

APPLICATION OF NEAREST-NEIGHBOR RESAMPLING FOR HOMOGENIZING TEMPERATURE RECORDS ON A DAILY TO SUB-DAILY LEVEL

T. BRANDSMA* and G. P. KÖNNEN

Royal Netherlands Meteorological Institute (KNMI), De Bilt, The Netherlands

Received 9 September 2004

Revised 6 June 2005

Accepted 6 June 2005

ABSTRACT

Nearest-neighbor resampling is introduced as a means for homogenizing temperature records on a daily to sub-daily level. Homogenization refers here to the problem of calculating daily mean and sub-daily temperatures from a time series subject to irregular observation frequencies and changing observation schedules. The method resamples diurnal temperature cycles from an observed hourly temperature subrecord at the station. Unlike other methods, the technique maintains the variance in a natural way. This property is especially important for the analysis of trends and variability of extremes. For a given day, the resampling technique does not generate a single-valued solution but this peculiarity is of no effect in the applications considered here. The skills of the nearest-neighbor resampling technique, in terms of bias, RMSE, and variability, are compared with those of four other methods: a sine-exponential model, a model that uses the climatological mean daily cycle, a regression model for calculating daily values, and a deterministic version of the nearest-neighbor technique. The series used in the tests is the 1951–2000 meteorological record of De Bilt (The Netherlands). The emphasis in the comparisons is on the reconstruction of daily mean temperatures. The analysis shows important differences in performance between the models. The regression-based method performs best with respect to the calculation of the individual daily mean temperatures; the day-to-day variability is best reproduced with the nearest-neighbor resampling technique. The performance of the models improves when cloudiness is used as an extra predictor. The improvement is, however, small compared to the intermodel differences. The type of model that should be used depends on the desired application. For trend and variability studies, the nearest-neighbor resampling technique performs best. Nearest-neighbor resampling can successfully be performed even in situations where the length of the hourly subrecord is an order of magnitude less than the length of the series to be homogenized. Copyright © 2005 Royal Meteorological Society.

KEY WORDS: temperature; homogenization; climate change; The Netherlands

1. INTRODUCTION

For the analysis of trends and variability of extreme weather events, datasets with a daily resolution or better are needed. Nowadays these data sets become more and more available to the scientific community (Peterson *et al.*, 1997; Lavery *et al.*, 1997, Vincent *et al.*, 2000; Klein Tank *et al.*, 2002). However, with the introduction of daily data sets and the accompanying analysis of extremes (e.g. Klein Tank and Können, 2003), the differences between methods used for homogenizing meteorological time series become more apparent.

In general, homogenization techniques focus on means of the climate variables. For the analysis of changes in extreme weather events and variability, this is not sufficient. This type of analysis requires that other characteristics than just the mean of the probability density functions be preserved. The development of methods to achieve this is still in its infancy (Wijngaard *et al.*, 2003). In this paper, we propose that

* Correspondence to: T. Brandsma, Royal Netherlands Meteorological Institute (KNMI), PO Box 201, 3730 AE De Bilt, The Netherlands; e-mail: theo.brandsma@knmi.nl

nearest-neighbor resampling (see e.g. Young, 1994; Lall and Sharma, 1996) can be used to deal with this problem. We apply the technique to the well-known problem of calculating daily mean and sub-daily temperature values for temperature time series subject to irregular observation frequencies and changing observation schedules. The technique has the potential to be applied to other homogenization problems as well.

Nearest-neighbor resampling is a nonparametric statistical simulation or estimation technique. In the nearest-neighbor resampling procedure, variables of interest are sampled (simultaneously) with replacement from the observed data. In our case this refers to the temperatures belonging to the diurnal temperature cycle, which are then taken from a day that is randomly chosen from a set of days with analogue properties and having hourly observations. An interesting feature of the resampling technique is that no assumptions have to be made about the underlying distributions of each of the variables and of the dependencies between those variables. Especially in the hydrological literature, nearest-neighbor resampling has received much attention, mainly as a technique for simulating time series of weather variables (Lall and Sharma, 1996; Brandsma and Buishand, 1998; Buishand and Brandsma, 2001; Harold *et al.*, 2003).

Irregular observation frequencies and changing observation schedules at a station are an important cause of inhomogeneities in daily temperature time series (e.g. Mitchell, 1958; Baker, 1975; Schaal and Dale, 1977; Karl *et al.*, 1986). Daily mean temperatures (T_m), for instance, derived for a station that has been subject to such changes, are often used for trend analysis. Before the analysis, the series are homogenized by reducing the observation-hour averaged temperatures to T_m (often taken equal to the 24-h mean). The effectiveness of the corrections for a particular day is, however, highly dependent on the sensitivity of the daily temperature cycle to weather situations. Particularly on days with pronounced or extreme temperatures, the diurnal cycle differs largely from its mean shape and amplitude. As a result, correction factors based on regression or climatology underestimate the variance in the reconstructed T_m series. This problem can be bypassed with the nearest-neighbor resampling technique, as, in contrast to other methods, the resampling technique is able to preserve the variance in a natural way by resampling diurnal temperature cycles from the observed temperature record. This property is especially important for trend and variability analysis of extremes.

Because the resampling technique implies an insertion of sub-daily values originating from a day that is chosen out of the set of nearest neighbors by a random selection procedure, it may lead to more than one answer for each day and is therefore not single-valued. Hence, each resampling run may provide another value. Although being multivalued is an uncommon property in the climate literature on homogenization, it does not need to be a disadvantage. It has to be regarded in the context of the application. For instance in trend analysis multivalued solutions for specified days have no effect on the result, while for the calculation of a particular day, it obviously has. If the requirement is homogenization without loss of variance, nearest-neighbor resampling may do the job. Apparently, the price to be paid is nonuniqueness of the individual values.

To illustrate the benefits of statistical selection procedures in homogenization problems, we also present a 'deterministic version' of the nearest-neighbor technique. This version uses a weighted mean of the nearest neighbors and is therefore not a resampling technique but rather a nearest-neighbor smoothing technique (Härdle, 1990). We further denote the deterministic version as the nearest-neighbor *averaging* technique.

Nearest-neighbor resampling is compared with a sine-exponential model (referred as SinExp), with a model that fixes the climatological daily cycle model (ClimDay), with a linear regression model (LinReg), and with nearest-neighbor averaging. The skills of the various methods, in terms of bias, root mean squared errors (RMSE), and variability, are tested using the hourly record of the Dutch meteorological station De Bilt (1951–2000). The emphasis is on the reconstruction of daily mean temperatures, but the reproduction of hourly and minimum and maximum temperatures is also considered. We also study the added value of using cloudiness as an extra predictor in the statistical models.

In Section 2, we describe the data of De Bilt and we introduce the nearest-neighbor resampling technique, nearest-neighbor averaging, and the three other models. Section 3 presents the results of the comparison of the methods. Section 4 closes the paper with a discussion and conclusion.

2. DATA AND METHODS

2.1. Data

For the analysis in this paper, the hourly temperature series 1951–2000 of De Bilt are used. De Bilt is situated in the center of the Netherlands at (52°06'N, 05°11'E) with the ground surface 2 m above mean sea level. The major observational changes that occurred during the 1951–2000 period are: (1) a relocation of the thermometer screen on 27 August 1951 to a grass field 210 m southwards; (2) a lowering of the measurement height on 29 June 1961 from 2.2 to 1.5 m; (3) a change of screen type on 5 August 1980 from a wooden to a synthetic Stevenson screen (artificially ventilated); and (4) a change of screen type on 26 June 1993 to a KNMI multiplated screen (naturally ventilated). There is a linear trend of 1.25 °C over the 1951–2000 period. This is, however, of no effect to the mutual comparison of the models in this paper.

For the test of sensitivity of the methods to cloud cover, hourly and daily mean cloud information is needed. In the hourly 1951–2000 De Bilt series, visually observed cloud cover on a scale between zero (clear) to eight (overcast) is continuously available. For the nearest-neighbor methods, we used cloudiness directly on the observation hours, while for ClimDay and LinReg we defined three cloud classes for the daily mean cloudiness: few cloud (0–4, occurring in 28.6% of the data), medium cloud (5–6, 35.6% of the data), and much cloud (7–8, 35.8% of the data). Daily mean cloudiness is calculated here as the 24-h mean cloudiness rounded to the nearest integer.

2.2. Methods for calculating the diurnal cycle, T_m , T_{\min} and T_{\max}

In this section, we describe the five methods considered that can be used to estimate T_m from a few observations per day (the average over 24 hourly observations is considered here to represent the 'true' daily mean temperature T_m). The methods are nearest-neighbor resampling (k -NN₁), nearest-neighbor averaging (k -NN₂), SinExp, ClimDay and LinReg. The nearest-neighbor techniques and SinExp stand out from the other models in the sense that they also provide an internally consistent estimate of the full diurnal cycle, including the values of T_{\min} and T_{\max} . LinReg can also be used to directly estimate T_m , as well as T_{\min} and T_{\max} , but these estimates occur independently of each other. ClimDay cannot be expected to provide reasonable values for sub-daily values or for T_{\min} and T_{\max} , particularly if the deviation of T_m from climatology is large. In fact, owing to their construction, k -NN₂ and SinExp may also *a priori* be expected to yield biased estimates of T_{\min} and T_{\max} , but for these methods it is less obvious than for ClimDay.

2.2.1. k -nearest neighbor techniques (k -NN₁ and k -NN₂). The method of nearest-neighbor resampling (Lall and Sharma, 1996; Rajagopalan and Lall, 1999) is used here to search for analog diurnal temperature cycles in an observational record of hourly temperature observations. We denote the nearest-neighbor techniques by k -NN, the k referring to the number of nearest neighbors on which the selection is based and NN referring to nearest neighbors. We use two versions of the method. The first version (k -NN₁), refers to nearest-neighbor resampling, in which for each day, one of the k -NN is randomly selected according to some predefined weight-function to construct a diurnal temperature cycle. The second version (k -NN₂), referred here as the nearest-neighbor averaging, calculates the diurnal cycle as a weighted average of all k -NN using the same weight-function as in k -NN₁. We now give a concise description of how the nearest neighbors are determined; for more details of the method we refer to Brandsma and Buishand (1998) and Buishand and Brandsma (2001).

To find an analog diurnal temperature cycle for day t , a state vector \mathbf{D}_t is formed consisting of q characteristics for that day (e.g. temperature and cloudiness at certain times of day t or even day $t - 1$ or $t + 1$). Next, a distance δ between \mathbf{D}_t of the day of interest and the state vector \mathbf{D}_u of the days in the observed record (within a certain window centered on the calendar day of interest) has to be determined. The Mahalanobis metric (e.g. Kendall *et al.*, 1983, p. 290) is used here to determine this distance:

$$\delta(\mathbf{D}_t, \mathbf{D}_u) = ((\mathbf{D}_t - \mathbf{D}_u)^T \mathbf{B}^{-1} (\mathbf{D}_t - \mathbf{D}_u))^{1/2} \quad (1)$$

where \mathbf{B} is the covariance matrix of \mathbf{D}_t , i.e. the $q \times q$ matrix containing the covariances between the q variables in \mathbf{D}_t (\mathbf{B} is calculated for each calendar day separately, yielding 365 matrices). A discrete probability

distribution is then used for resampling one (k -NN₁) or all (k -NN₂) of the k -NN. Here we use the decreasing kernel recommended by Lall and Sharma (1996) that gives higher weight to the closer neighbors. The probability p_j that the j th closest neighbor is resampled is then given by:

$$p_j = \frac{1/j}{\sum_{i=1}^k 1/i} \quad (2)$$

For instance, for $k=20$, the probability p_1 that the closest NN is selected equals $1/(\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{20}) = 0.278$ and the probability p_{20} that the farthest NN is selected equals $\frac{1}{20}/(\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{20}) = 0.014$. The day itself is excluded here from the NN to make a more fair comparison between the method in the calibration period possible.

To reduce the effect of seasonal variation of the variables, two provisions are made here. First, for all variables x in the state vector \mathbf{D} we calculated anomalies x^* by subtracting an estimate $m_{d,h}$ of the mean of the variable for each calendar day d ($d = 1, 2, \dots, 365$) and hour h ($h = 1, 2, \dots, 24$); after the resampling the x^* is backtransformed into x . To reduce the effect of sampling variability, smooth approximations for successive calendar days of $m_{d,h}$ are used instead of the raw values. Smoothing was done with the so-called *supersmoother* (Härdle, 1990; compare also Brandsma and Buishand, 1998). Second, the search for nearest neighbors is restricted to days within a moving window of width w , centered on the calendar day of interest. Accordingly, the covariance matrix \mathbf{B} is calculated for each calendar day using the days in the window. Although the method is nonparametric, values must be specified for k and the window width w .

2.2.2. Sine-exponential model (SinExp). SinExp is an analytical model of the diurnal cycle proposed by Parton and Logan (1981). This model describes the diurnal cycle by a sine function during daytime connected to a decreasing exponential function during nighttime. For a given day, it has three free parameters, T_{\min} and T_{\max} on day t and T_{\min} on day $t + 1$ denoted as T'_{\min} . In an uninterrupted series of days, two free parameters for each day remain because of double use of T_{\min} . The model assumes that T_{\max} occurs somewhere during the daytime hours and T_{\min} occurs within a few hours before or after sunrise. The phase shift α between the time of the maximum temperature h_{\max} and midday is defined as $\alpha = h_{\max} - \frac{1}{2}(h_r + h_s)$, where h_r and h_s are the times of sunrise and sunset respectively. The phase shift β between the time of the minimum temperature h_{\min} and h_r is defined by $\beta = h_{\min} - h_r$. Then, the SinExp model is given by:

$$T(h) = \begin{cases} T_{\min} + (T_{\max} - T_{\min}) \sin \left[\frac{\pi(h - h_r - \beta)}{l + 2(\alpha - \beta)} \right] & h_{\min} \leq h \leq h_s \\ \tau + [T(h_s) - \tau] \exp \left[-\frac{\gamma(h - h_s)}{24 - l + \beta} \right] & h_s \leq h \leq h'_{\min} \end{cases} \quad (3a)$$

where day length l is defined as $l = h_s - h_r$, γ is an exponential decay coefficient, and h'_{\min} is the time of the minimum temperature at day $t + 1$. All times are in hours. The variables h_{\min} , h'_{\min} , h_{\max} are climatological parameters, calculated per month, and h_r and h_s are astronomical quantities. The decay coefficient γ can also be treated as a climatological parameter, but better results are obtained if γ is adjusted to the observations. This approach, which we follow in the present paper, adds one more degree of freedom to the model, bringing the total (for uninterrupted series) to three.

The parameter τ deserves special attention. In Parton and Logan's (1981) paper, as well as in follow-up papers (e.g. Van Engelen and Geurts, 1983; Wann *et al.*, 1985), τ in Equation (3b) is taken as $\tau = T'_{\min}$. This substitution is, however, not correct, as the actual meaning of τ should be the asymptotic value of T of the nighttime decay, hence Equation (3b) for $h \rightarrow \infty$. Consequently, the erroneous insertion $\tau = T'_{\min}$ results in physical inconsistencies, which manifest themselves, among other things, in a discontinuity in the diurnal temperature cycle at the time of the minimum temperature. The correct value of τ follows from Equation (3b)

and the condition of continuity of the daily temperature curve at h'_{\min} :

$$\tau = \frac{T'_{\min} - T(h_s) \exp\left[-\frac{\gamma(h'_{\min} - h_s)}{24 - l + \beta}\right]}{1 - \exp\left[-\frac{\gamma(h'_{\min} - h_s)}{24 - l + \beta}\right]}. \quad (4)$$

So far, no extension of the SinExp model that described cloud-cover dependence of the diurnal cycle has been formulated.

2.2.3. Climatological daily cycle model (ClimDay). ClimDay uses the mean climatological diurnal temperature cycle, known from a period for which hourly temperature readings are available. This cycle can then be used to take into account the observational hours in calculating T_m . Here we follow the approach suggested by Moberg *et al.* (2002) and Bergström and Moberg (2002) using a slightly different notation:

$$T_m = \frac{1}{N} \sum_{i=1}^N (T_{h(i)} - \Delta_{h(i)}), \quad N \leq 24; \quad h(i) \in [1, 2, \dots, 24] \quad (5)$$

where N is the number of observation hours on a day, $h(i)$ is the observation time of the i th observation, and $\Delta_{h(i)}$ denotes the deviation of the climatological temperature at time $h(i)$ from the climatological T_m for the calendar day to which the observation day belongs. Note that Equation (5) is equivalent to climatologically correcting the arithmetic average of a restricted number of sub-daily observations to the arithmetic average of an hourly set (see e.g. Können *et al.*, 1998).

For the inclusion of cloud, we also follow the approach suggested by Moberg *et al.* (2002) and Bergström and Moberg (2002):

$$T_m = \frac{1}{N} \sum_{i=1}^N \left(T_{h(i)} - \frac{\overline{\text{DTR}(p)}}{\overline{\text{DTR}}} \Delta_{h(i)} \right), \quad N \leq 24; \quad h(i) \in [1, 2, \dots, 24] \quad (6)$$

where $\overline{\text{DTR}}$ and $\overline{\text{DTR}(p)}$ are the mean diurnal range $\overline{T_{\max}} - \overline{T_{\min}}$ for all days and for days in cloud class p , respectively. The calendar-day stratified values of $\Delta_{h(i)}$, $\overline{\text{DTR}(p)}$, and $\overline{\text{DTR}}$, are derived for a period for which hourly measurements are available.

2.2.4. Linear regression model (LinReg). LinReg models may be used to calculate T_m , T_{\min} , or T_{\max} from a restricted number of observations per day. The method used here is described as follows:

$$T_m = a_0 + \sum_{i=1}^N a_i T_{h(i)} + \varepsilon, \quad h \in [\dots, -1, 0, 1, 2, \dots, 24, 25, \dots] \quad (7)$$

where the notation is the same as in Equation (5) and ε is an error term. Cloudiness may be included as a predictor by making the coefficients a_i cloud-deck dependent according to climatology. For the calculation of T_{\min} or T_{\max} , Equation (7) is also used, by replacing T_m in Equation (7) with T_{\min} and T_{\max} , resulting in estimates of T_{\min} and T_{\max} that are independent of each other and of T_m .

2.3. Model calibration and verification

Reconstructions of T_m , T_{\min} or T_{\max} from the 1951–2000 observations of De Bilt at 0800, 1400, and 2000 GMT (about 20 min difference with local solar time) are calculated with each of the five methods (with exceptions of those cases for which a particular model cannot be used). Both reconstructed and observed values of T_{\min} and T_{\max} are for each day taken to be the minimum and maximum, respectively, of the 24

hourly temperatures. The reconstructed values are then compared with the observed values of T_m , T_{\min} , or T_{\max} in the series. The 1971–2000 period is used for model calibration and the 1951–1970 period for model verification.

For the k -NN techniques, values have to be specified for the moving window width w and the number of nearest neighbors k used in the selection/averaging procedure. For w we used a value of 31 days, which is considered small enough to minimize the effect of seasonal variation of the variables, but large enough to permit a large sample population. Thus for the 1971–2000 observed record, the k -NN for a given day are selected from $30 \times 31 - 1 = 929$ days. For k , a sensitivity calculation showed that adequate results, in terms of annual mean RMSE of the hourly data, could be obtained for a rather broad range of values (about 5–50). It is known that the optimal value of k depends on the type of application (see e.g. Buishand and Brandsma, 2001). In general, the optimal k will be smaller when the purpose is time series simulation than when the purpose is prediction. Here we chose $k = 20$ for resampling from the 30-year observed record. Thus the nearest-neighbor resampling technique (k -NN₁) randomly selects one of the 20 nearest neighbors, with the probability of a NN being selected specified by Equation (2), whereas the nearest-neighbor averaging technique (k -NN₂) calculates a weighted average of the 20 nearest neighbors according to the weights taken from Equation (2). The NN techniques are applied to the anomalies T^* of the temperatures in the state vector. The resulting simulated daily cycle for a given day is linearly adjusted to fit the known temperature observations of the day in consideration. Then, jumps between days are removed by linearly adjusting between the closest known observation before midnight and the closest known observation after midnight. After these procedures, T^* is backtransformed into T . Figure 1 gives an illustration of the k -NN₁ technique for 7–9 July 1971, using T_8 and T_{20} values as variables in the state vector.

For the SinExp model, the h_r , h_s are inserted from astronomical formulas. Climatological values of the parameters h_{\max} , h_{\min} , and an initial value of γ were determined for each month from the 1971–2000 hourly data. Thereafter, the values of the parameters h_{\max} and h_{\min} were optimized for each month in terms of the RMSE of the hourly temperatures. For each day t , the values of T_{\min} and T_{\max} are determined by solving them from two versions of Equation (3a), in which the first one in the left-hand side of the observed value T_8 of day t has been inserted and in the second one the observed value T_{14} (see also Van Engelen and Geurts, 1983). After this, τ for each day t can be found from Equation (4) using the value of $T(h_s)$ that follows from the calculated values of T_{\min} and T_{\max} of day t with Equation (3a), and the minimum temperature T'_{\min} ,

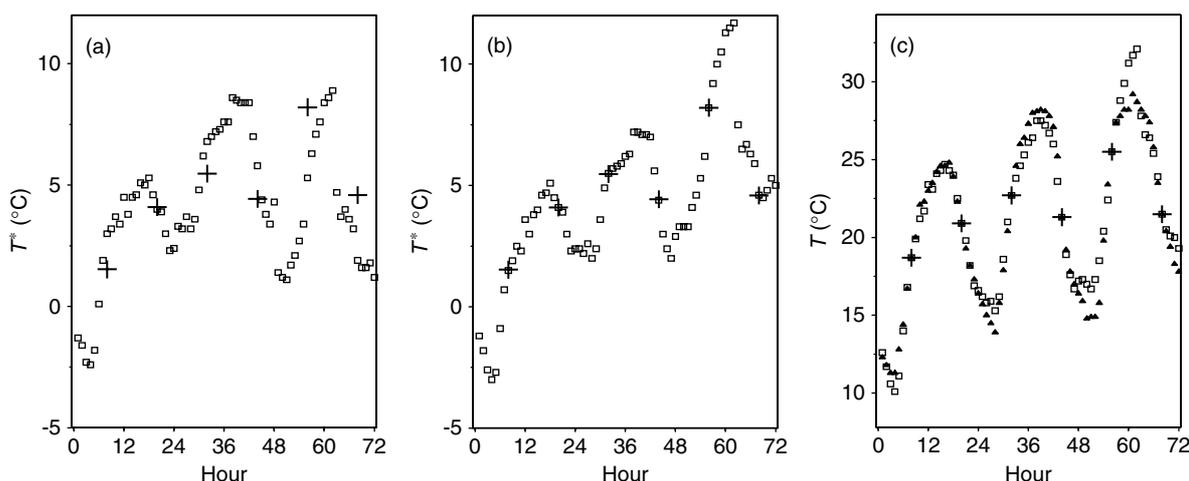


Figure 1. Illustration of the homogenization and interpolation by the nearest-neighbor resampling technique (k -NN₁) for 7–9 July 1971, using T_8 and T_{20} as variables in the state vector. Plot (a) gives the anomaly values T^* (with respect to climatology) of the three pairs of variables (T_8 and T_{20}) in the state vector (crosses) and the three uncorrected resampled diurnal cycles (open squares). In plot (b), the resampled diurnal cycles are linearly adjusted such that the T_8 and T_{20} values equal those of the variables in the state vector and such that jumps between successive days are removed. In plot (c), the resampled anomaly values (plot (b)) are backtransformed into temperature and compared with the measured hourly values (filled triangles). Note the difference in vertical scales

as calculated for day $t + 1$. With T_{\min} , T_{\max} , and τ of a given day known, and γ put on its climatological value, an initial solution to the whole diurnal cycle (Equation (3)) of that day is fixed. The T_{20} value is then used to adjust the diurnal cycle in its exponential part (Equation (3b)), which implies an adjustment of γ to the observations of day t .

For the application of ClimDay, no specific climatological parameters had to be calculated. Smoothed daily cycles are calculated for each calendar day from the observed hourly temperatures (1971–2000). The smooth approximations for successive calendar days are obtained with the supersmoother (Härdle, 1990).

In the LinReg model, Equation (7) is fitted for each month separately using least-squares regression for T_m , T_{\min} , and T_{\max} . The parameters, a_i , are thus determined for each of the 12 calendar months. We also fitted the model for a moving window, giving parameters for each of the 365 calendar days. However, the sensitivity of RMSE to window size is found to be small.

3. RESULTS

In this section, we compare the five models with respect to their ability to accurately calculate hourly (k -NN₁, k -NN₂ and SinExp) and daily (all models) temperatures from a few temperature readings per day. In Section 3.1, we first compare the models for the calibration 1971–2000 and verification period 1951–1970 for the following three combinations of input values: (a) T_{20} , (b) T_8 and T_{20} , (c) T_8 , T_{14} and T_{20} , where the index refers to GMT hour. Note that ClimDay and LinReg are not able to simulate complete hourly diurnal temperature cycles that are internally consistent. As a result, ClimDay is only used to provide estimates of T_m , while LinReg only provides independent estimates of T_m , T_{\min} , and T_{\max} . Note further that SinExp can only run on choice (c). As a consequence of these limitations, some of the cells in our tables are empty.

In Section 3.2, we compare the skill of the models when cloudiness is used as an extra predictor variable. Here SinExp is excluded, as no cloud-dependent version exists. In Section 3.3, we compare the models (without the addition of cloudiness) with respect to their ability to simulate extreme daily mean temperatures.

To compare the output of the models with the observed values, we calculated the bias (BIAS) and RMSE per month and averaged these values to obtain annual mean values. In addition, we considered a measure for the hour-to-hour variability (HHVAR) and day-to-day variability (DDVAR) defined as:

$$\text{HHVAR} = \frac{1}{N} \sum_{h=1}^N |T_h - T_{h-1}| \quad (8)$$

$$\text{DDVAR} = \frac{1}{M} \sum_{t=1}^M |T_{x,t} - T_{x,t-1}| \quad (9)$$

where h is the index for hour number and N is the total number of hours in the time series of hourly T values, t is the index for day number, and M is the total number of days in the daily time series. T_x in Equation (9) may refer to T_m , T_{\min} , or T_{\max} . In Tables I–VI, the differences between Δ_{HHVAR} and Δ_{DDVAR} with respect to climatology are averaged per month and presented as percentage differences (reconstructed minus observed, relative to the observed).

3.1. Comparison of the five models for the calibration and verification period

The results for the five models for the three defined sets of input variables: (a) T_{20} , (b) T_8 and T_{20} , and (c) T_8 , T_{14} , and T_{20} are presented in Tables I–III. It should be emphasized that the results for k -NN₁, here and in the rest of the paper, refer to one (randomly selected) realization only. Although other realizations may give other values for individual days, the influence on the statistics in the tables is negligible. We note that this feature naturally follows from the definition of the method. In Sections 3.1.1 and 3.1.2, we discuss the results for the calibration period and those for the verification period respectively.

Table I. Comparison of models with T_{20} as input for the calibration period 1971–2000 and the verification period 1951–1970 using the hourly temperature data of the De Bilt. The k -NN₁ and k -NN₂ stand for the k nearest-neighbor technique with, respectively, randomly sampling one of the k -nearest neighbors (nearest-neighbor resampling) and weighted averaging of all k -nearest neighbors (nearest-neighbor averaging); ClimDay stands for the climatological daily cycle model and LinReg stands for the multiple linear regression model. T_h , T_m , T_{\min} , and T_{\max} represent the hourly temperatures, daily mean, minimum, and maximum temperatures, respectively. Results are presented for the bias (BIAS), root mean squared errors (RMSE), hour-to-hour variability (Δ_{HHVAR}), and day-to-day variability (Δ_{DDVAR}), the latter two expressed as a percentage difference with respect to the observed record. The number of nearest neighbors k is chosen to be 20 and the window width equals 31 days

Model	$T_{h,(h=1,\dots,24)}$			T_m			T_{\min}			T_{\max}		
	BIAS (°C)	RMSE (°C)	Δ_{HHVAR} (%)	BIAS (°C)	RMSE (°C)	Δ_{DDVAR} (%)	BIAS (°C)	RMSE (°C)	Δ_{DDVAR} (%)	BIAS (°C)	RMSE (°C)	Δ_{DDVAR} (%)
Calibration 1971–2000												
k -NN ₁	0.002	2.122	−0.7	0.002	1.136	4.8	0.069	2.088	−2.5	−0.005	2.337	18.2
k -NN ₂	0.001	1.607	−26.7	0.001	0.846	−10.5	0.617	1.804	−25.2	−0.515	1.796	−19.4
ClimDay	–	–	–	0.000	1.340	40.1	–	–	–	–	–	–
LinReg	–	–	–	0.000	0.812	−11.8	0.000	1.718	−31.9	0.000	1.666	−28.9
Verification 1951–1970												
k -NN ₁	0.119	2.262	0.6	0.119	1.222	8.2	0.114	2.310	−3.2	0.066	2.415	22.0
k -NN ₂	0.109	1.695	−25.6	0.109	0.893	−9.3	0.749	1.960	−29.5	−0.466	1.859	−17.5
ClimDay	–	–	–	0.109	1.354	37.7	–	–	–	–	–	–
LinReg	–	–	–	0.117	0.843	−11.8	0.136	1.813	−31.1	0.020	1.723	−28.8

Table II. As in Table I but now with T_8 and T_{20} as input

Model	$T_{h,(h=1,\dots,24)}$			T_m			T_{\min}			T_{\max}		
	BIAS (°C)	RMSE (°C)	Δ_{HHVAR} (%)	BIAS (°C)	RMSE (°C)	Δ_{DDVAR} (%)	BIAS (°C)	RMSE (°C)	Δ_{DDVAR} (%)	BIAS (°C)	RMSE (°C)	Δ_{DDVAR} (%)
Calibration 1971–2000												
k -NN ₁	−0.011	1.698	−1.8	−0.011	0.696	3.8	0.075	1.783	0.2	−0.077	2.003	16.5
k -NN ₂	−0.009	1.304	−25.2	−0.009	0.528	−2.6	0.499	1.508	−18.3	−0.508	1.605	−11.0
ClimDay	–	–	–	−0.002	0.621	10.8	–	–	–	–	–	–
LinReg	–	–	–	0.000	0.518	−4.1	0.000	1.574	−28.3	0.000	1.542	−20.2
Verification 1951–1970												
k -NN ₁	−0.033	1.816	−0.9	−0.033	0.755	3.9	0.076	1.887	−1.5	−0.164	2.196	18.8
k -NN ₂	−0.033	1.390	−24.5	−0.033	0.574	−2.5	0.498	1.563	−20.9	−0.619	1.771	−9.8
ClimDay	–	–	–	−0.035	0.652	9.3	–	–	–	–	–	–
LinReg	–	–	–	0.001	0.557	−4.1	0.100	1.641	−27.6	−0.132	1.658	−20.2

3.1.1. *Calibration period (1971–2000).* As expected, the output statistics for all models improve when we increase the number of predictors. Consider, e.g. the statistics of the hourly temperatures in the first three columns of Tables I–III. For each addition of an observation hour, there is a reduction in RMSE for both the k -NN₁ and k -NN₂ methods. RMSE decreases about linearly with the number of observation hours. For the daily data (T_m , T_{\min} , T_{\max}) the reduction in RMSE is largest for T_{\max} when T_{14} is added as a predictor. The obvious reason is that T_{14} is generally close to the maximum temperature.

For the reproduction of hourly temperatures, k -NN₂ performs better than k -NN₁ and SinExp (only in Table III) in terms of RMSE. BIAS of all three models is close to zero ($<0.04^\circ\text{C}$). The k -NN₁ model

Table III. As in Table I but now with T_8 , T_{14} , and T_{20} as input. The table includes the results of the sine–exponential model (SinExp), which needs three sub-daily temperature values as input

Model	$T_{h,(h=1,\dots,24)}$			T_m			T_{\min}			T_{\max}		
	BIAS (°C)	RMSE (°C)	Δ_{HHVAR} (%)	BIAS (°C)	RMSE (°C)	Δ_{DDVAR} (%)	BIAS (°C)	RMSE (°C)	Δ_{DDVAR} (%)	BIAS (°C)	RMSE (°C)	Δ_{DDVAR} (%)
Calibration 1971–2000												
k -NN ₁	0.002	1.257	−1.7	0.002	0.495	0.5	0.089	1.658	−0.8	−0.028	0.771	2.3
k -NN ₂	0.000	0.972	−22.7	0.000	0.386	−2.2	0.476	1.403	−16.5	−0.348	0.701	1.1
SinExp	−0.036	1.168	−30.5	−0.036	0.457	4.1	0.097	1.648	3.3	−0.418	0.778	2.9
ClimDay	–	–	–	0.000	0.593	12.3	–	–	–	–	–	–
LinReg	–	–	–	0.000	0.381	−4.0	0.000	1.495	−23.0	0.000	0.760	−5.2
Verification 1951–1970												
k -NN ₁	−0.008	1.333	1.4	−0.008	0.522	0.3	0.011	1.798	−1.1	−0.036	0.798	2.2
k -NN ₂	−0.006	1.026	−19.9	−0.006	0.403	−2.9	0.420	1.452	−18.8	−0.355	0.731	1.7
SinExp	−0.062	1.216	−24.3	−0.062	0.479	4.1	−0.077	1.686	3.5	−0.430	0.795	2.9
ClimDay	–	–	–	0.006	0.577	10.6	–	–	–	–	–	–
LinReg	–	–	–	0.021	0.400	−4.0	0.061	1.564	−22.4	−0.040	0.804	−5.2

Table IV. As in Table I but with cloudiness added as predictor/state variable

Model	$T_{h,(h=1,\dots,24)}$			T_m			T_{\min}			T_{\max}		
	BIAS (°C)	RMSE (°C)	Δ_{HHVAR} (%)	BIAS (°C)	RMSE (°C)	Δ_{DDVAR} (%)	BIAS (°C)	RMSE (°C)	Δ_{DDVAR} (%)	BIAS (°C)	RMSE (°C)	Δ_{DDVAR} (%)
Calibration 1971–2000												
k -NN ₁	−0.009	1.999	−2.2	−0.009	1.107	4.4	0.073	2.001	−0.2	−0.079	2.056	10.0
k -NN ₂	−0.007	1.491	−27.2	−0.007	0.822	−9.4	0.602	1.673	−20.8	−0.549	1.607	−18.5
ClimDay	–	–	–	−0.001	1.336	39.4	–	–	–	–	–	–
LinReg	–	–	–	0.000	0.796	−11.6	0.000	1.496	−19.4	0.000	1.393	−20.0
Verification 1951–1970												
k -NN ₁	0.107	2.051	−1.3	0.107	1.133	3.4	0.207	2.078	−5.6	−0.012	2.122	12.2
k -NN ₂	0.104	1.558	−26.6	0.104	0.861	−10.7	0.742	1.804	−27.3	−0.502	1.652	−18.2
ClimDay	–	–	–	0.110	1.350	36.5	–	–	–	–	–	–
LinReg	–	–	–	0.119	0.841	−11.6	0.114	1.580	−19.0	0.038	1.478	−19.9

is clearly superior in reproducing Δ_{HHVAR} . This is because this model resamples complete diurnal cycles without averaging. Figure 2 presents the monthly values of the output statistics for the hourly temperatures of k -NN₁, k -NN₂, and SinExp for the calibration period in Table III (input: T_8 , T_{14} and T_{20}). The figure shows large differences between the three methods. Note for instance, the annual cycle in BIAS for SinExp, which is absent for the k -NN methods.

The results for T_m show that LinReg always has the lowest RMSE, closely followed by k -NN₂. The RMSE of ClimDay is largest, which is probably because this method does not maximize the regression coefficient as in LinReg. The fact that ClimDay does not use information of the previous or next day may also play a role. Δ_{DDVAR} of k -NN₁ is closest to zero. BIAS is close to zero for all models ($<0.01^\circ\text{C}$). Figure 3 presents the monthly values of the output statistics of T_m for all five models for the calibration period in Table III (input: T_8 , T_{14} , and T_{20}). It illustrates, among other things, that k -NN₁ performs best with respect to reproducing the annual cycle of DDVAR (ranging between 1.9°C in January and 1.3°C in September) and that k -NN₂ and LinReg have about the same annual cycle for the RMSE. The shape of the annual cycle in the RMSE (large values in summer and small values in winter) is a result of the diurnal temperature range, which is small in

Table V. As in Table II but with cloudiness added as predictor/state variable

Model	$T_{h,(h=1,\dots,24)}$			T_m			T_{\min}			T_{\max}		
	BIAS (°C)	RMSE (°C)	Δ_{HHVAR} (%)	BIAS (°C)	RMSE (°C)	Δ_{DDVAR} (%)	BIAS (°C)	RMSE (°C)	Δ_{DDVAR} (%)	BIAS (°C)	RMSE (°C)	Δ_{DDVAR} (%)
Calibration 1971–2000												
$k\text{-NN}_1$	-0.013	1.520	-3.6	-0.013	0.651	2.0	0.126	1.562	0.0	-0.136	1.770	9.7
$k\text{-NN}_2$	-0.025	1.147	-26.5	-0.025	0.489	-3.0	0.489	1.334	-13.7	-0.580	1.428	-11.8
ClimDay	-	-	-	-0.002	0.592	9.1	-	-	-	-	-	-
LinReg	-	-	-	0.000	0.481	-4.6	0.000	1.366	-15.6	0.000	1.248	-14.3
Verification 1951–1970												
$k\text{-NN}_1$	-0.038	1.594	-3.0	-0.038	0.708	3.1	0.099	1.636	-2.5	-0.219	1.915	13.4
$k\text{-NN}_2$	-0.045	1.210	-25.5	-0.045	0.529	-2.6	0.463	1.365	-15.0	-0.667	1.548	-10.5
ClimDay	-	-	-	-0.034	0.617	7.6	-	-	-	-	-	-
LinReg	-	-	-	-0.004	0.514	-4.6	0.044	1.425	-15.3	-0.105	1.367	-14.2

Table VI. As in Table III but with cloudiness added as predictor/state variable

Model	$T_{h,(h=1,\dots,24)}$			T_m			T_{\min}			T_{\max}		
	BIAS (°C)	RMSE (°C)	Δ_{HHVAR} (%)	BIAS (°C)	RMSE (°C)	Δ_{DDVAR} (%)	BIAS (°C)	RMSE (°C)	Δ_{DDVAR} (%)	BIAS (°C)	RMSE (°C)	Δ_{DDVAR} (%)
Calibration 1971–2000												
$k\text{-NN}_1$	0.015	1.211	-3.1	0.015	0.479	1.0	0.139	1.575	0.3	-0.047	0.763	1.0
$k\text{-NN}_2$	0.017	0.925	-23.5	0.017	0.368	-1.3	0.517	1.337	-14.4	-0.352	0.694	0.6
ClimDay	-	-	-	0.002	0.522	12.1	-	-	-	-	-	-
LinReg	-	-	-	0.000	0.356	-2.8	0.000	1.358	-15.6	0.000	0.718	-5.6
Verification 1951–1970												
$k\text{-NN}_1$	0.014	1.255	-0.7	0.014	0.492	0.4	0.085	1.660	-0.7	-0.051	0.798	2.0
$k\text{-NN}_2$	0.014	0.967	-21.0	0.014	0.386	-2.2	0.469	1.365	-15.3	-0.361	0.721	1.2
ClimDay	-	-	-	0.003	0.512	10.6	-	-	-	-	-	-
LinReg	-	-	-	0.014	0.375	-2.8	0.033	1.420	-15.2	-0.046	0.777	-5.6

Table VII. BIAS in the annual number of days with daily mean temperatures $T_m < 5\text{th percentile}$ (left) and $T_m > 95\text{th percentile}$ (right) for all five models for the 1951–2000 period for three combinations of input variables (a) T_{20} ; (b) T_8 , T_{20} ; and (c) T_8 , T_{14} , T_{20} . The values in brackets give the corresponding standard errors

	BIAS (annual number of days $T_m < 5\text{th percentile}$)			BIAS (annual number of days $T_m > 95\text{th percentile}$)		
	T_{20}	T_8, T_{20}	T_8, T_{14}, T_{20}	T_{20}	T_8, T_{20}	T_8, T_{14}, T_{20}
$k\text{-NN}_1$	-0.68 (0.52)	-0.50 (0.49)	0.04 (0.37)	0.88 (0.47)	-0.20 (0.37)	-0.48 (0.35)
$k\text{-NN}_2$	-3.00 (0.42)	-1.52 (0.36)	-0.72 (0.36)	-1.22 (0.40)	-1.30 (0.35)	-0.68 (0.30)
SinExp	-	-	0.02 (0.40)	-	-	-1.00 (0.29)
ClimDay	6.38 (0.55)	0.22 (0.43)	3.20 (0.42)	7.12 (0.56)	0.12 (0.39)	2.28 (0.42)
LinReg	-3.04 (0.47)	-1.56 (0.38)	-2.14 (0.38)	-1.56 (0.46)	-1.82 (0.33)	-1.06 (0.30)

winter and large in summer. A favorable feature of $k\text{-NN}_1$ is that the dependence of Δ_{DDVAR} on the number of predictors is weakest among the presented methods.

For T_{\min} and T_{\max} , there is a large difference between the results. With respect to the BIAS, the $k\text{-NN}_1$ performs better than the other models (the BIAS of LinReg is zero by definition), while for the RMSE, both

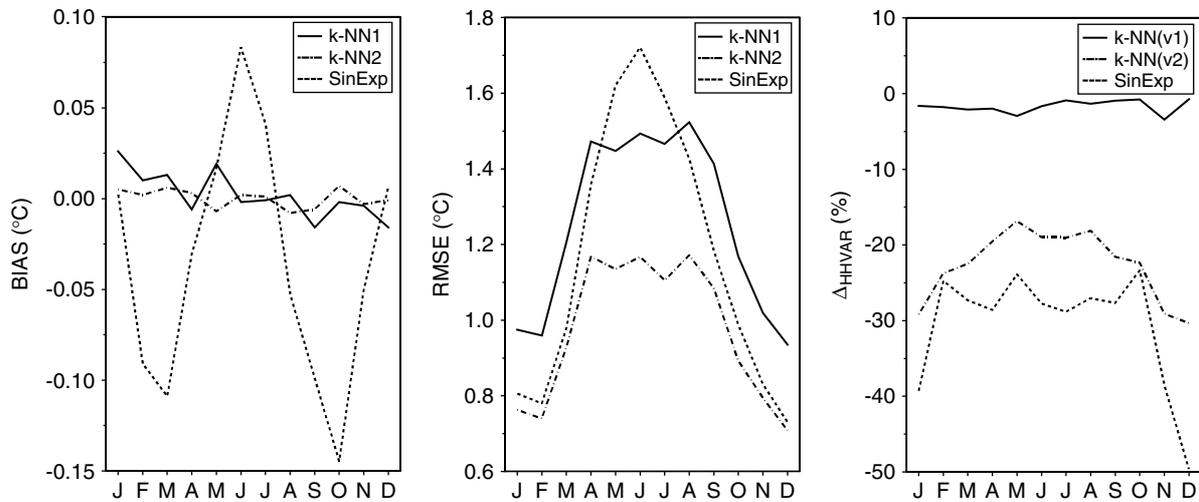


Figure 2. Comparison of monthly output statistics for the hourly data produced by the three models that yield complete diurnal temperature cycles using T_8 , T_{14} , and T_{20} as input for the 1971–2000 calibration period. See the caption of Table I for an explanation of the abbreviations

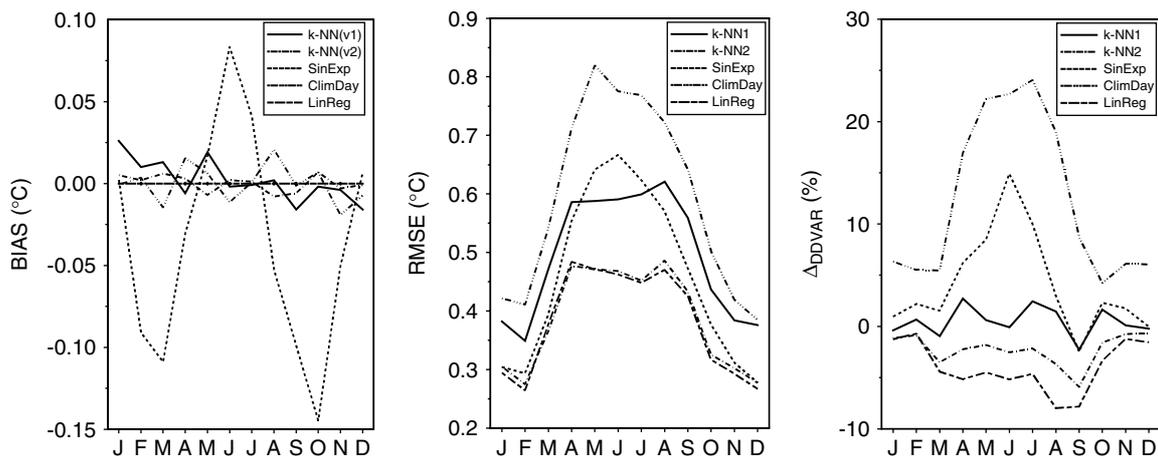


Figure 3. Comparison of monthly output statistics for the daily mean temperature data calculated by all five models using T_8 , T_{14} , and T_{20} as input for the 1971–2000 calibration period. See the caption of Table I for an explanation of the abbreviations

k -NN₂ and LinReg perform well. The k -NN₁ method is most suitable to reproduce the DDVAR, especially for T_{\min} . It is somewhat surprising that Δ_{DDVAR} for k -NN₁ has values as large as 18.2% (T_{\max} in Table I) and 16.5% (T_{\max} in Table II) for the T_{20} and T_8 , T_{20} sets of input variables. Again, the variation of Δ_{DDVAR} with the number of predictors is smallest for k -NN₁.

3.1.2. *Verification period (1951–1970).* Besides results for the calibration period, Tables I–III also contain results for the verification period (1951–1970). Large differences between the statistics for the calibration and verification period would indicate that models are over-fitted. For all cases, the differences between the statistics for the two periods are sufficiently small.

3.1.3. *Conclusions of the comparisons.* Summarizing, for the reproduction of hourly data (k -NN₁, k -NN₂ and SinExp), k -NN₂ is preferred when the objective is to obtain a good estimation of the individual hourly

temperatures, whereas k -NN₁ is preferred when it is important that the variance of the hourly temperatures is maintained. For the daily data (all models), a similar conclusion can be drawn: the best choice for estimating individual values is k -NN₂ followed by LinReg; k -NN₁ is the best if variance should be maintained. SinExp and ClimDay, being the models that lean most heavily on climatology, perform less on both aspects.

3.2. Cloudiness as an extra predictor variable

The k -NN, ClimDay, and LinReg models allow for incorporating the effect of including, besides temperature, cloudiness as predictor. The results for both the calibration and verification period are presented in the following text.

3.2.1. Calibration period (1971–2000). As in Section 3.1, Tables IV–VI present the results for the input variables: (a) T_{20} , (b) T_8 and T_{20} , (c) T_8 , T_{14} , and T_{20} , but now with the addition of cloud. The most distinct feature in comparison with Tables I–III is that RMSE is always lower when cloud is included (for LinReg with least-squares estimation this is automatically true), ranging from a few percent lower to about 20% lower (compare RMSE of LinReg for T_{\max} in Tables V and II). The inclusion of cloud also has a positive effect on the reproduction of HHVAR and DDVAR but this is less clear than for the reduction in RMSE. The largest improvement is found for the DDVAR of T_{\max} for k -NN₁ and LinReg in Tables IV and V, ranging between 5.9 and 8.9%. The difference between the calculations with and without clouds are, however, much smaller than the mutual differences between the models. There is a tendency that BIAS becomes somewhat larger when cloudiness is included, but overall, the effect is marginal.

For the two versions of the k -NN model, we also studied the inclusion of wind direction and speed and relative humidity at the observation hours (not shown). Although RMSE, Δ_{HHVAR} and Δ_{DDVAR} further reduce, the reduction is small even when compared to the reduction introduced by the inclusion of cloudiness alone. Furthermore, the inclusion of more variables causes BIAS to become slightly larger.

3.2.2. Verification period (1951–1970). As in Tables I–III, the differences between the statistics for the calibration and verification period in Tables IV–VI are small. In contrast to the situation without cloud, the differences for the BIAS and RMSE of T_{\max} for LinReg in Table VI are also small.

3.2.3. Conclusions of the comparisons. In conclusion, the inclusion of clouds as predictor in the models improves the fit of the models in terms of RMSE, HHVAR, and DDVAR. However, in general, this improvement is marginal and small compared to mutual differences between the models.

3.3. Calculation of extreme daily mean temperatures

An important test of the models is their ability to calculate extreme temperatures. Here we calculated the annual number of days with daily mean temperatures $T_m < 5\text{th percentile}$ and $T_m > 95\text{th percentile}$ and compared them with the observed values. The percentile values were calculated for each calendar day for the 1971–2000 period and thereafter were smoothed with the supersmoother. Table VII presents the BIAS in the annual number of extreme temperatures for the 1951–2000 period. The table shows that the BIAS for k -NN₁ is small and differs not more than two times the standard error from zero. For all other models, the BIAS is mostly larger than for k -NN₁ and is generally significantly different from zero. In all cases, the BIAS for k -NN₂ is smaller than that for LinReg. ClimDay shows the strongest variation of the BIAS with the number of predictors, with a BIAS close to zero for the (T_8 , T_{20}) input variables. Note that SinExp performs surprisingly well for $T_m < 5\text{th percentile}$.

In conclusion, k -NN₁ performs well in reproducing the 5th or 95th percentiles of T_m , and its results vary only slightly with the number of predictors. This implies that k -NN₁ is best capable of removing inhomogeneities in these percentiles introduced by changing observation schedules and/or irregular observation frequencies.

4. DISCUSSION AND CONCLUSIONS

We introduced nearest-neighbor resampling as a new technique for homogenizing temperature records on a daily to sub-daily level. We compared the technique with other methods for the problem of calculating daily and sub-daily temperature values from temperature time series subject to irregular observation frequencies and changing observation schedules. A striking feature of the nearest-neighbor resampling technique (k -NN₁) is that it provides nonunique solutions, which is uncommon in the literature on homogenization of temperature time series. Because the method resamples complete diurnal temperature cycles, the variance is preserved in a natural way, which is important in the analysis of trends and variability of extremes. Another advantage of k -NN₁ is its low BIAS for T_{\min} and T_{\max} . The merits of k -NN₁ in homogenizing series with preservation of variance is apparent from the small bias, regardless of the number of predictors, in the 5th and 95th percentiles of the daily temperatures, which demonstrates its power in trend analysis of extremes.

Because of its nondeterministic nature, the k -NN₁ method is not suitable for all problems, e.g. when the interest is in the exact dates of temperature records. In that case, k -NN₁ will not be able to provide single-valued dates because the results depend on the random seed used. The nearest-neighbor averaging technique (k -NN₂) or the LinReg model are then more appropriate. Note that for most test statistics, SinExp and ClimDay produce results inferior to k -NN₁, k -NN₂ and LinReg. In this context, it should be underlined that the k -NN techniques do not rely on the presence of a climatological diurnal cycle; on the contrary, they should work equally well in reproducing T_m for records with no such cycle, e.g. for midwinter polar stations.

We have tested the nearest-neighbor techniques to the temperature of De Bilt in the Netherlands in the current paper. There are no reasons to expect that the technique works less successfully for other stations. Like other homogenization or interpolation methods, the nearest-neighbor technique relies on the assumption that the shape of the diurnal temperature cycle does not change significantly between the calibration period and the period of interest. The importance of this assumption increases with the decreasing number of temperature observations per day.

The main advantage of the nearest-neighbor resampling compared to other methods is that the technique maintains the variance in a natural way. Other techniques are proposed to artificially 'inflate' the variance (Von Storch, 1999; Huth, 2002). We checked one approach for the LinReg model, in which we 'inflated' the variance by means of adding randomized daily errors for each month separately. Although Δ_{DDVAR} becomes smaller, it is still larger than Δ_{DDVAR} of k -NN₁. However, RMSE increases and becomes larger than that of k -NN₁. Improvement may be obtained by using more sophisticated procedures for the addition of errors. If the sophistication is pushed to a high level, the LinReg starts to resemble the k -NN₁ so much that its benefits with respect to the latter become unclear. And even then, an important drawback of the LinReg model remains, namely, that T_{\min} and T_{\max} are independently estimated from each other and from T_m , whereas the estimates of k -NN₁ (but also k -NN₂) are internally consistent.

An obvious limitation of the nearest-neighbor resampling technique (which it shares with other methods) is that it requires that a subperiod of hourly observations is available in the record. The required length of this high-frequency observational record is, however, shorter than intuitively thought. In addition to the 1971–2000 sampling period, we experimented with various other sampling periods with lengths of 10, 5, and 2 years (1971–1980, 1981–1990, 1991–2000, 1971–1975 and 1971–1972). We compared the statistics for the verification period with those in Tables I–III and found that the results only slowly deteriorate with decreasing length of the sampling period. Even the 2-year sampling period seems to be of sufficient length to reproduce the 20-year T_m record without any perceptual loss in skill. A related phenomenon can be observed in applications of nearest-neighbor resampling for simulation of daily rainfall, where 1000-year series with realistic extreme-value properties can successfully be generated from observational series of only 30-year length (see, e.g. Brandsma and Buishand, 1998; Buishand and Brandsma, 2001). This result has important consequences. First, it widens the potentials of the method significantly; second, it provides an important argument to augment existing long once-, twice-, or thrice-daily observational records with an hourly extension of a couple of years on the same spot. We stress that, because of the construction of the method and its application to anomalies, a series of a nearby station of hourly observations may also be suited to do the job.

We anticipate that nearest-neighbor resampling can be applied successfully to other homogenization problems and elements other than temperature as well. Consider e.g. a situation where there is a break in a time series due to station relocation. In that case, weather-dependent corrections may be resampled from an overlapping interval or, if that does not exist, from an overlap with a nearby third station. Thereafter, these corrections can be added to the series forwards or backwards. Another application, e.g. in data preprocessing for trend studies of spatial patterns of extremes of daily temperature percentiles (see, e.g. Klein Tank *et al.*, 2005), is the estimation of daily means from daily minimum and maximum temperatures. If for part of the stations the daily temperatures are not directly available but have to be estimated from the daily minimum and maximum temperatures instead, nearest-neighbor resampling may be invoked to obtain homogeneity between the various stations' series.

An advantage of nearest-neighbor resampling to some other techniques is that extra weather variables can easily be included in the state vector. Although the added value of the inclusion of extra weather variables was not obvious for the examples presented in this paper, it may be different for other applications.

In conclusion, in this paper we have showed that nearest-neighbor resampling may be a promising alternative to existing homogenization methods, especially where the interest is changes in extreme weather events and subannual properties. The method has the capability of deriving weather-dependent corrections and the potential to be applied to a wide range of problems.

ACKNOWLEDGEMENTS

We are grateful to our colleague T. A. Buishand and to an anonymous referee for their constructive comments.

REFERENCES

- Baker DG. 1975. Effect of observation time on mean temperature estimation. *Journal of Applied Meteorology* **14**: 471–476.
- Bergström H, Moberg A. 2002. Daily air temperature and pressure series for upsala (1722–1998). *Climatic Change* **53**: 213–252.
- Brandsma T, Buishand TA. 1998. Simulation of extreme precipitation in the rhine basin by nearest-neighbour resampling. *Hydrology and Earth System Sciences* **2**: 195–209.
- Buishand TA, Brandsma T. 2001. Multi-site simulation of daily precipitation and temperature in the rhine basin by nearest-neighbor resampling. *Water Resources Research* **37**: 2761–2776.
- Härdle W. 1990. *Applied Nonparametric Regression*. Cambridge University Press: Cambridge, 333.
- Harold TI, Sharma A, Sheater SJ. 2003. A nonparametric model for stochastic generation of rainfall occurrence. *Water Resources Research* **39**: SWC10–1–SWC10–11.
- Huth R. 2002. Statistical downscaling of daily temperature in Central Europe. *Journal of Climate* **15**: 1731–1742.
- Karl TR, Williams CN, Young PJ, Wendland WM. 1986. A model to estimate the time of observation bias associated with monthly mean maximum, minimum and mean temperatures for the United States. *Journal of Climate and Applied Meteorology* **25**: 145–160.
- Kendall M, Stuart A, Ord JK. 1983. *The Advanced Theory of Statistics, Vol.3: Design and Analysis, and Time Series*. Charles Griffin: London.
- Klein Tank AMG, Können GP. 2003. Trends in indices of daily temperature and precipitation extremes in Europe, 1946–1999. *Journal of Climate* **16**: 3665–3680.
- Klein Tank AMG, Wijngaard JB, Können GP, Böhm R, Demarée G, Gocheva A, Mileta M, Pashiardis S, Hejkrlik L, Kern-Hansen C, Heino R, Bessemoulin P, Müller-Westermeier G, Tzanakou M, Szalai S, Pálsdóttir T, Fitzgerald D, Rubin S, Capaldo M, Maugeri M, Leitass A, Bukantis A, Aberfeld R, van Engelen AFV, Forland E, Miletus M, Coelho F, Mares C, Razuvaev V, Nieplova E, Cegnar T, Antonio López J, Dahlström B, Moberg A, Kirchhofer W, Ceylan A, Pachaliuk O, Alexander LV, Petrovic P. 2002. Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *International Journal of Climatology* **22**: 1441–1453.
- Klein Tank AMG, Können GP, Selten FM. 2005. Signals of anthropogenic influence on European warming as seen in the trend patterns of day-to-day temperature variance. *International Journal of Climatology* **25**: 1–16.
- Können GP, Jones PD, Kaltofen MH, Allan RJ. 1998. Pre-1866 extensions of the Southern oscillation index using early Indonesian and Tahitian meteorological readings. *Journal of Climate* **11**: 2325–2339.
- Lall U, Sharma A. 1996. A nearest neighbor bootstrap for resampling hydrologic time series. *Water Resources Research* **32**: 679–693.
- Lavery BM, Joung G, Nicholls N. 1997. An extended high-quality historical rainfall dataset for Australia. *Australian Meteorological Magazine* **46**: 27–38.
- Mitchell JM. 1958. Effect of changing observation time on mean temperature. *Bulletin American Meteorological Society* **39**: 83–89.
- Moberg A, Bergström H, Krigsman JR, Svanered O. 2002. Daily air temperature and pressure series for Stockholm (1756–1998). *Climatic Change* **53**: 171–212.
- Parton WJ, Logan JA. 1981. A model for diurnal variation in soil and air temperature. *Agricultural Meteorology* **23**: 205–216.
- Peterson T, Daan H, Jones P. 1997. Initial selection of a GCOS surface network. *Bulletin American Meteorological Society* **78**: 2145–2152.
- Rajagopalan B, Lall U. 1999. A *k*-nearest-neighbor simulator for daily precipitation and other weather variables. *Water Resources Research* **35**: 3089–3101.

- Schaal LA, Dale RF. 1977. Time of observation temperature bias and "climatic change". *Journal of Applied Meteorology* **16**: 215–222.
- Van Engelen AFVvan, Geurts HAM. 1983. *Historische Weerkundige Waarnemingen, Deel III: Een Rekenmodel Dat Het Verloop van de Temperatuur Over Een Eemaal Berekent Uit Drie Termijnmetingen Van Temperatuur*. KNMI: De Bilt, 44, (in Dutch).
- Von Storch H. 1999. On the use of "inflation" in statistical downscaling. *Journal of Climate* **12**: 3505–3506.
- Vincent LA, Zhang X, Bonsal BR, Hogg WD. 2000. Homogenized daily temperatures for trend analysis in extremes over Canada. In *Preprints, 12th Conference on Applied Climatology*. American Meteorological Society: Asheville, NC, 86–89.
- Wann M, Yen D, Gold HJ. 1985. Evaluation and calibration of three models for daily cycle of air temperature. *Agricultural and Forest Meteorology* **34**: 121–128.
- Wijngaard JB, Klein Tank AMG, Können GP. 2003. Homogeneity of 20th century European daily temperature and precipitation series. *International Journal of Climatology* **23**: 679–692.
- Young KC. 1994. A multivariate chain model for simulating climatic parameters from daily data. *Journal of Applied Meteorology* **33**: 661–671.