SPLIDHOM: a method for homogenization of daily temperature observations

Olivier Mestre[1]

Ecole Nationale de la Météorologie, Météo-France, Toulouse, France and Institut Mathématique de Toulouse, Laboratoire Statistique et Probabilités, Toulouse, France

Christine Gruber

Zentralanstalt für Meteorologie und Geodynamik, Wien, Austria

Clémentine Prieur

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

Henri Caussinus

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

Sylvie Jourdain

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

Recived Revised Accepted

[1]Corresponding Author's Address : Olivier Mestre, ENM, Météo-France, 42, avenue Coriolis, 31057 Toulouse cedex, France; e-mail: olivier.mestre@meteo.fr

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 etc. Since observed series are likely to be affected by changes in the measurement conditions, adapted homogenization procedures are required. While 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 SPLIDHOM, relies on an indirect non-linear regression method, estimation being ensured by cubic smoothing splines. This method is able to correct 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, thanks to an optimal selection of the smoothing parameter. Applications on the Toulouse temperature series are shown as 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 (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 quite easy to estimate and remove using linear techniques (Caussinus and Mestre, 2004). But this is no longer the case when daily temperature data are processed, where adjustments should vary according to the meteorological situation of each day. Differences in shelter radiative properties may dramatically influence observations, as shown in shelter inter comparison experiments (Lefèvre, 1998). For example, on average, the difference between a standard French BMO 1050 shelter and a "CIMEL" shelter, that was provided to non-professional observers is of around +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 radiative heating. A recent inter comparison study of 9 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 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 earlier site positions environment. Using this unique material, Theo Brandsma et al. (2002) were able to carefully plan parallel measurement experiments, not only for temperature measurements, but also for windspeed and sunshine duration. But the conditions in which the De Bilt series was homogenized are rather unique. Windspeed or sunshine duration data are extremely rare when considering older data, where 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, 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 assess 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 or Aguilar et al., 2005).

If there is a need for daily data adjustment, the most simple adjustment method relies on interpolation of monthly adjustment coefficients (Vincent et al., 2002 – denoted Vincent Method in the following), a procedure also applied by Moberg et al. (2002), Brunet et al. (2006) to obtain a better performance in the calculation of extreme indices based on daily-temperature. But this method provides adjustments only for the mean of an inhomogeneity, not for its higher order moments. Note that in Brunet et al. (2006), data are "pre-homogenized" by means of transfer functions obtained through 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 (when there is a change in shelter type or location for example) to build a transfer function between the Probability Density Function (PDF) of the old and new measurement system. 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, however it does not need overlap observations and instead uses information from nearby reference stations. The main improvement of this method, called HOM, compared to Trewin and Trevitt (1996) is the use of a non-linear model making it capable to deal with inhomogeneities in higher moments. This method has been applied to summer daily maximum temperature at 26 western European stations (Della-Marta et al., 2006).

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 quite similar, relying on the definition of homogeneous sub periods, we propose a very different direct non-linear spline regression approach rather than a adjustment based on quantiles. Our proposed method is then referred as SPLIDHOM (SPLIne Daily HOMogenization).

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

2. Methodology

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 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 sub periods of at least two years around each break affecting the candidate series, since i) fitting spline models require a minimum amount of data and ii) data has to cover a range of situations large enough, in order to avoid extrapolation of the functions. Note that definition of homogeneous sub periods provided in the notation section is exactly the same as in Della-Marta and Wanner (2006).

a. Notation

In the following, we denote Y the candidate series, and X the reference series. Let j=1,…,k be the set of change-points affecting Y. For practical algorithmic reasons, we introduce dummy change-points j=0, corresponding to the last observation of Y, and k+1 corresponding to the day before the first observation of Y. Note that 1 refers to the most recent non-dummy change-point, while k is the most ancient one. Let us denote HSP_{Xjaft} the homogeneous subperiod of X after the j^{th} change-point on Y and HSP_{Xjbef} the homogeneous subperiod of X before (see figure 1). The homogeneous subperiod on Y between change-points j and j-1 is denoted HSP_{Yj}. Since X may be affected by change-points also, homogeneous subperiods HSP_{Xjaft}, $HSP_{Xj-1bef}$ may be shorter than HSP_{Yj}. Let m_{YXjaft} be the non-linear regression function of Y on X after the j^{th} change-point, and m_{YXjbef} the non-linear regression function of Y on X before the j^{th} change-point, while m_{XYjbef} is the non-linear regression function of X on Y before the j^{th} change-point.

b. Model

The change-point effects are adjusted sequentially, from the most recent (1) to the most ancient one (k). The last period HSP_{Y1} remains unchanged. Adjustment is first applied to HSP_{Y2}, then HSP_{Y3} up to HSP_{Yk+1}. For adjustment of the whole sub period HSP_{Yj+1} (corresponding to the effect of the j^{th} change-point) the first step is to estimate m_{YXjbef} (respectively m_{YXjaft}), that is the regression of Y on X before (resp. after) the break on HSP_{Xjbef} (resp. HSP_{Xjaft}) subperiods (Fig 1.).

Fig. 1 about here

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

$$Y_t^* = Y_t + \hat{m}_{YXjaft-YXjbef}(X_t) \quad (1)$$

were Y_t* is the adjusted value according to (1). At this stage, if reference X is homogeneous on HSP_{Yj+1}, (that is, HSP_{Yj+1} and HSP_{Xjbef} coincide) adjustments can be directly applied to Y before the j^{th} change-point using (1). But in the general case, reference X itself might be inhomogeneous, or missing, on parts of HSP_{Yj+1}. So an additional step is performed. The m_{XYjbef}

regression function is estimated. This is the regression of X on Y for subperiod HSP_{Xjbef}. It allows the substitution of Y_t into "pseudo" X_t values: $\hat{X}_t = \hat{m}_{XYjbef}(Y_t)$ in equation (1), $\hat{m}_{XYjbef}$ denoting the estimation of m_{XYjbef} on HSP_{Xjbef}. Finally, the SPLIDHOM adjusted observations $\hat{Y}_t$ are given by:

$$\hat{Y}_t = Y_t + \hat{m}_{YXjaft-YXjbef}\left(\hat{m}_{XYjbef}(Y_t)\right) \quad (2)$$

In the following, the term $\hat{m}_{YXjaft-YXjbef}\left(\hat{m}_{XYjbef}(Y_t)\right)$ is called adjustment or adjustment function.

While based on the same definition of sub periods than HOM, the adjustment proposed by SPLIDHOM differs in its principle. SPLIDHOM is based on regression only, while 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, while a large proportion of the useful information is hidden in the non linear part of the regressions. For estimating the regression function, several techniques have been tested: kernel smoothers (Brockman *et al.*, 1993, too noisy at the edge for data scarcity reasons), wavelet thresholding (Nason, 2008, too sensitive to small outliers) and LOESS (Cleveland and Grosse, 1991, too computationally demanding when applying for cross-validation techniques). Our final choice relies on classical cubic smoothing spline that does not have the previously mentioned drawback for our application. In the following we recall the basics of smoothing spline. Readers may refer to Hastie and Tibshirani (1990) for a more complete overview of this technique.

Cubic smoothing spline are the solution of the following optimization problem: let (Xi, Yi) for i=1…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” the second derivative of m and λ the smoothing parameter) that minimizes the penalized residual sum of squares:

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

Interval [a,b] corresponds to the range of X. This problem has a unique (and explicit) solution which is a natural cubic spline with knots at the values X_i. This model may seem over parameterized, 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 $\lambda \to \infty$, $m''(t) \to 0$ and the minimiser is the least squares line. The smoothing parameter is estimated for each regression by means of a standard cross-validation technique, in order to avoid over fitting. Let $\hat{m}_\lambda^{(-i)}$ be the solution for a given value λ, obtained leaving out observation i – which mimics training and test sample procedures. Estimated λ is the value that minimizes the cross-validation sum of squares:

$$CV\left(\lambda\right) = \sum_{i=1}^{n}\left(Y_i - \hat{m}_\lambda^{(-i)}\left(X_i\right)\right)^2$$

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 degree 1 to 3 polynomials. This is a significant difference to HOM, where the LOESS smoothing parameter is fitted rather empirically, as stated by the authors themselves.

Since the range of the data within different HSPs can be different, we often face an additional extrapolation problem. Linear extrapolation of m_{XYjbef} is easy to achieve, but extrapolation of $m_{YXjaft-YXjbef}$ may lead to incorrect results. So, we also choose to bound adjustments at the edges, as in HOM method. Practically, adjusting values greater (resp. lower) that the largest (lowest) observed value of X on the estimation interval is performed using adjustment computed for the largest (resp. lowest) observed value of X on the estimation interval.

3. Results

a. Simulation study

This experiment has two purposes: first, establish the correlation necessary to obtain good results with HOM and SPLIDHOM methods, then show SPLIDHOM improvements compared to Vincent's and HOM results on 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 (JJA) mean, Q05 and Q95 quantiles, and annual absolute maximum temperature.

Data are simulated according to the following scheme: Toulouse daily maximum temperature (TX) series is decomposed into seasonal, trend and noise component using moving averages of width equal to one year, according to a classical additive model (Brockwell and Davis, 2006). Result of this decomposition is shown in figure 2. The random component is then modeled as an AR(1) process. The estimation of first order autocorrelation is equal to 0.672, while the noise component of the AR(1) process is found to have variance equal to 8.6°C². 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 hereinafter:

$$\begin{cases} U_{1t} = \varphi_1 U_{1t-1} + \varepsilon_{1t} \\ U_{2t} = \varphi_2 U_{2t-1} + \varepsilon_{2t} \end{cases} \quad \text{with vector } \varepsilon_t = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} \right)$$

that is the noise term $\{U_1, U_2\}$ of the process follows a centered bivariate normal distribution, correlation between ε_1 and ε_2 being controlled by parameter r. Practically we set $\varphi_1=\varphi_2=0.672$, $\sigma^2=8.6$, that are values estimated on real Toulouse temperature series. Pairs of series are created summing the same trend and seasonal (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, 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 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. Type III transfer function is given by $t \rightarrow t+(e^{t/10})/20+\xi$, ($\xi$ defined as in type II). Type III results in larger skewness. We applied type I to period 1966-1970, type II to periods 1951-1965 and 1986-1995, and type III to periods 1971-1985, 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 figure 3.

Fig. 3 about here

For r taking values 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. Candidate series ("truth") is then perturbed as described previously, to give the "raw" candidate. Raw candidate series is then adjusted using HOM and SPLIDHOM methods. A pseudo-Vincent method is also used: for each sub period, 12 monthly adjustment coefficients are computed, computing the monthly mean differences between "truth" and "raw" over the whole sub periods. Since those estimates are much more accurate than they would be in reality, noise is added, consisting in a random centered normal variable of standard deviation 0.3, which is roughly the standard error estimate observed on monthly adjustment coefficients computed using Caussinus and Mestre (2004) ANOVA model. The annual cycle of adjustments is then interpolated using spline as described in Vincent's method. Note the multivariate ANOVA model takes all available monthly series in a regional neighborhood. In this experiment, we consider that 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 and series adjusted by means of Vincent's method, HOM and SPLIDHOM, on a variety of indices: root mean square error of the adjusted daily values *vs* "truth" (RMSE), and annual indices, such as annual means (average of the 365 values), annual absolute minimum and maximum temperatures (respectively lowest and highest temperature that occurred during the year), annual quantiles Q95 and Q05 of the daily values of the considered year. For each correlation, we compute boxplots of the 50 corresponding RMSE, as well as boxplots 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 figures 4, 5 and 6. Perfect adjustments would result on null differences and RMSE.

Figures 4, 5 and 6 about here

From these results, a number of comments can be made:

- all three methods improve the data: inhomogeneities are reduced, when comparing adjusted series to raw series;
- Vincent method is able to correct the means (annual, 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 bias of the annual maxima, Q95 and Q05 is about 1.0, 0.4 and -0.4°C respectively using the Vincent Method (Figure 6 bottom panels) in our experiment.
- HOM and SPLIDHOM improvements compared to Vincent are hardly noticeable for $r<0.90$. For example, when $r=0.8$, the bias of annual maxima, Q95 is 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.
- SPLIHOM clearly performs better than HOM in terms of RMSE.

Since r is really a crucial parameter, we plot median and inter quartile range of RMSE for the three methods as a function of r, for each of the scores.

Figure 7 about here

This confirms that both HOM and SPLIDHOM need well correlated series (r>0.90) to outperform Vincent method, in terms of RMSE and bias reduction for extreme quantiles. Performances of Vincent Method are less sensitive to r value, at least for the range of correlations we tested. Regarding comparison of HOM and SPLIDHOM, 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. Regarding Q95, SPLIDHOM is more biased for r≤0.90 but gets the best results for r>0.96. If we roughly consider that SPLIDOM is superior to Vincent Method for a correlation of 0.90, and delivers trustful results at a correlation of 0.95, those correlation thresholds are not anecdotic. For maximum temperatures, on a flat terrain region such as Paris region, a correlation of 0.95 (respectively 0.90) is achieved for an approximate station distance of around 75km (resp. around 150 km). In the more mountainous area around Lyon, those distances are respectively 18km and 60 km (not shown here).

b. Application on Toulouse-Blagnac temperature series

Toulouse-Blagnac (Toulouse civil airport, professional station, index number 31069001) minimum (TN) and maximum (TX) temperatures series are affected by several abrupt change-points. Those changes are detected using PRODIGE software (Caussinus and Mestre, 2004) that 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 position of changes together with an adapted penalized likelihood criterion (Caussinus and Lyazrhi, 1997) assessing significance of changes. Metadata allows validate those detections and provide causes and precise days for changes: 1962/06/20 (new instrumental park), 1968/10/15 (relocation, new shelter), 1972/05/01 (sensor change), 1986/06/17 and 1991/11/08 (for both, relocations of instrumental park, due to construction of new runways). The reference data is provided by Toulouse-Francazal series (French “Armée de l’Air” station, military airport), situated 12km south of Toulouse-Blagnac airport. This series is affected by a large change-point in 1955/11/14 (relocation and shelter change). Toulouse-Blagnac series starts in 1951. Correlation of the series is high: r=0.98 (at a daily time scale, seasonal cycle removed), justifying the use of SPLIDHOM technique. Change-point effects are adjusted sequentially, for each season, from period before the most recent change-point (1991) to the most ancient one (1962). On this example, we choose seasonal estimations, instead of monthly, since the results appeared to be more stable. Let us analyze in detail period 1986-1991, for autumn season (SON for September-October-November months). Figure 8a shows the scatterplot of observed daily Toulouse-Blagnac TN (candidate Y) as a function of daily Toulouse-Francazal TN (reference X), for homogeneous subperiod 01/09/1986−08/11/1991, for SON season. The solid grey line corresponds to the smoothing spline estimation of regression function m_{YXjbef}. Similarly, Figure 8b shows the scatterplot of daily data and estimation of regression function m_{YXjaft}, after the target change on 8/11/1991, for SON season, over sub period 08/11/1991−30/11/2009. Fig. 8c 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 $m_{YXjaft\text{-}YXjbef}$ in equation (1). In this example, $m_{YXjaft\text{-}YXjbef}$ can be considered linear. Estimation of the m_{XYjbef} additional transfer function used in equation (2) is also provided in Fig. 8d. Fig. 9 shows the estimated SPLIDHOM adjustment function $\hat{m}_{YXjaft-YXjbef}\left(\hat{m}_{XYjbef}\left(Y_t\right)\right)$. Note that, given the precision of the original database, the final adjustment function is rounded to a precision of 0.1°C, which explains its staircase behavior. For autumns 1986 to 1991, large adjustments are observed for low temperatures (up to +0.8°C), being almost null for warmer temperatures.

Fig. 8, 9 about here

When analyzing the correction of the previous break (1986) for the same SON autumn season, we find a rather different shape (Fig. 10 and 11). The estimation of $m_{YXjaft-YXjbef}$ is not linear (Fig. 10c), resulting in a non-linear adjustment function (Fig. 11), thus justifying the use of non-linear models in SPLIDHOM. When analyzing adjustments of Toulouse series for every breakpoint and every season, roughly half of the adjustment functions have a linear or quasi-linear (including constant) shape, the other half exhibiting a non-linear shape.

Fig. 10, 11 about here

The examples given above show that adjustments of our method are sensitive to temperature itself, thus taking into account in a crude way meteorological situation of each day. 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 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 experiment 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 software (monthly homogenization, Caussinus and Mestre, 2004), we find very close results (Fig. 12). This is a remarkable result, since PRODIGE method relies on a completely different principle, where mean biases are estimated using an ANOVA (ANalysis Of Variance) model, applied on a large set of monthly series in the same climatic area. When considering annual averages, we get very similar results using by two completely different methods applied independently.

Fig. 12 about here

4. Conclusion

Although part of the principle involved in HOM and SPLIDHOM are quite similar, especially the definition of sub periods, SPLIDHOM adjustments differ: they are based on non-parametric regression (by means of cubic smoothing spline) while HOM involves fitting data to several candidate distributions. The use of a smoothing parameter set by means of cross-validation avoids over fitting during the estimation process. On simulated examples, our SPLIDHOM technique is shown to improve HOM (especially in terms of RMSE) and Vincent's method for the correction of extreme quantiles if correlation is high enough, since application of the latter should not be neglected when correlation of involved series is lower than 0.90. A very important result of our study is that correlation of the candidate is the essential parameter that drives performances 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 "temperature dependant" (Type I errors in our simulation), SPLIDHOM still removes the main biases.

Finally, SPLIDHOM should be compared to new emerging techniques recently developed, such as an improved version of HOM, HOMAD (Toreti *et al.*, 2010) and a quantile matching technique (Wang *et al.*, submitted). Performances of those methods will be investigated further using various benchmarks and more types of inhomogeneities, during last phase of COST Action ES0601 “HOME”.

Acknowledgements

This work was partly funded by EU COST Action ES0601 “HOME”. The authors wish to thank Paul-Della Marta for fruitful discussions and for providing his HOM algorithm, and three anonymous reviewers for very useful suggestions.

References

Aguilar, E., T.C. Peterson, P. Ramírez Obando, R. Frutos, J.A. Retana, M. Solera, I., González Santos, R.M. Araujo, A. Rosa Santos, V.E. Valle, M. Brunet-India, L. Aguilar, L. Álvarez, M. Bautista, C. Castañón, L. Herrera, E. Ruano, J.J. Siani, F. Obed, G.I. Hernández Oviedo, J.E. Salgado, J.L. Vásquez, M. Baca, M. Gutiérrez, C. Centella, J. Espinosa, D., Martínez, B. Olmedo, C.E. Ojeda Espinoza, M. Haylock, R. Núnez, H. Benavides, and R. Mayorga, 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., 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.

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

Brandsma, T., 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.

Brockmann, M., Gasser, T. and E. Herrmann, 1993: Locally Adaptive Bandwidth Choice for Kernel Regression Estimators. *J. Am. Stat. Assoc.*, **88**, 424, 1302-1309.

Brockwell, J. and R. A. Davis, 2006: *time Series: Theory and methods (second edition)*. Springer Series in Statistics, Springer, pp. 577.

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. In *La meteorología y el clima atlánticos,* Publicaciones de la Asociación Española de Meteorología: Badajoz; 93-103.

Brunet, M., P.D. Jones, J. Sigró, O. Saladié, E. Aguilar, A. Moberg, P. Della-Marta, D. Lister, A. Walther, and D. López, 2007: Spatial and temporal temperature variability and change over Spain during 1850-2005. *J. Geophys. Res.*, 112, D12117, doi:10.1029/2006JD008249.

Brunet, M., O. Saladié, P.D. Jones, J. Sigró, E. Aguilar, A. Moberg, D. Lister, A. Walther, D. López, and C. Almarza, 2006: The development of a new dataset of Spanish Daily Adjusted Temperature Series (SDATS) (1850-2003). *Int. J. Climatol.*, **26**, 1777-1802.

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

Caussinus, H. and O. Mestre, 2004: Detection and correction of artificial shifts in climate series. *Applied Statistics*, **53**, part 3, 405-425.

Cleveland, W. S. and E. Grosse, E., 1991: Computational Methods for Local Regression. *Statistics and Computing*, **1**, 47–62.

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

Della-Marta, P.M., M. R. Haylock, J. Luterbacher, and H. Wanner, 2007: Doubled length of western European summer heatwaves since 1880. *J. Geophys. Res.*, **112**:D15103, doi:10.1029/2007JD008510.

Klein Tank, A. M. G., Zwiers F. W. 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-No. 72, WMO-TD No. 1500, 56pp.

Hastie, T.J. and R.J. Tibshirani, 1990: *Generalized Additive Models*. Monographs on Statistics and Applied Probability 43, Chapman and Hall/CRC, pp 335.

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

Lefèvre, G., 1998: Comparison of Meteorological Screens for Temperature Measurement, presented at TECO-98 (Casablanca), *Instruments and Observing Methods Report No. 70*, WMO/TD–No. 877, pp. 315, Geneva.

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, pp. 257.

Neumaier, A., and T. Schneider, 2001: Estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Transactions on Mathematical Software*, **27**, 1, 27-57.

Peterson, T.C., M.A. Taylor, R. Demeritte, D.L. Duncombe, S. Burton, F. Thompson, A. Porter, M. Mejia, E. Villegas, R.S. Fils, A. Klein Tank, A. Martis, R. Warner, A. Joyette, W. Mills, L. Alexander, and B. Gleason, 2002: Recent Changes in Climate Extremes in the Caribbean Region. *J. Geophys. Res.*, **107**, (D21), 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. Description and relevance for climate change analysis. *J. Climate*, **23**, 5325-5331, DOI: 10.1175/2010JCLI3499.1.

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

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

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 detection and adjustment of shifts in daily precipitation data series. *J. Appl. Meteor. Climatol.* (submitted).

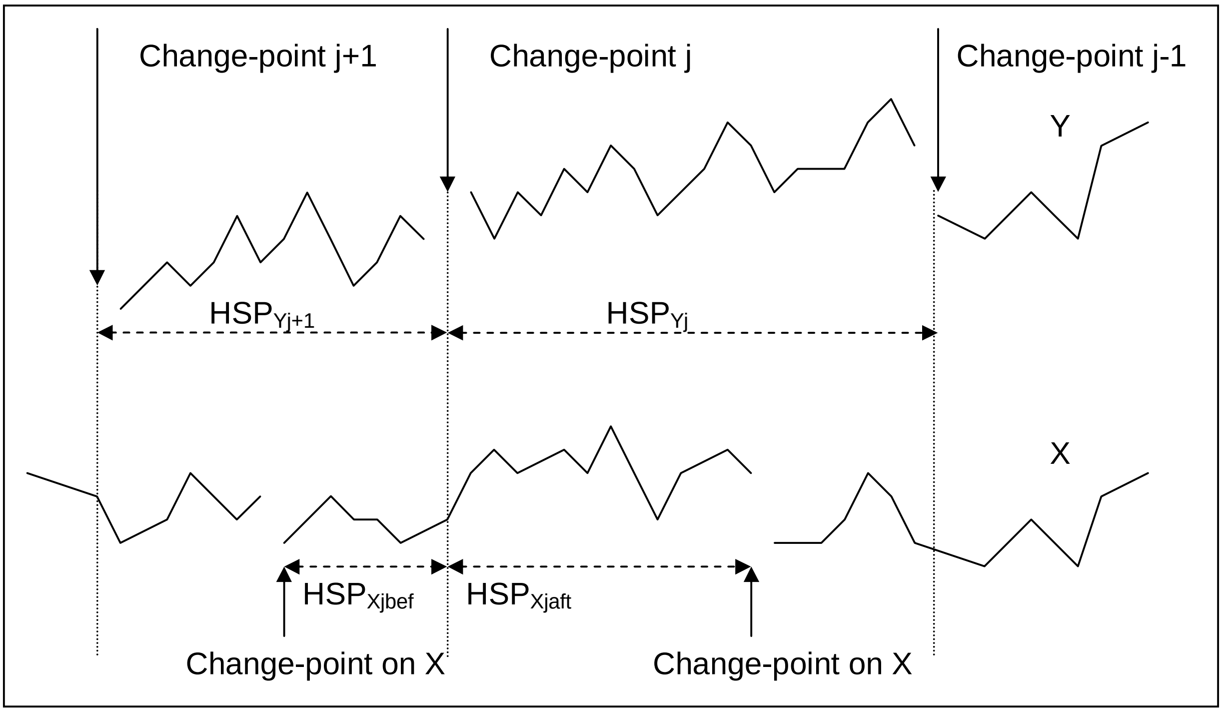

Fig.1. Definition of Homogeneous Sub Periods (HSPs)

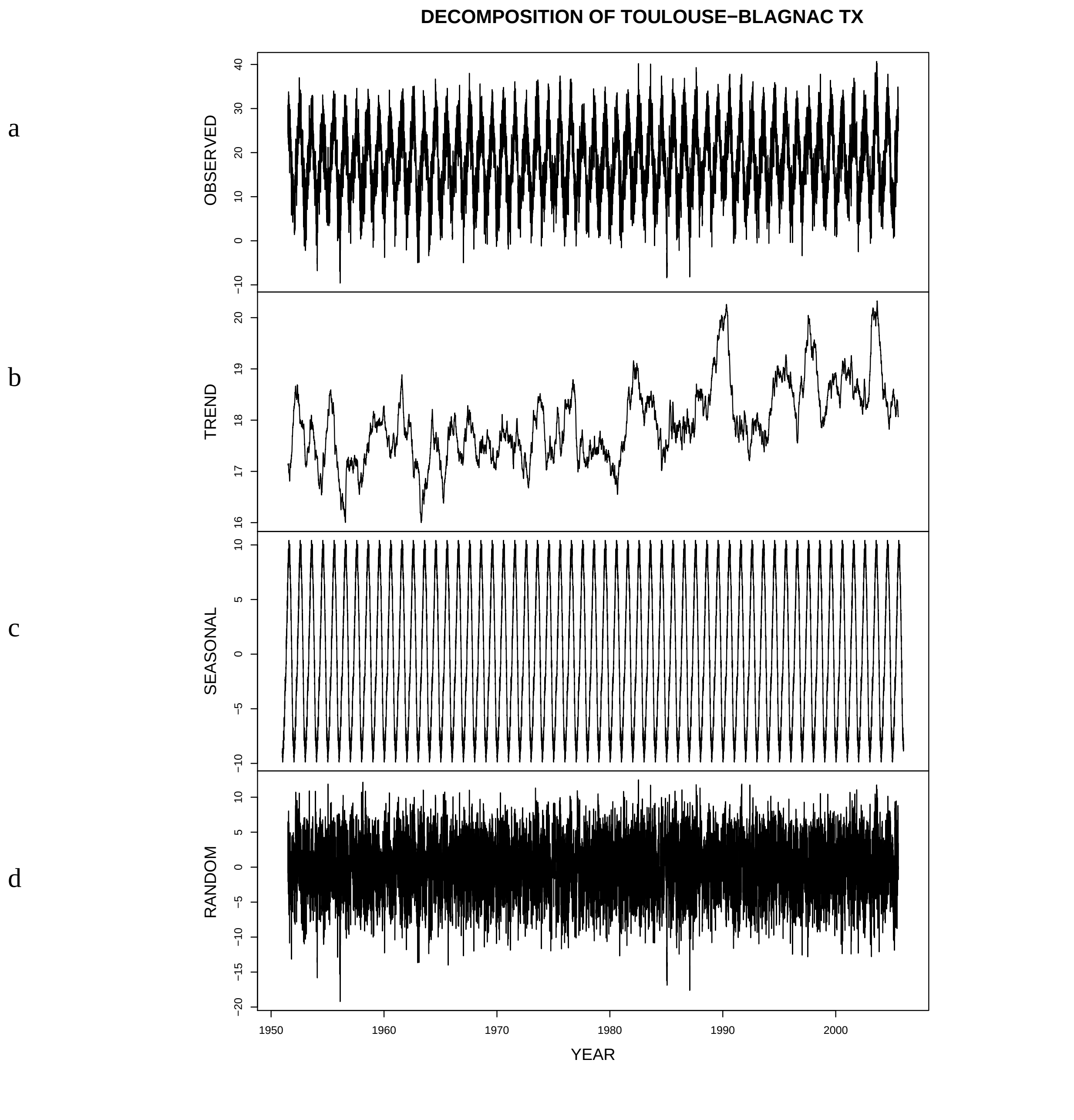

Fig. 2. Decomposition of observed (a) Toulouse daily maximum temperature series into trend (b), seasonal (c) and random noise (d) components.

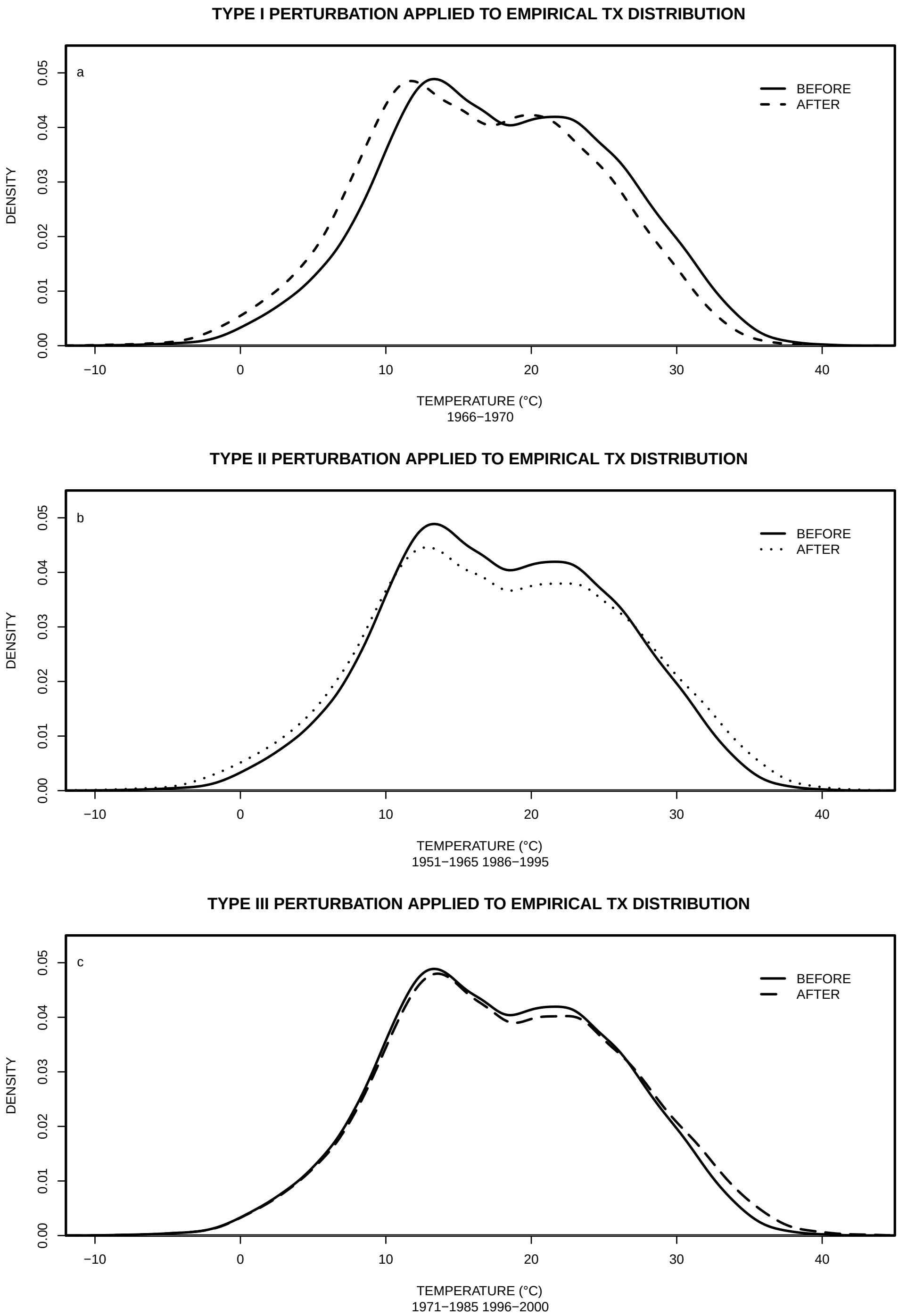

Fig. 3. Histogram of daily TX distribution of Toulouse data, before (solid) and after (dotted or dashes) application of type I (a), type II (b) and type III (c) inhomogeneities.

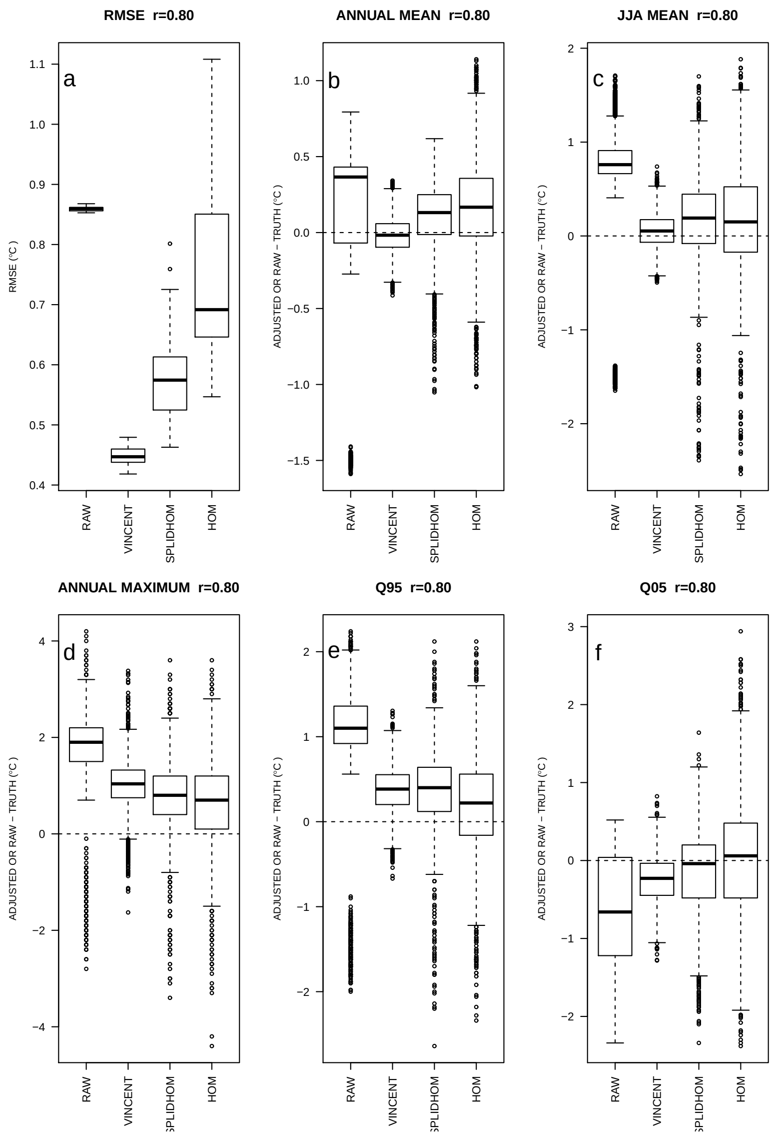

Fig. 4. Boxplots of root mean square error (RMSE) of daily raw (RAW) and adjusted values (Vincent, HOM and SPLIDHOM) compared to "truth" (a), and boxplots of differences between original unperturbed "true" series and raw ("RAW") series or adjusted series ("Vincent", "HOM" and "SPLIDHOM" methods) on a variety of annual indices, computed for each year: annual means (average of the 365 values, b), summer means (c), annual absolute maximum temperature (d) (highest temperature that occurred during the year), Q95 (resp. Q05) quantile of distribution of daily values (e) (resp. f), for correlation r=0.80 and for 50 computed series.

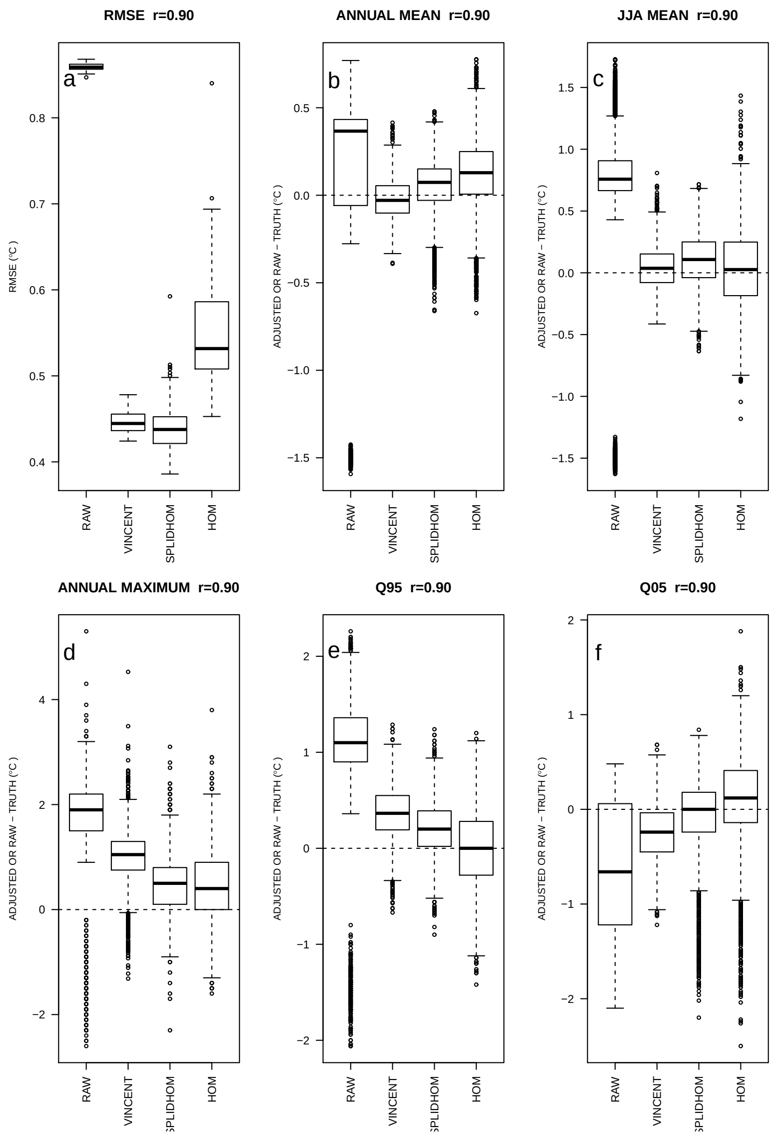

Fig. 5. Boxplots of root mean square error (RMSE) of daily raw (RAW) and adjusted values (Vincent, HOM and SPLIDHOM) compared to "truth" (a), and boxplots of differences between original unperturbed "true" series and raw ("RAW") series or adjusted series ("Vincent", "HOM" and "SPLIDHOM" methods) on a variety of annual indices, computed for each year: annual means (average of the 365 values, b), summer means (c), annual absolute maximum temperature (d) (highest temperature that occurred during the year), Q95 (resp. Q05) quantile of distribution of daily values (e) (resp. f), for correlation r=0.90 and for 50 computed series.

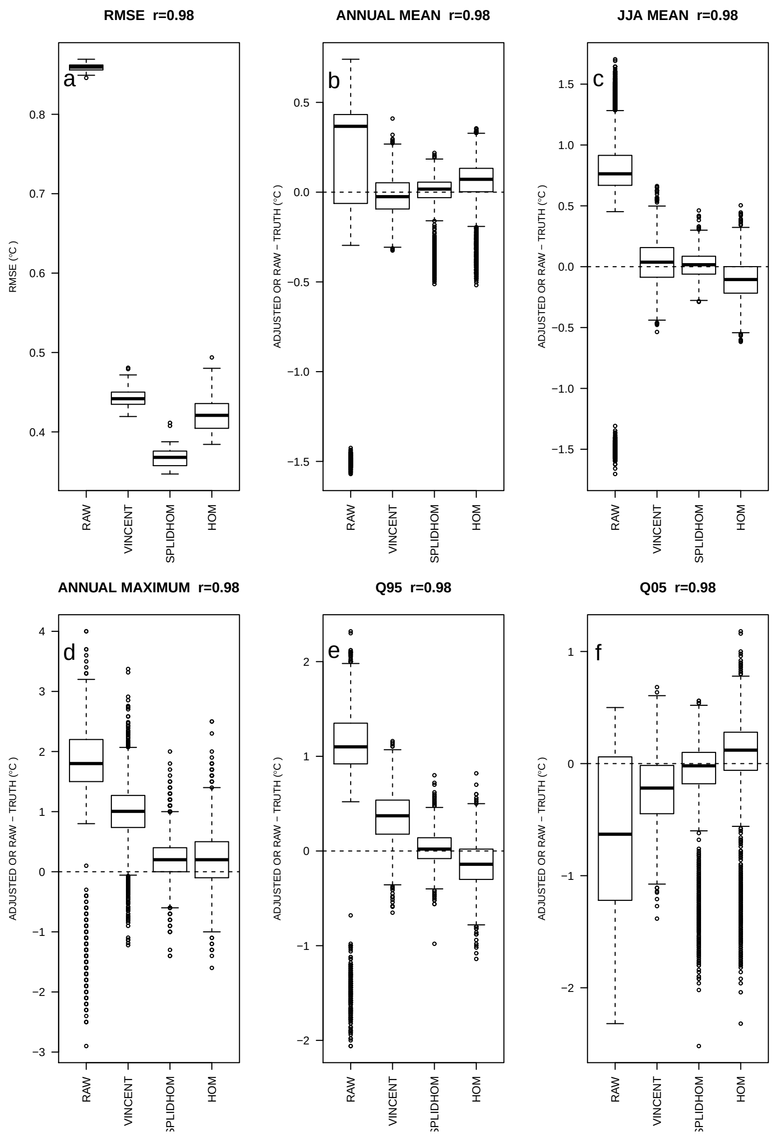

Fig. 6. Boxplots of root mean square error (RMSE) of daily raw (RAW) and adjusted values (Vincent, HOM and SPLIDHOM) compared to "truth" (a), and boxplots of differences between original unperturbed "true" series and raw ("RAW") series or adjusted series ("Vincent", "HOM" and "SPLIDHOM" methods) on a variety of annual indices, computed for each year: annual means (average of the 365 values, b), summer means (c), annual absolute maximum temperature (d) (highest temperature that occurred during the year), Q95 (resp. Q05) quantile of distribution of daily values (e) (resp. f), for correlation r=0.90 and for 50 computed series.

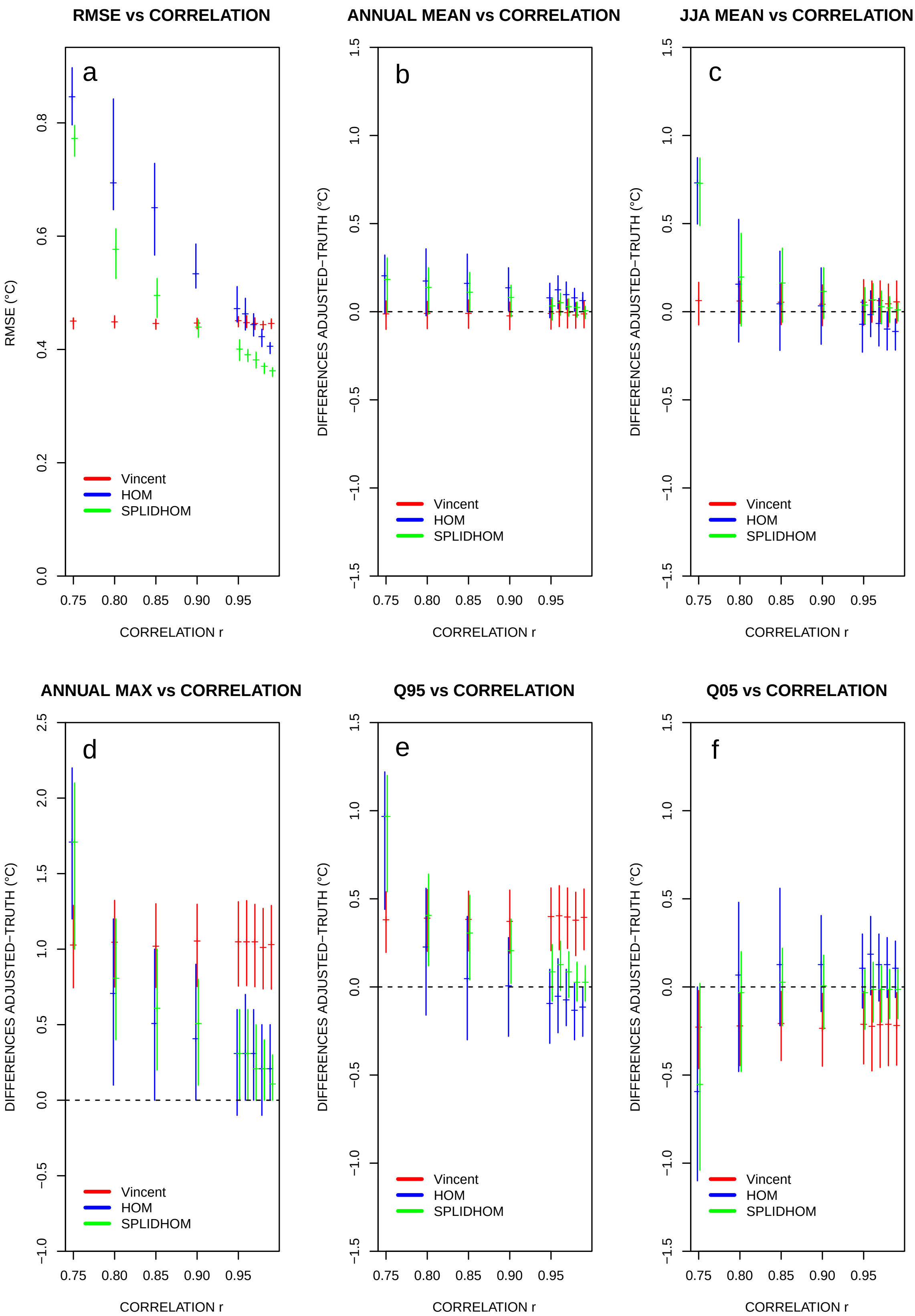
RMSE vs CORRELATION
ANNUAL MEAN vs CORRELATION
JJA MEAN vs CORRELATION
ANNUAL MAX vs CORRELATION
Q95 vs CORRELATION
Q05 vs CORRELATION
a
b
c
d
e
f
RMSE (°C)
DIFFERENCES ADJUSTED−TRUTH (°C)
CORRELATION r
Vincent
HOM
SPLIDHOM

Fig. 7. Interquartile range (vertical bars) and median (short horizontal bar) of root mean square error (RMSE) of daily raw (RAW) and adjusted values (Vincent, HOM and SPLIDHOM) compared to “truth” (a), and boxplots 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: annual means (average of the 365 values, b), summer means (c), annual absolute maximum temperature (d) (highest temperature that occurred during the year), Q95 (resp. Q05) quantile of distribution of daily values (e) (resp. f), for correlation r=0.75, 0.80, 0.85, 0.90, 0.95, 0.96, 0.97, 0.98, 0.99 and for 50 computed series.

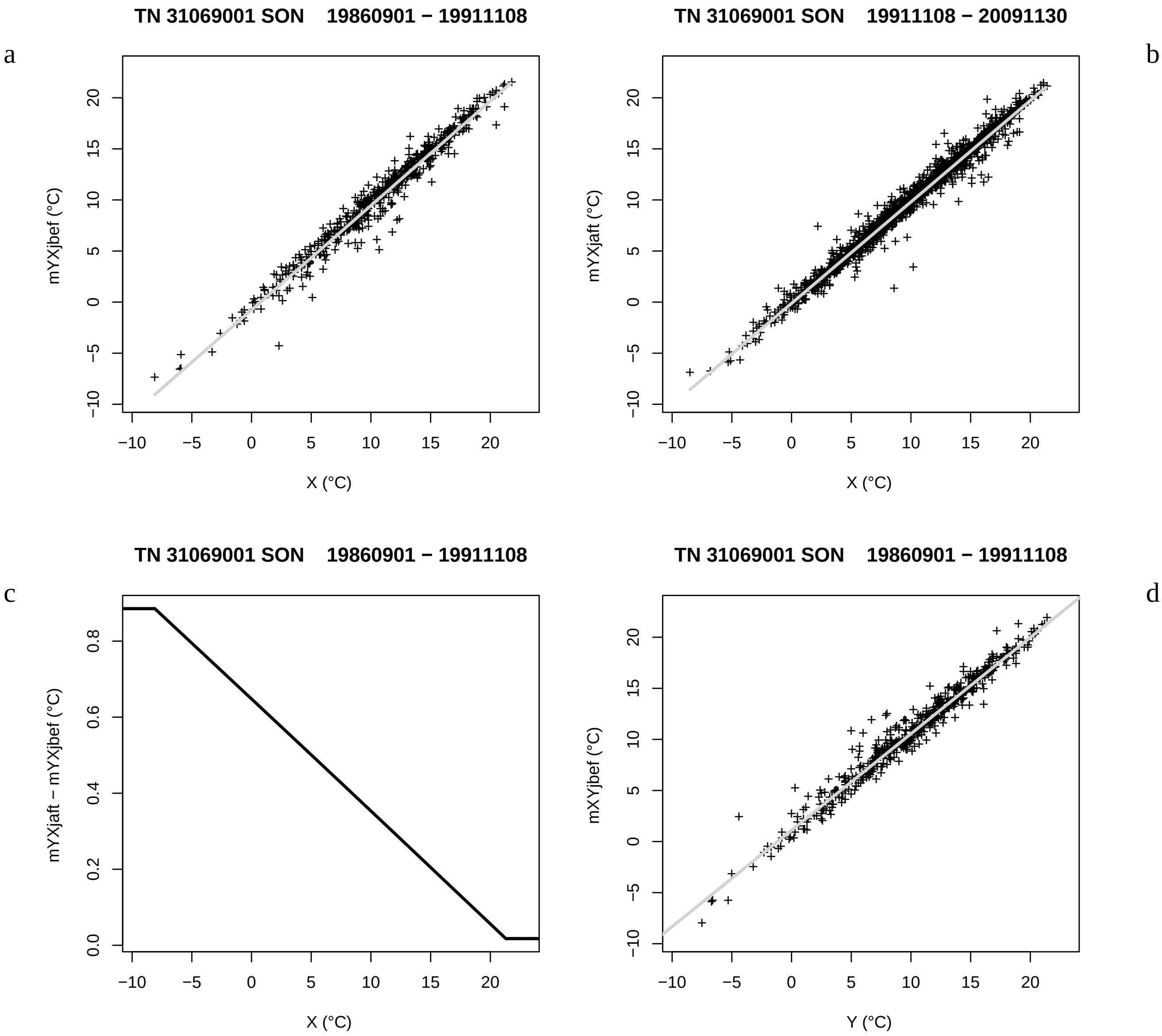

Fig. 8. Regression estimations for adjustment of HSP between 1986/06/17 and 1991/11/06, for Toulouse-Blagnac daily minimum temperature (Y) and autumn season (SON), using Toulouse-Francazal (X) as a reference. Scatter plot of Y versus X before the 1991 shift, together with corresponding cubic spline estimation of m_{YXjbef} (a), scatter plot of Y versus X after the 1991 shift with corresponding cubic spline estimation of m_{YXjaft} (b), estimation of m_{YXjaft}-$_{YXjbef}$ (c),

scatter plot of X versus Y, together with cubic spline estimation of m_{XYjbef} (d). Note that data are split according to the definition of HSPs and position of detected change-points.

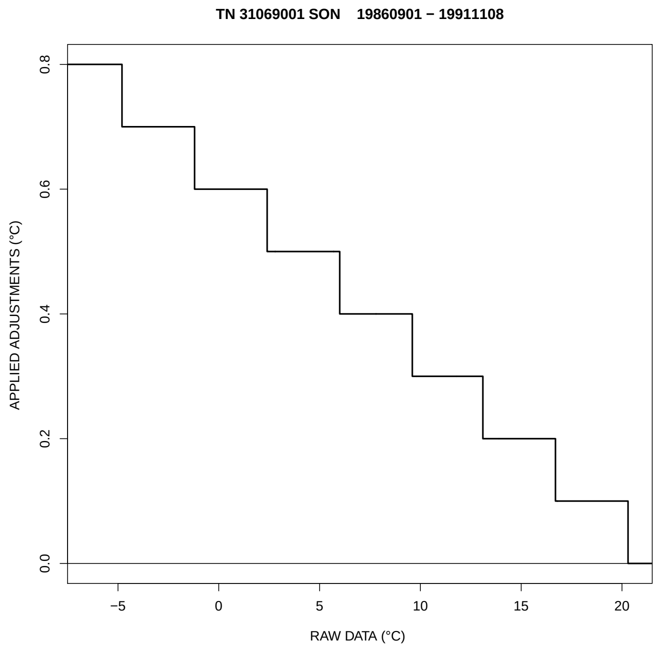

Fig. 9. Estimation of the adjustment function $m_{YXjaft-YXjbef}[m_{XYjbef}(\bullet)]$ for HSP between 1986/06/17 and 1991/11/08, for Toulouse-Blagnac daily minimum temperature (Y) and autumn season (SON). This function gives the adjustment to be applied to Y as a function of Y itself (in °C). Note that this estimation is rounded to a precision of one tenth of a degree – which is the precision of data itself in the database.

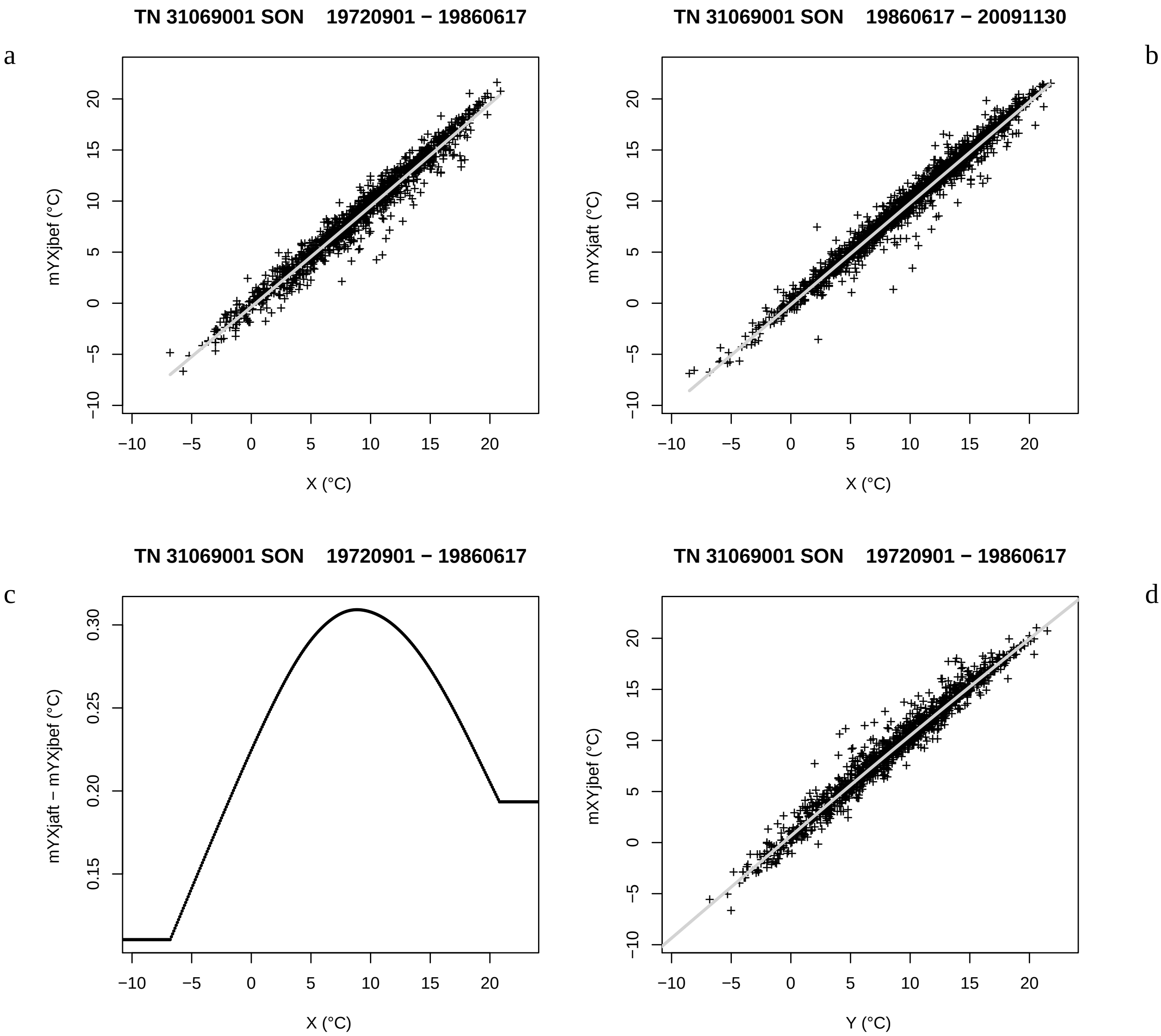

Fig. 10. Regression estimations for adjustment of HSP between 1972/05/01 and 1986/06/17, for Toulouse-Blagnac daily minimum temperature (Y) and autumn (SON), using Toulouse-Francazal (X) as a reference. Scatter plot of Y versus X before the 1986 shift, together with corresponding cubic spline estimation of m_{YXjbef} (a), scatter plot of Y versus X after the 1986 shift with corresponding cubic spline estimation of m_{YXjaft} (b), estimation of m_{YXjaft}-$_{YXjbef}$ (c), scatter plot of

X versus Y, together with cubic spline estimation of m_{XYjbef} (d). Note that data are split according to the definition of HSPs and position of detected change-points.

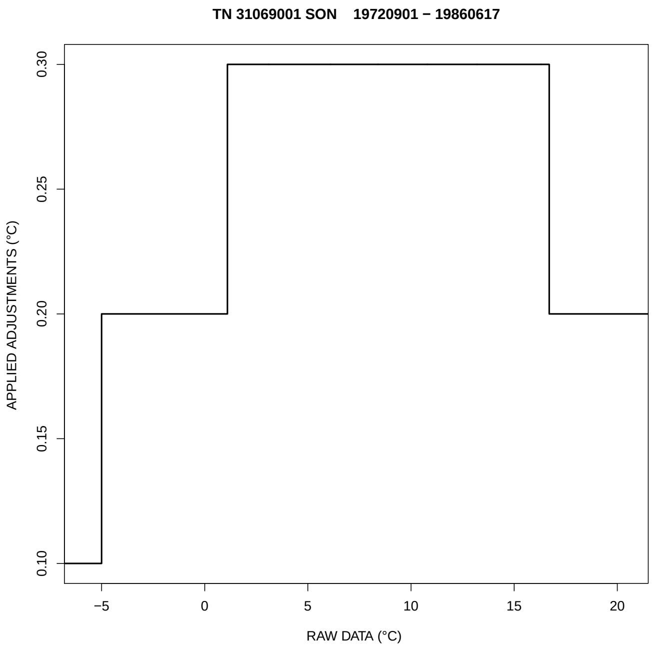

Fig. 11. Estimation of the adjustment function $m_{YXjaft-YXjbef}[m_{XYjbef}(\bullet)]$ for HSP between 1972/05/01 and 1986/06/17, for Toulouse-Blagnac daily minimum temperature (Y) and autumn season (SON). This function gives the adjustment to be applied to Y as a function of Y itself (in °C). Note that this estimation is rounded to a precision of one tenth of a degree – which is the precision of data itself in the database.

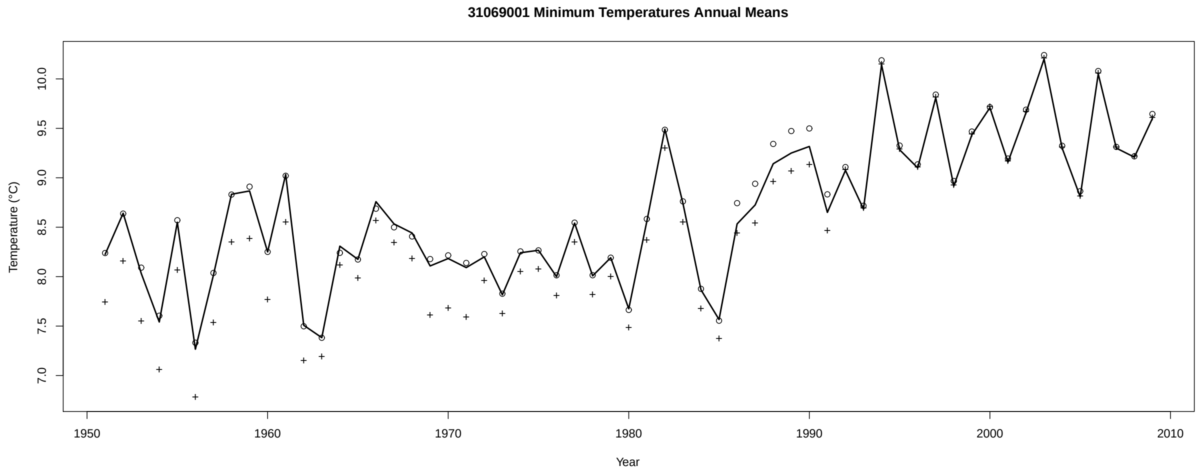

Fig. 12. Annual averages of daily adjusted TN series (o) compared to annual averages of raw (+) series and annual averages of monthly homogenized series (solid).