SPLIDHOM: A Method for Homogenization of Daily Temperature Observations

OLIVIER MESTRE

Ecole Nationale de la Météorologie, Météo-France, and Equipe Statistique et Probabilités,
Institut de Mathématiques de Toulouse, Toulouse, France

CHRISTINE GRUBER

Zentralanstalt für Meteorologie und Geodynamik, Vienna, Austria

CLÉMENTINE PRIEUR

Equipe Projet INRIA MOISE, Laboratoire Jean Kunzmann, Université Joseph Fourier, Grenoble, France

HENRI CAUSSINUS

Equipe Statistique et Probabilités, Institut de Mathématiques de Toulouse, and Université Paul Sabatier, Toulouse, France

SYLVIE JOURDAIN

Direction de la Climatologie, Météo-France, Toulouse, France

(Manuscript received 31 August 2010, in final form 31 March 2011)

ABSTRACT

One major concern of climate change is the possible rise of temperature extreme events, in terms of occurrence and intensity. To study this phenomenon, reliable daily series are required, for instance to compute daily-based indices: high-order quantiles, annual extrema, number of days exceeding thresholds, and so on. Because observed series are likely to be affected by changes in the measurement conditions, adapted homogenization procedures are required. Although a very large number of procedures have been proposed for adjustment of observed series at a monthly time scale, few have been proposed for adjustment of daily temperature series. This article proposes a new adjustment method for temperature series at a daily time scale. This method, called spline daily homogenization (SPLIDHOM), relies on an indirect nonlinear regression method. Estimation of the regression functions is performed by cubic smoothing splines. This method is able to adjust the mean of the series as well as high-order quantiles and moments of the series. When using well-correlated series, SPLIDHOM improves the results of two widely used methods, as a result of an optimal selection of the smoothing parameter. Applications to the Toulouse, France, temperature series are shown as a real example.

1. Introduction

Extreme indices have recently been used by a greater part of the climatological community to assess the impacts of extreme events on our society (e.g., Klein Tank et al. 2009). Computing extreme indices requires reliable daily data. Thus the development of suitable techniques to homogenize daily data is necessary.

Homogenization of temperatures at a daily time scale is much more difficult than at monthly or annual scales. This is not due to the detection of shifts, since this information may be provided by the analysis of annual or monthly series. Thus this is mainly an adjustment problem. When considering annual or monthly data, the effect of the changes affecting the series can be assumed to be a bias that may vary according to the season. These biases are very easy to estimate and remove using linear techniques (e.g., Caussinus and Mestre 2004). This is no longer the case when daily temperature data are processed, however, where adjustments should vary according to the meteorological situation of each day. Differences in

Corresponding author address: Olivier Mestre, ENM, Météo-France, 42, Avenue Coriolis, 31057 Toulouse CEDEX, France.
E-mail: olivier.mestre@meteo.fr

shelter radiative properties may dramatically influence observations, as shown in shelter intercomparison experiments (Lefèvre 1998). For example, on average, the difference between a standard French BMO 1050 shelter and a CIMEL Electronique shelter, which was provided to nonprofessional observers, is approximately +0.5°C, but for individual days this difference may rise up to 1.8°C. This occurs especially during hot sunny days with little wind, where the natural ventilation of this small shelter fails to compensate for radiative heating. A recent intercomparison study of nine widely used screens also shows increasing absolute temperature differences with decreasing cloud cover and wind speed (Brandsma and Van der Meulen 2008).

For temperature adjustment, multiple regression models—including other parameters such as wind speed and direction, sunshine duration, and parallel measurements—are the best way to proceed, as was achieved for the De Bilt series (Brandsma et al. 2002; Brandsma 2004). The Netherlands Meteorological Institute (KNMI) has kept all original instruments as well as complete metadata and photographic archives of the environment of earlier site positions. Using this unique material, Brandsma et al. (2002) were able to carefully plan parallel measurement experiments, not only for temperature measurements, but also for wind speed and sunshine duration. The conditions in which the De Bilt series was homogenized are unique, however. Wind speed or sunshine duration data are extremely rare when considering older data, for which usually only precipitation and temperature were observed. Furthermore, metadata simply do not exist in many cases. Reproducing the old measurement conditions (Brandsma et al. 2002; Brandsma 2004; Brunet et al. 2004, 2007) is a way to correct the series, but this approach is expensive, is time consuming, and requires waiting a long time to get a sufficient archive after the experiment has started. For these reasons, some authors have limited themselves to assessing homogeneity using graphical analysis of time series of annual indices derived from daily data to suppress inhomogeneous stations from any further analyses (Peterson et al. 2002; Aguilar et al. 2005).

If there is a need for daily data adjustment, the simplest adjustment method relies on interpolation of monthly adjustment coefficients (Vincent et al. 2002; denoted as the Vincent method in the following), a procedure also applied by Moberg et al. (2002) and Brunet et al. (2006) to obtain a better performance in the calculation of extreme indices based on daily temperature. This method provides adjustments only for the mean of an inhomogeneity and not for its higher-order moments. Note that in Brunet et al. (2006) data are "prehomogenized" by means of transfer functions obtained through a shelter intercomparison experiment before applying Vincent's method.

Other methods characterize the changes of the entire distribution function using overlapping data between observing systems. Trewin and Trevitt (1996) use overlapping observations between temperature observing systems (e.g., when there is a change in shelter type or location) to build a transfer function between the probability density function (PDF) of the old and new measurement systems. Their method was used to homogenize Australian daily temperature measurements (Trewin 2001). Della-Marta and Wanner (2006) use a similar approach that models the changes to PDFs; it does not need overlap observations, however, and instead uses information from nearby reference stations. The main improvement of this homogenization method, called higher-order moments (HOM), in comparison with that of Trewin and Trevitt (1996) is the use of a nonlinear model, making it capable of handling inhomogeneities in higher moments. This method has been applied to summer daily maximum temperature at 26 western European stations (Della-Marta et al. 2007).

In the following, we propose a variation of the HOM method for homogenization of daily measurement temperature series. Although part of the principle involved is similar, relying on the definition of homogeneous subperiods, we propose a very different direct nonlinear spline regression approach rather than an adjustment that is based on quantiles. Our proposed method is referred to as spline daily homogenization (SPLIDHOM).

The SPLIDHOM model and the cubic smoothing spline estimation are described in section 2. In section 3, a simulation study is realized, by means of bivariate autoregressive models. This simulation allows one to compare SPLIDHOM, HOM, and Vincent's adjustments. Advantages and drawbacks of each method are then discussed. In section 4, the example of Toulouse, France, daily minimum temperature (TN) series demonstrates the usefulness of the SPLIDHOM method.

2. Method

Our goal is to provide realistic adjustments of individual temperature measurements of a candidate series Y (the series to be adjusted), given the temperature of the series itself, by means of an estimated transfer function. The estimation of this function has to be possible even in the absence of overlapping parallel measurements. Like in Della-Marta and Wanner (2006), we rely on the existence of a close and well-correlated reference series X. This reference series does not necessarily need to be totally homogeneous but should be homogeneous on subperiods of at least 2 yr around each break affecting the candidate series, since 1) fitting spline models requires a minimum amount of data and 2) data have to cover a range of situations that is large enough to avoid extrapolation of the
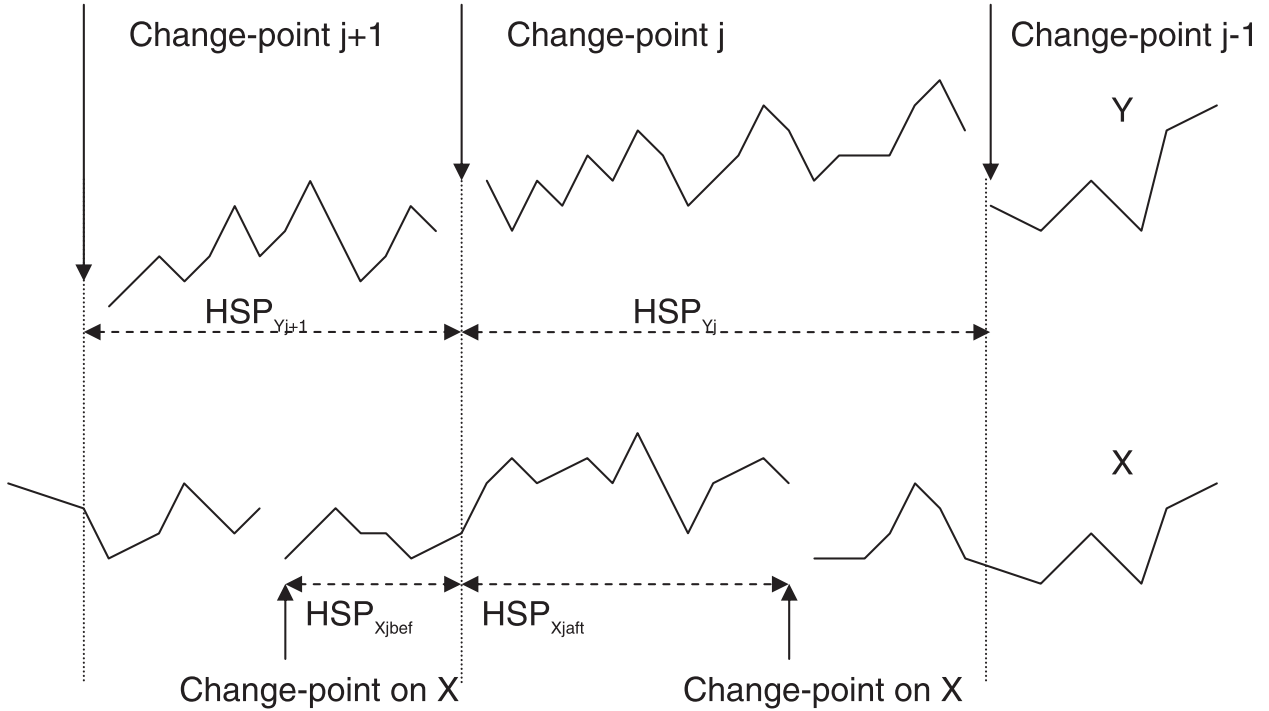
FIG. 1. Definition of homogeneous subperiods (HSPs).

functions. Note that the definition of homogeneous subperiods provided in section 2a is exactly the same as that in Della-Marta and Wanner (2006).

a. Notation

In the following, we denote Y as the candidate series and X as the reference series. Let $j = 1, \ldots, k$ be the set of changepoints affecting Y. For practical algorithmic reasons, we introduce dummy changepoints $j = 0$ that corresponds to the last observation of Y and $k + 1$ that corresponds to the day before the first observation of Y. Note that 1 refers to the most recent nondummy changepoint and that k indicates the oldest one. Let us denote $HSP_{Xj\text{aft}}$ as the homogeneous subperiod of X after the jth changepoint on Y and $HSP_{Xj\text{bef}}$ as the homogeneous subperiod of X before (see Fig. 1). The homogeneous subperiod on Y between changepoints j and $j - 1$ is denoted as HSP_{Yj}. Because X may be affected by changepoints also, homogeneous subperiods $HSP_{Xj\text{aft}}$ and $HSP_{Xj-1\text{bef}}$ may be shorter than HSP_{Yj}. Let $m_{YXj\text{aft}}$ be the nonlinear regression function of Y on X after the jth changepoint and $m_{YXj\text{bef}}$ be the nonlinear regression function of Y on X before the jth changepoint; $m_{XYj\text{bef}}$ is the nonlinear regression function of X on Y before the jth changepoint.

b. Model

The changepoint effects are adjusted sequentially, from the most recent (denoted as 1) to the oldest (k). The

last period HSP_{Y1} remains unchanged. Adjustment is first applied to HSP_{Y2}, then to HSP_{Y3}, up to HSP_{Yk+1}. For adjustment of the whole subperiod HSP_{Yj+1} (corresponding to the effect of the jth changepoint), the first step is to estimate $m_{YXj\text{bef}}$ ($m_{YXj\text{aft}}$), that is, the regression of Y on X before (after) the break on $HSP_{Xj\text{bef}}$ ($HSP_{Xj\text{aft}}$) subperiods (Fig. 1).

If there is a change, $m_{YXj\text{bef}}$ and $m_{YXj\text{aft}}$ do not coincide, and their difference $m_{YXj\text{aft}-YXj\text{bef}} = m_{YXj\text{aft}} - m_{YXj\text{bef}}$ is not null, at least on parts of the data range. We adjust HSP_{Yj+1} so that the $m_{YXj\text{bef}}$ regression function matches the regression $m_{YXj\text{aft}}$ estimated on HSP_{Yj}. Thus, adjustments are given by the estimation of $m_{YXj\text{aft}-YXj\text{bef}}$ (denoted as $\hat{m}_{YXj\text{aft}-YXj\text{bef}}$). A straightforward calculation shows that, conditional to X, if estimates of $m_{YXj\text{bef}}$ and $m_{YXj\text{aft}}$ are unbiased, then their difference is an unbiased estimator of $m_{YXj\text{aft}-YXj\text{bef}} = m_{YXj\text{aft}} - m_{YXj\text{bef}}$. Any observed value Y_t may be adjusted using this function and the corresponding X_t value, according to

$$Y_t^* = Y_t + \hat{m}_{YXj\text{aft}-YXj\text{bef}}(X_t), \qquad (1)$$

where Y_t^* is the adjusted value according to (1). At this stage, if reference X is homogeneous on HSP_{Yj+1} (i.e., HSP_{Yj+1} and $HSP_{Xj\text{bef}}$ coincide) then adjustments can be directly applied to Y before the jth changepoint using (1). In the general case, however, reference X itself might be inhomogeneous, or missing, on parts of HSP_{Yj+1}. So

an additional step is performed. The $m_{XYj\text{bef}}$ regression function is estimated. This is the regression of X on Y for subperiod $\text{HSP}_{Xj\text{bef}}$. It allows the substitution of Y_t into "pseudo" X_t values: $\hat{X}_t = \hat{m}_{XYj\text{bef}}(Y_t)$ in (1), with $\hat{m}_{XYj\text{bef}}$ denoting the estimation of $m_{XYj\text{bef}}$ on $\text{HSP}_{Xj\text{bef}}$. Last, the SPLIDHOM adjusted observations $\hat{Y}_t$ are given by

$$\hat{Y}_t = Y_t + \hat{m}_{YXj\text{aft}-YXj\text{bef}}[\hat{m}_{XYj\text{bef}}(Y_t)]. \quad (2)$$

In the following, the term $\hat{m}_{YXj\text{aft}-YXj\text{bef}}[\hat{m}_{XYj\text{bef}}(Y_t)]$ is called adjustment or the adjustment function.

Although it is based on the same definition of subperiods as HOM is, the adjustment proposed by SPLIDHOM differs in its principle. SPLIDHOM is based on regression only, whereas HOM is based on distribution fitting. Note that, in the practical implementation of our algorithm, the model may be applied for each month or each season separately.

c. Fitting

In practice, the various regressions involved are almost linear, whereas a large proportion of the useful information might be hidden in the nonlinear part of the regressions. For estimating the regression function, several techniques have been tested: kernel smoothers (Brockmann et al. 1993; too noisy at the edge for data-scarcity reasons), wavelet thresholding (Nason 2008; too sensitive to small outliers), and locally weighted polynomial regression and scatterplot smoothing (LOESS; Cleveland and Grosse 1991; too computationally demanding when applied for cross-validation techniques). Our final choice relies on a classical cubic smoothing spline that does not have the previously mentioned drawbacks for our application. In the following we recall the basics of smoothing splines. Readers may refer to Hastie and Tibshirani (1990) for a more complete overview of this technique.

Cubic smoothing splines are the solution of the following optimization problem: let (X_i, Y_i) for $i = 1, \ldots, n$ be a sequence of observations, modeled by the relation $E(Y_i|X_i) = m(X_i)$. The smoothing-spline estimate is defined as the function $\hat{m}$ (over the class of twice-differentiable functions, denoting m'' as the second derivative of m and λ as the smoothing parameter) that minimizes the penalized residual sum of squares:

$$\sum_{i=1}^{n} [Y_i - m(X_i)]^2 + \lambda \int_a^b [m''(t)]^2 \, dt.$$

Interval $[a, b]$ corresponds to the range of X. This problem has a unique (and explicit) solution that is a natural cubic spline with knots at the values X_i. This model may seem overparameterized, but spline continuity constraints at knots bring down its dimension dramatically.

Smoothing parameter λ ($\lambda \geq 0$) controls the trade-off between fidelity to the data and roughness of the function estimate. Larger values of λ correspond to smoother solutions. If $\lambda \rightarrow \infty$, $m''(t) \rightarrow 0$ and the minimizer is the least squares line. The smoothing parameter is estimated for each regression by means of a standard cross-validation technique to avoid over-fitting. Let $\hat{m}_\lambda^{(-i)}$ be the solution for a given value of λ obtained by leaving out observation i, which mimics training and test-sample procedures. Estimated λ is the value that minimizes the cross-validation sum of squares:

$$\text{CV}(\lambda) = \sum_{i=1}^{n} [Y_i - \hat{m}_\lambda^{(-i)}(X_i)]^2.$$

This cross-validation technique gave satisfactory results in our application, selecting most of the time solutions having an equivalent degree of freedom from 2 to 4, roughly corresponding to polynomials of degree 1–3. This is a significant difference from HOM, in which the LOESS smoothing parameter is fitted empirically, as stated by the authors themselves.

Because the range of the data within different HSPs can be different, we often face an additional extrapolation problem. Linear extrapolation of $m_{XYj\text{bef}}$ is easy to achieve, but extrapolation of $m_{YXj\text{aft}-YXj\text{bef}}$ may lead to incorrect results. So, we also choose to bound adjustments at the edges, as in the HOM method. In practice, adjusting values greater (lower) than the largest (lowest) observed value of X on the estimation interval is performed using adjustment computed for the largest (lowest) observed value of X on the estimation interval.

3. Results

a. Simulation study

This experiment has two purposes: first, to establish the correlation necessary to obtain good results with the HOM and SPLIDHOM methods and, second, to show SPLIDHOM improvements relative to results from Vincent's method and HOM in a variety of situations. We show the influence of HOM, SPLIDHOM, and Vincent's method on several indices computed on daily maximum temperatures, including root-mean-square error (RMSE), annual mean, summer [June–August (JJA)] mean, 5% (Q05) and 95% (Q95) quantiles, and annual absolute maximum temperature.

Data are simulated according to the following scheme: the Toulouse daily maximum temperature (TX) series is decomposed into seasonal, trend, and noise components using moving averages of width equal to 1 yr,
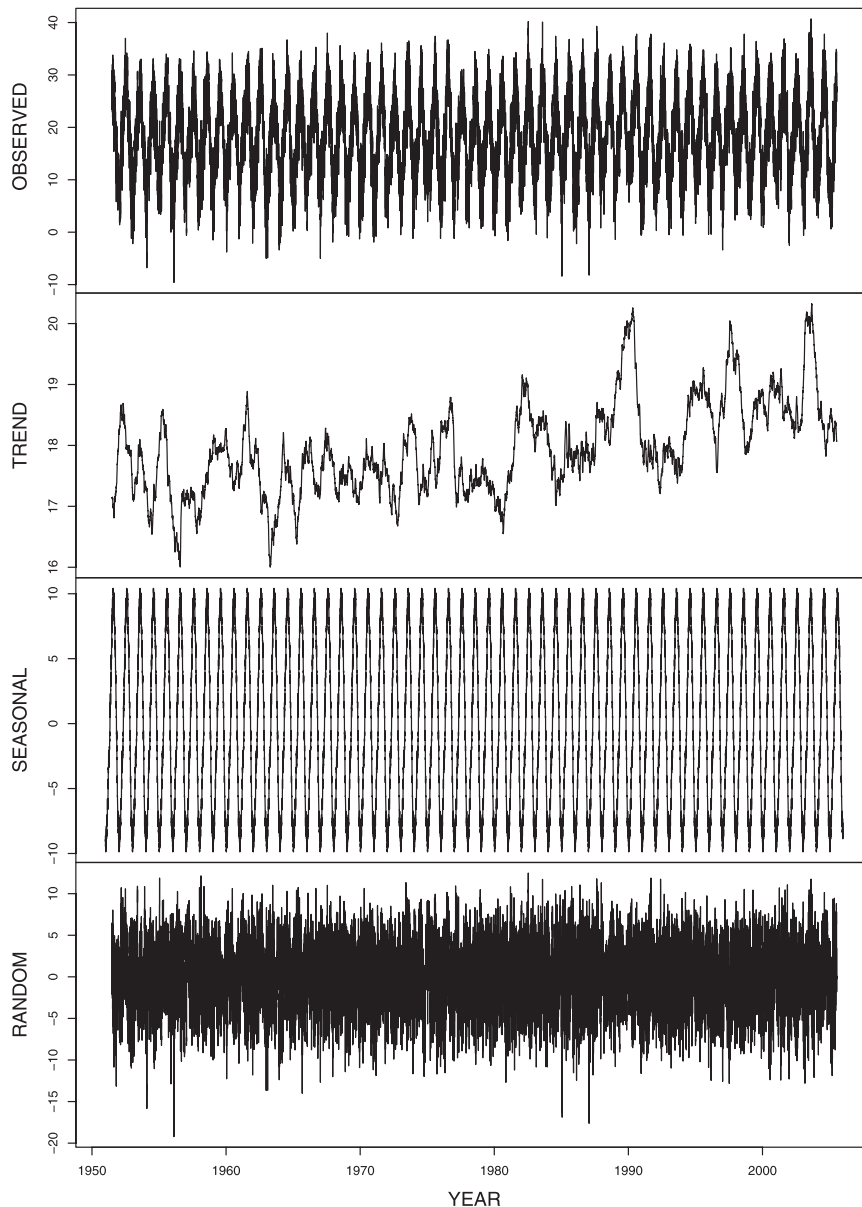
FIG. 2. Decomposition of (top) observed Toulouse daily maximum temperature series into (top middle) trend, (bottom middle) seasonal, and (bottom) random noise components.

according to a classical additive model (Brockwell and Davis 2006). The result of this decomposition is shown in Fig. 2. The random component is then modeled as a first-order (lag 1) autoregressive [AR(1)] process. The estimation of first-order autocorrelation is equal to 0.672, and the noise component of the AR(1) process is found to have variance equal to $8.6^\circ C^2$. Pairs of correlated candidate and reference series are then simulated using the following procedure. First, we generate correlated noise terms U_{1t} and U_{2t} by means of a bivariate AR(1) process $\{U_1, U_2\}$ (Neumaier and Schneider 2001) described as

$$\begin{cases} U_{1t} = \varphi_1 U_{1t-1} + \varepsilon_{1t} \\ U_{2t} = \varphi_2 U_{2t-1} + \varepsilon_{2t} \end{cases} \text{ with vector } \varepsilon_t = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}\right];$$
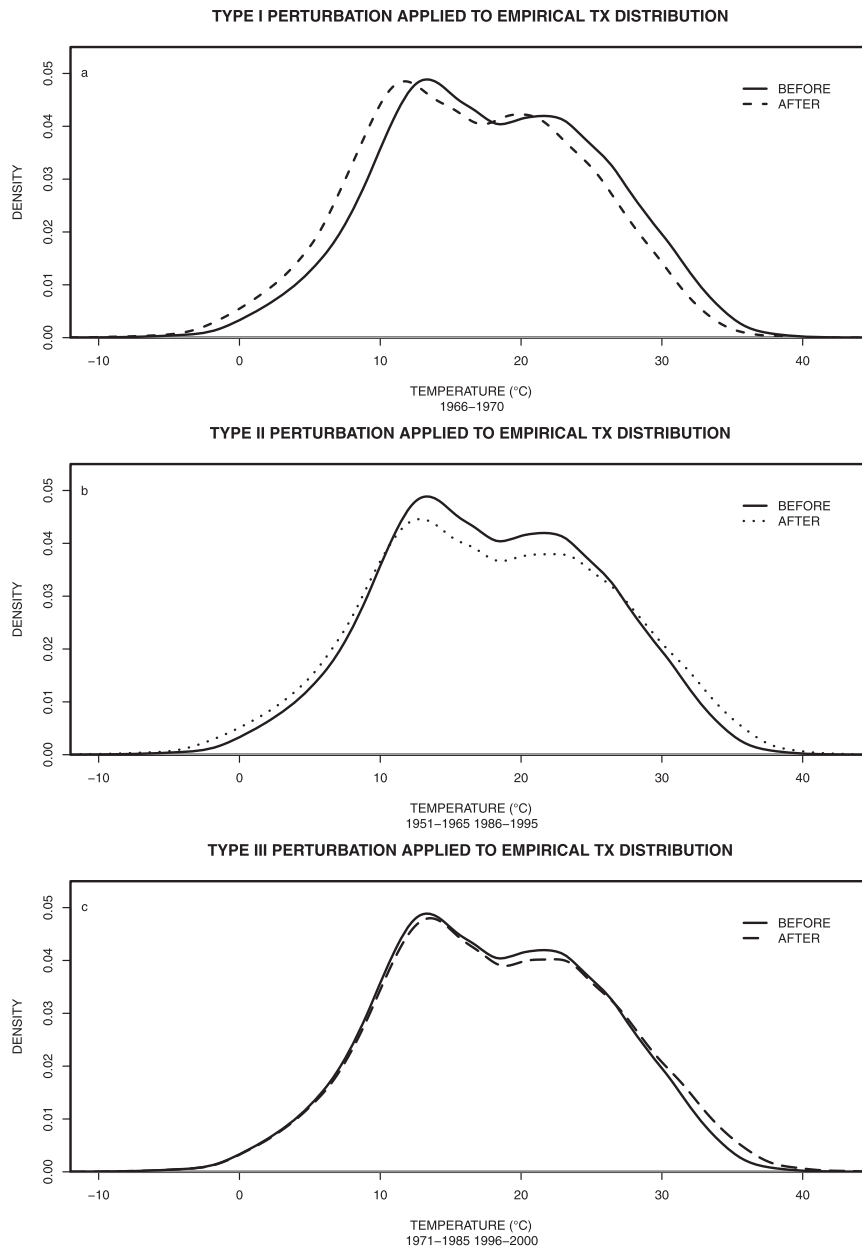
FIG. 3. Histogram of daily TX distribution of Toulouse data, before (solid) and after (dotted or dashed) application of (a) type-I, (b) type-II, and (c) type-III inhomogeneities.

that is, the noise term $\{U_1, U_2\}$ of the process follows a centered bivariate normal distribution, with the correlation between ε_1 and ε_2 being controlled by parameter r. In practice, we set $\varphi_1 = \varphi_2 = 0.672$ and $\sigma^2 = 8.6$, which are values estimated on the real Toulouse temperature series. Pairs of series are created by summing the same trend and seasonal cycle (estimated on Toulouse temperatures) to the noise terms U_1 (first series) and U_2 (second series). Inhomogeneities are added to the first series to create the candidate, with the second series being the reference.

We choose to add three different synthetic inhomogeneities to the candidate series to study a variety of situations: type-I inhomogeneity consists in adding a normal random variable of mean $-1.5°C$ and standard deviation $0.5°C$ to the daily data (pure noise). This "type I" inhomogeneity roughly reproduces temperature-independent errors. For example, an error related to sun exposure is likely independent of the actual observed temperature, since it may occur on hot days as well on cold late-winter days with snow cover. Type-II inhomogeneity
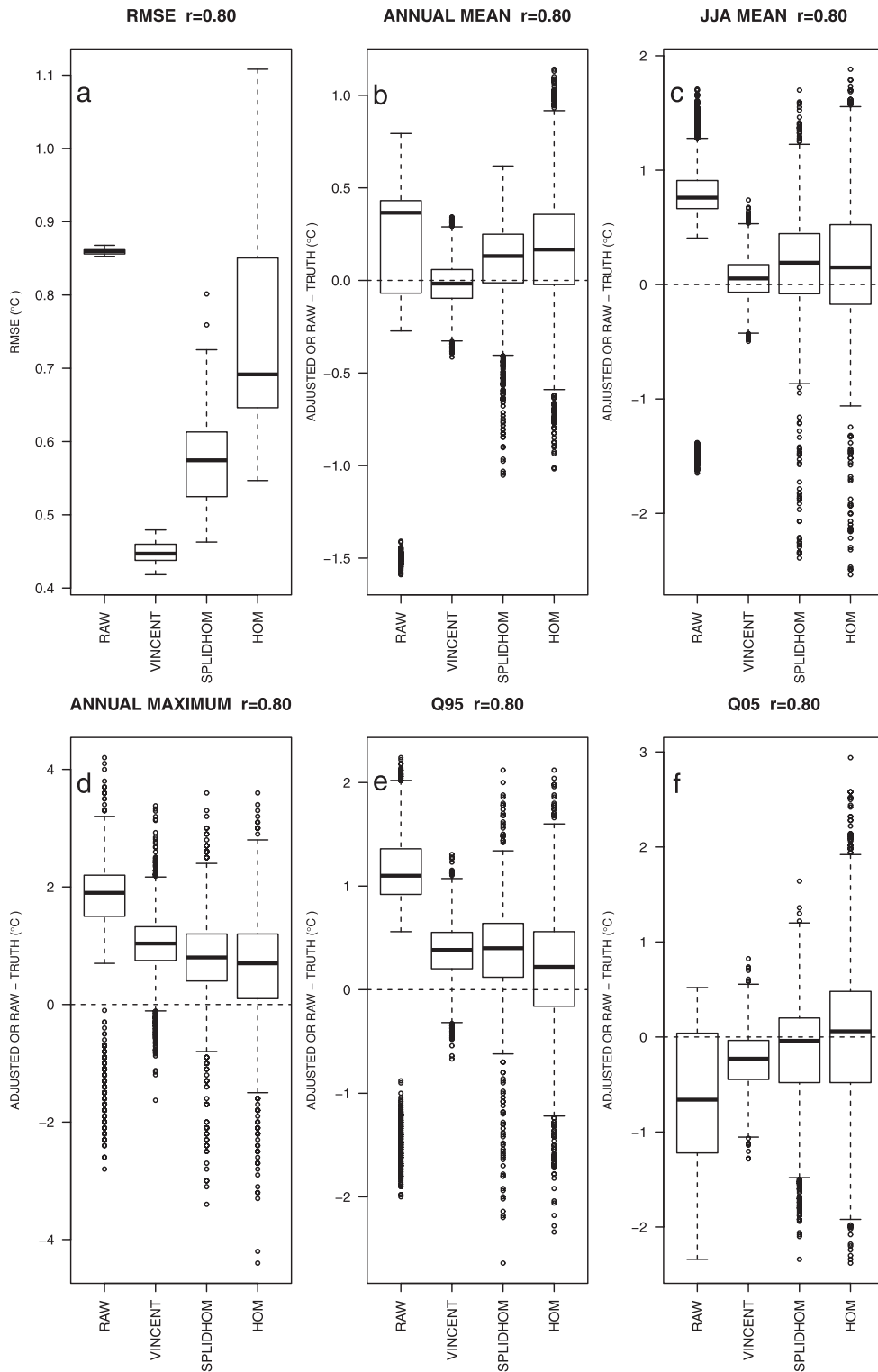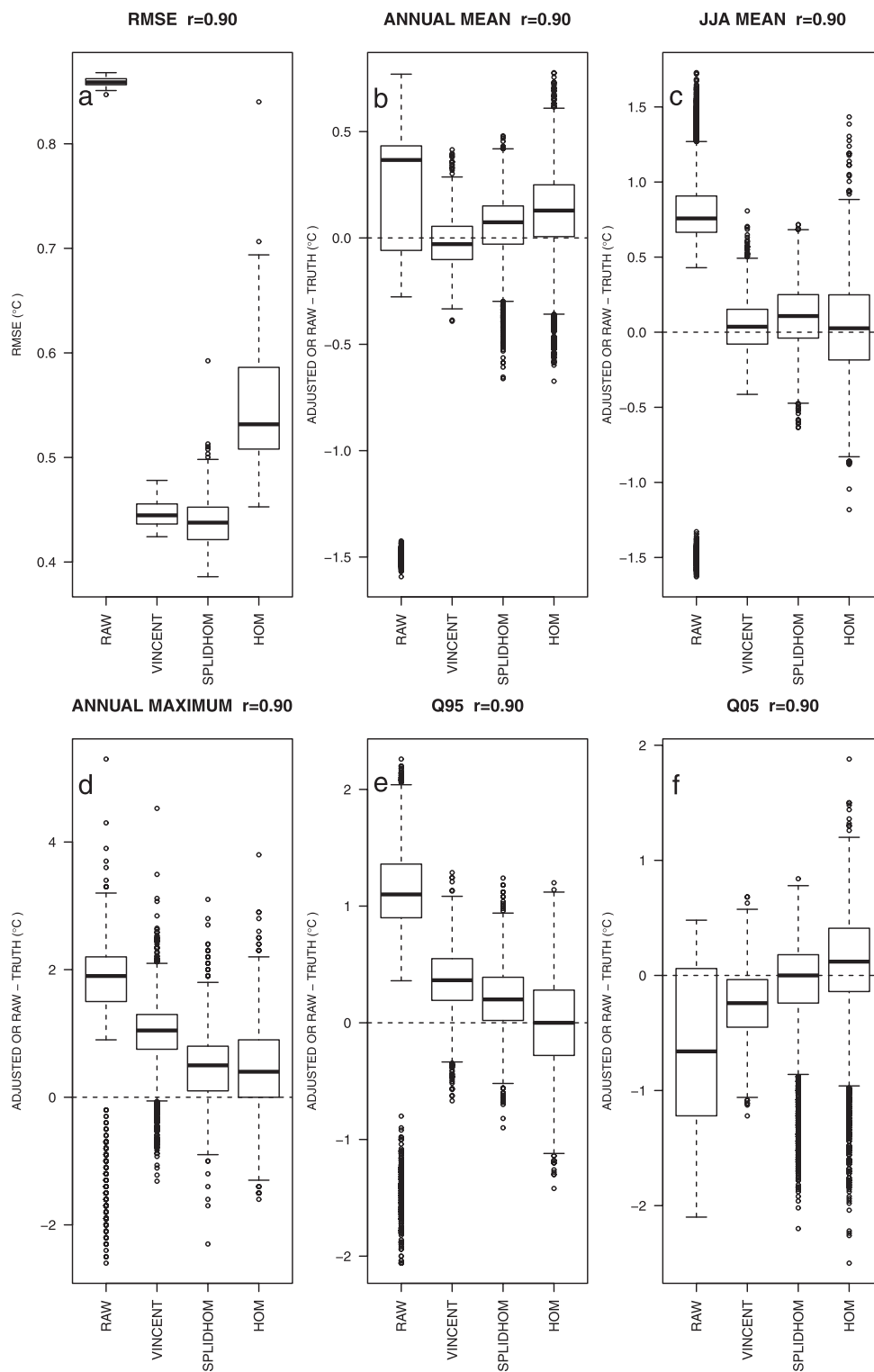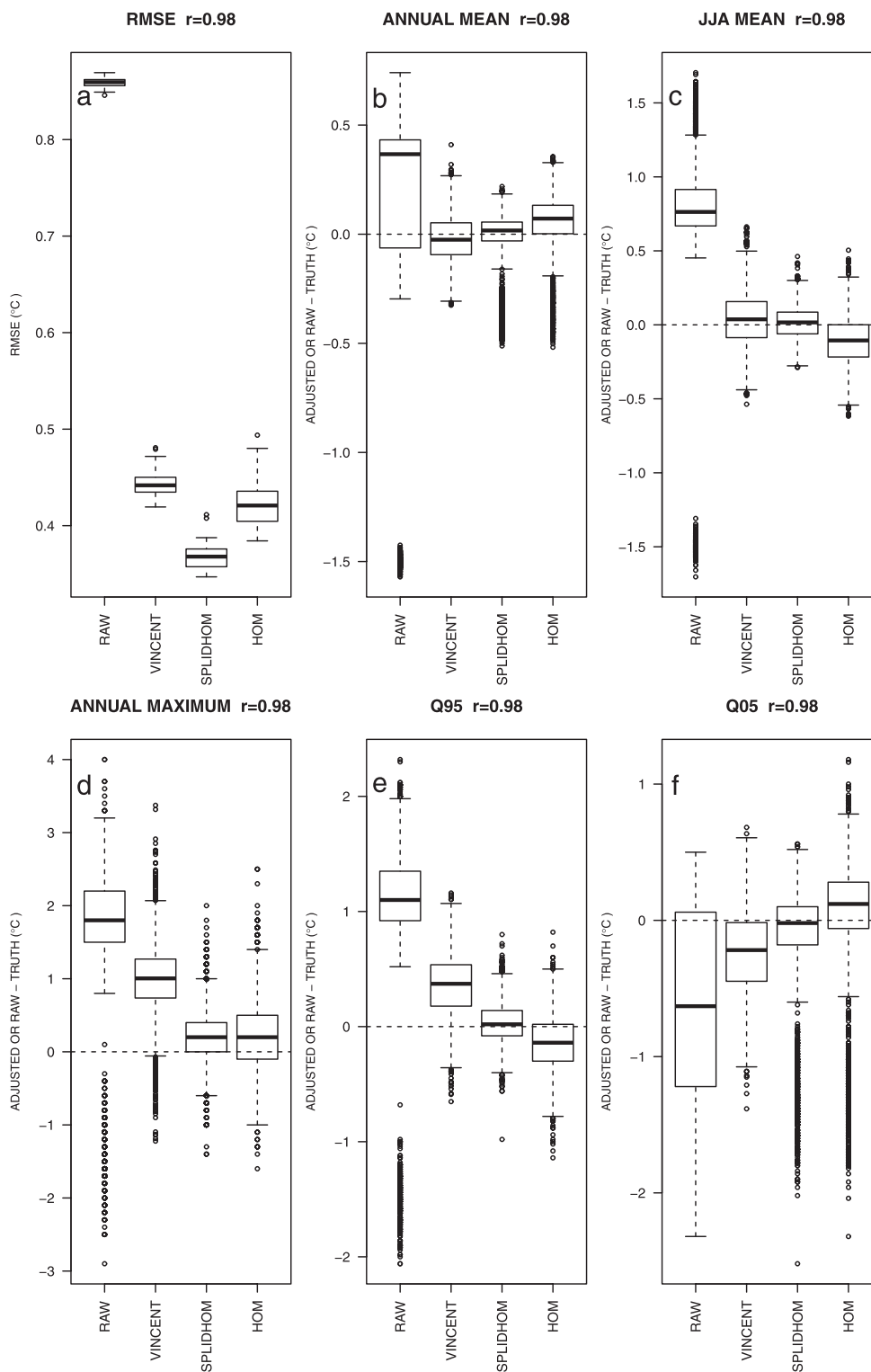
FIG. 4. Box plots of (a) RMSE of daily raw (labeled RAW) and adjusted values (labeled Vincent, HOM, and SPLIDHOM) in comparison with truth, and box plots of differences between original unperturbed true series and raw (RAW) series or adjusted series (Vincent, HOM, and SPLIDHOM methods) on a variety of annual indices, computed for each year: (b) annual means (average of the 365 values), (c) summer means, and (d) annual absolute maximum temperature (highest temperature that occurred during the year) and (e) Q95 and (f) Q05 quantiles of distribution of daily values, for correlation $r = 0.80$ and for 50 computed series.

FIG. 5. As in Fig. 4, but for $r = 0.90$.

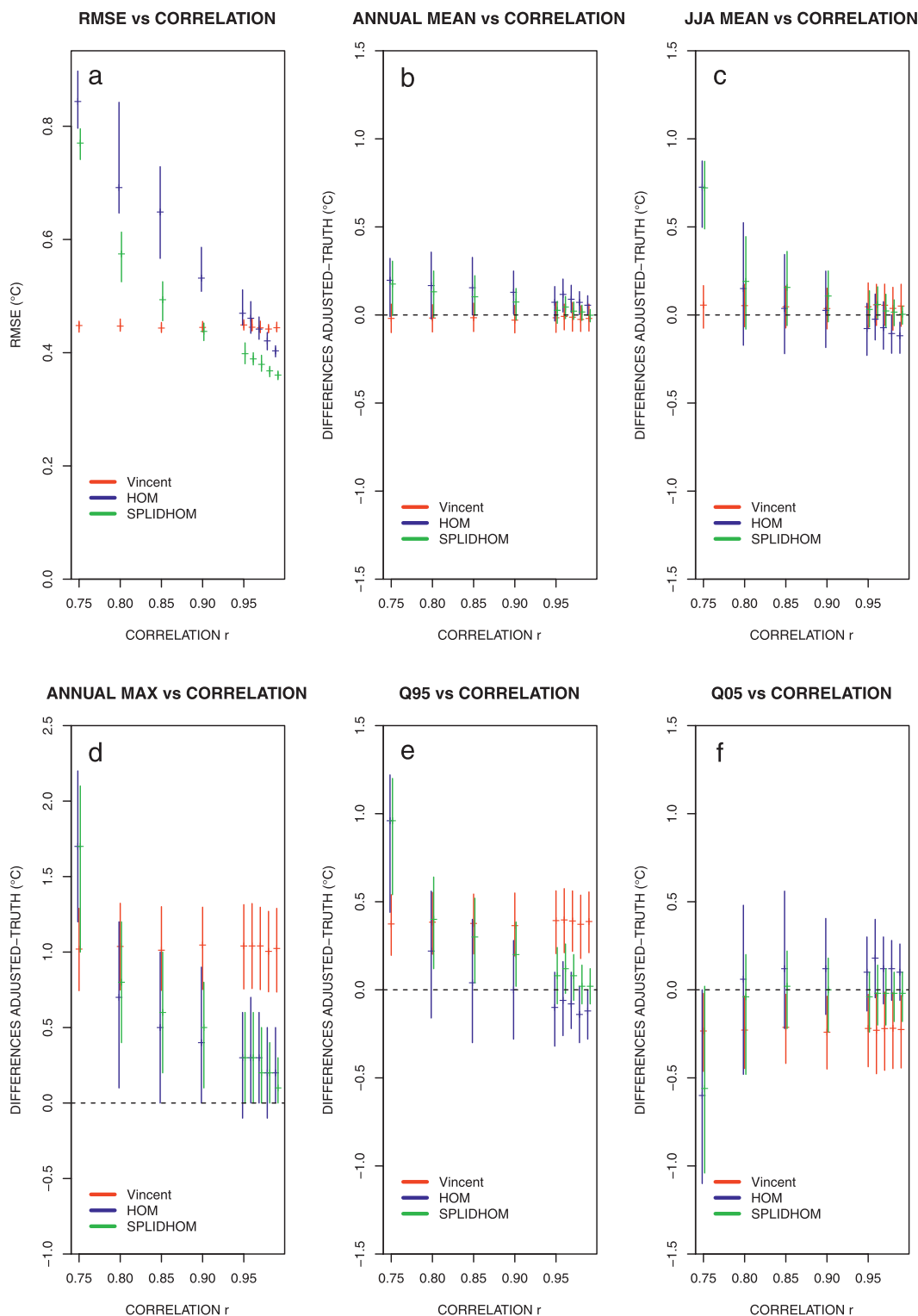FIG. 6. As in Fig. 4, but for $r = 0.98$.

FIG. 7. Interquartile range (vertical bars) and median (short horizontal bar) of (a) RMSE of daily raw (RAW) and adjusted values (Vincent, HOM, and SPLIDHOM) in comparison with truth, and box plots of differences between original unperturbed true series and raw (RAW) series or adjusted series (Vincent, HOM, and SPLIDHOM methods) on a variety of annual indices, computed for each year: (b) annual means (average of the 365 values), (c) summer means, and (d) annual absolute maximum temperature (highest temperature that occurred during the year) and (e) Q95 and (f) Q05 quantiles of distribution of daily values, for correlation $r = 0.75, 0.80, 0.85, 0.90, 0.95, 0.96,$ 0.97, 0.98, and 0.99 and for 50 computed series.
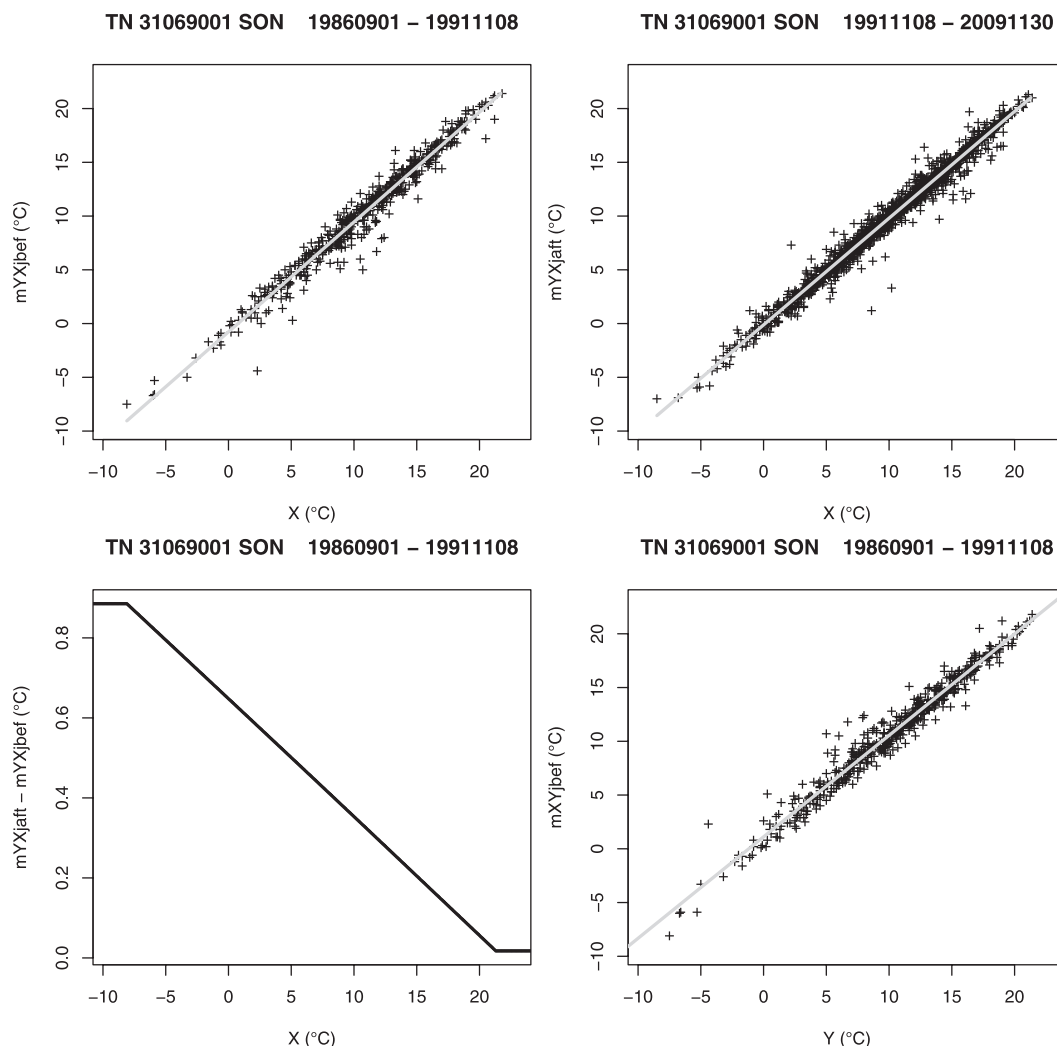
TN 31069001 SON   19860901 – 19911108

TN 31069001 SON   19911108 – 20091130

TN 31069001 SON   19860901 – 19911108

TN 31069001 SON   19860901 – 19911108

FIG. 8. Regression estimations for adjustment of HSP for Toulouse–Blagnac daily minimum temperature (Y) and autumn season (SON), using Toulouse–Francazal (X) as a reference, with (top left) scatterplot of Y vs X before the 1991 shift (1 Sep 1986–8 Nov 1991), together with corresponding cubic spline estimation of $m_{YXj\mathrm{bef}}$, and (top right) scatterplot of Y vs X after the 1991 shift (8 Nov 1991–30 Nov 2009) with corresponding cubic-spline estimation of $m_{YXj\mathrm{aft}}$. (bottom left) Estimation of $m_{YXj\mathrm{aft}-YXj\mathrm{bef}}$ and (bottom right) scatterplot of X vs Y, together with cubic-spline estimation of $m_{XYj\mathrm{bef}}$. Note that data are split according to the definition of HSPs and position of detected changepoints.

consists in transforming data using transfer function $t \rightarrow t + (t - 18)/10 + \xi$ (ξ being random normal noise with standard deviation 0.2°C). Type-II inhomogeneity enlarges the distribution of daily data. The type-III transfer function is given by $t \rightarrow t + (e^{t/10})/20 + \xi$ (ξ is defined as in type II). Type III results in larger skewness. The parameters of those functions are set empirically so that they roughly match the effect of potential shelter changes, as observed in shelter intercomparison experiments (not shown here). We applied type I to the period of 1966–70, type II to the periods of 1951–65 and 1986–95, and type III to the period of 1971–85 to study the adjustment of multiple inhomogeneities of various types in the data. The

effect of such transforms on Toulouse TX distribution is shown in Fig. 3.

For r taking values of 0.80, 0.85, 0.90, 0.95, 0.96, 0.97, 0.98, and 0.99, 50 pairs of candidate and reference series are generated. A candidate series (''truth'') is then perturbed as described previously to give the ''raw'' candidate. The raw candidate series is then adjusted using the HOM and SPLIDHOM methods. A pseudo-Vincent method is also used: for each subperiod, 12 monthly adjustment coefficients are computed, computing the monthly mean differences between truth and raw over the whole subperiods. Because those estimates are much more accurate than they would be in reality, noise is
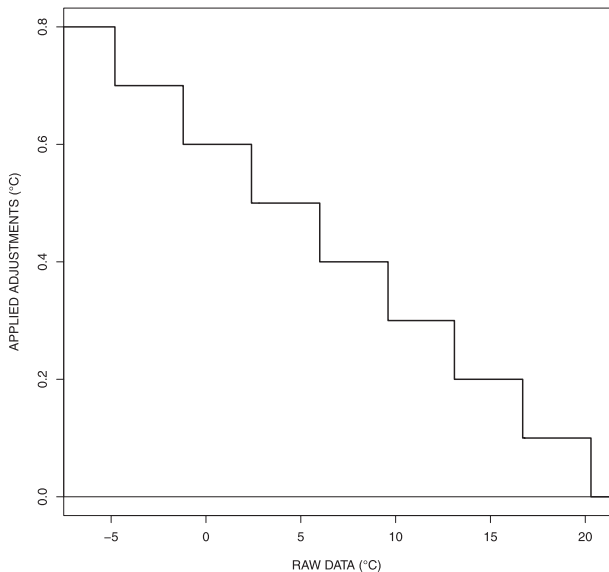
FIG. 9. Estimated adjustment function $\hat{m}_{YXj\text{aft}-YXj\text{bef}}[\hat{m}_{XYj\text{bef}}(Y_t)]$ for HSP between 17 Jun 1986 and 8 Nov 1991, for Toulouse–Blagnac daily minimum temperature (Y) and autumn season (SON). This function gives the adjustment to be applied to Y as a function of Y itself (°C). Note that this estimation is rounded to a precision of 0.1°C, which is the precision of the data themselves in the database.

added, consisting of a random centered normal variable of standard deviation 0.3, which is roughly the standard error estimate observed on monthly adjustment coefficients computed using the Caussinus and Mestre (2004) analysis-of-variance (ANOVA) model. The annual cycle of adjustments is then interpolated using a spline as described in Vincent's method. Note that the multivariate ANOVA model takes all available monthly series in a regional neighborhood. In this experiment, we consider that the average regional network density does not vary but that r can take a wide range of values within the regional network.

For each correlation and for 50 pairs of simulated series, we compare differences between "true" candidate and raw series, and differences between true candidate series and series adjusted by means of Vincent's method, HOM, and SPLIDHOM, on a variety of indices: RMSE of the adjusted daily values versus truth, and annual indices, such as annual means (average of the 365 values), annual absolute minimum and maximum temperatures (respectively the lowest and highest temperature that occurred during the year), and annual quantiles Q95 and Q05 of the daily values of the considered year. For each correlation, we compute box plots of the 50 corresponding RMSE as well as box plots of differences (raw minus true or "adjusted" minus true) observed on annual indices (for each simulated series,

and each year). Results for $r = 0.80$, 0.90, and 0.98 are provided in Figs. 4–6. Perfect adjustments would result in null differences and RMSE.

From these results, a number of comments can be made. 1) All three methods improve the data: inhomogeneities are reduced when comparing adjusted series to raw series. 2) The Vincent method is able to correct the means (annual and JJA) and outperforms both HOM and SPLIDHOM for lower correlations, in terms of RMSE, but is strongly biased regarding adjustment of annual maxima as well as extreme quantiles. The biases of the annual maxima, Q95, and Q05 are about 1.0°, 0.4°, and −0.4°C, respectively, using the Vincent method (Fig. 6, bottom panels) in our experiment. 3) The HOM and SPLIDHOM improvements relative to the Vincent method are hardly noticeable for $r < 0.90$. For example, when $r = 0.8$, the biases of annual maxima and Q95 are about 0.8° and 0.4°C, respectively, for SPLIDHOM, but higher correlations ensure for both HOM and SPLIDHOM a good adjustment of the means and significant improvements for extreme quantiles. 4) SPLIDHOM clearly performs better than HOM in terms of RMSE.

Because r is a really crucial parameter, we plot median and interquartile range of RMSE for the three methods as a function of r, for each of the scores (Fig. 7). This confirms that both HOM and SPLIDHOM need well-correlated series ($r > 0.90$) to outperform the Vincent method, in terms of RMSE and bias reduction for extreme quantiles. The performance of the Vincent method is less sensitive to r value, at least for the range of correlations we tested. In looking at a comparison of HOM and SPLIDHOM, it is seen that SPLIDHOM clearly exhibits lower RMSE. Adjustment of annual maxima is equivalent for both methods, but SPLIDHOM performs generally better than HOM for means (annual and JJA) and Q05. For Q95, SPLIDHOM is more biased for $r \leq 0.90$ but gets the best results for $r > 0.96$. If we roughly consider that SPLIDHOM is superior to the Vincent method for a correlation of 0.90 and delivers trustworthy results at a correlation of 0.95, those correlation thresholds are important to take into account. For maximum temperatures, in a region of flat terrain such as the Paris, France, region, a correlation of 0.95 (0.90) is achieved for an approximate station distance of ~75 km (~150 km). In the more-mountainous area around Lyon, France, those distances are, respectively, 18 and 60 km (not shown here).

b. Application to the Toulouse–Blagnac temperature series

Toulouse–Blagnac (Toulouse civil airport; professional station with index number 31069001) TN and TX series are affected by several abrupt changepoints. Those changes are detected using "PRODIGE" software (Caussinus and

TN 31069001 SON   19720901 – 19860617   TN 31069001 SON   19860617 – 20091130

TN 31069001 SON   19720901 – 19860617   TN 31069001 SON   19720901 – 19860617
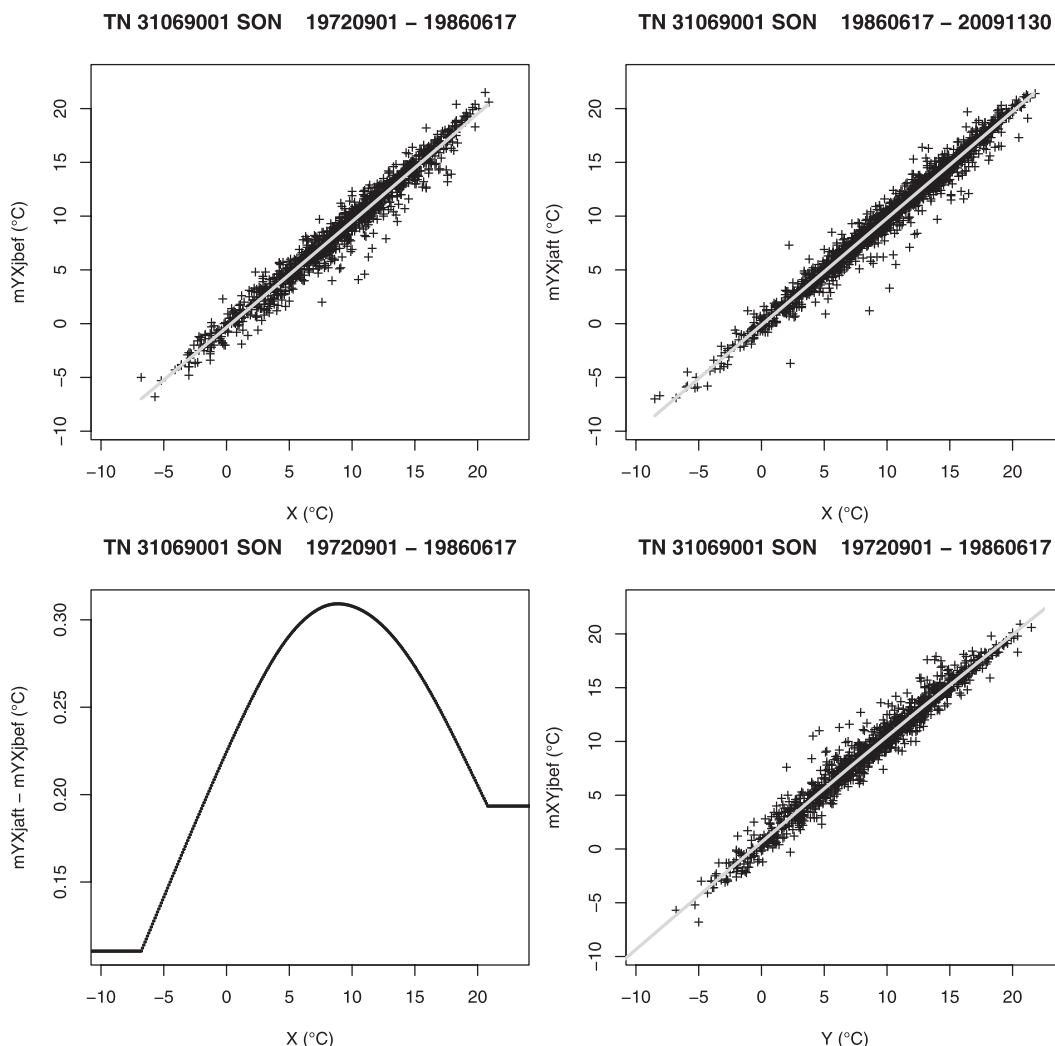
FIG. 10. As in Fig. 8, but (top left) before the 1986 shift (1 Sep 1972–17 Jun 1986) and (top right) after the 1986 shift (17 Jun 1986–30 Nov 2009).

Mestre 2004), which relies on multiple pairwise comparisons of annual Toulouse series with regional neighbors. Statistical detection itself is performed by means of a dynamic programming algorithm (Hawkins 2001) to find positions of changes together with an adapted penalized likelihood criterion (Caussinus and Lyazrhi 1997) assessing the significance of changes. Metadata allow validation of those detections and provide causes and precise days for changes: 20 June 1962 (new instrumental park), 15 October 1968 (relocation and new shelter), 1 May 1972 (sensor change), and 17 June 1986 and 8 November 1991 (for both, relocations of instrumental park caused by construction of new runways). The reference data are provided by the Toulouse–Francazal series (French "Armée de l'Air" station; military airport), situated 12 km south of the Toulouse–Blagnac airport. This series is affected by a large changepoint on 14 November 1955 (relocation and

shelter change). The Toulouse–Blagnac series starts in 1951. Correlation of the series is high: $r = 0.98$ (at a daily time scale, with seasonal cycle removed), justifying the use of the SPLIDHOM technique. Changepoint effects are adjusted sequentially, for each season, from the period before the most recent changepoint (1991) to the oldest one (1962). In this example, we choose seasonal instead of monthly estimates, because the results appeared to be more stable. Let us analyze in detail the period 1986–91 for the autumn season [September–November (SON)]. The top-left panel in Fig. 8 shows the scatterplot of observed daily Toulouse–Blagnac TN (candidate Y) as a function of daily Toulouse–Francazal TN (reference X), for homogeneous subperiod 1 September 1986–8 November 1991 for the SON season. The solid gray line corresponds to the smoothing-spline estimation of regression function m_{YXjbef}. In a similar way, the top-right panel of Fig. 8 shows the
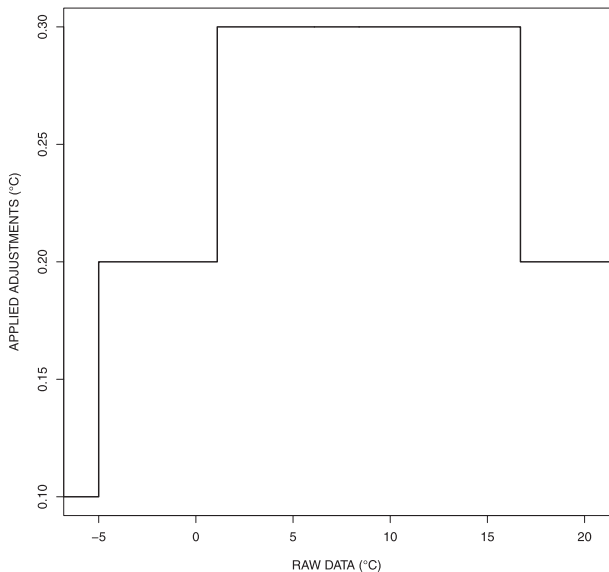
FIG. 11. As in Fig. 9, but between 1 May 1972 and 17 Jun 1986.

its staircase behavior. For autumns 1986–91, large adjustments are observed for low temperatures (up to +0.8°C), being almost null for warmer temperatures. Those values are not surprising for a change in location or exposure.

When analyzing the adjustments of the previous break (1986) for the same SON autumn season, we find a different shape (Figs. 10 and 11). The estimation of $m_{YXj\text{aft}-YXj\text{bef}}$ is not linear (Fig. 10, bottom-left panel), resulting in a non-linear adjustment function (Fig. 11) and thus justifying the use of nonlinear models in SPLIDHOM. In this example, the adjustments are less important. When analyzing adjustments of Toulouse series for every breakpoint and every season, roughly one-half of the adjustment functions have a linear or quasi-linear (including constant) shape, with the other one-half exhibiting a nonlinear shape.

The examples given above show that adjustments of our method are sensitive to temperature itself, thus taking into account the meteorological situation of each day in a crude way. This is the main contrast and improvement to the adjustments of the method provided by Vincent et al. (2002). This method depends on the seasonal variations of monthly adjustments and of the position of the day in the year. It is indeed simple to apply, and it keeps coherency between the usual homogenization methods (applied to annual and monthly data) and daily adjusted data, but adjustment of higher quantiles is a bit less realistic, as shown by the experimental study.

In addition, it can be shown that our method reaches also a good agreement with standard homogenization procedures: comparing time series of annual means of TN, homogenized using SPLIDHOM (daily homogenization) and by means of PRODIGE (monthly homogenization; Caussinus and Mestre 2004), we find very good agreement (Fig. 12). This is a remarkable result, since the PRODIGE method relies on a completely

scatterplot of daily data and estimation of regression function $m_{YXj\text{aft}}$ after the target change on 8 November 1991 for the SON season over subperiod 8 November 1991–30 November 2009. The bottom-left panel of Fig. 8 shows the estimation of this difference of the latter two functions, as a function of X (Toulouse–Francazal). This corresponds to the estimation of the function $m_{YXj\text{aft}-YXj\text{bef}}$ in (1). In this example, $m_{YXj\text{aft}-YXj\text{bef}}$ can be considered to be linear. An estimation of the $m_{XYj\text{bef}}$ additional transfer function used in (2) is also provided in Fig. 8 (bottom-right panel). Figure 9 shows the estimated SPLIDHOM adjustment function $\hat{m}_{YXj\text{aft}-YXj\text{bef}}[\hat{m}_{XYj\text{bef}}(Y_t)]$. Note that, given the precision of the original database, the final adjustment function is rounded to a precision of 0.1°C, which explains
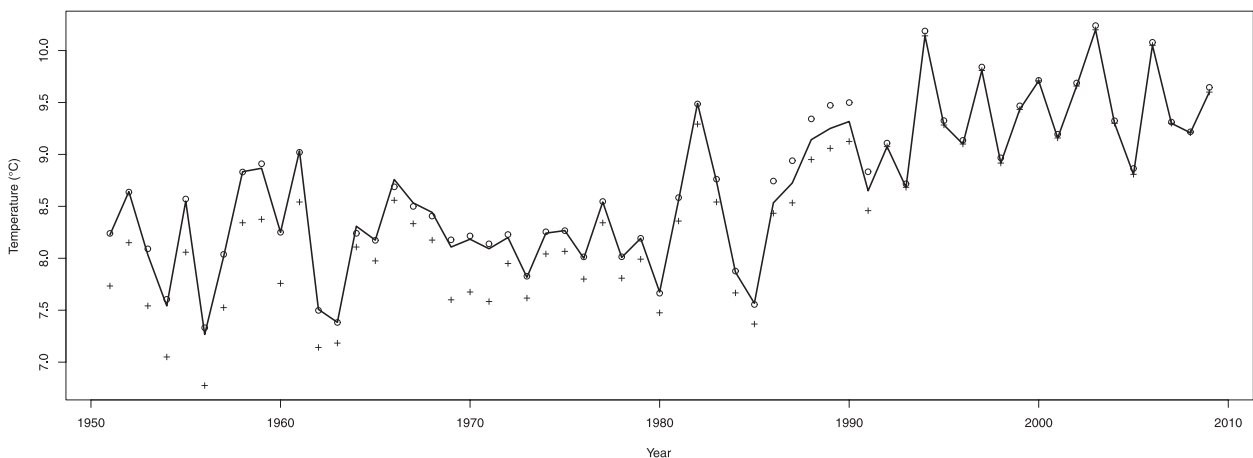

FIG. 12. Annual averages of daily adjusted TN series (open circles) in comparison with annual averages of raw (plus signs, generally seen below the solid line until 1992) series and annual averages of monthly homogenized series (solid line).

different principle in which mean biases are estimated using an ANOVA model, applied on a large set of monthly series in the same climatic area. When considering annual averages, we get very similar results by using two completely different methods applied independently.

4. Conclusions

Although parts of the principles involved in HOM and SPLIDHOM are very similar, especially the definition of subperiods, SPLIDHOM adjustments differ: they are based on nonparametric regression (by means of a cubic smoothing spline) whereas HOM involves fitting data to several candidate distributions. The use of a smoothing parameter set by means of cross-validation avoids overfitting during the estimation process. On simulated examples, our SPLIDHOM technique shows an improvement over HOM (especially in terms of RMSE) and Vincent's method for the correction of extreme quantiles if the correlation is high enough (above 0.90). When the correlation is lower than 0.9, Vincent's method is often superior and thus should not be neglected. Vincent's method is not designed for adjusting extreme values, however. Therefore, the adjustment of extreme quantiles is only sensible if a highly correlated reference station exists. A very important result of our study is that correlation of the candidate is the essential parameter that drives performance of both HOM and SPLIDHOM.

On practical examples, SPLIDHOM adjustments are compatible with more-classical homogenization techniques applied to monthly or annual series, which is a highly desirable feature. Also, when the individual errors cannot be considered to be "temperature dependent" (type-I errors in our simulation), SPLIDHOM still removes the main biases.

SPLIDHOM should be compared with new emerging techniques that have been recently developed, such as an improved version of HOM known as higher-order moments for autocorrelated data (HOMAD; Toreti et al. 2010), and a quantile-matching technique (Wang et al. 2010). Performance of those methods will be investigated further using various benchmarks and more types of inhomogeneities, during the last phase of European Cooperation in Science and Technology (COST) Action ES0601 "Advances in homogenization methods of climate series: An integrated approach (HOME)."

SPLIDHOM software is freely available through a simple request to the corresponding author.

REFERENCES

Aguilar, E., and Coauthors, 2005: Changes in precipitation and temperature extremes in Central America and northern South America, 1961–2003. *J. Geophys. Res.,* **110,** D23107, doi:10.1029/2005JD006119.

Brandsma, T., 2004: Parallel air temperature measurements at the KNMI-terrain in De Bilt (The Netherlands) May 2003–April 2004, interim report. KNMI Publ. 207, 29 pp.

——, and J. P. Van der Meulen, 2008: Thermometer screen intercomparison in De Bilt (The Netherlands), part II: Description and modelling of mean differences and extremes. *Int. J. Climatol.,* **28,** 389–400.

——, G. P. Können, and H. R. A. Wessels, 2002: Empirical estimation of the effect of urban heat advection on the temperature series of De Bilt (The Netherlands). *Int. J. Climatol.,* **23,** 829–845.

Brockmann, M., T. Gasser, and E. Herrmann, 1993: Locally adaptive bandwidth choice for kernel regression estimators. *J. Amer. Stat. Assoc.,* **88,** 1302–1309.

Brockwell, J., and R. A. Davis, 2006: *Time Series: Theory and Methods.* 2nd ed. Springer Series in Statistics, Springer, 577 pp.

Brunet, M., M. Bañón, F. García, E. Aguilar, O. Saladié, J. Sigró, J. Asín, and D. López, 2004: Una aproximación experimental tendente a la minimización del sesgo artificial asociado al tipo de garita meteorológica mediante la observación dual de la temperatura del aire (An experimental approach directed toward the minimization of the artificial bias associated with the type of meteorological shelter by means of dual observations of air temperature). *La Meteorología y el Clima Atlánticos* (*The Weather and Climate of the Atlantic*), La Asociación Española de Meteorología: Badajoz, 93–103.

——, and Coauthors, 2006: The development of a new dataset of Spanish Daily Adjusted Temperature Series (SDATS) (1850–2003). *Int. J. Climatol.,* **26,** 1777–1802.

——, and Coauthors, 2007: Temporal and spatial temperature variability and change over Spain during 1850–2005. *J. Geophys. Res.,* **112,** D12117, doi:10.1029/2006JD008249.

Caussinus, H., and F. Lyazrhi, 1997: Choosing a linear model with a random number of change-points and outliers. *Ann. Inst. Stat. Math.,* **49,** 761–775.

——, and O. Mestre, 2004: Detection and correction of artificial shifts in climate series. *Appl. Stat.,* **53,** 405–425.

Cleveland, W. S., and E. Grosse, E., 1991: Computational methods for local regression. *Stat. Comput.,* **1,** 47–62.

Della-Marta, P. M., and H. Wanner, 2006: A method for homogenizing the extremes and mean of daily temperature measurements. *J. Climate,* **19,** 4179–4197.

——, M. R. Haylock, J. Luterbacher, and H. Wanner, 2007: Doubled length of western European summer heat waves since 1880. *J. Geophys. Res.,* **112,** D15103, doi:10.1029/2007JD008510.

Hastie, T. J., and R. J. Tibshirani, 1990: *Generalized Additive Models. Stat. Appl. Probab. Monogr.,* Vol. 43, Chapman and Hall/CRC, 335 pp.

Hawkins, D. M., 2001: Fitting multiple change-points to data. *Comput. Stat. Data Anal.,* **37,** 323–341.

Klein Tank, A. M. G., F. W. Zwiers, and X. Zhang, 2009: Guidelines on analysis of extremes in a changing climate in support

of informed decisions for adaptation. Climate Data and Monitoring WCDMP 72, WMO-TD 1500, 56 pp.

Lefèvre, G., 1998: Comparison of meteorological screens for temperature measurement. Instruments and Observing Methods Rep. 70, WMO/TD 877, 315 pp.

Moberg, A., H. Bergstrom, J. R. Krigsman, and O. Svanered, 2002: Daily air temperature and pressure series for Stockholm (1756–1998). *Climatic Change,* **53,** 171–212.

Nason, G. P., 2008: *Wavelet Methods in Statistics with R.* Springer, 257 pp.

Neumaier, A., and T. Schneider, 2001: Estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Trans. Math. Software,* **27,** 27–57.

Peterson, T. C., and Coauthors, 2002: Recent changes in climate extremes in the Caribbean region. *J. Geophys. Res.,* **107,** 4601, doi:10.1029/2002JD002251.

Toreti, A., F. G. Kuglitsch, E. Xoplaki, and J. Luterbacher, 2010: A novel method for the homogenization of daily temperature series and its relevance for climate change analysis. *J. Climate,* **23,** 5325–5331.

Trewin, B. C., 2001: Extreme temperature events in Australia. Ph.D. thesis, School of Earth Sciences, University of Melbourne, Australia, 239 pp.

——, and A. C. F. T. Trevitt, 1996: The development of composite temperature records. *Int. J. Climatol.,* **16,** 1227–1242.

Vincent, L. A., X. Zhang, B. R. Bonsal, and W. D. Hogg, 2002: Homogenization of daily temperatures over Canada. *J. Climate,* **15,** 1322–1334.

Wang, X. L., H. Chen, Y. Wu, Y. Feng, and Q. Pu, 2010: New techniques for the detection and adjustment of shifts in daily precipitation data series. *J. Appl. Meteor. Climatol.,* **49,** 2416–2436.