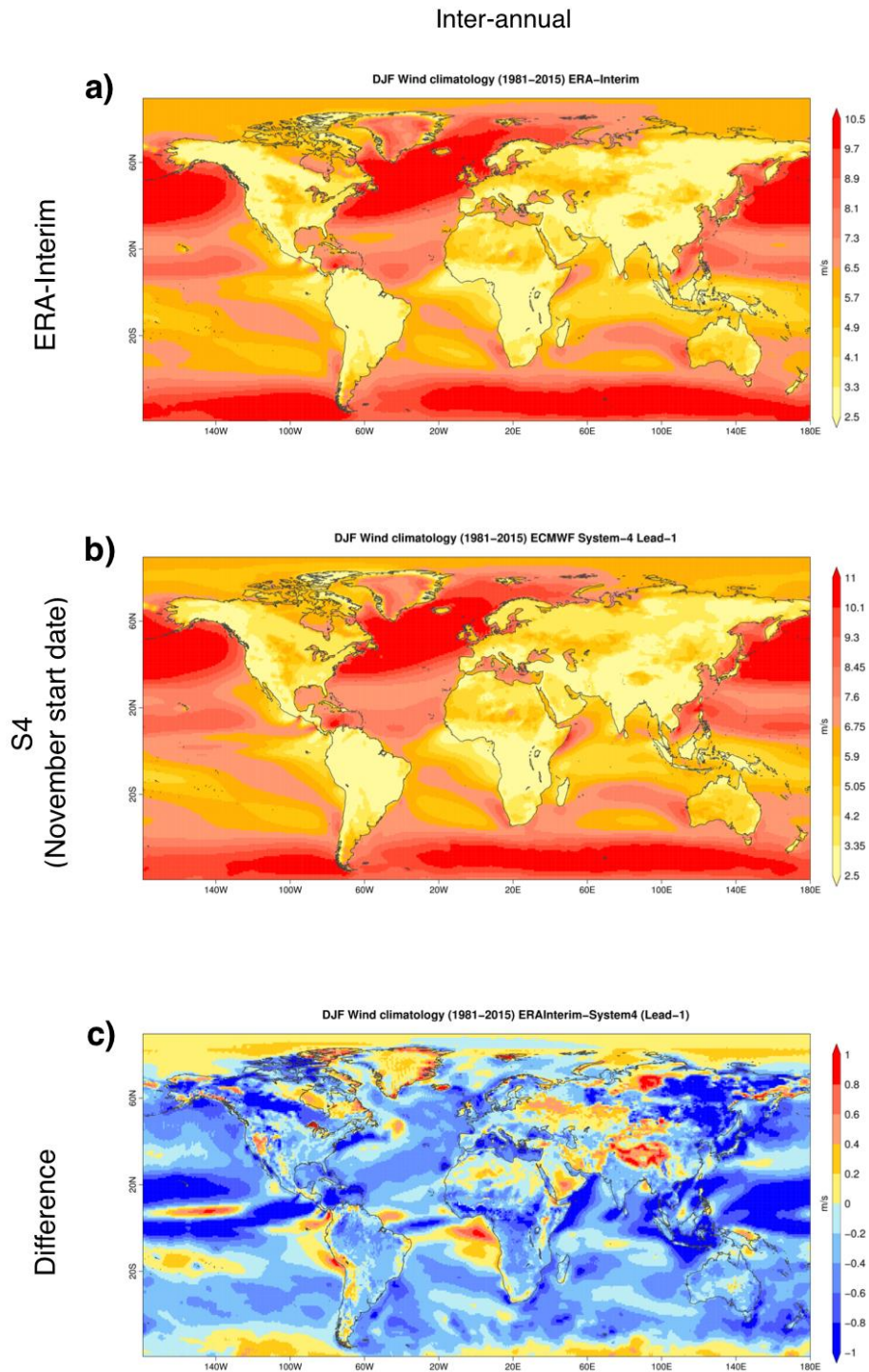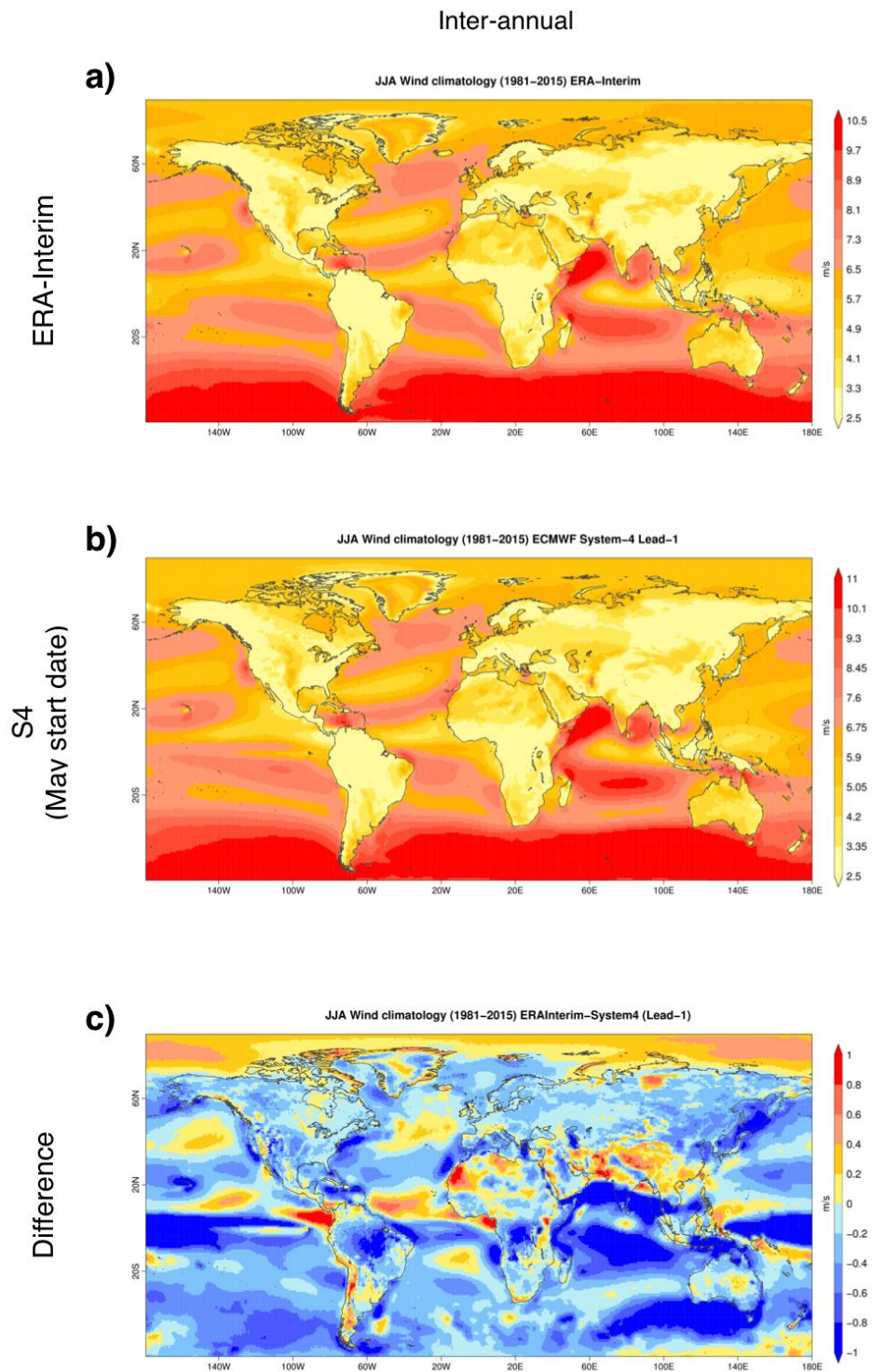
Figure 3. *10m wind speed climatology for JJA.*

*This figure depicts 10m wind speed climatology for JJA spanning the period 1981-2015 **a)** ERA-Interim **b)** S4 November start date **c)** Result of comparing a) minus b) maps. Note these figures have different colorbars.*

3.2. Standard deviation

3.2.1. DJF

The variability patterns of the ERA-Interim and the S4 are similar, both at inter-annual and intra-seasonal level (**Figure 4**a, **Figure 4**b, **Figure 4**c and **Figure 4**d). The strongest variability is over the oceans. In the Northern Hemisphere there are two remarkable high variability regions, one in the eastern northern Atlantic and another covering the north-Pacific basin. Moving to the southern oceans, there is a strong variability region in the maritime continent (Indonesi**a)** that propagates northwards.

The differences between the ERA-Interim and S4 range between 0.5 and -0.5 m/s for the 95% of the grid-points. These differences have no dominant sign, either for the inter-annual or the intra-seasonal time frame. From an inter-annual point of view, these differences are less pronounced than for the intra-seasonal (**Figure 4**e and **Figure 4**f), being the structure of the corresponding differences comparable. We can identify three regions of special interest: the Pacific Ocean inter-tropics (S4 underestimates wind speed values), the maritime continent (S4 overestimation) and the Scandinavian and Siberian regions (S4 underestimation).

3.2.2. JJA

The ERA-Interim and S4 show very similar structures at inter-annual and intra-seasonal time-scales (**Figure 5**a, **Figure 5**b, **Figure 5**c and **Figure 5**d) and similar to those obtained for DJF. The inter-annual variability is slightly smaller than the intra-seasonal (around 0.5 m/s less), and higher over the oceans than inland. Conversely to DJF, in JJA the maximum variability is centered in the northern tropical regions and the southern Antarctic Ocean. In the tropical Atlantic, around 10 degrees north, we observe a remarkable disparity between the inter-annual and intra-seasonal aggregation, since the latter is much more variable than the former. In the other zones, though there are some differences, they are not that important.

The differences between the ERA-Interim and S4 range between 0.5 and -0.5 m/s for the 95% of the grid-points with no predominant global sign, either for the inter-annual or the intra-seasonal time frame (**Figure 5**e and **Figure 5**f). At inter-annual level the differences are less pronounced than for the latter, but holding similar structure in both. In this case the interesting areas are in the inter-tropics, especially in the maritime continent (where there is a combination of over and underestimation from the S4), the Bengal Bay (where S4 overestimates the wind speed values) and the equatorial Atlantic and Pacific Oceans (S4 shows mixed behavior in the Pacific and overestimates in the Atlantic).

Figure 4. *10m wind speed standard deviation for DJF.*

This figure depicts 10m wind speed standard deviation for DJF spanning the period 1981-2015 ***a)*** *ERA-Interim inter-annual* ***b)*** *ERA-Interim intra-seasonal* ***c)*** *S4 November start date inter-annual* ***d)*** *S4 November start date intra-seasonal* ***e)*** *Result of comparing* ***a) minus c)*** *maps* ***f)*** *Result of comparing b) minus d) maps. Note these figures have different colorbars.*

Standard deviation - JJA

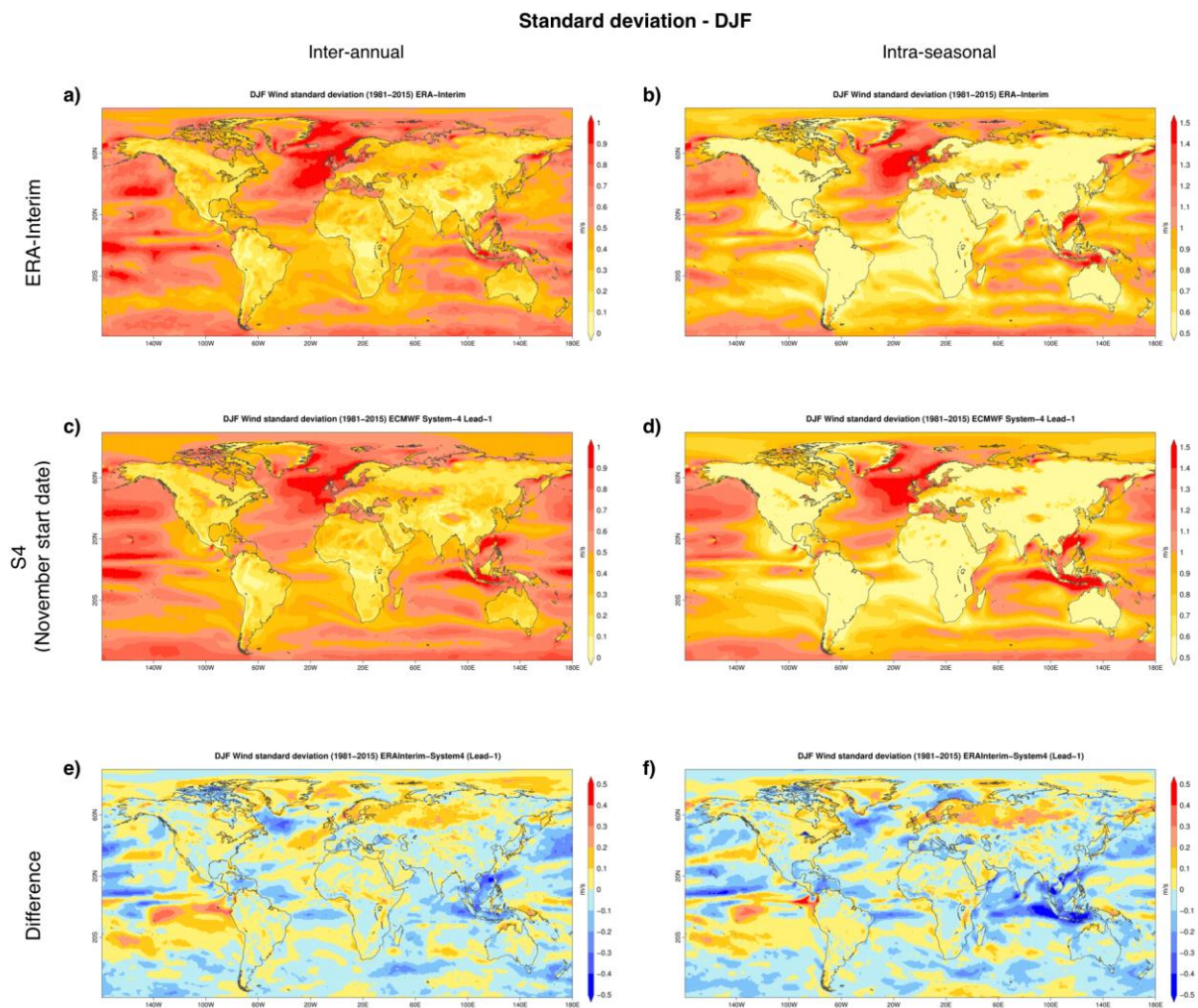Inter-annual   Intra-seasonal

Figure 5. 10m wind speed standard deviation for JJA.

*This figure depicts 10m wind speed standard deviation for JJA spanning the period 1981-2015 **a)** ERA-Interim inter-annual **b)** ERA-Interim intra-seasonal **c)** S4 May start date inter-annual **d)** S4 May start date intra-seasonal **e)** Result of comparing **a) minus c)** maps **f)** Result of comparing b) minus d) maps. Note these figures have different colorbars.*
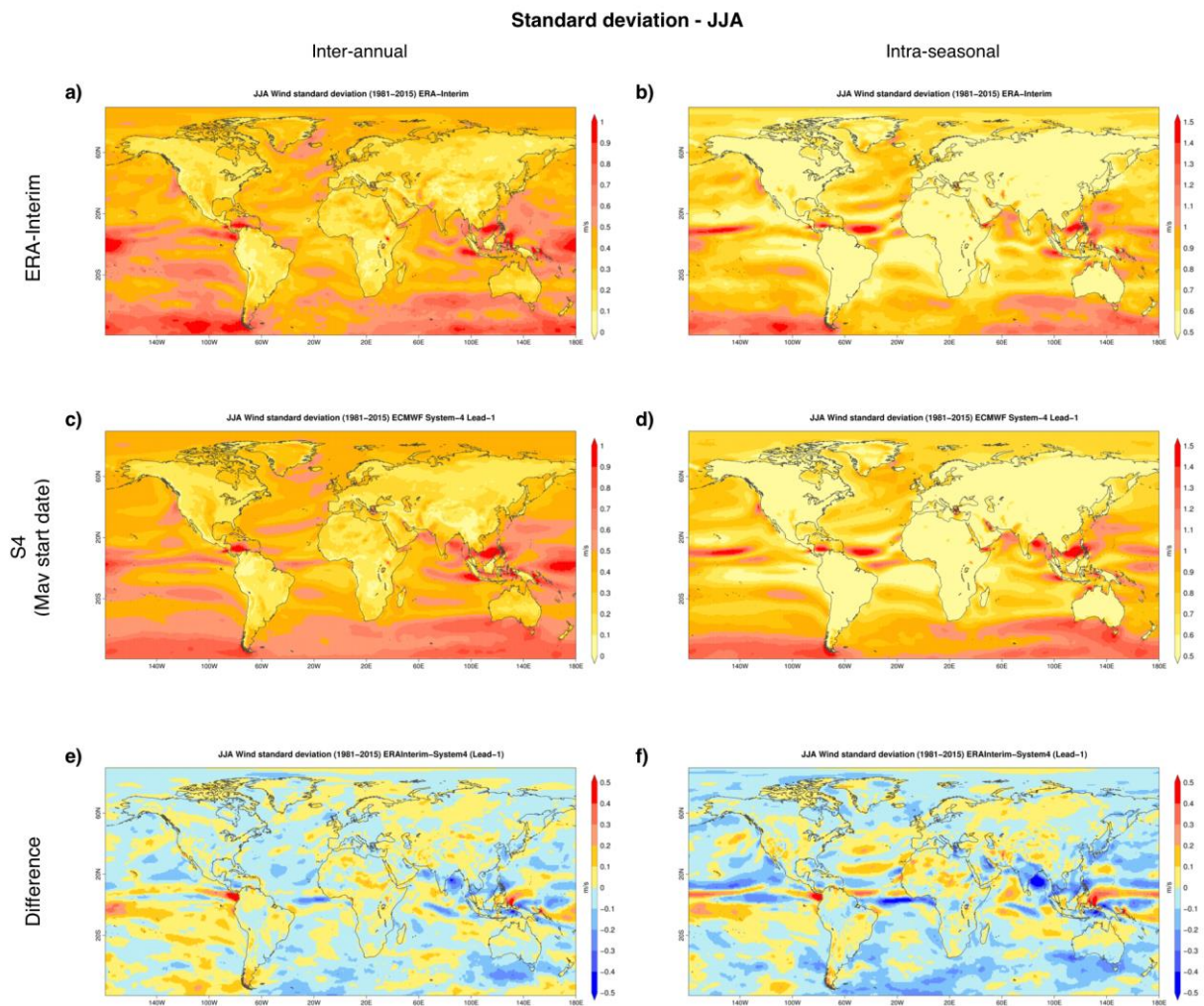
3.3. Coefficient of variation

The coefficient of variation comprises the information about the climatology and standard deviation showed in the 3.1. and 3.2. sections respectively in one single measure.

3.3.1. DJF

The spatial patterns of the ERA-Interim and S4 are similar (**Figure 6**a, **Figure 6**b, **Figure 6**c and **Figure 6**d). Looking at the inter-annual and intra-seasonal coefficients for 95% of the grid-points, we see that intra-seasonally the values range from 8 to 20, whereas at inter-annual level they are restricted from 4.5 to 11.5. This is normal since the variability between months within the same season tends to be higher than the inter-annual. The greatest values are seen in the southern tropical Pacific and Indian Oceans and, also, in Europe, the Greenland coast and western European regions.

The differences between the ERA-Interim and S4 range between -2.5 and 2.5 for the 95% of the grid-points in the inter-annual time-scale, and between -3.5 and 3.5 for the intra-seasonal time frame (**Figure 6**e and **Figure 6**f). Over land the S4 underestimates the coefficient of variation. We can identify two remarkable regions. On one side, the inter-tropics are where the greatest discrepancies can be observed. Actually, in the Pacific Ocean the S4 overestimates wind speeds at inter-annual and intra-seasonal level, whereas in the Atlantic and Indian Oceans there is an S4 overestimation. The other significant area is a strip of land around 60 degrees north, from Scandinavia to Siberia, where the S4 underestimates wind speeds.

3.3.2. JJA

In this season the patterns of the ERA-Interim and S4 are also comparable (**Figure 7**a, **Figure 7**b, **Figure 7**c and **Figure 7**d). Looking at the inter-annual and intra-seasonal coefficients we see that for the 95% of the grid-points the former oscillates from 4.5 to 11.5 whereas the latter show much higher values, ranging from 8 to 20. In JJA, conversely to DJF, the largest values are found only in the inter-tropical oceans, generally to the north of the Equator.

The differences between the ERA-Interim and S4 range between -2.5 and 2.5 for the 95% of the grid-points in the inter-annual time-scale, and between -3.5 and 3.5 for the intra-seasonal time frame (**Figure 7**e and **Figure 7**f). Over land the S4 underestimates the coefficient of variation. In the inter-tropics where the magnitude differences are greater, there is an alternation of areas where the S4 overestimates and underestimates wind speeds. However, in the Pacific there is a dominant tendency to underestimate wind from the S4, whereas in the Atlantic the behavior is the opposite. In the continents, though, S4 is prone to underestimate wind speeds, except in the Indian area, where it overestimates them.
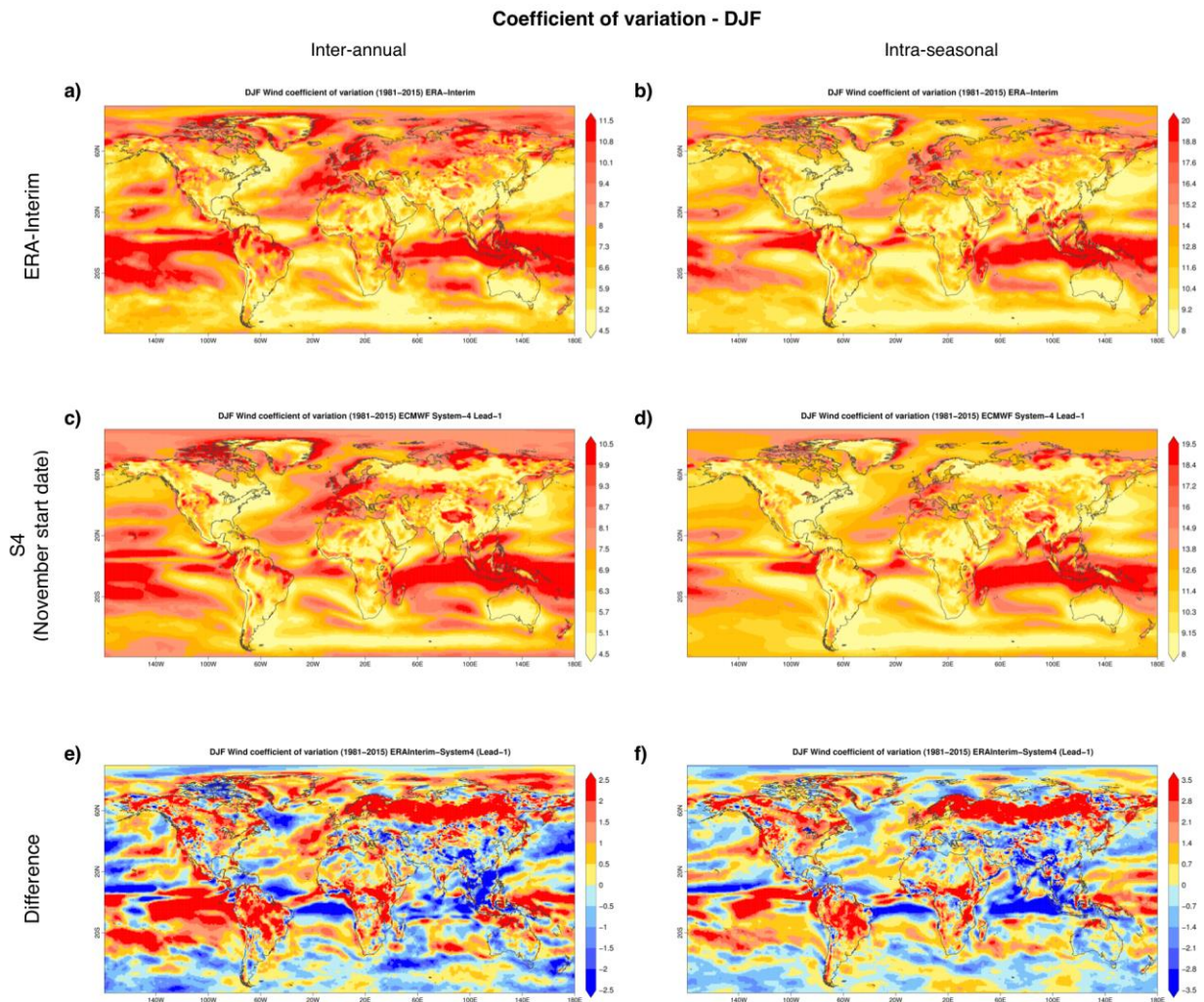
Figure 6. *10m wind speed coefficient of variation for DJF.*

This figure depicts 10m wind speed coefficient of variation for DJF spanning the period 1981-2015 **a)** *ERA-Interim inter-annual* **b)** *ERA-Interim intra-seasonal* **c)** *S4 November start date inter-annual* **d)** *S4 November start date intra-seasonal* **e)** *Result of comparing* **a) minus c)** *maps* **f)** *Result of comparing b) minus d) maps. Note these figures have different colorbars.*
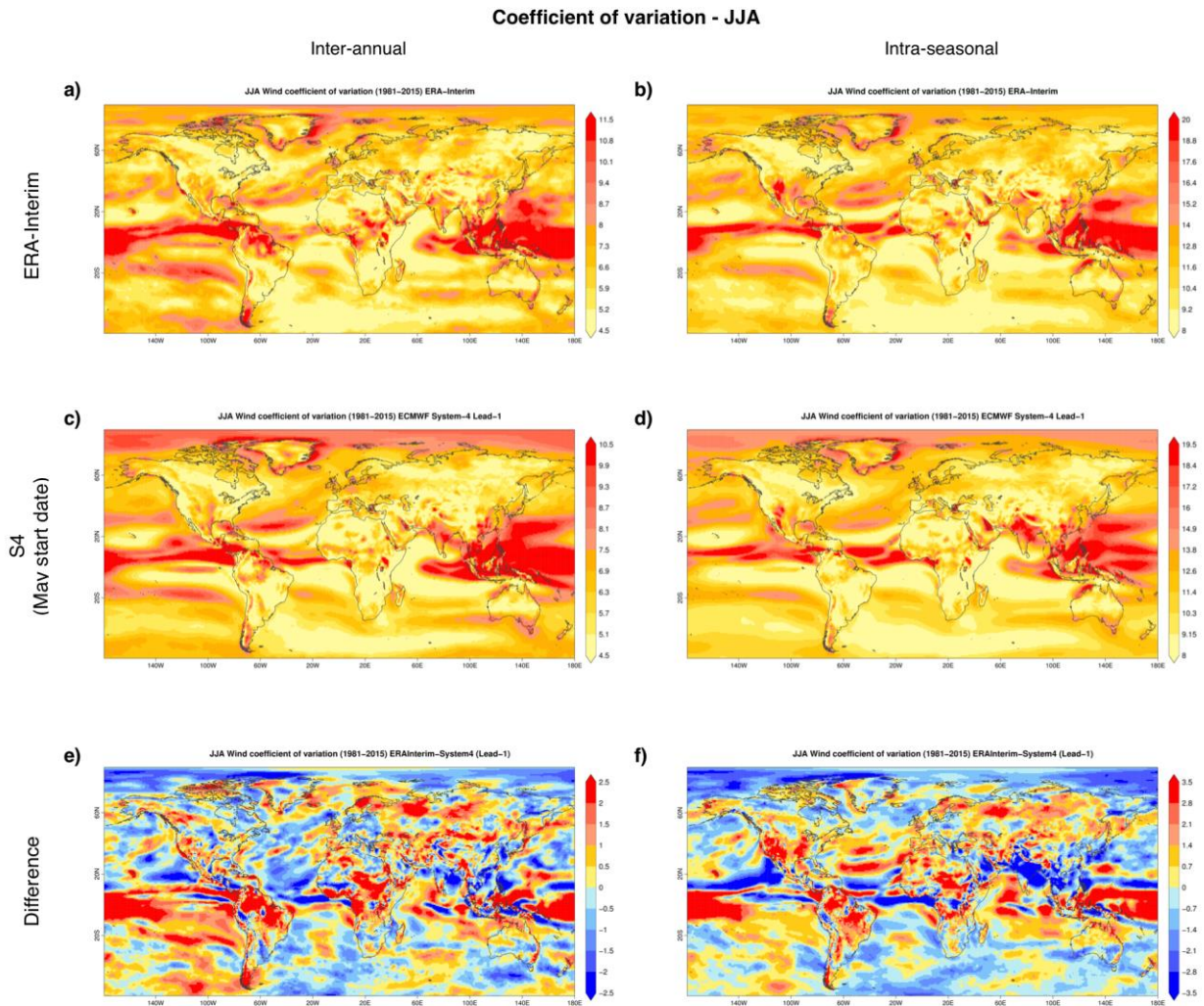
Figure 7. *10m wind speed coefficient of variation for JJA.*

This figure depicts 10m wind speed coefficient of variation for JJA spanning the period 1981-2015 ***a)*** *ERA-Interim inter-annual* ***b)*** *ERA-Interim intra-seasonal* ***c)*** *S4 May start date inter-annual* ***d)*** *S4 May start date intra-seasonal* ***e)*** *Result of comparing* ***a) minus c)*** *maps* ***f)*** *Result of comparing b) minus d) maps. Note these figures have different colorbars.*

3.4. Skewness

3.4.1. DJF

The skewness patterns of ERA-Interim and S4 scarcely match, especially at inter-annual level (**Figure 8**a, **Figure 8**b, **Figure 8**c and **Figure 8**d). They only resemble each other in the intra-seasonal time-scale. Besides, in ERA-Interim there is generally more noise than in the S4 (**Figure 8**e and **Figure 8**f). Positive (negativ**e)** skewness means that the median is lower (greater) than the mean. Inter-annually there is no predominant sign in skewness, while intra-seasonally it seems there is certain prevalence of positive values, especially over land (**Figure 8**f). The differences between ERA-Interim and S4 range between -0.5 and 0.5 for the 95% of the grid-points both at inter-annual and intra-seasonal time frames. The patterns are very noisy, being these differences greater inter-annually than intra-seasonally.

3.4.2. JJA

In this season ERA-Interim and S4 does not match either (**Figure 9**a, **Figure 9**b, **Figure 9**c and **Figure 9**d). They only resemble each other in the intra-seasonal time-scales (**Figure 9**b and **Figure 9**d). Besides, in ERA-Interim there is generally more noise than in the S4, and also, more noise inter-annually than intra-seasonally. At inter-annual level there is no predominant sign in skewness (**Figure 9**a and **Figure 9**c). Intra-seasonally, although there is no prevalent worldwide sign, there is also some predominance of positive skewness over land (**Figure 9**b and **Figure 9**d). The differences between ERA-Interim and S4 range between -0.5 and 0.5 for the 95% of the grid-points at both aggregations (**Figure 9**e and **Figure 9**f). The patterns are still very noisy, being those differences greater inter-annually than intra-seasonally.
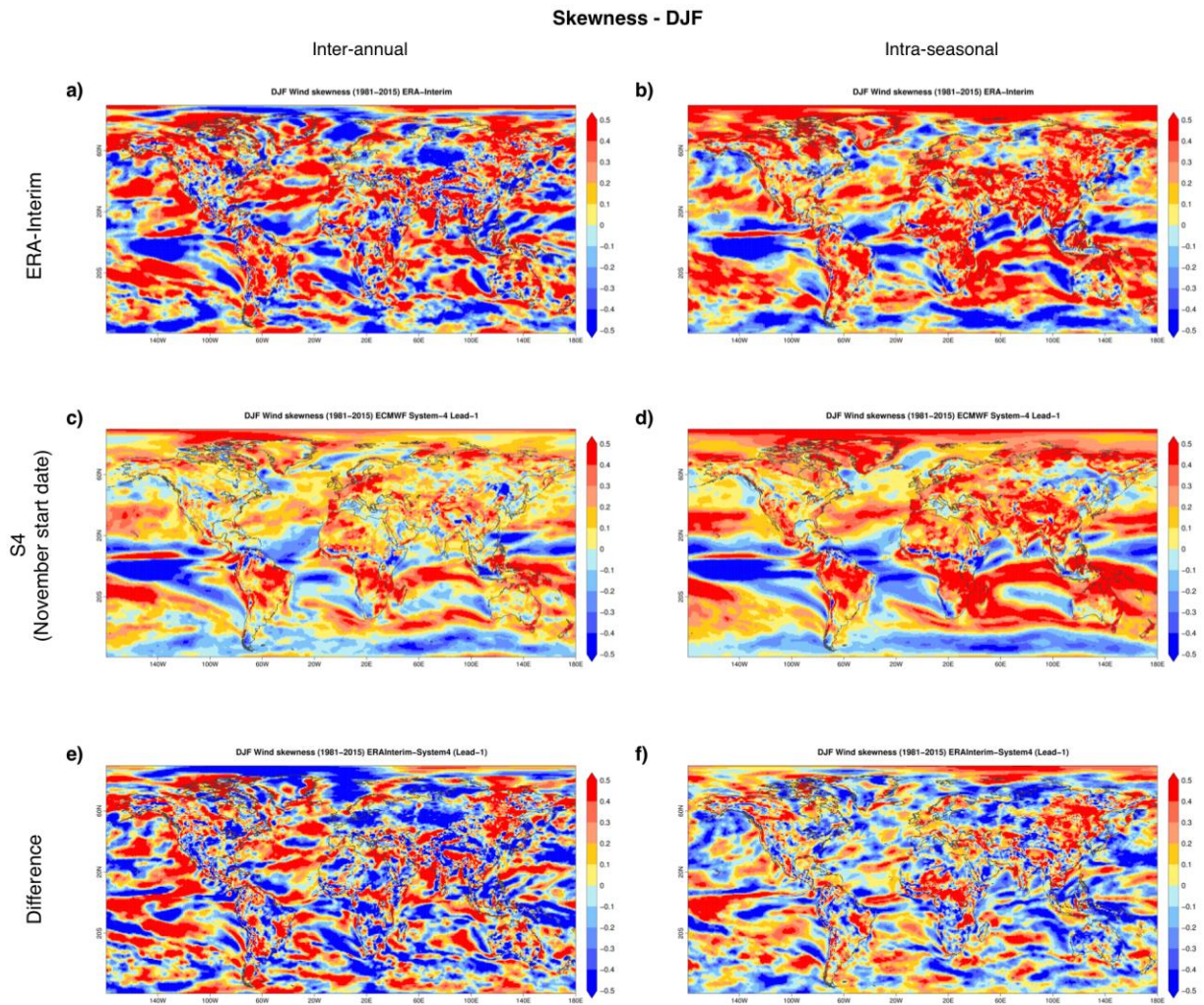
Figure 8. *10m wind speed skewness for DJF.*

*This figure depicts 10m wind speed skewness for DJF spanning the period 1981-2015 **a)** ERA-Interim inter-annual **b)** ERA-Interim intra-seasonal **c)** S4 November start date inter-annual **d)** S4 November start date intra-seasonal **e)** Result of comparing **a) minus c)** maps **f)** Result of comparing b) minus d) maps. Note these figures have different colorbars.*
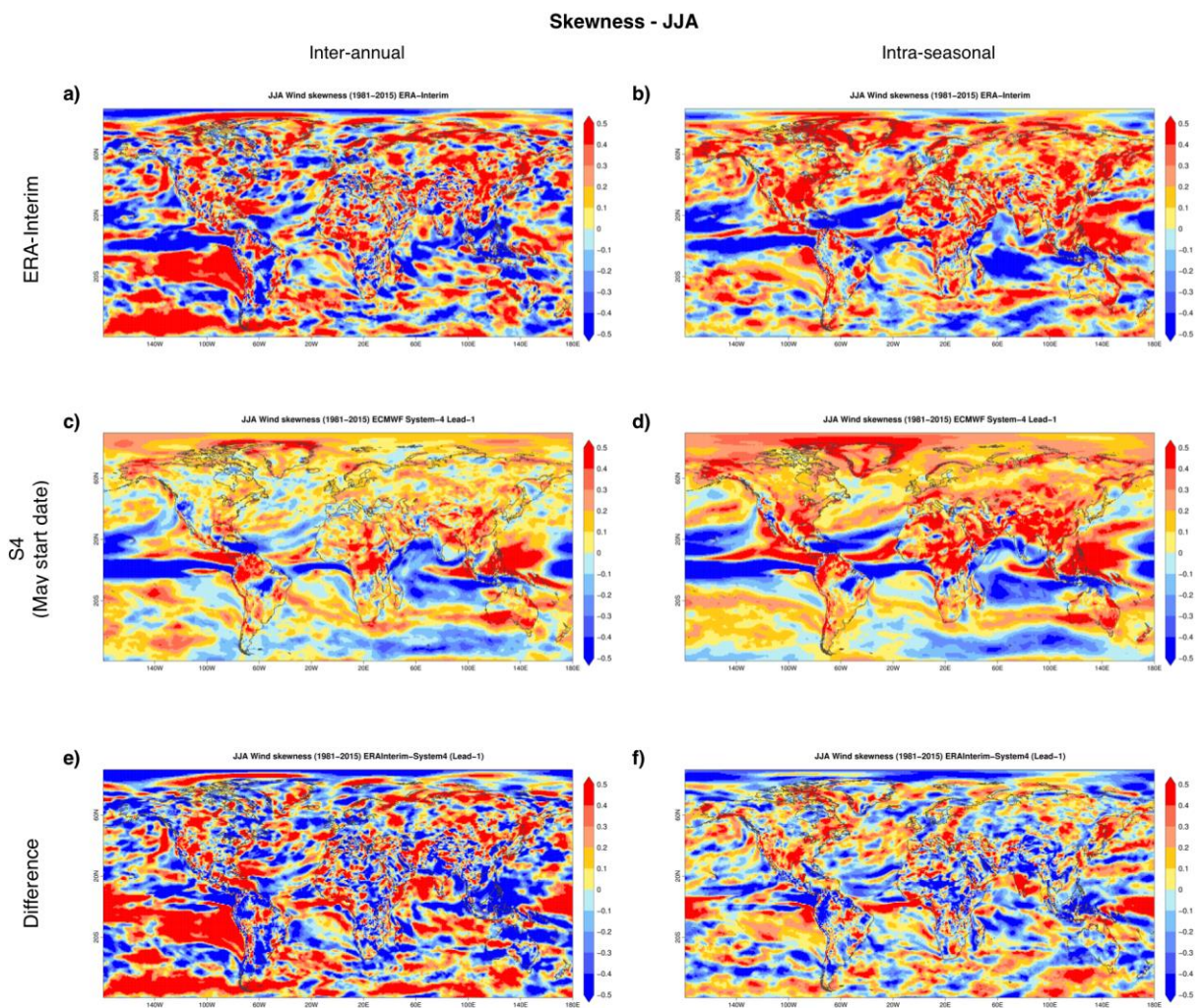
Figure 9. *10m wind speed skewness for JJA.*

*This figure depicts 10m wind speed skewness for JJA spanning the period 1981-2015 **a)** ERA-Interim inter-annual **b)** ERA-Interim intra-seasonal **c)** S4 May start date inter-annual **d)** S4 May start date intra-seasonal **e)** Result of comparing **a) minus c)** maps **f)** Result of comparing b) minus d) maps.*

3.5. Kurtosis

3.5.1. DJF

When analyzing the excess kurtosis of ERA-Interim and S4 there is an obvious disparity between them (**Figure 10**a, **Figure 10**b, **Figure 10**c and **Figure 10**d). Although they both show noisy maps, in ERA-Interim there is generally more noise than in the S4, and there is also more noise at inter-annual level than intra-seasonally. In neither time scale the excess kurtosis shows a predominant global sign. We have to remember that the excess kurtosis refers to the weight of the tails of the distribution. In a positive (negativ**e)** excess kurtosis the weight of the tails is lower (higher) than for the Gaussian distribution. That is, the probability of an extreme is higher (lower) than for the Gaussian distribution. Negative values of the kurtosis obtained in the tropical Atlantic for S4 at interannual scales indicates that the wind speed distribution is sharper than the normal distribution so most of the values are very close to the median therefore the probability of wind extremes occurrence in that region is lower than in a typical Gaussian distribution.

The differences between the ERA-Interim and S4 range between -1.5 and 1.5 for the 95% of the grid-points inter-annually, and between -1 and 1 intra-seasonally (**Figure 10**e and **Figure 10**f). This is because the S4 has a more restricted range of excess kurtosis values at inter-annual and intra-seasonal time frames than ERA-Interim. Besides, there are no predominant signs or structures in these differences.

3.5.2. JJA

In this season the analysis of the excess kurtosis is similar to DJF. There is a lot of noise both for ERA-Interim and S4 and, additionally, they do not show the same structures (**Figure 11**a, **Figure 11**b, **Figure 11**c, and **Figure 11**d). They only resemble each other a little in the inter-tropics intra-seasonally (**Figure 11**c, and **Figure 11**d). Besides, in ERA-Interim there is generally more noise than in the S4, and also, more noise at inter-annual level than intra-seasonally. The differences between the ERA-Interim and S4 range between -1.5 and 1.5 for the 95% of the grid-points inter-annually, and between -1 and 1 intra-seasonally (**Figure 11**e and **Figure 11**f). Again this is because the S4 has a more restricted range of excess kurtosis values at inter-annual and intra-seasonal time frames than ERA-Interim.
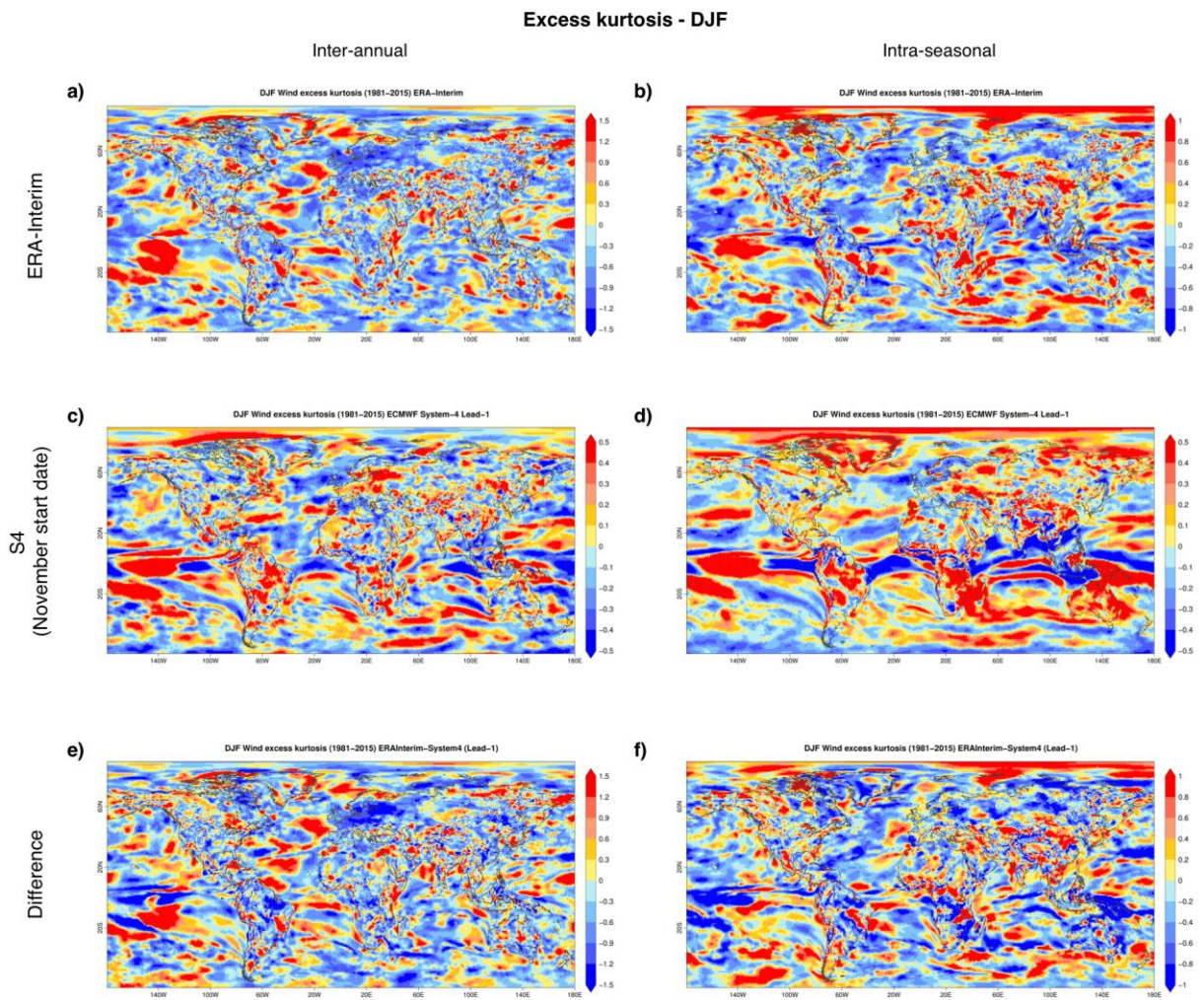
Excess kurtosis - DJF

Figure 10. 10m wind speed excess kurtosis for DJF.

*This figure depicts 10m wind speed excess kurtosis for DJF spanning the period 1981-2015 **a)** ERA-Interim inter-annual **b)** ERA-Interim intra-seasonal **c)** S4 November start date inter-annual **d)** S4 November start date intra-seasonal **e)** Result of comparing **a) minus c)** maps **f)** Result of comparing b) minus d) maps. Note these figures have different colorbars.*
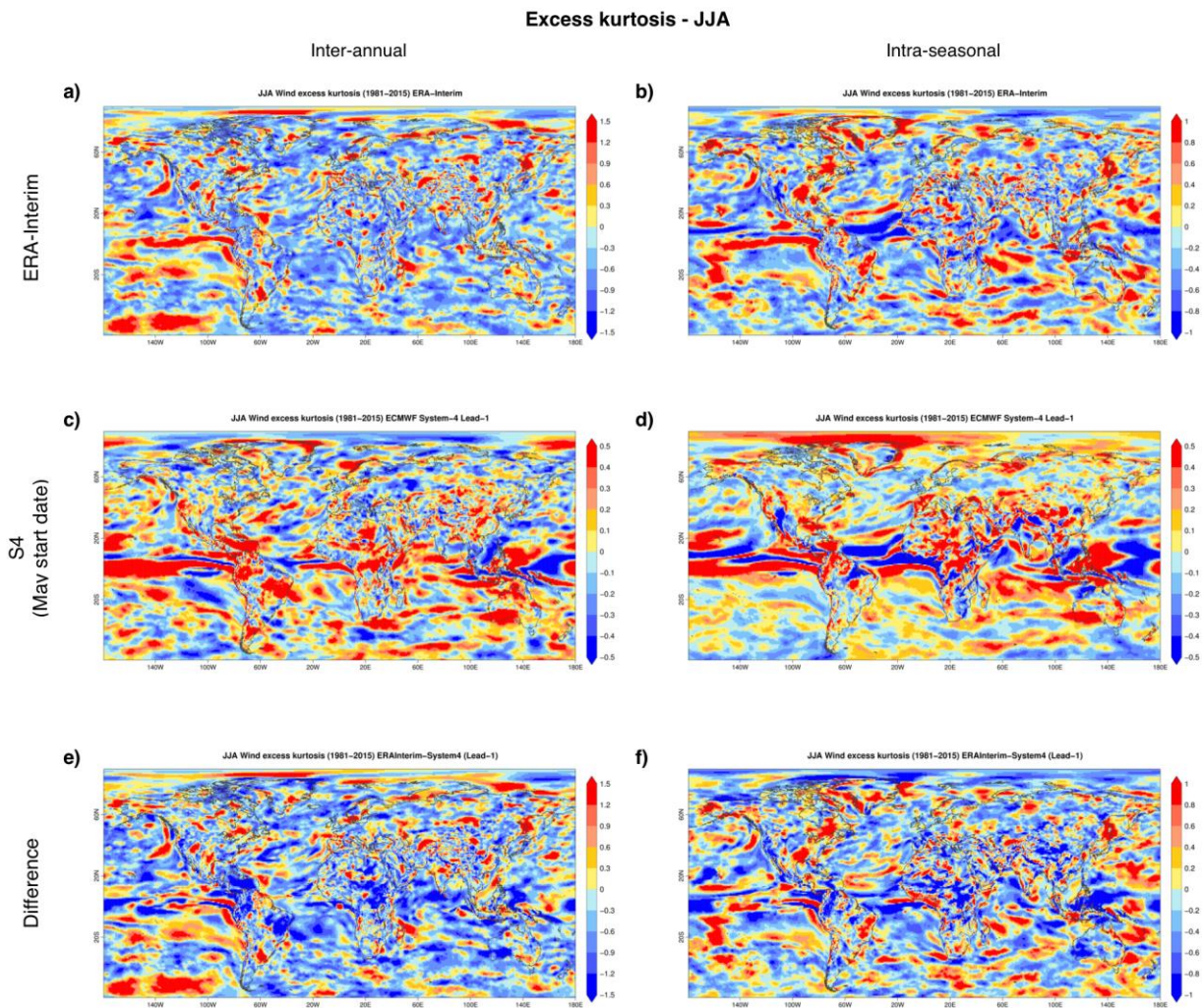
Excess kurtosis - JJA

Figure 11. 10m wind speed excess kurtosis for JJA.

This figure depicts 10m wind speed excess kurtosis for JJA spanning the period 1981-2015 **a)** ERA-Interim inter-annual **b)** ERA-Interim intra-seasonal **c)** S4 May start date inter-annual **d)** S4 May start date intra-seasonal **e)** Result of comparing **a) minus c)** maps **f)** Result of comparing b) minus d) maps. Note these figures have different colorbars.

3.6. Goodness of fit test

We have evaluated whether the 10m wind speed follows a Gaussian distribution by applying a Shapiro-Wilks normality test. In this test the null hypothesis is that the underlying distribution is of Gaussian type. We have established our significance level at 95% meaning that when the p_{val} is equal or below 0.05 we have to reject the null hypothesis.

In ERA-Interim the regions that we cannot regard Gaussian are the inter-tropical, especially at intra-seasonal time scales and for JJA (**Figure 12**b and **Figure 13**a and **Figure 13**b). In DJF these areas are more scarce and scattered around the globe (**Figure 12**a and **Figure 12**b). In the remaining regions we cannot discard normality. In S4 there are much more territories with p_{val} under 0.05, that is, where we can reject the null hypothesis (**Figure 12**c, **Figure 12**d and **Figure 13**c and **Figure 13**d). This is especially true for the intra-seasonal time scales where only some areas in the northern and southern extra-tropics do not reject the null hypothesis (**Figure 12**d and **Figure 13**d). At an inter-annual level, these regions are wider in the extra-tropics (**Figure 12**c and **Figure 13**c).

The number of values (sample size) entering the Shapiro-Wilks test can explain the observed differences between ERA-Interim/S4 and inter-annual/intra-seasonal time-scales. This value affects the power of the test and, thus, the p_{val}. For instance, in the intra-seasonal configuration there is 3 fold more data entering the test than for the inter-annual case. Furthermore, when considering ERA-Interim and S4 there are 15 times more values in the latter (one for each ensemble member).
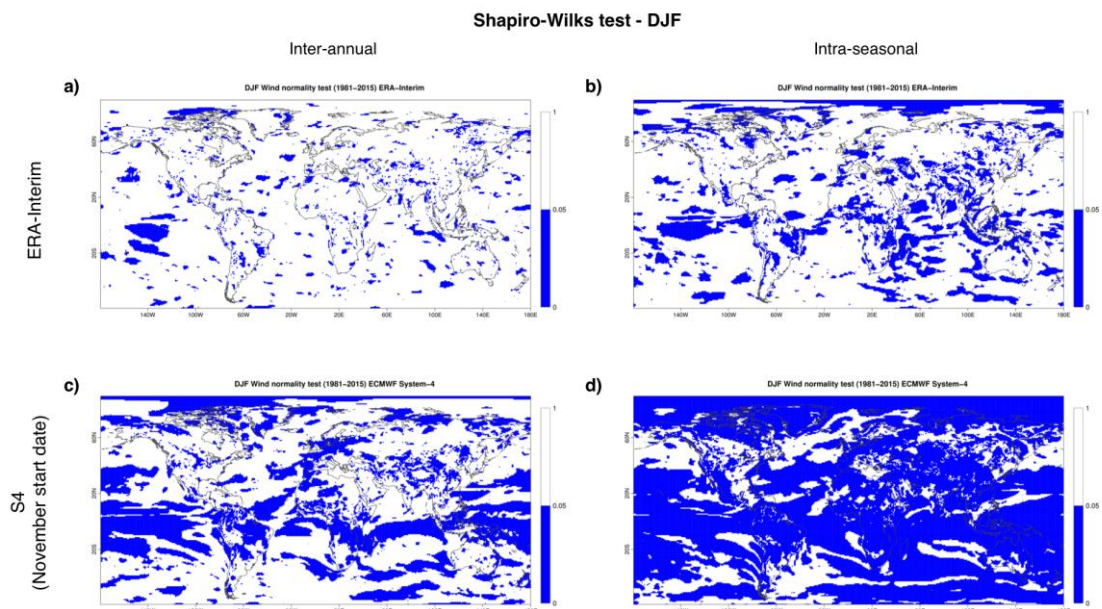

Figure 12. *10m wind speed Shapiro-Wilks goodness of fit test for DJF.*

*This figure depicts 10m wind speed Shapiro-Wilks goodness of fit test for DJF spanning the period 1981-2015 **a)** ERA-Interim inter-annual **b)** ERA-Interim intra-seasonal **c)** S4 November start date inter-annual **d)** S4 November start date intra-seasonal.*
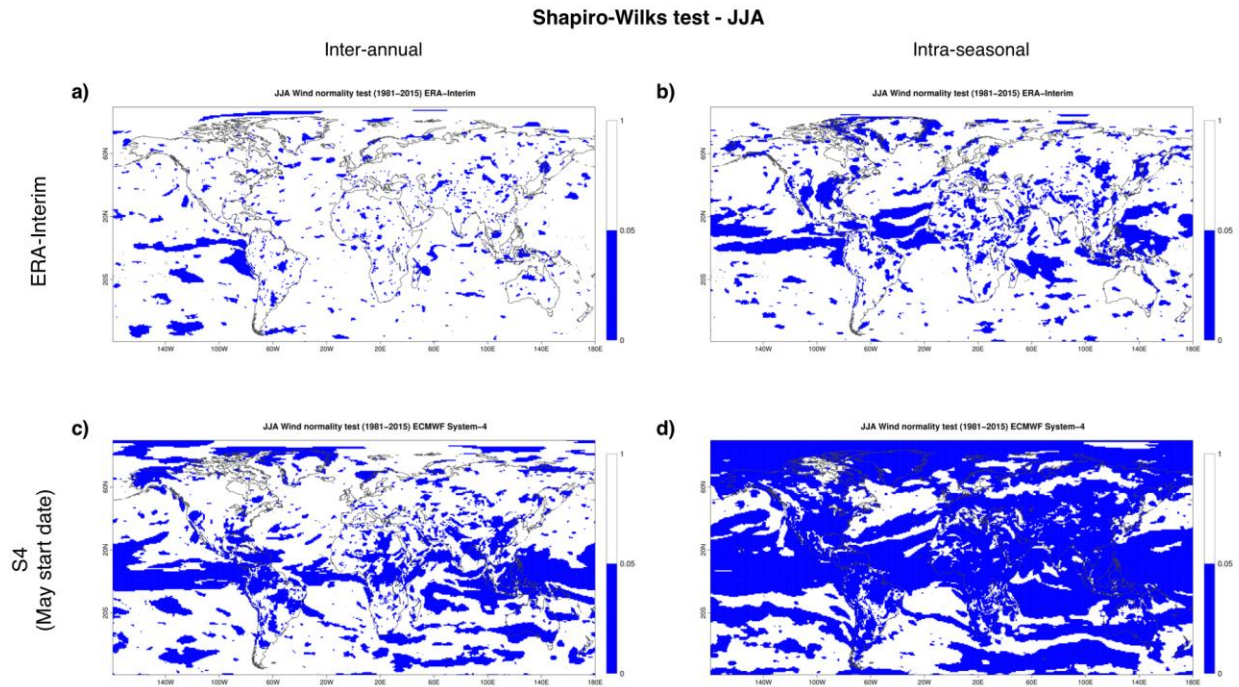
Figure 13. *10m wind Shapiro-Wilks goodness of fit test for JJA.*

*This figure depicts 10m wind speed Shapiro-Wilks goodness of fit test for JJA spanning the period 1981-2015 **a)** ERA-Interim inter-annual **b)** ERA-Interim intra-seasonal **c)** S4 May start date inter-annual **d)** S4 May start date intra-seasonal*

4. Conclusions

In this work we have outlined the main characteristics of the 10m wind speed distribution from ERA-Interim reanalysis and ECMWF System-4 seasonal prediction system. We have applied a set of six statistical diagnostics: mean value, standard deviation, kurtosis, skewness, coefficient of variation, and goodness of fit Shapiro-Wilks test; and studied them seasonally at inter-annual and intra-seasonal time scales. Additionally, we have computed differences between the ERA-Interim and S4 parameters to locate discordant behaviours. Even though the assessment covers all the seasons, in this report we have only commented the results for DJF and JJA for 10m wind speed. The remaining maps for 10m wind speed and 2m temperature are available the Earth System Services Catalogue (www.bsc.es/ESS/catalogue).

This diagnostic work is part of the studies aimed to characterize 10m wind speed and 2m temperature in the RESILIENCE project. Hence, it helps to locate the hot-spot regions for the study of wind from a variability standpoint, which differ depending on the parameter considered. It also aids in the identification of areas that may need special attention from a bias correction perspective. Besides, it warns us about locations where normality dependent methods should be either carefully applied or completely avoided.

We have found the climatologies represented by the ERA-Interim reanalysis and S4 are very close to each other (95% of the grid points have differences between 1 and -1 m/s), correctly attributing largest speeds over the oceans and the lowest values over landmasses. Actually, the speed over land is generally less than half the observed over the oceans. The bigger discrepancies have been located in the inter-tropical regions. It is worth mentioning that S4 systematically overestimates wind speed at global scales, a somewhat expected result since seasonal forecasting models often show bias-related difficulties.

Regarding wind speed variability, we have seen that there are higher spreads over the ocean than inland, both in DJF and JJA, and for the two used datasets. The variability is also larger at intra-seasonal scales, especially for the inter-tropics. As for the regions with the greater variability differences, we can outline the inter-tropical areas, although the extra-tropics should also be highlighted depending on the season. In this case, however, there is no clear prevailing sign in the differences.

The coefficient of variation is an interesting parameter because it relates the magnitude of the variability to the mean wind value. Thus, not only lets us compare distinct wind regime regions, but also it is an indicator for identifying areas with lower variability compared to the mean, which is important when studying appropriate spots for wind-farm installation. Similarly to the former statistical parameters, the coefficient of variation also presents its larger values in the inter-tropics and over the oceans for DJF/JJA (in ERA-Interim and S4). Concerning inter-annual/intra-seasonal differences, we find much higher values in the intra-seasonal maps than in their inter-annual counterparts, and additionally, greater in DJF than in JJA. The disparity of values between ERA-Interim and S4 is larger in DJF than in JJA, and in the inter-tropical regions than over the continents.

Finally, regarding the goodness of fit to a normal distribution and the related kurtosis and skewness moments, the inter-tropical areas and the intra-seasonal time-scales cannot be regarded normally fitted. For the extra-tropical, this is only true at intra-seasonal scale. However, the way in which this normality is violated is different depending on the season, the time scale and the dataset.

It is possible that some of these differences are consequence of the characteristics of the S4 or the number of values entering the Shapiro-Wilks goodness-of-fit test. In fact, in the case of skewness and kurtosis the S4 tends to smooth out the noise found in ERA-Interim, probably due to the relatively limited variability of the distributions depicted by climate models. Besides, in the intra-seasonal configuration there is 3 fold more data entering the Shapiro-Wilks test than for the inter-annual case and when considering ERA-Interim and S4 there are 15 times more values in the latter due to the ensemble size.

5. Acknowledgements

6. References

Achuthavarier, D. & Krishnamurthy, V., 2010. Relation between intraseasonal and interannual variability of the South Asian monsoon in the National centers for environmental predictions forecast systems. *Journal of Geophysical Research Atmospheres*, 115(8).

Andres, N. et al., 2014. Water resources and climate change impact modelling on a daily time scale in the Peruvian Andes. *Hydrological Sciences Journal*, 59(11), pp.2043–2059.

Archer, C.L., 2005. Evaluation of global wind power. *Journal of Geophysical Research*, 110(D12), pp.1–20.

Cornes, R.C. & Jones, P.D., 2013. How well does the ERA-Interim reanalysis replicate trends in extremes of surface temperature across Europe? *Journal of Geophysical Research Atmospheres*, 118(18), pp.10262–10276.

Dee, D.P. et al., 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656), pp.553–597.

Giannakopoulos, C. & Psiloglou, B.E., 2006. Trends in energy load demand for Athens, Greece: Weather and non-weather related factors. *Climate Research*, 31(1), pp.97–108.

Hekkenberg, M. et al., 2009. Indications for a changing electricity demand pattern: The temperature dependence of electricity demand in the Netherlands. *Energy Policy*, 37(4), pp.1542–1551.

Kiss, P., Varga, L. & Jánosi, I.M., 2009. Comparison of wind power estimates from the ECMWF reanalyses with direct turbine measurements. *Journal of Renewable and Sustainable Energy*, 1(3), p.33105.

Lorenz, T. & Barstad, I., 2016. A dynamical downscaling of ERA-Interim in the North Sea using WRF with a 3km grid-for wind resource applications. *Wind Energy*.

Lu, X., McElroy, M.B. & Kiviluoma, J., 2009. Global potential for wind-generated electricity. *Proceedings of the National Academy of Sciences of the United States of America*, 106(27), pp.10933–10938.

Luo, D., Diao, Y. & Feldstein, S.B., 2011. The Variability of the Atlantic Storm Track and the North Atlantic Oscillation: A Link between Intraseasonal and Interannual Variability. *Journal of the Atmospheric Sciences*, 68(3), pp.577–601.

Marcos, R. et al., 2016. Seasonal predictability of water resources in a Mediterranean freshwater reservoir and assessment of its utility for end-users. *Science of The Total Environment*, 575, pp.681–691.

Molteni, F., Stockdale, T., Balmaseda, M.A., et al., 2011. The new ECMWF seasonal forecast system (System 4). *ECMWF Technical Memorandum*, 656(November), p.49.

Molteni, F., Stockdale, T., Balmaseda, M., et al., 2011. *The new ECMWF seasonal forecast*

system (System 4), European Centre for Medium-Range Weather Forecasts.

Ogutu, G., Supit, I. & Hutjes, R., 2016. Probabilistic maize yield simulation over East Africa using ensemble seasonal climate forecasts. In *Geophysical Research Abstracts EGU General Assembly*. pp. 2016–17111.

Rose, S. & Apt, J., 2015. What can reanalysis data tell us about wind power? *Renewable Energy*, 83, pp.963–969.

Ruffault, J. et al., 2013. Projecting future drought in Mediterranean forests: Bias correction of climate models matters! *Theoretical and Applied Climatology*, 117(1), pp.113–122.

Shapiro, S.S. & Wilk, M.B., 1965. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3), pp.591–611.

Škerlak, B., Sprenger, M. & Wernli, H., 2014. A global climatology of stratosphere-troposphere exchange using the ERA-Interim data set from 1979 to 2011. *Atmospheric Chemistry and Physics*, 14(2), pp.913–937.

Stockdale, T.N. et al., 2011. ECMWF seasonal forecast system 3 and its prediction of sea surface temperature. *Climate Dynamics*, 37(3), pp.455–471.

Tompkins, A.M. & di Giuseppe, F., 2015. Potential predictability of malaria in Africa using ECMWF monthly and seasonal climate forecasts. *Journal of Applied Meteorology and Climatology*, 54(3), pp.521–540.

Torralba-fernandez, V., Davis, M. & Doblas-reyes, F.J., 2015. State-of-the-Art Climate Predictions for Energy Climate Services. In *BSC International Doctoral Symposium - 2nd BSC International Doctoral Symposium, Barcelona, 5th-7th May, 2015*. Barcelona: BSC, p. 5468.

Trenberth, K.E., Koike, T. & Onogi, K., 2008. Progress and prospects for reanalysis for weather and climate. *Eos*, 89(26), pp.234–235.

Uppala, S.M. et al., 2005. The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, 131(612), pp.2961–3012.

Wilks, D.S., 2006. *Statistical methods in the atmospheric sciences*, London: Academic Press.