Adaptive Smoothing of the Ensemble Mean of Climate Model Output for Improved Projections of Future Rainfall

Stephen Jewson[1], Giuliana Barbato[2], Paola Mercogliano[2], Maximiliano Sassi[3]

[1]Lambda Climate Research Ltd, W4 3HR, London, UK

5   [2]Euro-Mediterranean Center on Climate Change (CMCC) Foundation, Via Augusto Imperatore, 16, 73100, Lecce, Italy

[3]RMS Ltd, London, UK

Correspondence to: Stephen Jewson (stephen.jewson@gmail.com)

Abstract. Ensemble simulations of future climate can be described as consisting of a forced climate change response and noise, where the noise arises from internal variability and errors in the different models. In the ensemble mean the noise is

10   reduced, making it easier to identify the mean of the forced response. The noise in the ensemble mean can potentially be reduced further by spatial smoothing, and this potential has been explored by previous authors. Depending on the variable, the resolution and the size of the ensemble it has been reported that the benefit of spatial smoothing of the ensemble mean may be small, and that spatial smoothing may have the unwanted side-effect that it modifies genuine features in the forced response. However, the spatial smoothing methods that have been tested previously used the same degree of smoothing at all

15   locations, which limits their effectiveness. We derive a novel adaptive smoothing methodology for the ensemble mean that utilizes ensemble information with respect to signal, uncertainty and spatial correlations in order to vary the degree of smoothing in space. The methodology corresponds to simple intuitive concepts, such as the idea that locations with higher signal to noise ratio should be smoothed less. We apply the method to EURO-CORDEX simulations of future annual mean rainfall, and by using cross-validation within the ensemble are able to demonstrate a three times greater increase in potential

20   predictive accuracy than from the non-adaptive smoothing methods we compare with. The adaptive smoothing method also preserves sharp features in the ensemble mean to a greater extent than the non-adaptive methods. We conclude that adaptive smoothing may be a useful post-processing tool for improving the potential accuracy of climate projections.

1 Introduction

In this article we revisit the question of whether it may be worth applying spatial smoothing to climate model ensemble

25   means to try and derive more precise estimates of the mean of the forced climate response. This question has been considered in detail by previous authors, in particular Masson and Knutti (2011), henceforth MK, and Raisanen and Ylhaisi (2010), henceforth RY. Based on MK and RY we know that applying smoothing to the ensemble mean may make sense, and may lead to estimates with greater precision (potential accuracy), for two reasons. First, climate models cannot be expected to simulate realistic spatial variability at the scale of the model resolution, which leads to errors at that scale. These errors

30   can potentially be reduced with appropriate smoothing. Second, climate model ensemble means are affected by internal variability and errors in the individual climate model simulations in the ensemble, and spatial smoothing can potentially remove some of the impact of this variability and these errors from the ensemble mean.

MK and RY also discuss reasons why applying spatial smoothing to the ensemble mean may not be beneficial. Calculating the ensemble mean already reduces the effects of both errors and internal variability. Also, since it is difficult for spatial

35   smoothing methods to distinguish between the forced climate response and the noise, the smoothing may inadvertently degrade the estimate of the forced climate response to some extent. This is an unwanted side-effect and may offset any potential benefit from the smoothing.

Whether or not smoothing is beneficial has been tested in practice by MK and RY. MK applied Gaussian and circular top-hat smoothers to CMIP3 (Meehl et al. 2007) temperature and rainfall, for both present and future climate. For present climate

40   they compared smoothed model results with observations. They reported that one downside of smoothing is that it creates undesirable fuzziness around features in the ensemble mean. RY also applied Gaussian smoothers to present and future climate and compared with observations. For future climate, they tested the performance of their smoothing methods using cross-validation within the ensemble. They reported that optimal smoothing length-scales were much smaller for the ensemble mean than for individual models and found only small increases in potential accuracy when smoothing the

45   ensemble mean. They discussed fundamental limits that apply to how well smoothing can improve potential accuracy and concluded that there is perhaps no great benefit to be obtained from spatial smoothing of the ensemble mean for large ensembles of independent models.

The results presented in MK and RY were based on spatial-smoothing methods in which the same amount of smoothing is used in all locations, which we will call *non-adaptive* methods, since nothing about the amount of smoothing adapts to the

50   data. However, the ensemble itself contains information that we can potentially use to determine how to vary the type or amount of smoothing used from one location to another. In particular, there are four factors that can be derived from the ensemble that, intuitively, one would think might be useful as a way to determine what amount of smoothing to use. The first factor is the ensemble spread at the location at which the forced climate response is being estimated (the target location), which can be estimated using the ensemble members. If the ensemble spread at the target location is narrow then the signal

55   there is already estimated relatively accurately by the ensemble mean, and less spatial smoothing might be needed than for other target locations where the spread is wider. The second factor is the ensemble spread at the locations surrounding the target location, that are being used as predictors (the predictor locations). If the ensemble spread at the predictor locations is narrow then the signal is estimated relatively accurately at those locations and hence putting more weight on those predictors might be beneficial. The third factor is the correlation between the variability in the ensemble at the target location and the

60   predictor locations, which can also be estimated using the ensemble members. If the correlation between the target location and the predictor locations is high, that perhaps suggests that putting more weight on the predictor locations might be beneficial. The fourth factor is the size of the ensemble mean signal at the target location relative to the size of the ensemble mean signal at the predictor locations. If there is a large difference in the size of the two signals then that suggests that less

smoothing would be beneficial. One might imagine that these factors should be combined in some way e.g., the size of the

65　difference between the target and predictor locations should perhaps be converted into a signal to noise type of parameter using the ensemble spread. However, exactly how to combine all four factors is unclear.

In this article we derive and test a novel *adaptive* smoothing methodology, where adaptive means that a different amount of smoothing is used at each location, depending on the ensemble data. It is based on a simple method that naturally derives and combines the four factors listed above and leads to an equation that specifies the amount of smoothing to be used. The

70　statistical method is novel, and comes from an area of statistics known as Frequentist Model Averaging (FMA: see text-books such as Burnham and Anderson (2002), Claeskens and Hjort (2008), Fletcher (2019)). The method can potentially break the trade-off inherent in non-adaptive smoothing, discussed in RY, in which increasing the smoothing length-scale may improve potential accuracy but will destroy details. It may be able break this trade-off by smoothing less when there are details that the ensemble indicates should be preserved.

75　The methodology is based on an equation that is derived by minimising predictive root mean-squared error (PRMSE) in a bias-variance trade-off. Smoothed estimates are made by estimating the forced climate response as a linear combination of the ensemble means at the target location and the predictor locations. The weights in the linear combination are the adaptive feature of the method and are derived separately at each location from the statistics of the ensemble at that location and nearby locations. We apply the methodology with four different shapes of smoothing kernel. We test the methodology using

80　high resolution EURO-CORDEX projections of annual mean rainfall over Europe, for two RCPs and three points in time. We use cross-validation within the ensemble, following RY, to evaluate the PRMSE performance of non-adaptive and adaptive smoothing methods based on each of the four smoothing kernels. We also evaluate the results visually, to assess the extent to which the different methods preserve sharp features in the ensemble mean.

In section 2 we describe the EURO-CORDEX data, and the cross-validation framework we will use to test the smoothing

85　methodologies. In section 3 we describe the smoothing methodologies we will test. In section 4 we present results in terms of RMSE performance and spatial maps of smoothed fields. In section 5 we conclude.

2 Data and Cross-validation Methodology

2.1 EURO-CORDEX Data

The data we use for this study is annual mean rainfall data extracted from the EURO-CORDEX set of ensemble projections

90　of future climate (Jacob and authors 2014). The EURO-CORDEX ensembles have been used for a large number of studies of European climate in the last decade (e.g., Colmet-Daage et al. (2018), Dalelane et al. (2018), Dyrrdal, Stordal and Lussana (2018), Foley and Kelman (2018), Hosseinzadehtalaei, Tabari and Willems (2018), Mascaro, Viola and Deidda (2018), Soares et al. (2017)). We use data from 10 models, each of which is a different combination of a global model and a regional model. The models are listed in Table 1. Further details on EURO-CORDEX and the models we use are given in the EURO-

95　CORDEX guidance report (Benestad et al. 2017). The model output is at 0.11 degree resolution (about 12km). This is

roughly an order of magnitude higher resolution that the climate model output considered in MK and RY. In order to focus on changes in climate, we consider differences between temporal means over future time periods and the baseline time period 1981-2010. We use six scenarios, based on combining results for two different RCPs (RCP4.5 and RCP8.5) and three different future time periods (2011-2040, 2041-2070 and 2071-2100).

100

Model	Driving GCM	GCM Member	RCM
M1	CNRM-CM5	r1i1p1	ALADIN53
M2	IPSL-CM5A-MR	r1i1p1	RCA4
M3	CNRM-CM5	r1i1p1	RCA4
M4	CNRM-CM5	r1i1p1	CCLM4-8-17
M5	EC-EARTH	r12i1p1	CCLM4-8-17
M6	EC-EARTH	r12i1p1	RACMO22E
M7	EC-EARTH	r12i1p1	RCA4
M8	EC-EARTH	r1i1p1	RACMO22E
M9	EC-EARTH	r3i1p1	HIRHAM5
M10	IPSL-CM5A-MR	r1i1p1	WRF331F

Table 1: Models used in this study.

2.2 Cross-validation Methodology

105 We will derive our smoothing methodology using frequentist statistical ideas. These frequentist arguments should not taken as reflecting a particular philosophy, or set of beliefs, about climate models and how they should be interpreted, or, indeed, any preference for frequentist statistical methods over any other statistical methods. They are simply a convenient mathematical device for deriving a smoothing equation, that may be a useful solution for this particular problem. Once the smoothing equation has been derived, the frequentist assumptions used to derive it can be forgotten. The equation must then
110 be tested and should be judged purely according to how well it works in practice. It may, in fact, be possible to derive the same equation using different arguments.

We are ultimately interested in deriving means, variances and distributions that give projections for possible values of future climate. As part of the frequentist derivation, we will imagine an infinite ensemble of reasonable climate models, covering all sources of uncertainty, and we will write the mean of this ensemble as μ. We do not assume that this infinite ensemble is
115 unbiased, relative to observations. The fact that this infinite ensemble is entirely fictional, and that the word reasonable is impossible to define, is of no concern: all that matters is the effectiveness of the equation we derive based on these abstract concepts.

We will assume that the ensemble members in the EURO-CORDEX ensemble we use are independent samples from the infinite ensemble of models. It is possible to address issues of dependency between models, rather than assume

120   independence, and there has been a large amount of research into how dependencies between the models in an ensemble
       might be estimated and modelled using weighting (e.g., Abramowitz & Bishop (2015), Annan & Hargreaves (2016), Rauser
       et al. (2015), Sanderson et al. (2015b), Knutti et al. (2017)). However, understanding model dependencies is difficult, and
       our impression is that this research has so far been somewhat inconclusive. Since our aim is simply to test whether adaptive
       smoothing is useful, we do not apply weighting. If weighting were included, that would, if anything, reduce the effective

125   sample size of our ensemble and make smoothing more effective.

       The EURO-CORDEX ensemble is a multi-model ensemble. Relative to single model ensembles, multi-model ensembles are
       more complex statistical entities. They are, in some ways, more difficult to interpret, e.g., are more likely to need weighting,
       but may sample model uncertainty more thoroughly. Both single and multi-model ensembles can be used to create estimates
       of future climate. Given all the challenges involved in converting ensemble output to estimates of future climate, such

130   estimates are perhaps best considered as subjective views of what future climate might be, conditional on all the subjective
       assumptions used in the post-processing of the models.

       Our analysis is restricted to working with climate model results and involves no comparison with observations. This setup is
       sometimes known as a perfect model experiment. Within this setup our goal can be described as trying to estimate the mean
       of the infinite ensemble as well as possible. Some aspects of the larger goal, of estimating what *real* climate may be in the

135   future, cannot be addressed in the absence of observations. For example, there is no way to estimate possible bias versus the
       real climate, or the extent to which our smoothing methods might reduce such biases. However, we make the assumption that
       developing a better understanding of climate model ensembles, in the perfect model framework, is nevertheless a step
       towards this larger goal.

       We will evaluate the methods that we test using cross-validated predictive root mean squared error (PRMSE), following RY.

140   In the frequentist framework, this can be considered as a way of evaluating how well we can estimate the infinite ensemble
       mean. Alternatively, the infinite ensemble concept can be avoided, and cross-validation can be justified by simply assuming
       that if we have a method that can improve how well we can predict climate model output, in an out-of-sample way, it may be
       a better method for predicting the real climate.

       For each smoothing methodology we will compute the PRMSE by using a standard leave-one-out cross-validation scheme.

145   In this scheme, we will drop each of the ten EURO-CORDEX models from the ensemble in turn and apply the smoothing to
       the remaining nine models. The prediction for μ from the smoothing will then be compared with the dropped model, on a
       grid-point by grid-point basis. We will calculate the squared error of the prediction at each grid-point, and then calculate the
       mean squared error (MSE) where the mean is calculated over all grid-points, and by dropping each of the ten models in turn.
       We then subtract the variance over all grid-points and models from the MSE value in order to convert the MSE from a value

150   that represents the MSE around an individual model to a value that represents the MSE around the unknown mean μ. The
       downside of using cross-validation on ensembles of high-resolution model output over large domains is that it is
       computationally expensive, which limits the number of numerical experiments that can be performed.

Nonlinear Processes
in Geophysics
Discussions

Open Access

EGU

3 Smoothing Methodologies

We now describe the smoothing methodologies we will apply to the EURO-CORDEX data. There are eight methodologies
155 in all: four non-adaptive methodologies and their four adaptive counterparts. Following the notation in MK and RY, we will
say that we are trying to estimate μ at the target location (k, l) using predictor locations (i, j), taking into account the
distance between the target and predictor locations $d_{i,j}(k, l)$.

3.1 Non-adaptive Smoothing Methodologies

We first describe the four non-adaptive smoothing methodologies that we will test. They are all non-adaptive because, within
160 each test, nothing about the methodology varies by location except the input rainfall data at the target and predictor
locations. The four methods can all be described as creating a smoothed prediction for μ, written as $\hat{\mu}$, at location (k, l) using
a weighted average over ensemble mean values given by:

$$\hat{\mu} = \frac{\sum_{(i,j) \in S_n} m_{i,j} w_{i,j}\left(k, l, d_{i,j}(k,l)\right)}{\sum_{(i,j) \in S_n} w_{i,j}\left(k, l, d_{i,j}(k,l)\right)} \tag{1}$$

where $m_{i,j}$ is the ensemble mean at location (i, j), $w_{i,j}(k, l, d_{i,j}(k, l))$ is the corresponding weight, and S_n is the set of spatial
165 points used to make the prediction, which consists of the target location and all the predictor locations. If one wishes to avoid
the use of the mean of the infinite ensemble μ in the derivation then the methods can be described as predicting the excluded
ensemble member in the cross-validation: the end result of the derivation would be the same. We consider different sizes for
the set of spatial points used to make the prediction, as indicated by the subscript n. The four non-adaptive methods differ
from each other in terms of the formulation of the weight, and the set of points S_n. Where some of the points in the set of
170 predictor locations S_n are missing, because of coastlines or because of the edge of the domain, we simply eliminate those
points from the set. We now describe each of the four non-adaptive smoothing methods in detail.

3.1.1 Method 1: Square Top-hat

The first smoothing method we apply considers all the points in a square around the target location (k, l), for squares of
different sizes. The target location (k, l) is at the centre of the square, and the squares have an odd number of points along
175 each side. The different squares are labelled using $n = 1, 2, 3 ...$, the length of the side of the square in each case is given by
$2n - 1$, and the number of points in the square is given by $(2n - 1)^2$. The square corresponding to $n = 1$ involves no
smoothing at all and defines a baseline prediction which the smoothing methods can be compared against. We then calculate
the average rainfall over all the points in the square, and that serves as the prediction $\hat{\mu}$. The weight function for this method
is given by:

180 $$w_{i,j}\left(k, l, d_{i,j}(k,l)\right) = 1 \tag{2}$$

One of the reasons this method is relevant is because it relates to the commonly used aggregation method in which data is
aggregated within squares.

3.1.2 Method 2: Circle Top-hat

The second smoothing method we apply is similar to the square top-hat, and uses the same weight function, but only

185 considers predictor locations in a circle of diameter n points around the target location (k, l). The difference between this method and the square top-hat method is that points in the corners of the squares are now eliminated because they lie outside the circle. This method was also considered by MK.

3.1.3 Method 3: Exponential

The third smoothing method we apply considers the same circular set of predictor locations as in the circle top-hat method,

190 but now down-weights them as a function of distance from the point (k, l) using an exponential function with a lengthscale λ. We write the weights as:

$$w_{i,j}\left(k, l, d_{i,j}(k, l)\right) = \exp\left(-d_{i,j}(k, l)/\lambda\right) \tag{3}$$

The length-scale is chosen so that the unnormalized weight at the edge of the circle is 0.1, which gives $\lambda = -\frac{1-2n}{2\ln(0.1)}$.

3.1.4 Method 4: Gaussian

195 The fourth smoothing method we apply considers the same circular set of predictor locations as in the circle top-hat and exponential methods, but now down-weights them using a Gaussian function with lengthscale λ. Very similar methods were used in MK and RY. We write the weights as:

$$w_{i,j}\left(k, l, d_{i,j}(k, l)\right) = \exp\left(-d_{i,j}^2(k, l)/\lambda^2\right) \tag{4}$$

where the length-scale is again chosen so that the weight at the edge of the circle is 0.1, which in this case gives $\lambda =$

200 $-\frac{(1-2n)^2}{4\ln(0.1)}$.

3.2 Adaptive Smoothing Derivation

We now derive the adaptive smoothing methodology that we will use. We will again write our smoothed estimate for the unknown ensemble mean μ at the target location (k, l) as μ̂. This estimate will be based on a linear combination of the ensemble mean m at location (k, l) and a predictor based on a weighted sum of ensemble means at the surrounding points

205 (not including the target location (k, l)), which we write as p. The two are combined using a factor α, and an expression for α will be derived below. The prediction μ̂ is therefore given by:

$$μ̂ = m(1 - \alpha) + \alpha p = m + \alpha d \tag{5}$$

Where $d = p - m$. The parameter α will be adaptive, i.e., will vary in space according to information in the ensemble, and this is where the adaptivity of the method comes in. The predictor p will be non-adaptive and will be constructed from the

210 surrounding predictor locations using smoothing kernels very similar to those used in the non-adaptive methods, in ways that are explained below.

We are only considering a single predictor p, with a single adaptive parameter α. One could also imagine more complex adaptive smoothing methodologies that would consider multiple predictors with multiple adaptive parameters. One could also consider methodologies in which the predictor p is adaptive e.g., in which the length-scales in the smoothing kernels

215 that determine the predictor p are adaptive and vary as a function of information in the ensemble.

In order to determine an appropriate expression for α we can consider the statistical properties of the predictor $\hat{\mu}$. The following derivation makes a bias-variance trade-off, similar to that used to derive methods that have been shown to increase the accuracy of estimates of temperature trends by Jewson and Penzer (2006), and increase the potential accuracy of future rainfall projections by Jewson et al. (2021). We write the unknown ensemble mean at the target location as μ_m, and for the

220 predictor locations as μ_p. The prediction error is then given as:

$$\text{Prediction error} = e = \hat{\mu} - \mu_m \tag{6}$$

We now derive expressions for the bias and variance of this prediction error, as a function of α. We will then combine the bias and the variance to give the predictive mean square error (PMSE). The bias in the prediction is given as:

$$\text{bias} = E(e) = E(\hat{\mu} - \mu_m) = E(m + \alpha d - \mu_m) = E(m) + \alpha E(d) - \mu_m$$

225
$$= \mu_m + \alpha(\mu_p - \mu_m) - \mu_m = \alpha(\mu_p - \mu_m) = \alpha\,\delta \tag{7}$$

where $\delta = \mu_p - \mu_m$.

The error variance is given as:

$$\text{error variance} = V(e) = V(\hat{\mu} - \mu_m) = V(m(1-\alpha) + \alpha p - \mu_m) = (1-\alpha)^2 V(m) + \alpha^2 V(p) + 2(1-\alpha)\alpha\, C(m, p)$$

$$= (1-\alpha)^2 V_m + \alpha^2 V_p + 2(1-\alpha)\alpha\, C \tag{8}$$

230 Where $C = C(m, p)$ is the covariance between the ensemble mean m at the target location (k, l) and the predictor p, and we have written $V_m = V(m)$ and $V_p = V(p)$.

Combining equations (7) and (8), the PMSE, as a function of α, is given by:

$$\text{PMSE} = \text{bias}^2 + \text{error variance} = \alpha^2 \delta^2 + (1-\alpha)^2 V_m + \alpha^2 V_p + 2(1-\alpha)\alpha\, C \tag{9}$$

This equation is a version of Eq. (3) from RY but evaluated for the particular prediction model given by Eq. (5). It is a

235 quadratic equation in α, and a proof that it has a minimum is given in Appendix A.

In order to find the α value at the minimum PMSE, we differentiate Eq. (9) by α, giving:

$$\frac{d\text{PMSE}}{d\alpha} = 2\alpha\delta^2 - 2(1-\alpha)V_m + 2\alpha V_p + 2(1-2\alpha)C \tag{10}$$

Setting this derivative equal to zero and solving for α gives:

$$\alpha = \frac{V_m - C}{\delta^2 + V_m + V_p - 2C} \tag{11}$$

240 This gives the value of α that minimises the PMSE and the PRMSE. We can see that it contains all four factors that were discussed in the introduction as potentially relevant for determining the optimal degree of smoothing: the ensemble variance

of the location being predicted, captured by the variance of the ensemble mean at that location (V_m); the ensemble variances of the surrounding points, captured by the variance of the predictor (V_p); the ensemble correlations between the location being predicted and the surrounding points, captured by the covariance between the ensemble mean at the location being predicted and the predictor (C); and the difference between the size of the signal at the location being predicted and the predictor (δ). Dividing the top and bottom of equation (11) by V_m gives a different form for this expression, in which the δ^2 term becomes a kind of signal to noise ratio.

If we could determine the exact value of α, then, based on the derivation above, we would be guaranteed to be able to produce predictions that would have a lower PMSE than using either the local ensemble mean m or the predictor p alone. However, in practice we can never know the exact value of α, and the best we can do is to estimate it from the sample data that we have. The most obvious (although not necessarily the best) way to do this is by using sample estimates for the terms on the right-hand side of Eq. (11) above. Using sample estimates $\hat{V}_m$ for V_m, $\hat{V}_p$ for V_p, $\hat{C}$ for C and d^2 for δ^2 gives an estimate $\hat{\alpha}$ which is:

$$\hat{\alpha} = \frac{\hat{V}_m - \hat{C}}{d^2 + \hat{V}_m + \hat{V}_p - 2\hat{C}} \tag{12}$$

Using this estimated value of α may or may not give improved estimates, and whether it does or not can only be resolved by testing. The sample estimates $\hat{V}_m$, $\hat{V}_p$, and $\hat{C}$ in this expression can be estimated in the usual ways:

$$\hat{V}_m = \frac{\hat{\sigma}_m^{\,2}}{N}, \hat{V}_p = \frac{\hat{\sigma}_p^{\,2}}{N} \text{ and } \hat{C} = \frac{\hat{c}_{mp}}{N} \tag{13}$$

where $\hat{\sigma}_m^{\,2}$ is the observed ensemble variance at the target point (k, l), $\hat{\sigma}_p^{\,2}$ is the observed ensemble variance for the predictor, $\hat{c}_{mp}$ is the observed covariance between ensemble members at points (k, l) and the ensemble members for the predictor and N is the number of members in the ensemble.

For our purposes $\hat{\alpha}$ only makes sense if it is in the range [0,1]. When $\hat{C}$ is non-zero, values from Eq. (12) could lie outside this range, and so we restrict $\hat{\alpha}$ to this range in our calculations by increasing values that are below 0 to 0 and reducing values that are above 1 to 1.

3.3 Adaptive Smoothing Methodologies

To define the four adaptive smoothing methodologies it just remains to define the predictor p in each case. In all four cases, we define p as a weighted average of ensemble means using the expression:

$$p = \frac{\sum_{(i,j) \in T_n} m_{i,j} w_{i,j}\left(k, l, d_{i,j}(k, l)\right)}{\sum_{(i,j) \in T_n} w_{i,j}\left(k, l, d_{i,j}(k, l)\right)} \tag{14}$$

This is very similar to Eq. (1), which defines the non-adaptive predictions. The only difference is that the set of points T_n differs from S_n in that it does not contain the target location (k, l). The target location is not included in T_n since it is already included in the prediction given by Eq. (5). Apart from this difference, the weights for the four adaptive smoothing methodologies are calculated in exactly the same way as they are for the non-adaptive methodologies, creating adaptive

Nonlinear Processes
in Geophysics
Discussions

square top-hat, adaptive circle top-hat, adaptive exponential and adaptive Gaussian methods. In this way, the adaptive methodologies combine the unsmoothed prediction based on just the ensemble mean at the target location (k, l) with information from surrounding points, and these two sources of information are combined together in a way that varies from location to location, taking into account what we can learn from the ensemble about how best to combine the two. Whether the method works is a matter for testing.

4 Results

We now show results from the application of the smoothing methods described in Sect. 3 above to the EURO-CORDEX rainfall projection data described in Sect. 2 above. We first present PRMSE performance of the various methods, and then give examples of some smoothed fields.

4.1 PRMSE Performance

Figure 1 shows the PRMSE performance of the four non-adaptive smoothing methods, versus the parameter n that defines the size of the region over which the field is smoothed, for each of the six scenarios (two RCPs and three future time periods). We show the six scenarios separately, rather than aggregated together, to give an idea of the variability around the results when the methodology is applied to different datasets. The results are normalized so that for $n = 1$ (which corresponds to no smoothing) the PRMSE is 100. For all four methods for all six scenarios as n increases the errors reduce initially, reach a minimum value, and then increase. The minimum errors, marked by circles, occur at values of n of 2, 3 or 4, corresponding to squares of dimension 3, 5 or 7 grid points. The results for the six different scenarios are similar: for the square top-hat method the minimum always occurs at $n = 2$, for the circle top-hat method it occurs at $n = 2$ or $n = 3$ and for the exponential and Gaussian methods it occurs at $n = 3$ or $n = 4$. The top-hat methods reach their minima for lower values of n because the shape of the smoothing kernel puts a greater weight on data further from the target location.

Nonlinear Processes
in Geophysics

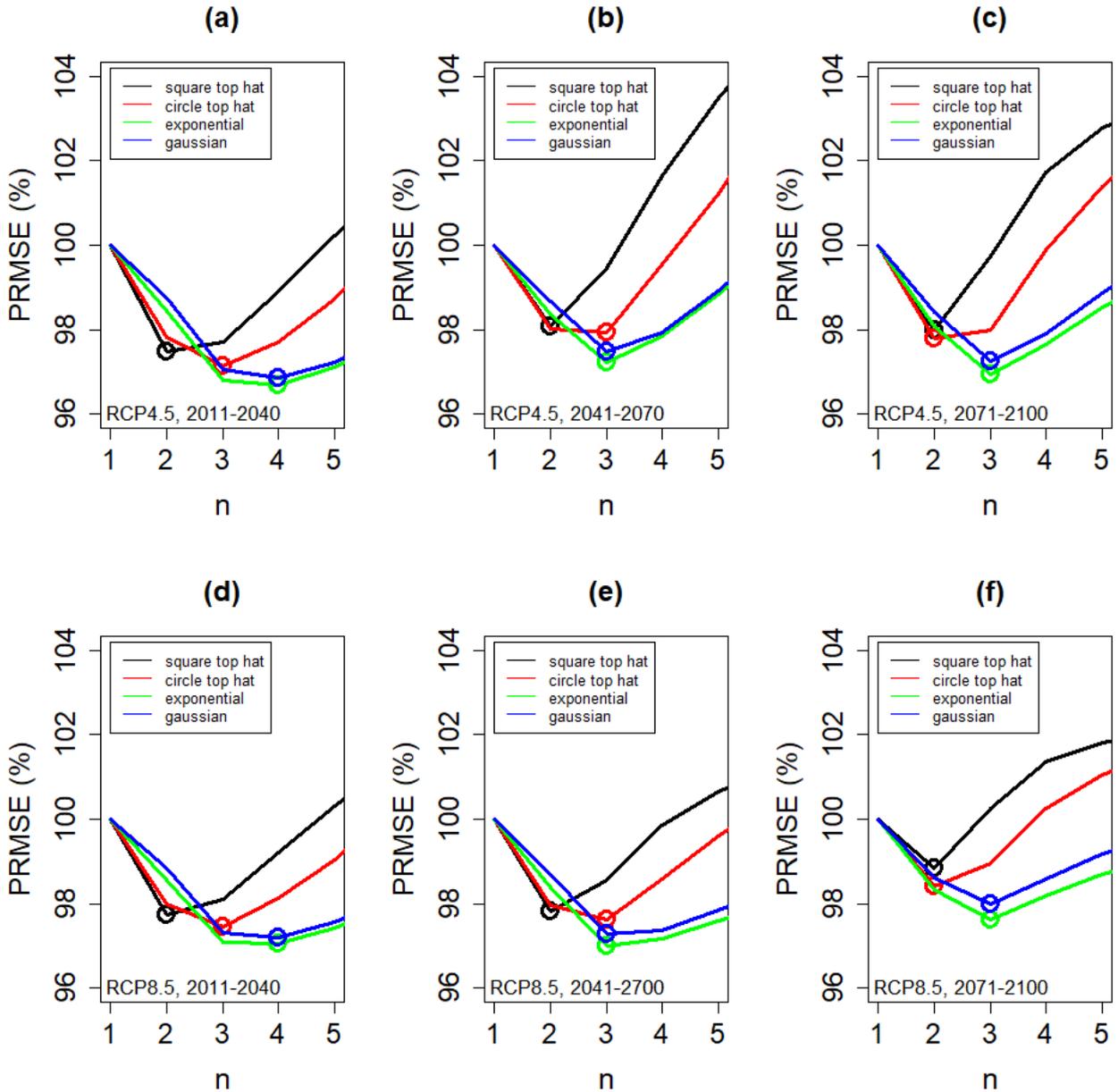Open Access

EGU

Discussions

Figure 1: PRMSE curves for 4 non-adaptive smoothing methods, for 6 future climate scenarios (panels (a) to (f)). The minimum PRMSE value of each curve is indicated with a circle.

295

In all cases exponential smoothing achieves the lowest minimum PRMSE, followed by Gaussian, followed by circle top-hat, followed by square top-hat. The minimum values achieved by the exponential smoothing are all between 96% and 98%,

corresponding to estimates of the mean rainfall in the notional infinite ensemble that are between 4% and 2% more precise than the unsmoothed ensemble mean. We would therefore say that the reduction in error due to the smoothing is rather small.

300 The optimal values of n correspond to very small optimal smoothing length-scales, relative to those derived by MK and RY: the optimal smoothing regions being used here would lie entirely within one grid box of the model output considered in those studies. We are therefore considering smoothing of different, and much smaller scale, aspects of model error and model internal variability.

Figure 2 shows the performance of all eight smoothing methods (four non-adaptive and four adaptive) on a larger horizontal 305 scale up to $n = 40$. We limit to $n = 40$ because the large number of spatial points in our domain, the use of cross-validation and the number of methods compared make the calculations computationally intensive. Nevertheless, we are able to find optimal values of n for all eight smoothing methods when we consider all scenarios together (see below). The PRMSE results for the adaptive methods have a completely different character to the results from the non-adaptive methods: they decrease monotonically up to $n = 18$ for all 24 cases (four methods and six scenarios). In 16 of the 24 cases they reach a 310 minimum and start to increase again before $n = 40$ while in eight of the 24 cases they are still decreasing slightly at $n = 40$. They show lower errors than the non-adaptive methods at all values of the parameter n, and the minima are in the range from 88% to 94%, corresponding to estimates of the mean rainfall that are between 12% and 6% more precise than the unsmoothed ensemble mean. These increases in precision are about three times greater than the increases in precision achieved by the non-adaptive methods, and are being achieved in very different way, by incorporating information from a 315 much wider area.

Figure 3 shows the performance of all eight smoothing methods aggregated over the six scenarios. All methods now reach their minima within the range of values of n that we are considering. The best of the non-adaptive methods is exponential, which achieves a minimum of 97.1% at $n = 3$, corresponding to a 2.9% increase in precision. This is equivalent to the increase in precision that would be achieved by adding 0.5 extra independent climate models to the ensemble. The best of the 320 adaptive methods is Gaussian, which achieves a minimum of 91.4% at $n = 37$, corresponding to a 8.6% increase in precision, which is 2.97 times the increase in precision due to the best of the non-adaptive methods. This is equivalent to the average increase in precision that would be achieved by adding 1.9 extra independent climate models to the ensemble, although adding extra climate models increases the precision everywhere, while adaptive smoothing increases the precision in a targeted fashion at certain locations, and so may be considered more effective than adding extra models.

325 Between the four adaptive methods the differences in the minima achieved are rather small, and can probably be considered immaterial. Once again, the top-hat methods achieve lower errors at lower values of n because they put more weight on more distant data points.

We now consider the resulting smoothed fields.
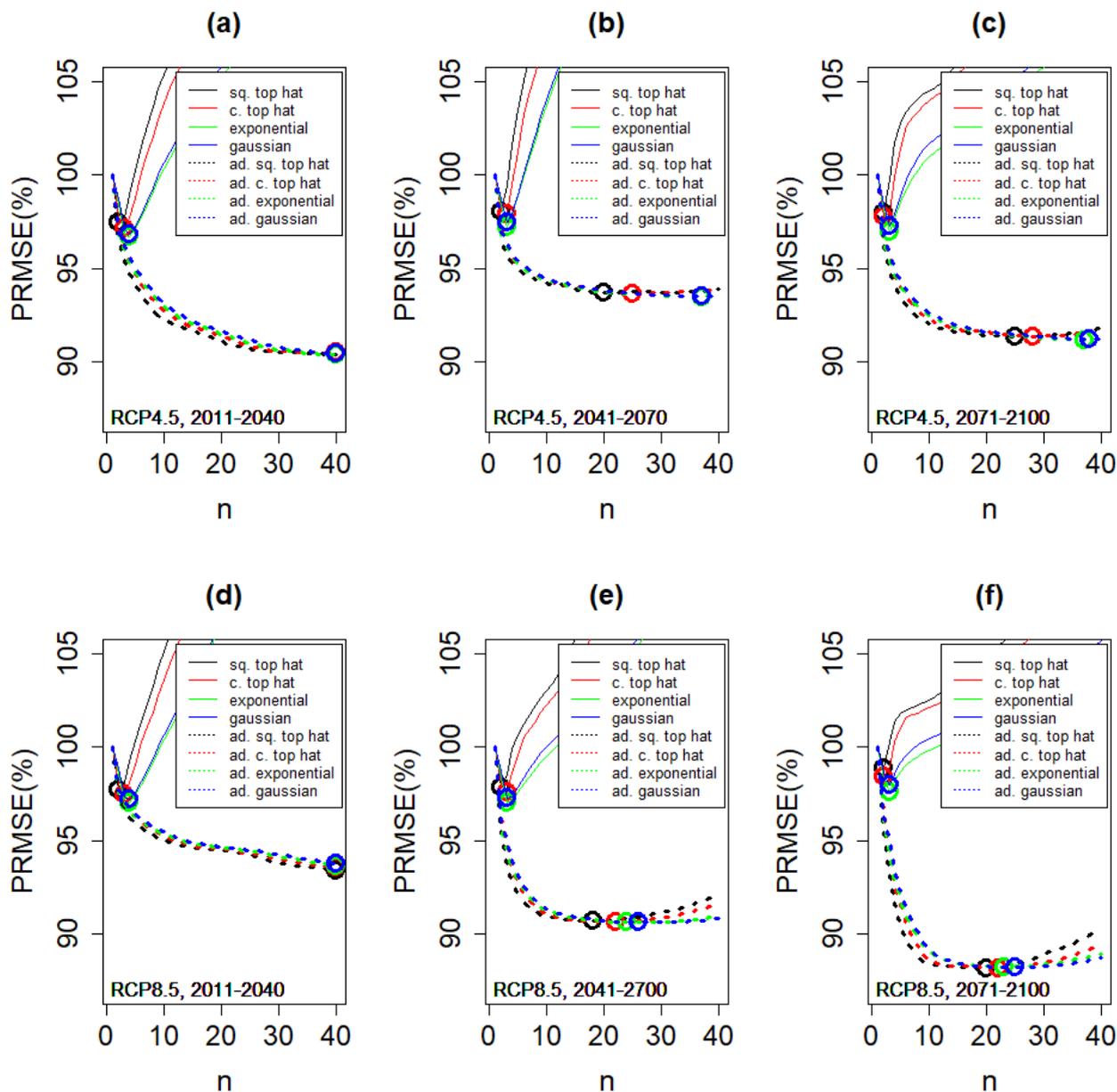
330   *Figure 2: PRMSE curves for 4 non-adaptive smoothing methods, and 4 adaptive smoothing methods, for 6 future climate scenarios (panels (a) to (f)). The minimum PRMSE value of each curve is indicated with a circle.*
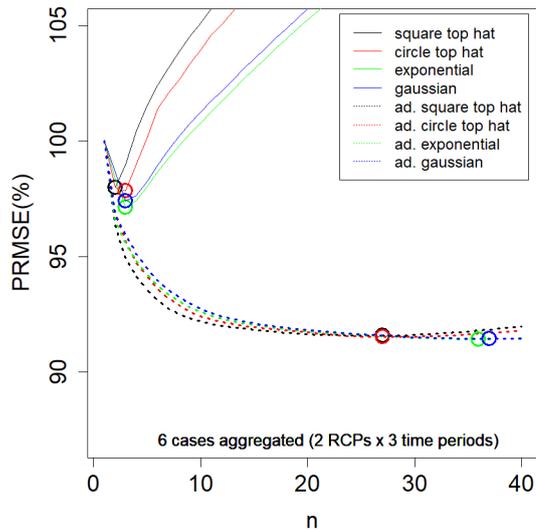
Figure 3: PRMSE curves for 4 non-adaptive smoothing methods, and 4 adaptive smoothing methods, for 6 future climate
335 *scenarios aggregated together. The minimum PRMSE value of each curve is indicated with a circle.*

4.2 Europe-wide Smoothed Fields

Figure 4 shows Europe-wide smoothed ensemble mean fields for the non-adaptive and adaptive Gaussian methods for the
first of our six scenarios (RC4.5, 2011-2040). Panel (a) shows the ensemble mean change in rainfall prior to smoothing,
340 which shows increases in rainfall in most of the domain, and decreases in the south of the domain. Panel (b) shows the value
of α used by the adaptive Gaussian smoothing based on $n = 37$, which is the value of n which gives the lowest PRMSE
overall. Panel (c) shows the results of non-adaptive Gaussian smoothing using the optimal value of $n = 3$, and panel (d)
shows the results of adaptive Gaussian smoothing using the value of $n = 37$. Panels (e) and (f) show the changes created by
the non-adaptive and adaptive smoothing methods, respectively. All the plots except panel (b) use non-linear colour-scales in
345 order to emphasize spatial variations in the values shown.

It is somewhat difficult to see differences between the unsmoothed ensemble mean and the smoothed fields (i.e., how panels
(c) and (d) differ from panel (a)). The smoothing is only changing the fields on small spatial scales and certainly does not
materially affect the large-scale pattern of rainfall changes, although it may be having large effects locally. The α values
show a high level of variability in space and take values that cover the range from 0 to 1. There are regions with more low α
350 values (corresponding to less weight being applied to the smoothing), such as Iceland (which is very mountainous) and there
are regions with more high α values (corresponding to more weight being applied to the smoothing), such as Belarus (which

14

is very flat). Overall, however, it is hard to interpret the α values visually because they are so variable. This is perhaps not surprising since they depend in a rather complex way on four different factors from the ensemble, each of which varies considerably in space. The difference field for the non-adaptive smoothing (panel (e)) show very small-scale features,

355 corresponding to the small n value used, while the difference field for the adaptive smoothing (panel (f)) is larger scale, corresponding to the larger n value used. Both methods show a greater impact of smoothing in mountainous regions (Iceland, Norway and the Alps, for instance).
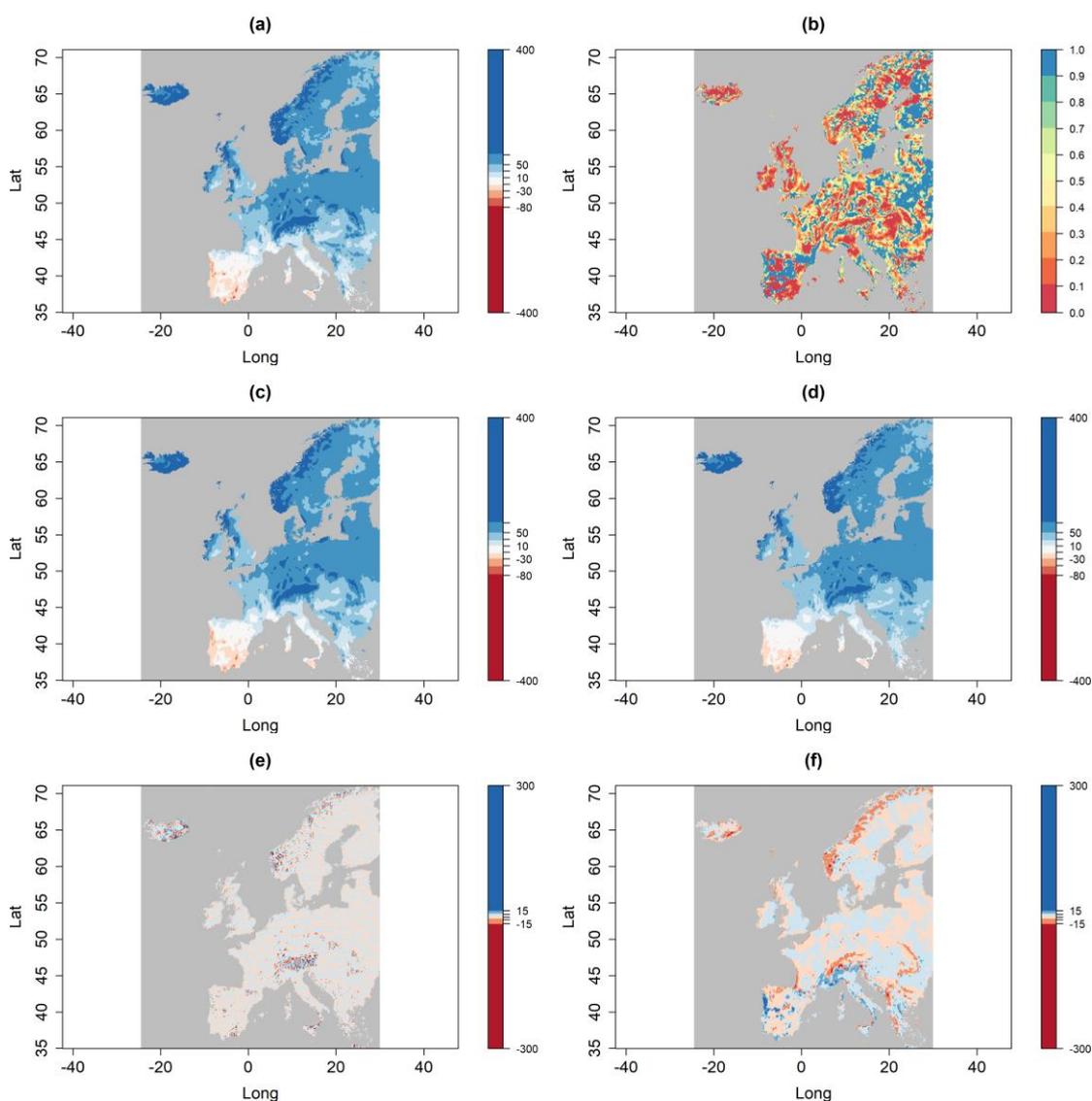
360 *Figure 4: (a): EURO-CORDEX ensemble mean rainfall change 2011-2040, RCP4.5. (b): the adaptive smoothing parameter α for adaptive Gaussian smoothing with n = 37. (c): the data from panel (a), but smoothed using non-adaptive Gaussian*

smoothing with $n = 3$. (d): the data from panel (a) but smoothed using adaptive Gaussian smoothing with $n = 37$. (e): the difference between panels (a) and (c), showing the change due to non-adaptive Gaussian smoothing. (f): the difference between panels (a) and (d), showing the change due to adaptive Gaussian smoothing.

365

4.3 Regional Smoothed Fields

To illustrate in more detail the effect of the different smoothing methods, we now show the same results as shown in Fig. 4 but for the Alps (Fig. 5), Norway and Sweden (Fig. 6) and Iceland (Fig. 7). These regions were chosen as regions where both smoothing methods have relatively larger impacts.

370 **4.3.1 The Alps**

For the Alps (Fig. 5) the unsmoothed ensemble mean shows increases in rainfall everywhere. The increases vary in space. The adaptive smoothing parameter α shows spatial structure that roughly corresponds to the spatial structures in the ensemble mean. For the smoothed ensemble means, careful inspection of panels (c) and (d) shows that the non-adaptive Gaussian smoothing leads to slightly greater changes in the features in the ensemble mean, while the adaptive method leads

375 to smaller and barely noticeable changes. The difference fields show that the non-adaptive smoothing is having an impact at very small scales, while the adaptive smoothing is having an impact at larger scales, and only in certain regions.

4.3.2 Norway and Sweden, Iceland

The results for Norway and Sweden (Fig. 6) and Iceland (Fig. 7) show similar effects. The smoothing effect of the adaptive smoothing is such that small-scale features are preserved to a greater extent than by the non-adaptive smoothing. The

380 adaptive smoothing parameter takes values in the whole range from 0 to 1. The difference fields show that the non-adaptive smoothing makes changes on very small scales, while the impact of the adaptive smoothing is on larger scales.

4.4 Discussion

Figures 5, 6 and 7 show that the value of α varies between near to 0 and near to 1 over very short length-scales. When α is near to 0 then very little smoothing takes place: the reduction in PRMSE values shown in Fig. 2 and Fig. 3 must therefore be

385 due to smoothing that is taking place in locations where α is not close to zero. The improvements in PRMSE in the regions which are more strongly smoothed must on average be higher than the improvements shown in Fig. 2 and Fig. 3. That the adaptive smoothing methods have lower PRMSE values than the non-adaptive methods is presumably partly because the adaptive methods avoid detrimental smoothing, by using values of α close to zero in many locations, and partly because it allows more smoothing in regions where smoothing is beneficial. In this way, by combining information from the ensemble

390 means, variances and covariances at different locations the adaptive smoothing may help distinguish between those signals in the ensemble mean which are genuine forced signals and those which are not.
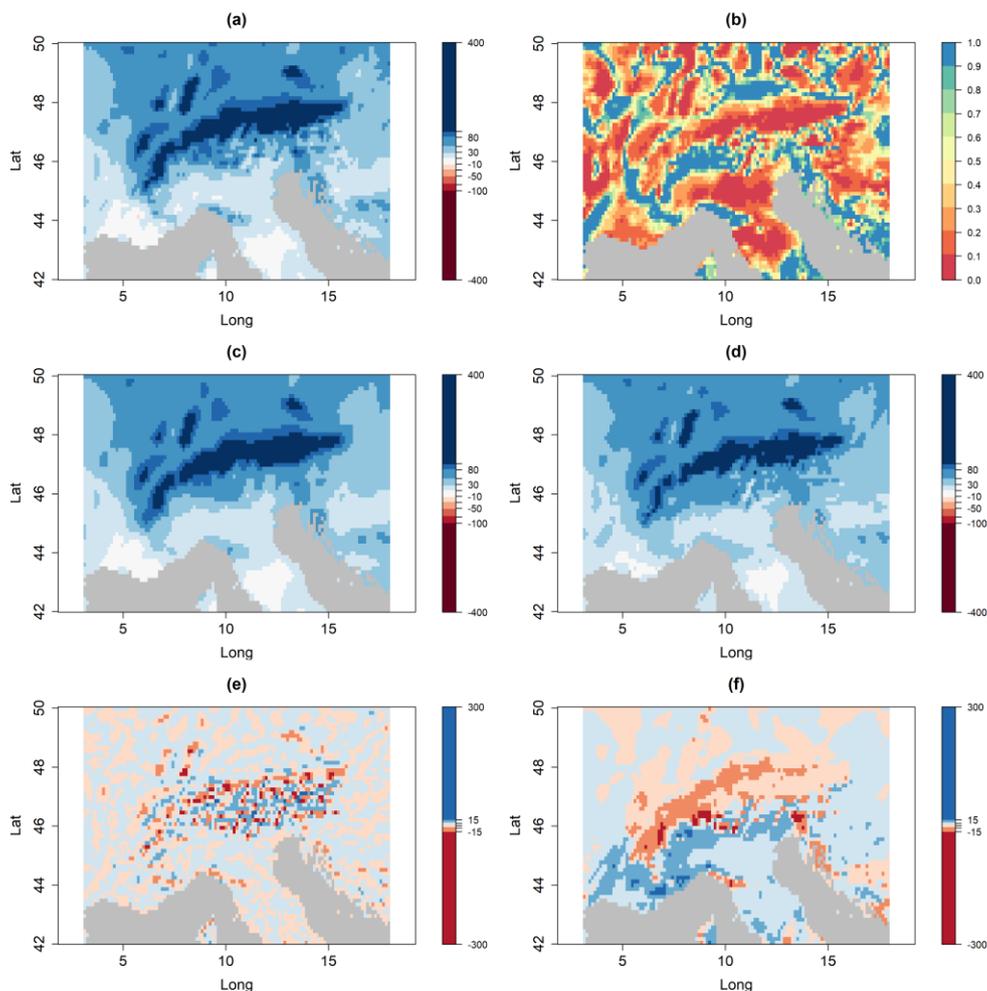
Figure 5: As Fig. 4, but for the Alps only.

395 We have also performed a visual comparison of the results from adaptive smoothing for different values of the smoothing parameter n in a range from $n = 10$ to $n = 40$, and for the other three adaptive smoothing methods. The results of these comparisons (not shown) show that there is almost no impact on the results of varying n within this range, and almost no difference between the results from the different methods. This is perhaps to be expected, given that the PRMSE values shown in Fig. 2 and Fig. 3 are very flat near the minimum, and are very similar for the four adaptive smoothing methods.

400 The lack of differences between the impact of the different methods is a useful result, since it suggests that future studies would not need to compare different smoothing kernels, but could choose any one of the four, and also do not need to consider every possible value of n, as we have done.

We have shown results for improvements in the estimate of the unknown ensemble mean. For many applications, probabilities would also be derived from the ensemble by fitting distributions. If the improved estimate for the ensemble

405   mean is used in the distribution fitting process, then the probabilities will also be improved.
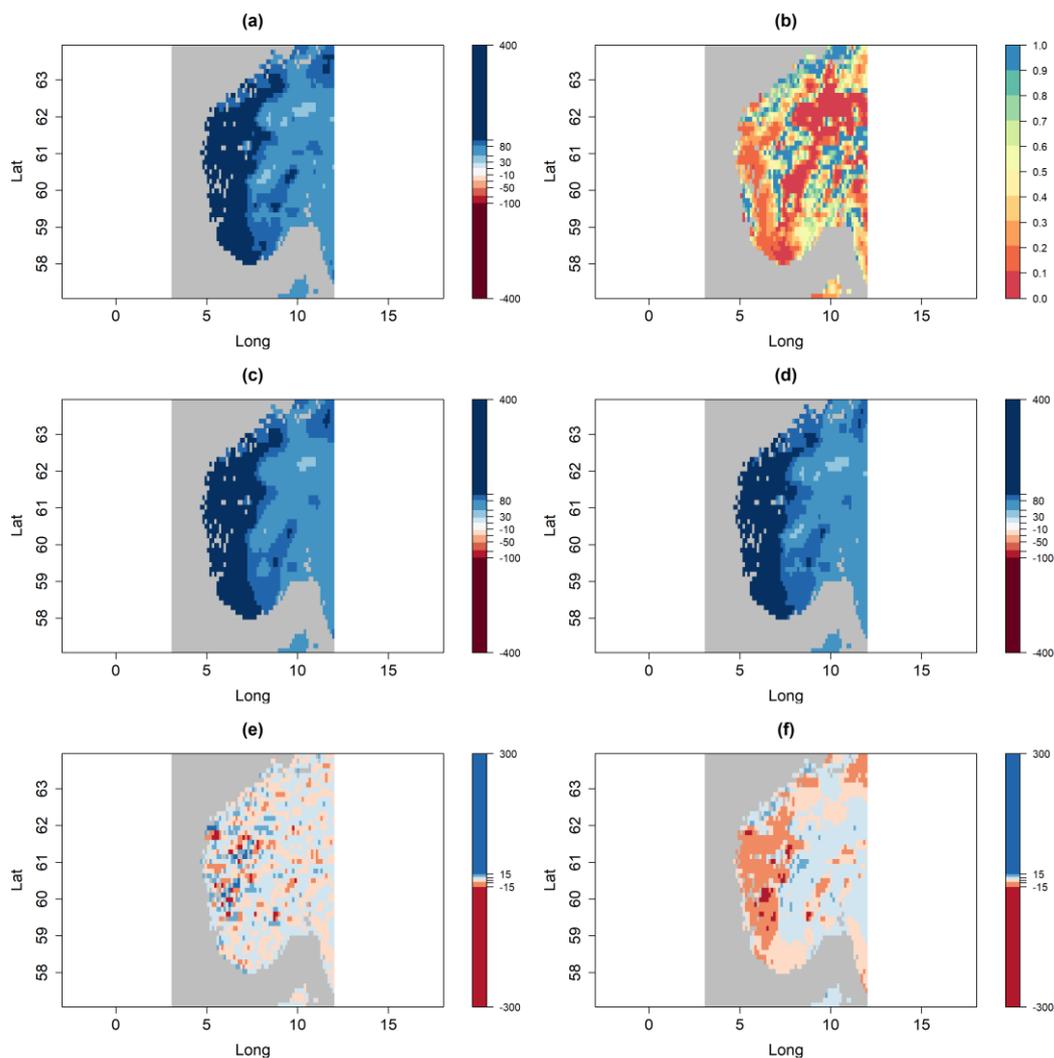

Figure 6: As Fig. 4, but for Norway and Sweden only.

410   **5 Summary and Conclusions**

One question that arises in the interpretation and usage of climate model ensembles is whether the ensemble mean should be smoothed in space order to give more precise estimates of the mean of the forced climate response in the ensemble. One of

the reasons this is of interest is that smoothing the ensemble is vastly cheaper in terms of resources than increasing the precision of the ensemble mean by building new climate models or improving existing climate models. Previous research using non-adaptive smoothing (i.e., smoothing that applies the same amount of smoothing at all locations, irrespective of the information contained in the ensemble) has found only a small benefit to smoothing, and has shown that smoothing tends to smooth away features in the ensemble mean that may be correct. We have tested four non-adaptive smoothing methods on EURO-CORDEX annual mean rainfall data, and our results for these methods tend to agree with these previous conclusions. However, there is no reason why all locations have to be smoothed in the same way, and the ensemble contains a wealth of information that can help determine how best to smooth each location. For instance, one might imagine that signals in the ensemble mean that have a large amplitude and low variance (i.e., a high signal-to-noise ratio) should not be smoothed as much as signals that have a low amplitude and high variance (i.e., a low signal-to-noise ratio). There are also other relevant factors, including the correlations in space in the ensemble, and the gradient of the signal in the ensemble mean. We have constructed a novel methodology that incorporates this idea of taking information from the ensemble into account in the smoothing algorithm, based on the idea of minimising predictive root-mean-squared error (PRMSE) via a bias-variance trade-off. We have tested the methodology using cross-validation within the ensemble, and we find that it gives much better results than the standard non-adaptive smoothing methods that we compare with. The reduction in the PRMSE from the adaptive smoothing methodologies is roughly three times as large as the reduction from the non-adaptive methodologies, and visual comparison shows that features in the ensemble mean are preserved to a greater extent by the adaptive methodologies. The four adaptive smoothing methods we have tested, which have different smoothing kernels, perform similarly. We conclude that the choice of smoothing kernel is not particularly important.

For the data we have tested we conclude that the using adaptive smoothing gives materially better estimates of the mean of the climate response in the ensemble than using either unsmoothed model output or non-adaptive smoothing. By better we mean higher precision, i.e., higher potential accuracy. The differences created by the smoothing are on small scales, up to around 200km, and do not affect the continental-scale features of the climate response. The changes are quite localized, and a small number of regions show large changes due to the adaptive smoothing. The predictions show greater potential accuracy in terms of PRMSE, and the smoothed fields may help distinguish genuine forced signals from variability more effectively than the unsmoothed ensemble mean, or the ensemble mean smoothed using non-adaptive methods. We cannot, however, make any statements about whether the smoothing methods we have tested would work for other datasets since the performance of the method is likely to depend on aspects of the behaviour of the ensemble (ensemble means, variances, length-scales and correlations) which would vary depending on the variable being considered, the model and the model resolution. In addition, the optimal amount of smoothing will depend on the size of the ensemble. Applying the adaptive smoothing methodology to different datasets than those we have considered would require testing using a cross-validation based methodology similar to the one that we have applied. One simplifying factor in doing so, relative to the study described here, is that our results show that it would not be necessary to test multiple versions of non-adaptive and adaptive smoothing, with different kernels, but just one of each.
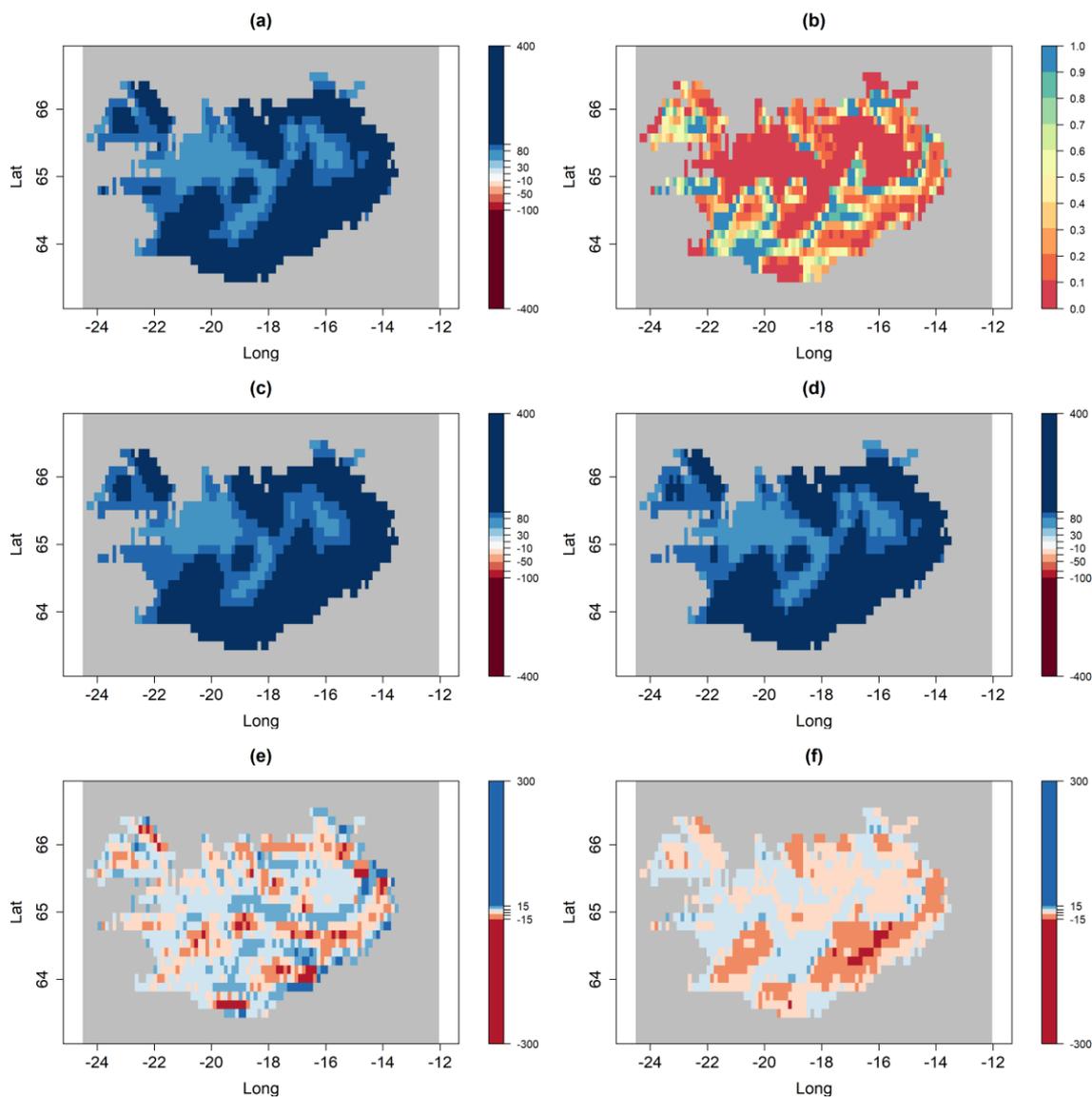
Figure 7: As Fig. 4, but for Iceland only.

450   The size of the beneficial impact of any spatial smoothing method, in terms of increased precision, is limited by the fact that we are improving the estimate of an ensemble mean that is already reasonably precisely estimated, especially for large ensembles based on independent models (as discussed in Raisanen and Ylhaisi (2010)). However, smoothing the ensemble mean is still vastly cheaper in terms of resources than increasing the precision of the ensemble mean by building or improving climate models. It is also possible that smoothing might reduce bias versus reality, although we have not been

455   able to evaluate that in this study.

This study is a first attempt to explore whether adaptive smoothing might be a useful tool for the post-processing of climate model output. The results suggest that it may be, depending on the application. If only the large-scale climate response is of interest, then adaptive smoothing does not appear to make much difference. If the small-scale response is of interest, as it would be for many applications, then adaptive smoothing does appear to make a difference, particularly in some regions.

460   However, further research is needed to assess the benefits in other situations than those we have considered i.e., for other variables and models.

The ideas presented here are generic methods for post-processing of ensembles, and could also be applied to weather, seasonal forecast and reanalysis ensembles. There may also be potential to improve the adaptive smoothing algorithm by e.g., also making the size of the smoothing region adaptive, or by combining the idea of adaptive spatial smoothing with the

465   idea of local model averaging as described in Jewson et al. (2021).

Appendix A

We now show that equation (9) has a minimum, when considered as a function of α.

Equation (9) is a quadratic in α, and the coefficient of α^2 is $\delta^2 + V_m + V_p - 2C$. Whether the extremum of the quadratic is a

470   maximum or minimum can be determined as follows. Writing $S_m = \sqrt{V_m}$ and $S_p = \sqrt{V_p}$, we can write the coefficient of α^2 as:

$$\delta^2 + V_m + V_p - 2C = \delta^2 + S_m{}^2 + S_p{}^2 - 2S_mS_pr = \delta^2 + (S_m - S_p)^2 + 2S_mS_p(1-r) \qquad (A1)$$

where r is a correlation coefficient. The final expression in Eq. (A1) is a sum of values that are all either zero or positive, and hence the coefficient of α^2 must itself be zero or positive. In all non-degenerate cases it will be positive, and the quadratic

475   equation for α will therefore have a minimum, which is as we would have hoped, since our goal is to minimize errors.

Author Contribution

SJ designed the study and the algorithms, wrote the smoothing code, ran the smoothing analyses, produced the graphics and wrote the text. GB, PM and JM extracted the EURO-CORDEX data. MS wrote the code to read the EURO-CORDEX data.

480   All the authors contributed to proof-reading.

Data Availability

The EURO-CORDEX data used in this study is freely available. Details are given at https://euro-cordex.net.

Competing Interests

MS works for RMS Ltd, a company that quantifies the impacts of weather, climate variability and climate change.

References

490

Abramowitz, G, and C Bishop. 2015. "Climate model dependence and the ensemble dependence transformation of CMIP projections." *J. Clim.* 28: 2332-2348.

Annan, J, and J Hargreaves. 2016. "On the meaning of independence in climate science." *Earth Syst. Dyn. Discuss.* 1-17.

495

Benestad, R, A Haensler, B Hennemuth, T Illy, D Jacob, E Keup-Thiel, S Kotlarski, et al. 2017. "Guidance for EURO-CORDEX." 1 08. Accessed 1 9, 2021. https://www.euro-cordex.net/imperia/md/content/csc/cordex/euro-cordex-guidelines-version1.0-2017.08.pdf.

Burnham, K, and D Anderson. 2002. *Model Selection and Multimodel Inference.* New York: Springer-Verlag.

500

Claeskens, G, and N Hjort. 2008. *Model Selection and Model Averaging.* Cambridge: CUP.

Colmet-Daage, A, E Sanchez-Gomez, S Ricci, C Llovel, V Estupina, P Quintana-Segui, M Carmen Llasat, and E Servat. 2018. "Evaluation of uncertainties in mean and extreme precipitation under climate change for northwestern Mediterranean watersheds from high-resolution Med and Euro-CORDEX ensembles." *Hydrology and Earth System Sciences* 22 (1): 673-687.

505

Dalelane, C, B Fruh, C Steger, and A Walter . 2018. "A Pragmatic Approach to Build a Reduced Regional Climate Projection Ensemble for Germany Using the EURO-CORDEX 8.5 Ensemble." *Appl. Met. Clim.* 57 (3): 477-491.

Dyrrdal, A, F Stordal, and C Lussana. 2018. "Evaluation of summer precipitation from EURO-CORDEX fine-scale RCM simulations over Norway." *Intl. J. Clim.* 38 (4): 1661-1677.

510

Fletcher, D. 2019. *Model Averaging.* Springer.

Foley, A, and I Kelman. 2018. "EURO-CORDEX regional climate model simulation of precipitation on Scottish islands (1971–2000): model performance and implications for decision-making in topographically complex regions." *Intl. J. Clim.* 38 (2): 1087-1095.

Hosseinzadehtalaei, P, H Tabari, and P Willems. 2018. "Precipitation intensity–duration–frequency curves for central Belgium with an ensemble of EURO-CORDEX simulations, and associated uncertainties." *Atmos. Res.* 200: 1-12.

515

Jacob, Daniela, and 39 other authors. 2014. "EURO-CORDEX: new high-resolution climate change projections for European impact research." *Reg Environ Change* 14: 563-578.

Jewson, S, B Giuliana, P Mercogliano, J Mysiak, and M Sassi. 2021. "Improving the Potential Accuracy and Usability of EURO-CORDEX Estimates of Future Rainfall Climate using Mean Squared Error Model Averaging." *Nonlinear Processes in Geophysics.*

Jewson, Stephen, and Jeremy Penzer. 2006. "Estimating Trends in Weather Series: Consequences for Pricing Derivatives." *Studies in Nonlinear Dynamics & Econometrics* 10 (3): 1-10.

Knutti, R, J Sedlacek, B Sanderson, R Lorenz, E Fischer, and V Eyring. 2017. "A climate model projection weighting scheme accounting for performance and interdependence." *GRL* 44: 1909-1918.

Mascaro, G, F Viola, and R Deidda. 2018. "Evaluation of Precipitation From EURO-CORDEX Regional Climate Simulations in a Small-Scale Mediterranean Site." *JGR Atmospheres* 123 (3): 1604-1625.

Masson, D, and R Knutti. 2011. "Spatial-Scale Dependence of Climate Model Performance in the CMIP3 Ensemble." *J. Clim* 24: 2680-2692.

Meehl, G, C Covey, T Delworth, M Latif, B McAvaney, J Mitchell, R Stouffer, and K Taylor. 2007. "The WCRP CMIP3 multimodel dataset - a new era in climate change research." *Bull. Amer. Meteor. Soc.* 88: 1383-1394.

Raisanen, J, and J Ylhaisi. 2010. "How Much Should Climate Model Output Be Smoothed in Space?" *J. Clim* 24: 867-880.

Rauser, F, P Gleckler, and J Marotzke. 2015. "Rethinking the default construction of multimodel climate ensembles." *Bull. Am. Meteorol. Soc.* 96 (6): 911-919.

Sanderson, B, R Knutti, and P Caldwell. 2015b. "Addressing interdependency in a multimodel ensemble by interpolation of model properties." *J. Clim.* 28: 5150-5170.

Soares, P, R Cardoso, D Lima, and P Miranda. 2017. "Future precipitation in Portugal: high-resolution projections using WRF model and EURO-CORDEX multi-model ensembles." *Clim. Dyn.* 49: 2503–2530.