Climate Dynamics manuscript No.
(will be inserted by the editor)

Estimating daily climatological normals in a changing climate

Alix Rigal · Jean-Marc Azaïs · Aurélien Ribes

Received: 06 April 2018 / Accepted: date

Abstract Climatological normals are widely used baselines for the description and the characterization of a given meteorological situation. The World Meteorological Organization (WMO) standard recommends estimating climatological normals as the average of observations over a 30-year period. This approach may lead to strongly biased normals in a changing climate. Here we propose a new method with which to estimate daily climatological normals in a non-stationary climate. Our statistical framework relies on the assumption that the response to climate change is smooth over time, and on a decomposition of the response inspired by the pattern scaling assumption. Estimation is carried out using smoothing splines techniques, with a careful examination of the selection of smoothing parameters. The new method is compared, in a predictive sense and in a perfect model framework, to previously proposed alternatives such as the WMO standard (reset either on a decadal or annual basis), averages over shorter periods, and hinge fits. Results show that our technique outperforms all alternatives considered. They confirm that previously proposed techniques are substantially biased – biases are typically as large as a few tenth to more than 1 degree by the end of the century – while our method is not. We argue that such "climate change corrected" normals might be very useful for climate monitoring, and that weather services could consider using two different sets of normals (i.e. both stationary and non-stationary) for different purposes.

Also funded by Météo France.

Alix Rigal
CNRM, UMR3589, Météo-France and CNRS, Toulouse, France
E-mail: alix.rigal@meteo.fr

Jean-Marc Azaïs
Institut de Mathématiques, Université Paul Sabatier (member of the federal University "Université de Toulouse"), 31062 Toulouse, cedex 09 France

Aurélien Ribes
CNRM, UMR3589, Météo-France and CNRS, Toulouse, France

Keywords daily climate normals · unbiased estimate · normals accounting for climate change · Smoothing splines

1 Introduction

Climatological normals are widely used baselines which describe and characterise a given meteorological situation. On the news, weather forecasts commonly refer to normals in order to compare a weather or seasonal forecast to its expectation. Retrospective climate monitoring also typically involves such a comparison. Climate normals are primarily meant to describe the mean seasonal cycle in standard meteorological variables such as temperature or precipitation. In the most common estimation techniques, normals are assumed to be stationary, i.e. the drift related to anthropogenic climate change is neglected, potentially leading to inaccurate or biased estimates. One issue with this approach is that, as pointed out by Arguez and Vose (2011): "climate normals are calculated retrospectively, but are often utilized prospectively". For instance, when they are compared to weather forecasts, it is assumed that normals provide an estimation of the expected weather to date. Neglecting on-going warming can prevent this.

For example, the current recommendation of the World Meteorological Organization (WMO) for the calculation of climatological normals, known as the Climatological Standard Normals, is to compute an average over a 30-year period (World Meteorological Organization, 2007; Baddour, 2011) (ref to `http://www.wmo.int/pages/prog/wcp/wcdmp/GCDS_1.php`). These normals are supposed to be updated every 30 years, with the current reference period being 1961-1990. Following these recommendations, the next generation of climatological normals would be available in 2021, based on the 1991-2020 average.

Several studies pointed out the limitation of such normals and their inaccuracy in a non-stationary climate (e.g Scherrer, Appenzeller, and Liniger (2006); Krakauer and Devineni (2015)). WMO itself advocated for a more frequent revision of climate normals, through updates every 10 years but still averaging over a 30-year period (Wright, 2014). This change was intended to reduce the bias, with a careful discussion of pros and cons in order to define a dual standard for normals. Other authors, e.g. Livezey, Vinnikov, Timofeyeva, Tinker, and van den Dool (2007); Wilks (2013); Wilks and Livezey (2013), proposed alternative methods for deriving climatological normals and assessed these methods across the US. Optimal Climatological Normals (OCNs) are averages calculated over periods shorter than 30-years. The length of the averaging period is then selected to maximize the accuracy of the estimation – the authors typically considered 15-year means for temperatures across the US. Hinge fits are break-point statistical models where the climatological mean is expected to be constant before a given date, and linearly growing after that date. Authors cited above suggest that 1975 is a good choice for the break point over the US. Some national meteorological services already use these alternative estimation techniques operationally (ref to https://www.ncdc.noaa.gov/normalsPDFaccess/).

Another important feature of climatological normals is their time-resolution. Most normals are calculated on a monthly time-scale. However, for specific applications, the estimation of daily normals is required (Arguez, Vose, and Dissen, 2013). A few techniques have been proposed and /or are routinely used to translate monthly into daily values (Arguez and Applequist, 2013; Arguez, Applequist, Vose, Durre, Squires, and Yin, 2011; Arguez, Durre, Applequist, Vose, Squires, Yin, Heim, and Owen, 2012). Another option is to estimate daily normals directly from raw data, assuming some type of regularity in the seasonal variations (i.e. normals do not vary much from one day to the next). Doing this in a non-stationary context will require smoothness both in the seasonal cycle and the climate change components. This is the method employed in this manuscript.

In this paper, we assess the accuracy of previously proposed techniques for the estimating of climatological normals. We outline their limitations if applied in the course of the 21st century We then introduce a new approach for the overcoming of these issues. With this approach, the drift related to climate change on the seasonal component is estimated, leading to daily estimates. All evaluations are made in a predictive sense, i.e. assessing whether normals calculated in the (recent) past provide a reliable estimation of current to near-future climates.

The manuscript is organized as follows. After presenting the dataset in Section 2 we elaborate on the methods used to estimate climatological normals, then introduce our new method. The predictive skills of the various techniques considered are assessed and discussed in Section 4. This is followed by a discussion along with some concluding remarks in the last section.

2 Data & existing methods

2.1 Data

In order to assess the accuracy of various techniques for estimating climatological normals during the 21st century – a period over which observations are not available – we use series of daily and annual mean temperatures. These are simulated by an ensemble of climate models from the Coupled Model Intercomparison Project Phase 5 (CMIP5) as realistic realizations of future observations. Estimation techniques are therefore compared in a perfect model framework (see more details in Section 4).

More specifically, we focus on four locations which are meant to be representative of a wide range of climates: Bengaluru (India) in the tropics, Alert (Canadian Arctic Archipelago) in high-latitude, Paris (France) and San Francisco (California, USA) in mid-latitude regions. Twenty one CMIP5 models were selected for the daily mean temperature and sixty for the annual mean temperature (see Appendix A for a detailed list of models).

The considered time-series cover a period of 238 years from 1862 to 2099. They consist of the concatenation of two types of experiments:

- historical runs (driven by observed radiative forcings) covering the period 1862-2005,
- RCP8.5 scenario (Representative Concentration Pathways 8.5, corresponding to a high increase in greenhouse gas emissions during the 21st) simulations covering the remaining of the 21^{St} century (2006-2100).

The choice of a RCP8.5 scenario involves a strong climate change signal in the coming decades, but results obtained with this scenario are expected to hold at least qualitatively with more moderated alternatives.

It must be noted that for daily calculations, all the 29^{Th} February were removed to facilitate processing. Also, extensions to other climate variables, such as precipitation, are beyond the scope of this paper.

2.2 Previously introduced methods considered within this study

Here we review methods proposed by various authors in order to estimate climatological normals. Some of these techniques have been introduced in order to cope with climate change, and/or build upon the standard WMO recommendation. First we explain how these methods can be used to estimate annual normals, then we discuss how they can be extended to the daily timescale. This list of techniques is not meant to be exhaustive, but instead representative of what has been proposed in the literature.

- *2.2.1* ***WMO standard***

  The WMO recommendation is to calculate climatological normals as a simple average over a 30-year period composed of 3 full decades:

  $$WMO(D+k) = \frac{1}{30} \sum_{i=D-29}^{D} T_i, \tag{1}$$

  where $D+k$ is the current year, D is the current decade (e.g 2010), $k \in [\![1, 10]\!]$ denotes the year within the decade, T_i is the mean temperature (or any other meteorological variable) of year i. This calculation is updated every 10 years which means that, after a decade is completed, the estimated normals are valid and can be used for the subsequent 10 years (as denoted by $D+k$ in (1)).

- *2.2.2* ***WMO reset***

  As a first very simple alternative, the same calculation can be made and updated every year (instead of every decade), leading to

  $$WMO(y) = \frac{1}{30} \sum_{i=y-30}^{y-1} T_i, \tag{2}$$

where y is the current year. This will be referred to as WMO reset in the following, and is expected to be less biased than WMO in a changing climate thanks to the more frequent update.

– *2.2.3 Optimal Climate Normals (OCN)*

Huang, van den Dool, and Barnston (1996); Wilks (2013); Wilks and Livezey (2013) argued that averaging over a 30-year period was non-optimal (too long) in a climate change context, and suggested tuning the length of the averaging period to improve the accuracy of the estimate. They suggested that averaging over the most recent 15 years was a good compromise for temperature normals. As follows, OCN therefore designates a 15-year average:

$$OCN(D+k) = \frac{1}{15} \sum_{i=D-14}^{D} T_i, \tag{3}$$

with $k \in [\![1, 10]\!]$. As for WMO, this average can be updated every 10 years (as assumed in the following), or every year. In the following, this 15-year average will be used as a benchmark for other operational normals using the mean of a different number of years.

– *2.2.4* ***Hinge fit***

In order to account for non-stationary climates, other authors proposed the use of a statistical model allowing for a trend in the estimation of climate normals. Among these, the most popular technique is the hinge fit (Livezey et al, 2007; Wilks and Livezey, 2013). This is a simple break-point model where the normals are assumed to be constant (i.e. non time-dependent) up to a given date, then linearly moving with time. The date of the break-point needs to be selected carefully – Livezey et al (2007), Wilks and Livezey (2013) suggested that 1975 was an appropriate choice for the continental US and this is the value used in this paper.

$$Hinge(D+k) = \beta_0 + \beta_1 I_{1975}(D+k), \tag{4}$$

where $I(x) = 0$ if $x \leq 1975$ and $I(x) = x - 1975$ if $x > 1975$. The coefficients β_0 and β_1 are estimated from the full observational record available up to year D (i.e. not restricted to a 30-year period) using simple linear regression. Again this type of estimate could be updated each decade or year.

– *2.2.5* ***Hinge fit reset***

The same calculation can be made and updated every year instead of every decade, leading to

$$Hinge(y) = \beta_0 + \beta_1 I_{1975}(y), \tag{5}$$

The coefficients β_0 and β_1 are estimated from the full observational record available up to year $y-1$ using simple linear regression. This will be referred to as Hinge fit reset in the following.

2.2.6 *From annual to daily normals*

All of the techniques listed above can be used to derive daily (instead of yearly or even monthly) normals. This requires an additional procedure first introduced by Arguez and Applequist (2013) and consisting of an expansion in a Fourier basis. This technique is described below using the WMO estimate as an example, but it can be applied to any other normal estimator, including the OCN and Hinge methods introduced above. Firstly, we compute normals for each single day within a year, i.e.

$$WMO_{raw}(D+k,d) = \frac{1}{30}\sum_{i=D-29}^{D} T_{i,d}, \tag{6}$$

where $d \in [\![1,365]\!]$ represents the day, while other notations are consistent with (1). These daily values are then fitted onto the thirteen first elements of the Fourier basis. Equivalently, we estimate the linear coefficients α_i, β_i involved in the statistical model

$$WMO_{raw}(D+k,d) = \alpha_0 + \sum_{k=1}^{6}\left(\alpha_k \cos\left(\frac{2k\pi}{365}d\right) + \beta_k \sin\left(\frac{2k\pi}{365}d\right)\right) + \varepsilon_d. \tag{7}$$

Finally the estimated daily normals $\mathrm{WMO}_{\mathrm{day}}$ for year $D+k$ and day d are

$$WMO_{\mathrm{day}}(D+k,d) = \widehat{\alpha}_0 + \sum_{k=1}^{6}\left(\widehat{\alpha}_k \cos\left(\frac{2k\pi}{365}d\right) + \widehat{\beta}_k \sin\left(\frac{2k\pi}{365}d\right)\right), \tag{8}$$

where $\widehat{\alpha}_i, \widehat{\beta}_i$ are the estimated regression coefficients. Through projection onto a Fourier basis, this technique ensures regularity in the estimated annual cycle.

3 New Method

All methods described above could be criticized for a certain lack of flexibility (e.g. Krakauer (2012)). Indeed, climate is either assumed to be stationary locally (computing averages) or moving linearly over time, with the linearity holding over a relatively long period of time, from 1975 onwards (hinge fit). In this section, we introduce an alternative method for computing climatological normals, which is somewhat more flexible for its being based on spline smoothing. Obviously, the increase in flexibility is at the cost of an increase in the variance of the estimator – this will be discussed through the investigation of the overall performance of our approach in subsequent sections.

3.1 Statistical framework

The general statistical model considered is inspired by and adapted from Azaïs and Ribes (2016). Let $T_{y,d}$ be the mean (i.e. statistical expectation of) temperature of day d in year y. Our statistical model assumes that the following decomposition holds:

$$T_{d,y} = f(d) + g(y)h(d) + \varepsilon_{d,y}, \qquad d \in [\![1, 365]\!], y \in [\![1, n]\!], \tag{9}$$

where:

- $f(), g(), h()$ are smooth functions ($f(d), g(y), h(d)$ being their trace on integer values),
- $f(), h()$ are, additionally, periodic functions with period 365,
- ε is assumed to be Gaussian white noise with unknown variance σ^2.

In addition we impose the constraints $\sum_{y=1}^{n} g(y) = 0$ and $\sum_{d=1}^{365} h(y) = 1$ in order to ensure model identifiability (i.e. to avoid any possible confusion between the terms f and gh). Note that another system of constraints is possible in order to facilitate interpretation (see Appendix B).

This statical model can be interpreted as follows. $f(d)$ represents a stationary seasonal cycle, which would be observed if the climate was stationary and the effect of climate change is described by the term $g(y)h(d)$. The key assumption is that this climate change response can be factorized into one component which describes how the shape of a seasonal cycle changes, $h(d)$, and another one which describes the variation of the magnitude of this change with time, in the long-term, $g(y)$.

This type of decomposition is an adaptation of the *pattern scaling* assumption (Mitchell, 2003; Tebaldi and Arblaster, 2014; Geoffroy and Saint-Martin, 2014) in a slightly different setup. Under *pattern scaling*, it is assumed that the spatial distribution of climate change does not vary with time – only the amplitude of the change does. It is thus possible to decompose climate change as the product between one spatial function, and one temporal function. In the present paper, the spatial component is replaced by the seasonal cycle. In both cases, the assumption can be thought of as a Taylor approximation of order one, which is valid as long as the change is small enough. This factorization assumption is obviously one of consequence but has already proven its descriptive capabilities on hourly surface air temperature observations (Vinnikov, Robock, Grody, and Basist, 2004). Its primary interest comes from the induced reduction in the model's complexity: estimating two univariate functions g and h is much easier than estimating a bivariate function (say $c(y, d)$). Its introduction therefore allows us to better constrain the estimation of the climate change component. Additionally, model (9) proved a very good capability across the entire time series considered and it can be at least partially validated by examining goodness-of-fit to model (9), as discussed below.

An illustration of this model and the typical outputs it can produce, is shown in Figure1. Next we will discuss how estimating the unknown functions

f, g, h within this model. Goodness-of-fit of this model is also discussed in Section 3.4.

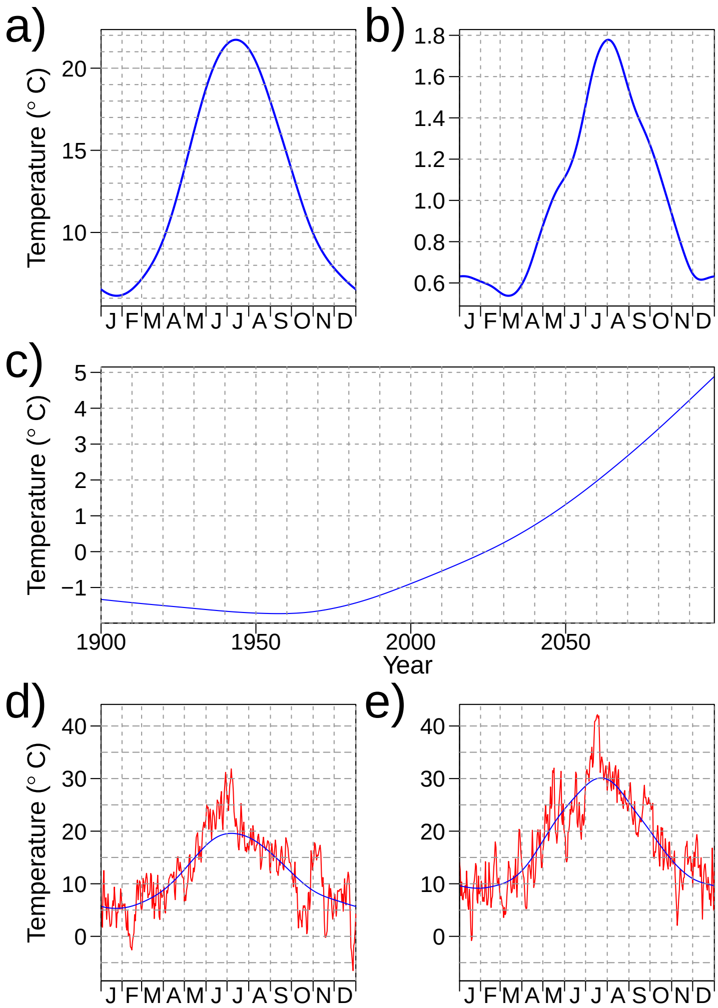

Fig. 1 Decomposition of a time series (Paris) by the spline model (9). a) represents the reference seasonal cycle f with *df=11* (see section 2.4), b) illustrates the seasonal drift h with *df=10*, and c) represents the annual trend g with *df=10*. The plots d) and e) show the estimation of the annual cycle in 1900 and 2030 respectively. Raw data are shown in red, while the fit of model (1) is in blue.

3.2 Estimation Algorithm

Our estimation procedure is a sequential, two-step procedure. Firstly, $g()$ is estimated using annual mean data only. Secondly, $f()$ and $h()$ are estimated assuming that $g()$ is known.

Both steps involve smoothing with cubic splines. For instance, denoting $T_{.y} = \frac{1}{365}\sum_{d=1}^{365} T_{d,y}$ the annual mean temperature of year y, the smoothing splines estimate that $\widehat{g}()$ of $g()$ can be defined as

$$\widehat{g}() = \underset{s()}{\operatorname{argmin}} \sum_{i=1}^{n} \left(T_{.y} - s(y)\right)^2 + \lambda \int_1^n \left(s''(x)\right)^2 dx, \tag{10}$$

where the minimum is taken over all possible function $s()$ belonging to the associated Sobolev space. A spline estimate thus performs a trade-off between closeness to input data (here $T_{.y}$), and roughness (last term in the right-hand side). λ is a regularization parameter determining the level of smoothness. The selection of λ is a common but difficult problem which is addressed in detail in Subsection 3.3. Remarkably, the solutions of (10) are known in closed forms and can be computed easily.

Furthermore, we attempt to provide a calculation which meets operational constraints, and which is thus computationally not too expensive so as to apply it to multiple grid points. For this reason we implemented the sequential algorithm described below.

Algorithm

---

1 **Estimation of $g()$:**
Calculate the annual means $T_{.y}$. From the $T_{.y}$ time-series, compute the smoothing spline estimate $\widehat{g}()$ of $g()$, with a given df_g. Note that this estimate has to be centered subsequently in order to satisfy the identifiability constrains.

2 **Linear regression on $\widehat{g}(y)$:**
For each day $d \in [\![1, 365]\!]$, the time-series $T_{y,d}$ is linearly regressed onto $\widehat{g}(y)$, i.e. we estimate the coefficients α_d, β_d involved in:

$$T_{d,y} = \alpha_d + \widehat{g}(y)\beta_d + \varepsilon_{y,d}. \tag{11}$$

Thanks to orthogonality, $\widehat{\alpha}_d = T_{d,.}$, where $T_{d.} = \frac{1}{n}\sum_{y=1}^{n} T_{y,d}$, and $\widehat{\beta}_d = \dfrac{\sum_{y=1}^{n} \widehat{g}(y) T_{d,y}}{\sum_{y=1}^{n} \widehat{g}(y)^2}$.

3 **Estimation of $f()$ and $h()$:**
From the series $\widehat{\alpha}_d$ and $\widehat{\beta}_d$, respectively, we calculate the estimates $\widehat{f}()$ and $\widehat{h}()$, as periodic cubic smoothing splines estimates, with given df_f and df_h.

As it is sequential and based on an orthogonal design in the regression step, this algorithm is very rapid. A more sophisticated, iterated version of the algorithm has also been studied, and is presented in Appendix C. This variant showed no real improvement however and was thus dismissed.

Predictions based on the model (9) can be derived by extrapolating the estimated spline $\widehat{g}()$ to the year in question. Note that, as natural splines are used to estimate g (i.e second derivatives are null at the terminating points), this extrapolation is linear.

3.3 Selecting degrees of freedom

The selection of the smoothing parameters λ (there is one parameter for each function $f(), g()$ and $h()$), will be discussed in terms of "equivalent degrees of freedom" (df), as in many spline papers. df is meant to be the equivalent of the number of parametric predictors involved in the estimation of the function. The smaller the df, the smoother the function estimate. Note that df is a complex one-to-one function of λ – but this correspondence will not be detailed further.

The determination of the different degrees of freedom (df) is performed using a variant of cross validation methods adapted to a prediction context. We use a multi-model ensemble of transient simulations covering the 1850-2100 period (see Section 2.1) as plausible realizations of the real world. From this dataset, we look for the value of df allowing the best prediction for the coming decade (e.g. 2011-2020) using data from previous years (e.g. 1850-2010).

This procedure is distinct from common cross-validation. Usual cross-validation would, in our case, consist of removing one or several years from the available observations (e.g. 1850-2017 if we are in 2018), and tuning the df coefficients to make the estimated normals as close as possible to the years removed. This procedure is then repeated by removing different years. If this type of cross-validation were used, then the df coefficients would be optimized to best estimate normals in the past – a period over which climate exhibits no or little change. Given that the non-stationary feature of climate is larger now than in the past, the best df for prediction might differ from the best df in the past – we checked that this was effectively the case.

Finally, the three coefficients involved, hereafter df_f, df_g, df_h, are estimated sequentially, instead of simultaneously. This makes the selection procedure computationally more affordable.

In each of the three cases, the observation is decomposed into a training sample and a testing sample. For various values of the number of degrees of freedom df, the considered function (f, g or h) is estimated on the training sample and then compared to the testing sample by measuring a Mean Square Error (MSE). Results are averaged over the available climate models. The df leading to the smallest MSE is then selected. In addition to the MSE value, we estimate its standard deviation which enables the computation of a plausible range of values for df, through the *one standard error rule* Hastie, Tibshirani, and Wainwright (2015).

Degree of freedom of the reference cycle $f()$

$f()$ is meant to represent the mean annual cycle in a stationary climate. In order to select df_f, we took the periods 1900-1930 as a training sample, and 1931-1940 as a testing sample. This somewhat subjective choice was motivated by the fact that climate in the early 20th century is almost stationary.

The selected df_f is typically between 10 and 20, depending on the location considered. Note that the signal-to-noise ratio is much higher for this stationary component $f()$ than for the remaining $g()$ and $h()$ functions, which explains why df_f is relatively large and well defined.

Degree of freedom of the annual trend $g()$

Unlike df_f, df_g depends on the decade considered. For a given decade D (for example 2001-2010), we use the data prior to D (i.e. the period 1862-2000) as a training sample, the decade itself being the testing sample. Again, we use the *one standard error rule* to assess a range of value for df_g.

Our selection procedure for df_g is illustrated in Figure 2a–b. Note that only annual mean values are used there. Focusing on the 2050 decade, the best value for df_g lies between 5 and 6 (panel a). Values smaller than 3 are clearly discarded, but the accuracy of the estimated normals is only slightly deteriorated if larger df_g are used (up to more than 15). Remarkably, the selected df_g is almost constant from 1990 to 2100 (panel b), with optimal values around 6. This applies to many other locations (not shown). Moreover, as the cross validation curve was very flat around its minimum, for all predictions made after 1990, we will use $df_g = 6$ in the following. Using such a constant value makes the algorithm easier to implement.

Degree of freedom of the delta cycle $h()$

Once df_f and df_g have been determined, estimates of $f()$ and $g()$ can be derived, and df_h is the only missing parameter to fit. In order to select df_h for a given year (2018 for example), we used *the past* (i.e. 1862–2017) as a training sample, then calculate the mean square error (MSE) over the next year (2018 in this case). Due to the strength of internal variability, we applied a smoothing over time. For each year, the selected df_h is the one minimising the smoothed MSE (see Figure 2c–d).

Our results suggest that df_h is the most sensitive (and therefore difficult to estimate) parameter in our statistical model. The selected values for df_h vary substantially both over space and time. In the case of Paris (Figure 2d), df_h increases with time from near 1 (i.e. the minimum possible value, corresponding to no change in the annual cycle) to 10 in 2100. This corresponds to the signal-to-noise increase across the 21st century. In 1990, climate change was limited, and it is unclear which season experienced the greatest warming. It is thus safer to assume a flat response (i.e. the same degree of warming throughout the year). During subsequent decades,this changing signal (including change in the annual cycle) becomes clearer and greater flexibility in $h()$ becomes effective.

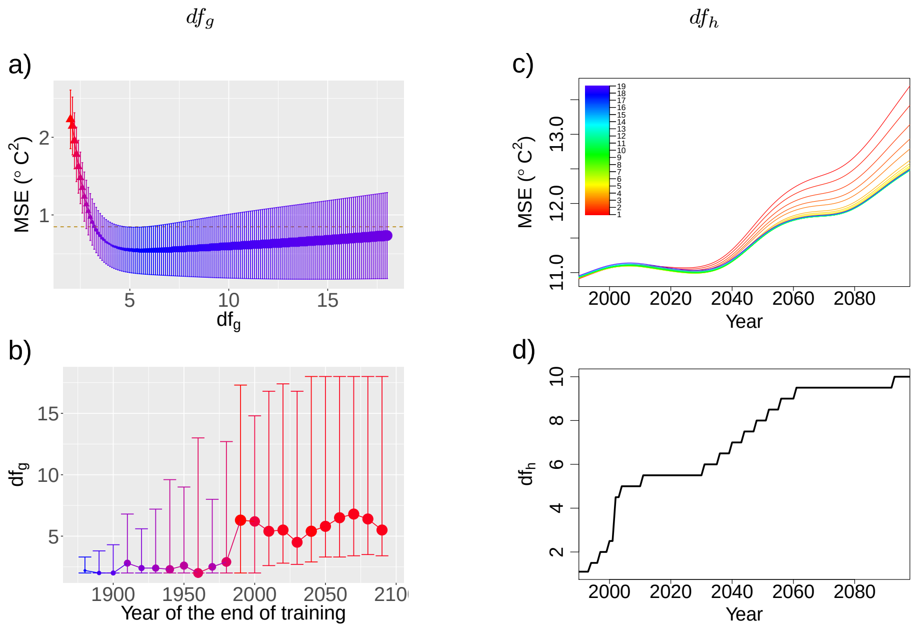

Fig. 2 **Selecting df_g and df_h. a)** error of annual mean temperature normals (points) for the decade 2050 (normals are estimated from 1862-2050 data, error is calculated over 2051-2060), and its standard deviation (bars), as a function of df_g. **b)** selected df_g (points), with the corresponding uncertainty according to the one standard error rule (bars), as a function of the predicted decade D. **c)** Mean square prediction error of daily normals as a function of time, for different values of df_h (df_g, df_f are given). **d)** Selected df_h as a function of the predicted year. All calculations in this figure are made for the Paris (France) grid-point.

3.4 Model goodness-of-fit

An important step in order to validate the use of our statistical model (9) and its underlying assumption is to assess the goodness-of-fit to this model. This can be done using climate model data, and fitting the model across the entire period considered (1862–2100). Such diagnoses are shown in Figure 3. Consistent with Figure 1, these diagnoses apply to Paris and the CNRM-CM5 climate model; they are representative of different locations and models.

Firstly, the determination coefficient $R^2 = 0.73$ is relatively high, and consistent with the internal climate variability. Residuals show no abnormal patterns: the Gaussian assumption is reasonably well-satisfied (2a), and they do not exhibit clear dependence on the fitted value (2b). Note that in the latter panel, the density of points depend strongly on the fitted values, thanks to the annual cycle and the fact that the climate is almost stationary over the first 100 years. For instance, the accumulation around $19°C$ is due to pre-industrial summer maxima.

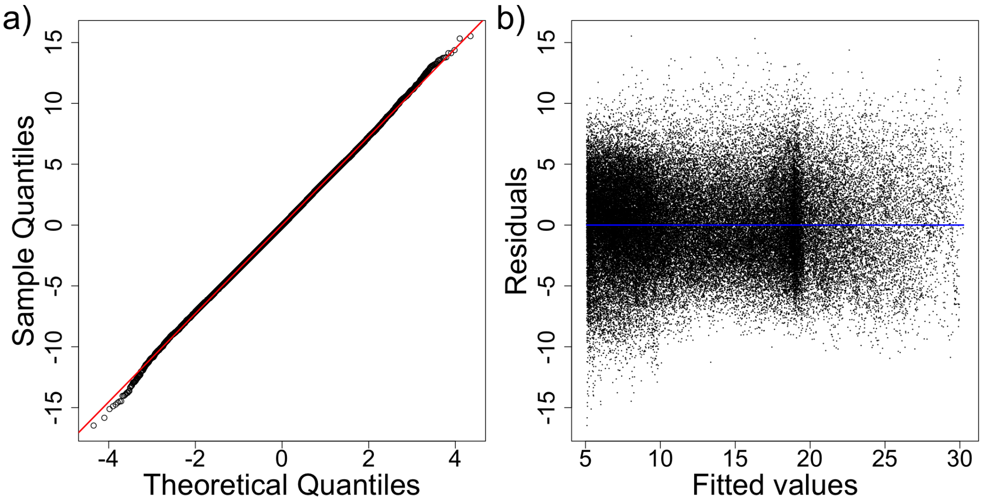

Fig. 3 Goodness-of-fit to our statistical model. **a)** normal QQ-plot of the residuals. **b)** residual vs fitted values plot.

4 Results

4.1 Scores on annual mean temperature

The results of the five methods introduced above are compared for annual mean temperatures in Figure 4. The comparison is performed for 4 distinct locations, corresponding to different climates (mean temperature ranges from −32°C to +27°C), amounts of warming (from 4°C to 8°C in 2100 under RCP8.5), and signal-to-noise ratio (internal variability being relatively smaller in the tropics).

Globally, for all locations, the methods have almost the same performance until the late 20th century (near 1990 or 2000, depending on the location). The hinge fit however seems to exhibit a larger variance after 1975 (see e.g. quick variations in Alert and Bengaluru). This is because very few points contribute significantly to fit the broken line's trend. This also applies to a lesser extent to OCN, given that the average is calculated on a smaller number of years than that of the WMO. The sampling margin of error is therefore larger.

During the early 21st century, methods based on averaging over past years (namely OCN, WMO, and WMO reset), are starting to depart from the reference, and show a negative bias. Hinge fit and our technique do a much better job and remain close to the reference.

Lastly, our method performs much better than any other in the second half of the 21st century. While this method remains continuously close to the reference, alternatives systematically underestimate the current state of the climate, by .5 to 1 degree.

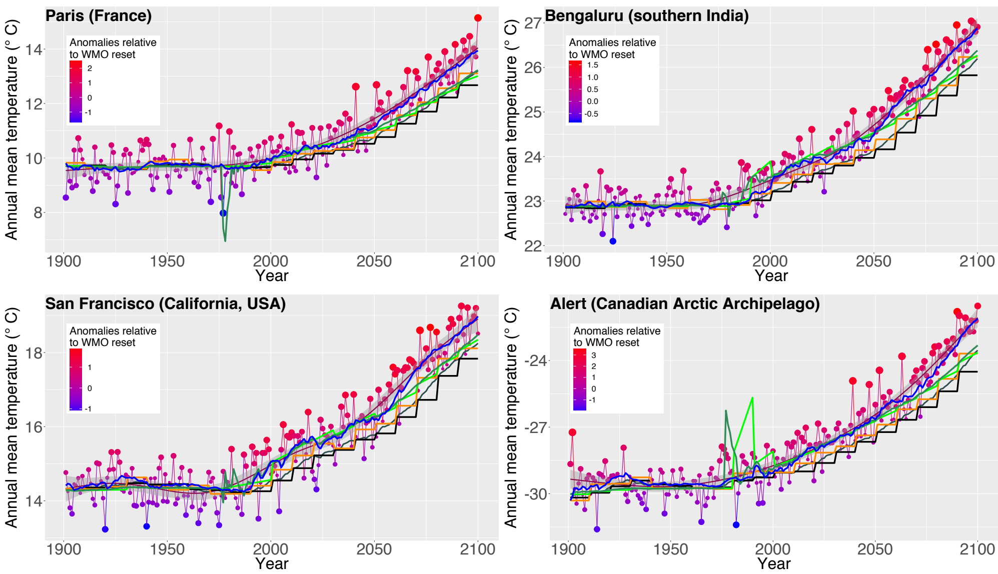

Fig. 4 Annual mean temperature and estimated normals. Temperature normals on an RCP8.5 scenario of the CNRM-CM5 model. Time-series of annual mean temperature (points) at four different locations (panels). Climatological normals are estimated using 6 different techniques: WMO standard (black line), WMO reset (grey), OCN (yellow), hinge fit (light green), hinge fit reset (dark green), and our method (blue). Normals for a given year (e.g. 2018) are estimated using data from previous years (e.g. 1900-2017). A smoothing spline of the entire time-series (1900-2100; purple line) can be considered as a reference. Anomalies of individual years are calculated with respect to the WMO reset, in order to further illustrate the bias related to this method. All calculations are based on one RCP8.5 simulation from CNRM-CM5.

Beyond the illustrative and qualitative assessment made in Figure 4, methods can be quantitatively compared using standard criterion such as MSE, bias and variance. Such a score-based comparison will be carried out in detail in the next section for daily normals. It is also appropriate on annual time-scales, and is illustrated in appendix. These quantitative results are consistent with those described above.

4.2 Scores on the daily timescale

Figure 5 compares the performance of the 5 considered methods, at the daily timescale, and for one grid point near San Francisco. Again, the estimation techniques are trained on all years prior to the one predicted, then the estimated normals are extrapolated to that year. Evaluation of the methods is based primarily on the mean square error (MSE)(Council et al, 2010). The latter is also decomposed as the sum of the $bias^2$ and the variance.

Bias varies from 0 to more than 1°C depending on the method and period of time. If all methods are nearly unbiased in the late 20th century, only our approach remains unbiased throughout the 21st century. Alternatives exhibit negative bias as large as .5 to .8°C, except for the standard WMO approach for which the bias is even larger, near or beyond 1°C. Even though hinge and hinge reset lie close to our method until 2040, their bias are slightly larger, on average. Overall, in the 2000-2100 period, methods can be sorted with respect to their bias (increasing order): our method, hinge reset, hinge, OCN, and WMO. These results are highly consistent with those obtained on annual mean temperature.

The variance of all estimation techniques are in fact very close to one another. Only the hinge and hinge reset estimators yield a slightly higher variance than others, especially near the beginning of the period. Our technique has the lowest variance on average over the entire period.

In terms of Mean Square Error (MSE), which is an aggregation of bias and variance and a very usual criterion, our technique performs muchbetter than all proposed alternatives. OCN and WMO approaches are reasonably accurate near the beginning of the period, for instance before 2020, when climate change remains slight. They are penalized by their large bias subsequently. The two variants of the hinge technique suffer from their large variance at the beginning, then rank second from 2010 to 2020.

Two additional remarks can be made. Firstly, results found for other locations were qualitatively similar. In particular, they confirm that our method outperforms the proposed alternatives, and remains almost unbiased across the 21st century. Secondly, all methods reset on a decadal basis exhibit some degradation of their scores at the end of each decade (WMO, OCN and hinge). This is particularly pronounced in the bias of OCN and WMO.

Overall, these results suggest that our method is more accurate than existing alternatives. This happens both in terms of bias and variance, which can be underlined. Furthermore, very low bias is revealed over the 21st century. This suggests that our technique has the appropriate level of flexibility to follow climate change, whilst not having too much variance. As our method exhibits almost no bias, potentially more sophisticated methods could improve on the variance (the bias is already minimal). This would probably lead to limited gain in terms of total MSE, as a large part of this variance is related to (irreducible) internal climate variability.

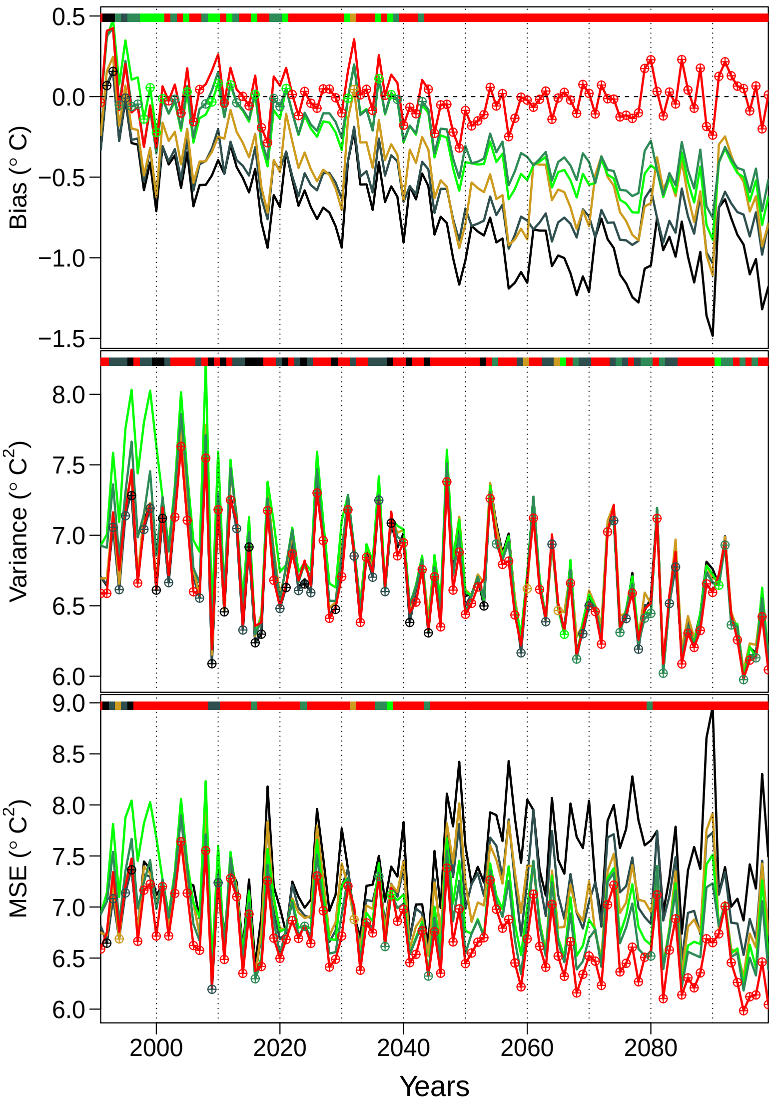

Fig. 5 Daily scores in San Francisco. Six techniques for estimating daily normals, namely WMO standard (black), WMO reset (grey), OCN (yellow), hinge fit (light green), hinge fit reset (dark green), and our method (red), are compared. Their evaluation is based on their bias (top), variance (middle), and mean square error (MSE, bottom). The year in the x-axis denotes the end of the training period; prediction is made for the following year. The coloured line (top of each panels) indicates which method performs best, for a given criterion and a given year. Calculations are made for one grid-point near San Francisco, using an ensemble of RCP8.5 simulation from the CMIP5 archive.

5 Conclusion and discussion

In this paper, we introduce a new method for estimating daily climatological normals, describing the corresponding model. This technique relies on the assumption that the response to climate change is smooth over time and nears the pattern scaling assumption. All terms can by estimated using smoothing splines. The proposed estimation algorithm is very fast and this is due to a two-step (as opposed to simultaneous) procedure. The main challenge is the tuning of the smoothing parameters which is done using an extension of cross validation specifically designed for prediction.

Our method is compared to previously proposed alternatives in a predictive sense: methods are used to estimate climatological normals for the next, unobserved year. Their accuracy is compared on that basis, using an ensemble of RCP8.5 simulation from the CMIP5 ensemble in a perfect model framework.

Results show that our method is more accurate than all considered alternatives on the yearly timescale. The gap is particularly large across the second part of the 21^{St} century. Additionally, on the daily timescale, our method was also shown to provide the best results in terms of bias, variance, and therefore mean square error. These good properties can be partly attributed to the flexibility of the method, adjusted through the selection of smoothing parameters.

Our results thus suggest that the proposed method brings a strong improvement in the estimation of climatological normals accounting for climate change. Such revised – with respect to the WMO recommendation – normals could be used to address several questions. Unbiased normals could be particularly useful for climate monitoring, e.g. qualify if a year or season is warmer or colder than *really* expected. It could also be used to produce climate change corrected times-series. This would be relevant e.g. to compare how anomalous different years or periods are. As a typical illustration, one might wonder if an extreme event like the 2003 European Event remains unprecedented after correction for the climate change effect. Additionally, our method could be used to provide a refined description of on-going climate change with respect to the annual cycle, i.e. beyond the annual mean warming.

These attractive features do not mean that the standard way of computing climatological normals is now obsolete. Having a stationary reference such as the WMO standard is still very valuable, e.g. in order to highlight climate change. We suggest therefore that weather or climate services in charge of climate monitoring could compute two different sets of normals – a stationary reference and a climate change corrected set of normals – and use one or the other depending on the application considered. Updating the revised set of normals on a regular annual basis seems to be something required for the delivery of an estimation as accurate as possible.

Future work on the method described in this paper could include the estimation of uncertainties in the estimated normals. This would be very valuable, e.g. for assessing the uncertainties in climate change corrected time-series. Future work could also include a pre-computation of smoothing parameters for a large number of locations, in order to make the method even easier to implement. This tuning step remains the most difficult in our procedure and has to be re-examined carefully for different places. Lastly, the selection of smoothing parameters could be re-examined for different emission scenarios for the shape of the time response (and therefore the optimal value of smoothing parameters) greatly depends on the emission pathway.

Appendix

The analysis in this article has been performed using the statistical software `R`.

A Computational and simulation details

The 21 simulations used for daily mean temperature were: ACCESS1-0, ACCESS1-3, CCSM4, CESM1-BGC, CMCC-CMS, CNRM-CM5, CSIRO-Mk3-6-0, CanESM2, GFDL-CM3, GFDL-ESM2G, GFDL-ESM2M, IPSL-CM5A-LR, IPSL-CM5A-MR, IPSL-CM5B-LR, MIROC-ESM-CHEM, MIROC-ESM, MPI-ESM-LR, MPI-ESM-MR, MRI-CGCM3, NorESM1-M, inmcm4

The simulations used for annual mean temperature were: ACCS0 _ r1i1p1, ACCS3 _ r1i1p1, BCCl _ r1i1p1, BCCm _ r1i1p1, BNU _ r1i1p1, CCCMA _ r1i1p1, CCCMA _ r2i1p1, CCCMA _ r3i1p1, CCCMA _ r4i1p1, CCCMA _ r5i1p1, CNRM _ r10i1p1, CNRM _ r1i1p1, CNRM _ r2i1p1, CNRM _ r4i1p1, CNRM _ r6i1p1, CSIRO _ r10i1p1, CSIRO _ r1i1p1, CSIRO _ r2i1p1, CSIRO _ r3i1p1, CSIRO _ r4i1p1, CSIRO _ r5i1p1, CSIRO _ r6i1p1, CSIRO _ r7i1p1, CSIRO _ r8i1p1, CSIRO _ r9i1p1, GFDLc _ r1i1p1, GFDLg _ r1i1p1, GFDLm _ r1i1p1, GISSr _ r1i1p1, IAPg _ r1i1p1, IAPs _ r1i1p1, IAPs _ r2i1p1, IAPs _ r3i1p1, INGVc _ r1i1p1, INGVe _ r1i1p1, INGVs _ r1i1p1, INM _ r1i1p1, IPSLal _ r1i1p1, IPSLal _ r2i1p1, IPSLal _ r3i1p1, IPSLal _ r4i1p1, IPSLam _ r1i1p1, IPSLb _ r1i1p1, MIROC5 _ r1i1p1, MIROC5 _ r2i1p1, MIROC5 _ r3i1p1, MIROCc _ r1i1p1, MIROCe _ r1i1p1, MPIMl _ r1i1p1, MPIMl _ r2i1p1, MPIMl _ r3i1p1, MPIMm _ r1i1p1, MRI _ r1i1p1, NCARc _ r1i1p1, NCARc _ r2i1p1, NCARc _ r3i1p1, NCARc _ r4i1p1, NCARc _ r5i1p1, NCARc _ r6i1p1, NCARe _ r1i1p1

B Another system of constraints for model (9)

Once we have obtained the decomposition of model (9), it is possible to make it more interpretable. Let $\tilde{g} = g - g(1)$, $\tilde{f} = f + g(1).h$. Then, the decomposition of model (1) can be rewritten as:

$$\begin{aligned} f(d) + g(y).h(d) &= (f(d) + g(1).h(d)) + (g(y) - g(1)).h(d) \\ &= \tilde{f}(d) + \tilde{g}(y).h(d) \end{aligned}$$

Thus, $\tilde{f}$ represents the annual reference cycle of the first year of the considered period and $\tilde{g}$ quantifies the annual mean temperature evolution. Therefore the first value, $\tilde{g}(1)$, is zero.

C Alternating least squares

Addition of a few steps to the sequential algorithm permitting an iterative procedure:

4 **Re-estimation of $g()$:**
We now fix $\hat{f}, \hat{h}$ and estimate g once again, the goal of the procedure being minimization of the total sum of squares
i.e $RSS = \sum_{y,d}(T_{y,d} - \hat{f}_d - \hat{g}_y.\hat{h}_d)^2$.
For a fixed y, let us define:
$RSS_y = \sum_d((T_{y,d} - \hat{f}_d) - g_y.\hat{h}_d)^2 = \sum_d(\widetilde{T}_{y,d} - g_y.\hat{h}_d)^2$
where $\widetilde{T}_{y,d} = T_{y,d} - \hat{f}_d$
let $g_{0,y}$ the mean square estimator $g_{0,y} = \frac{\sum_{j=1}^{365} \hat{h}_d.T_{y,d}}{\sum_{j=1}^{365} \hat{h}_d^2}$
Also by the Pythagorean theorem:

$$\begin{aligned}
\sum_{d=1}^{365}(\widetilde{T}_{y,d} - g_y.\hat{h}_d) &= \sum_{d=1}^{365}(\widetilde{T}_{y,d} - g_{0,y}.\hat{h}_d + (g_{0,y} - g_y).\hat{h}_d)^2 \\
&= \sum_{d=1}^{365}(\widetilde{T}_{y,d} - g_{0,y}.\hat{h}_d)^2 + \sum_{d=1}^{365}((g_{0,y} - g_y).\hat{h}_d)^2 \\
&= \sum_{d=1}^{365}(\widetilde{T}_{y,d} - g_{0,y}.\hat{h}_d)^2 + (g_{0,y} - g_y)^2.\sum_{d=1}^{365}\hat{h}_d^2
\end{aligned}$$

Finally,

$$\begin{aligned}
RSS &= \sum_{y=1}^{n} RSS_y \\
&= \sum_{d,y}(\widetilde{T}_{y,d} - g_{0,y}.\hat{h}_d)^2 + \sum_{y=1}^{n}(g_{0,y} - g_y)^2.\sum_{d=1}^{365}\hat{h}_d^2
\end{aligned}$$

Then, we compute the smoothing spline estimate $\widehat{g}()$ of $g_{0,y}$, with the given df_g.

5 We iterate steps 3 and 4 to minimize sum of squares RSS.

557 # D Annual scoring for normals

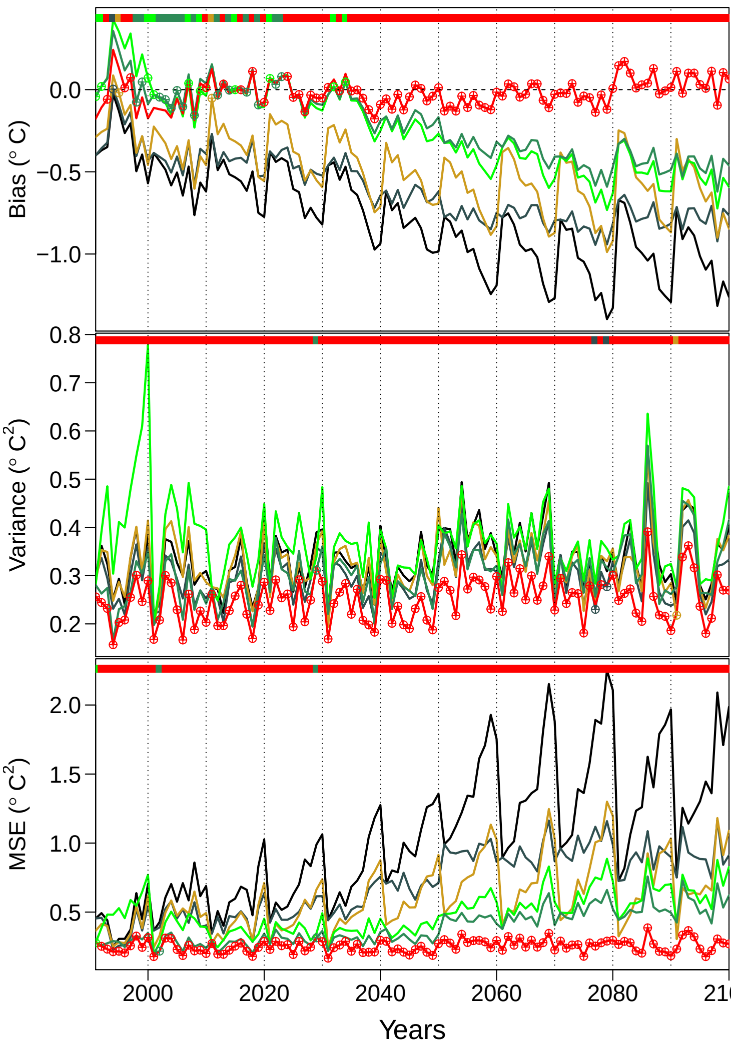

Fig. 6 The three plots illustrate scoring on the yearly mean temperature at San Francisco, for each year normal prediction occurs on all CMIP5 models. The horizontal axis represents the end of the training period and for each method, prediction occurs the following year. The upper line shows, for each score, the winning method for predicting the next year. The different calculations are WMO (black), WMO reset (grey), OCN (yellow), hinge (light green), hinge fit reset (green) and model(9) (blue). The upper figure shows the evolution of the bias, the middle one represents the variance of the prediction and the bottom plot illustrates the evolution of the mean square prediction error (MSE).

Acknowledgements he authors acknowledge Météo-France for supporting this study. They also wish to thank the climate modeling groups involved in CMIP5 for producing and sharing their simulations.

References

Arguez A, Applequist S (2013) A harmonic approach for calculating daily temperature normals constrained by homogenized monthly temperature normals. Journal of Atmospheric and Oceanic Technology 30(7):1259–1265, DOI 10.1175/JTECH-D-12-00195.1, URL https://doi.org/10.1175/JTECH-D-12-00195.1, https://doi.org/10.1175/JTECH-D-12-00195.1

Arguez A, Vose RS (2011) The definition of the standard WMO climate normal: The key to deriving alternative climate normals. Bulletin of the American Meteorological Society 92(6):699–704, DOI 10.1175/2010BAMS2955.1, URL http://journals.ametsoc.org/doi/pdf/10.1175/2010BAMS2955.1

Arguez A, Applequist S, Vose RS, Durre I, Squires MF, Yin X (2011) NOAA's 1981–2010 climate normals: Methodology of temperature-related normals. Tech. rep., URL https://www1.ncdc.noaa.gov/pub/data/normals/1981-2010/documentation/temperature-methodology.pdf

Arguez A, Durre I, Applequist S, Vose RS, Squires MF, Yin X, Heim RR, Owen TW (2012) Noaa's 1981-2010 U.S. climate normals. Bulletin of the American Meteorological Society 93(11):1687–1697, DOI 10.1175/BAMS-D-11-00197.1

Arguez A, Vose RS, Dissen J (2013) Alternative climate normals: Impacts to the energy industry. In: Bulletin of the American Meteorological Society, vol 94, pp 915–917, DOI 10.1175/BAMS-D-12-00155.1, URL http://journals.ametsoc.org/doi/pdf/10.1175/BAMS-D-12-00155.1

Azaïs JM, Ribes A (2016) Multivariate spline analysis for multiplicative models: Estimation, testing and application to climate change. Journal of Multivariate Analysis 144:38–53, DOI 10.1016/j.jmva.2015.09.026, URL http://ac.els-cdn.com/S0047259X15002560/1-s2.0-S0047259X15002560-main.pdf?_tid=b6d26b82-243a-11e7-95bb-00000aacb362&acdnat=1492522205_708816799da1e435d00bab291f7a5631

Baddour O (2011) Climate Normals, World Meteorological Organization Commission for Climatology Management Group Meeting , ITEM 10 URL http://www.wmo.int/pages/prog/wcp/ccl/mg/documents/mg2011/CCl-MG-2011-Doc_10_climatenormals1.pdf

Council NR, et al (2010) Assessment of intraseasonal to interannual climate prediction and predictability. National Academies Press

Geoffroy O, Saint-Martin D (2014) Pattern decomposition of the transient climate response 66:23,393

Hastie T, Tibshirani R, Wainwright M (2015) Statistical learning with sparsity: the lasso and generalizations. CRC press

Huang J, van den Dool HM, Barnston AG (1996) Long-lead seasonal temperature prediction using optimal climate normals. Journal of Climate 9(4):809–817

Krakauer NY (2012) Estimating climate trends: Application to united states plant hardiness zones. Advances in Meteorology DOI 10.1155/2012/404876

Krakauer NY, Devineni N (2015) Up-to-date probabilistic temperature climatologies. Environ Res Lett Environ Res Lett 10(10), DOI 10.1088/1748-9326/10/2/024014, URL http://iopscience.iop.org/1748-9326/10/2/024014

Livezey RE, Vinnikov KY, Timofeyeva MM, Tinker R, van den Dool HM (2007) Estimation and extrapolation of climate normals and climatic trends. Journal of Applied Meteorology and Climatology 46(11):1759–1776, DOI 10.1175/2007JAMC1666.1, URL https://doi.org/10.1175/2007JAMC1666.1, https://doi.org/10.1175/2007JAMC1666.1

Mitchell TD (2003) An Examination of the Accuracy of the Technique for Describing Future Climates. Climatic Change Volume 60(Issue 3):pp 217–242, URL https://link.springer.com/content/pdf/10.1023%2FA%3A1026035305597.pdf

Scherrer SC, Appenzeller C, Liniger MA (2006) Temperature trends in Switzerland and Europe: Implications for climate normals. International Journal of Climatology DOI 10.1002/joc.1270

Tebaldi C, Arblaster J (2014) Pattern scaling: Its strengths and limitations, and an update on the latest model simulations. Climatic Change 122:459–471, DOI 10.1007/s10584-013-1032-9

Vinnikov KY, Robock A, Grody NC, Basist A (2004) Analysis of diurnal and seasonal cycles and trends in climatic records with arbitrary observation times. Geophysical Research Letters 31(6):n/a–n/a, DOI 10.1029/2003GL019196, URL http://doi.wiley.com/10.1029/2003GL019196

Wilks DS (2013) Projecting "normals" in a nonstationary climate. Journal of Applied Meteorology and Climatology 52(2):289–302, DOI 10.1175/JAMC-D-11-0267.1

Wilks DS, Livezey RE (2013) Performance of alternative "normals" for tracking climate changes, using homogenized and nonhomogenized seasonal u.s. surface temperatures. Journal of Applied Meteorology and Climatology 52(8):1677–1687, DOI 10.1175/JAMC-D-13-026.1, URL https://doi.org/10.1175/JAMC-D-13-026.1, https://doi.org/10.1175/JAMC-D-13-026.1

World Meteorological Organization (2007) The Role OF Climatological Normals In a Changing Climate. WCDMP-No 61 WMO-TD

No.1377(World Climate Data and Monitoring Programme), URL https://www.wmo.int/datastat/documents/WCDMPNo61_1.pdf
Wright W (2014) Discussion paper on the calculation of the standard climate normals: A proposal for a dual system. World Climate Data and Monitoring Program, accessed 14