

Detecting and correcting sensor drifts in long-term weather data

Georg von Arx · Matthias Dobbertin ·
Martine Rebetez

Received: 27 March 2012 / Accepted: 9 August 2012 / Published online: 23 August 2012
© Springer Science+Business Media B.V. 2012

Abstract Quality control of long-term monitoring data of thousands and millions of individual records as present in meteorological data is cumbersome. In such data series, sensor drifts, stalled values, and scale shifts may occur and potentially result in flawed conclusions if not noticed and handled properly. However, there is no established standard procedure to perform quality control of high-frequency meteorological data. In this paper, we outline a procedure to remove sensor drift in high-frequency data series using the example of 15-year-long sets of hourly relative humidity (RH) data from 28 stations subdivided into 202 individual sensor operation periods. The procedure involves basic quality control, relative homogeneity testing, and drift removal. Significant sensor drifts were observed in 40.6 % of all sensor operation periods. The drifts varied between data series and depended in a complex, usually inconsistent way on absolute RH values; within single series for instance, a drift could be negative in the lower RH range and positive in the upper RH range. Detrending changed RH values by, on average, 1.96 %. For one fifth of the detrended data, adjustments were

2.75 % and more of the measured value, and in one tenth 4.75 % and more. Overall, drifts were strongest for RH values close to 100 %. The detrending procedure proved to effectively remove sensor drifts. The principles of the procedure also apply to other meteorological parameters and more generally to any time series of data for which comparable reference data are available.

Keywords Detrending · Homogeneity testing · Quality control · Relative humidity · Sensor drift · Swiss Long-Term Forest Ecosystem Research Programme LWF

Introduction

Monitoring the changes of target parameters over time is one of the most common method of data collection in environmental research, for instance when tracking growth of plants, concentrations of pollutants, and the course of weather conditions. Among data obtained from monitoring, meteorological data are particularly widely used, be it to investigate climate change, or, even more frequently, to serve as an influencing factor when studying environmental, ecological, and biological processes. In ecological research, meteorological data from the closest official weather stations are most often used for this purpose. When micro-climatic information that more accurately describes site-specific conditions is required, the use of in situ weather stations is common practice. However, most publications using in situ

G. von Arx (✉) · M. Dobbertin · M. Rebetez
Swiss Federal Institute for Forest, Snow and Landscape
Research WSL,
Birmensdorf, Switzerland
e-mail: georg.vonarx@wsl.ch

M. Rebetez
University of Neuchatel, Institute of Geography,
Neuchatel, Switzerland

weather data do not report whether and how quality control was performed; yet, improper quality control may impair or even invalidate results.

Meteorological sensors are usually exposed for long periods to outdoor conditions. The associated rigorous demands challenge the instruments and may result in systematic and stochastic inhomogeneities of the measured data, such as shifts, stalled values, and various kinds of drifts (Begert et al. 2005; Conrad and Pollack 1950; Trewin and Trevitt 1996). Particularly when dealing with monthly or daily time aggregates, there are established approaches for quality control (see e.g., Aguilar et al. 2003; Costa and Soares 2009; Begert et al. 2005, 2008) that official data providers usually employ. Where no redundant or overlapping measurements are available, the approaches for quality control are usually based on comparisons to data from reference stations and employing the following four basic steps (Aguilar et al. 2003): (1) meta-data analysis and basic quality control, (2) creation of reference time series, (3) breakpoint detection, and (4) data adjustment. For sub-daily meteorological data, there are no official recommendations of how to assess and improve data quality (Aguilar et al. 2003), although some methods have been suggested to control quality of daily minima and maxima (Trewin and Trevitt 1996; Della-Marta and Wanner 2006; Brandsma and Konnen 2006). However, some ecological studies demand a sub-daily time resolution, e.g., when investigating diurnal air temperature and relative humidity patterns or daytime vs. nighttime conditions (von Arx et al. 2012).

In this paper, we outline a procedure for quality control and data adjustment of hourly weather data. We elaborate the case with 28 long-term sets of relative humidity data. Therefore, we first introduce some routines for an initial basic quality control. We then explain the test for relative homogeneity to detect data trends over time (“sensor drift”), followed by a value-dependent detrending procedure and subsequent relative homogeneity testing. Finally, we evaluate the effect of the performed data processing on data quality and briefly discuss some implications for ecological research.

Materials and methods

Data sets

The relative humidity (RH) data processed in this study were collected within the Long-Term Forest

Ecosystem Research Programme LWF in Switzerland (Innes 1995). At 14 distinct sites in Switzerland, automated meteorological stations were established one each within and outside of the forest stand between 1996 and 2000. The stations have since been recording, among others, hourly RH data (see von Arx et al. 2012 for a more detailed description). Among others, the stations were equipped with combined RH and temperature sensors of the unventilated model MP100A (Rotronic AG, Switzerland) that were placed 2 m above ground. After every 2 to 3 years, corresponding here to a “sensor operation period,” each sensor was replaced by a spare sensor that was previously calibrated by the Rotronic company (uncertainty $\leq \pm 1.5\%$). With this maintenance scheme, the data series from the 28 stations were thus subdivided in a total of 202 individual sensor operation periods covering the period 1996–2011. Altogether, over 3.3 million individual data points were recorded and automatically transferred into an Oracle data base (Jakob et al. 2007).

Reference series for each of the 28 LWF RH data series (hereafter referred to as “candidate series”) were created by averaging hourly data from two MeteoSwiss stations. MeteoSwiss uses ventilated thermohygrometers of the type Thygan VTP 6. The much more expensive and energy-consuming ventilated Thygan sensors are expected to produce better quality data. In order to assure that the reference stations showed climatological characteristics comparable to the candidate stations (Begert et al. 2005), reference series were selected based on the greatest correlation (Peterson and Easterling 1994), which usually, but not always, was obtained when considering the two closest MeteoSwiss stations. Pearson’s correlation coefficients between candidate and reference series were, on average, $r=0.805$ and ranged from $r=0.941$ to $r=0.658$.

Basic quality control

The aim of basic quality control was to (conservatively) eliminate the most obvious erroneous data points. A couple of test routines written in Visual Basic for Application (VBA) were used to detect data points that were suspicious. In each case, it was decided individually whether to keep or delete the value(s). Among the test routines were such to detect outliers (values >110 and $<0\%$), abnormally low variability (six and more

identical values in a row), excessive variability (jumps of 30 % and more from one hourly record to the next), and extraordinarily high-spatial inconsistency (more than 50 % difference of concomitant values between within- and outside-stand station). In a few instances, these test routines revealed entire periods with suspicious data that were, if obviously erroneous, eliminated.

Homogeneity testing

A relative homogeneity test was designed to mainly detect a linear drift over time, which is one of the most frequent inhomogeneity after scale shifts (Costa and Soares 2009). The test is related to the frequency distribution matching method used by Trewin and Trevitt (1996) and assumed that, although the frequency of RH values within different RH bins may systematically differ between candidate and reference stations, the difference would remain constant over time, if there was no inhomogeneity. It further assumed that records at the beginning of each of the 202 sensor operation periods, when a newly calibrated sensor was installed, were correct and only thereafter may have deviated. Breakpoints (sudden shifts in time series) were assumed to mainly occur when a sensor was replaced, but not within a 2- to 3-year lasting sensor operation period.

In detail, the procedure included the following steps: first, the number of hourly records above consecutive 10 % bins (i.e., $\geq 90\%$ RH/ $\geq 80\%$ RH/ $\geq 70\%$ RH/ $\geq 20\%$ RH/ $< 20\%$) was counted for each month within a sensor operation period and then expressed as a percentage of the total monthly count to adjust for varying numbers of missing data points. Second, for each month and bin, the difference of the relative frequency between the candidate and reference series was calculated. Finally, linear regression was performed between the relative frequency differences of each bin and elapsed months of the corresponding sensor operation period. A significant drift within the respective bin was assumed when the slope of the regression line had a p value ≤ 0.05 (Conrad and Pollack 1950). Sensor operation periods of 6 months and less duration (for instance occurring when the sensor was replaced shortly before the end date of the covered period 1996–2011) were considered homogeneous no matter of the p value because linear regression with just a few data points yields unreliable results. To ease

the extensive computation and enhance reproducibility, the procedure was implemented as a VBA routine.

Detrending procedure

Sensor operation periods containing significant drifts in one or several RH bins were detrended using a value-dependent procedure that was similar to a previously published linear regression technique with neighboring reference stations (see Della-Marta and Wanner 2006; Trewin and Trevitt 1996). The procedure made the same assumption about initial data integrity after sensor replacement, value-dependent differences between candidate and reference stations, and occurrence of breakpoints as the relative homogeneity test (cf. above). It was implemented as a VBA tool.

In a first step, reference series were partitioned into 5 % bins (e.g., 50–55, 55–60 %, etc.). Based on the differences between reference and concomitant candidate values over time, a daily linear drift rate was calculated for each bin. No linear change over time thus resulted in a 0 daily drift rate. The daily drift rates for each bin were then plotted together with the corresponding numerical values and metadata such as the number of individual data points considered for the calculation of the respective daily drift rate, statistical significance of the respective linear trend, and the number of days from the first to last data point within each bin (Fig. 1). In a user dialog, the operator then selected and deselected individual data points (=daily drift rates) and chose the best fitting polynomial curve (up to fourth order) to yield a continuous, value-dependent function of daily drift rates.

The visual inspection step was designated to allow the operator to exclude data points that most likely represent outliers. As a default, only significant daily drift rates were used to determine the continuous daily drift rates (cf. Fig. 1). Exclusion of a significant daily drift rate occurred when it was clearly offset from the other significant daily drift rates; a common reason for this was when a daily drift rate was based on a few data points only or when it was covering a relatively short period of time. Inclusion of an insignificant drift rate mainly occurred if it was located close to 0 (close to 0 linear trends are by definition statistically insignificant) but contributed to stabilize and/or extend the range of the continuous polynomial drift rate function based on visual inspection.

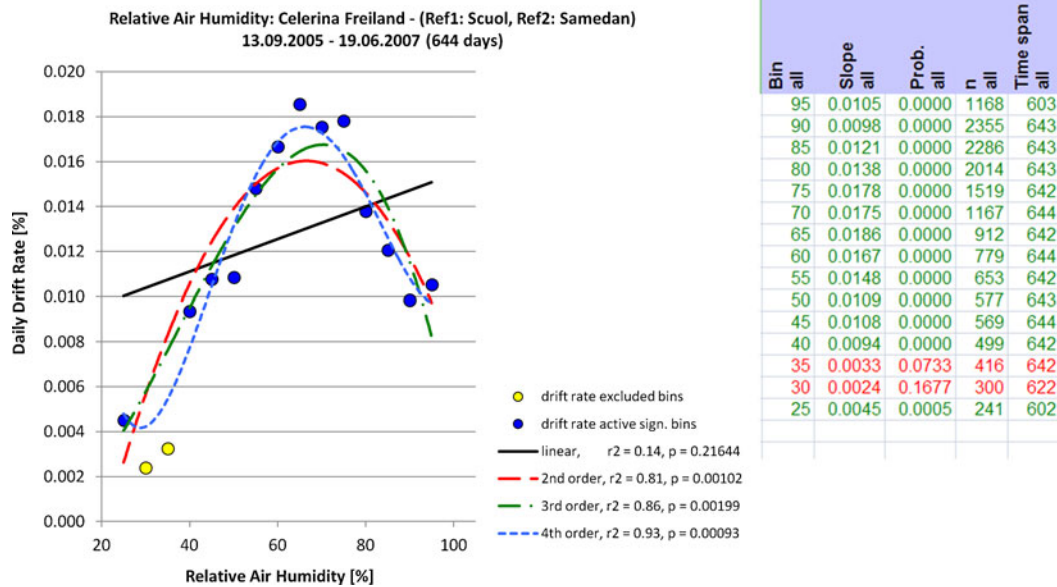


Fig. 1 Screenshot of the used RH detrending tool showing a chart with the daily drift rates (y axis) for the different RH bins (x axis) and the corresponding tabular data. The example shows data from a sensor operation period at the site “Celerina Freiland.” The user dialog for customizing the specific detrending options (see “Materials and methods”) is not shown. Continuous daily drift rates for each RH value were approximated by the best fitting polynomial function as judged from visual

inspection and r^2 values (see lower right corner of the chart). In this example, fourth-order polynomial was selected. The daily drift rates for two RH bins (30 and 35 %; yellow fillings) were not significant and excluded for calculation of the drift rate function. This example illustrates that drift rates depended on absolute RH values; drift was nearly absent in the lowest RH scale range and peaked for RH=65 %

The polynomial drift rate function was used to calculate for each value the corresponding transfer or detrending function as:

$$RH^* = RH - t \times d(RH)$$

where RH^* —detrended RH value, t —day of respective sensor operation period, and $d(RH)$ —daily value-dependent drift rate as obtained from the continuous polynomial drift rate function. Figure 2 shows how this detrending procedure removed sensor drift while maintaining the natural difference between candidate and reference station that is to be expected even if selecting highly correlated reference stations. Finally, candidate RH values >100 % were replaced by 100 % in all 202 series and the same relative homogeneity testing as described above was performed again on all sensor operation periods.

Evaluation

The performance of quality control and data adjustment was assessed by comparing different data

properties between the original and adjusted candidate series.

Results and discussion

Overall, 2,453 data points corresponding to 0.074 % of all data were excluded after basic quality control because they were most likely outliers. Significant drifts in one or several RH bins were observed in 82 of 202 (i.e., 40.6 %) sensor operation periods. In 2 of these 82 sensor operation periods, detrending did not remove inhomogeneity, thus suggesting that other causes than linear drifts (e.g., scale shifts or nonlinear drifts) produced the inhomogeneity. Drift rates depended strongly and in a complex way on absolute RH values (Figs. 1 and 2a). The used detrending procedure considered this fact and also maintained the natural difference between candidate and reference stations as seen by the different locations of the difference trend curves depending on RH range after detrending (Fig. 2b); this is an important characteristic when investigating higher order moments such as

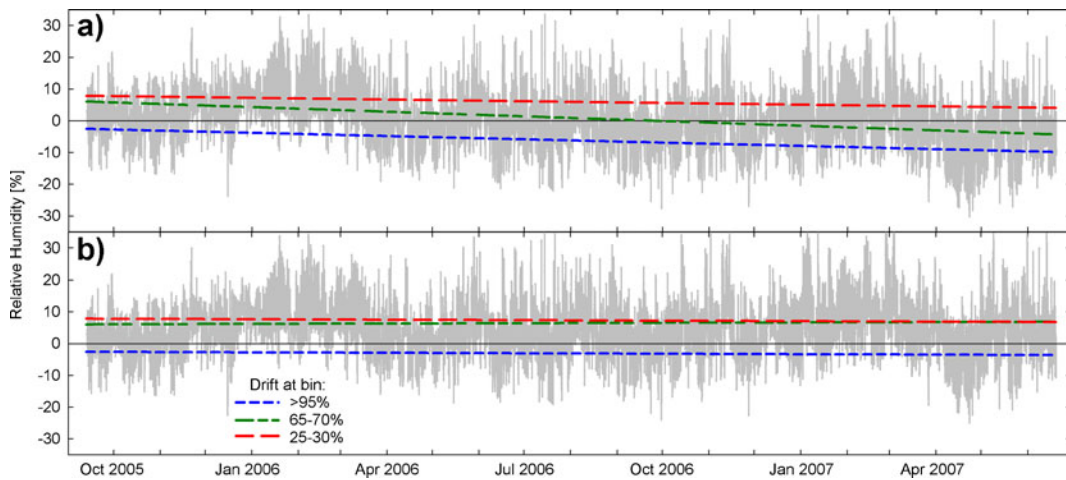


Fig. 2 Difference series between candidate and reference data **a** before and **b** after detrending for the same example as given in Fig. 1. *Colored dashed lines* show linear trends for different RH bins. The *solid line* shows the 0 difference reference line. Homogeneity test of the shown sensor operation period revealed

significant drift before detrending as also evidence by the inclined trend lines in **a**. The value-dependent detrending procedure maintained natural differences between candidate and reference station while removing sensor drift

diurnal minima and maxima (e.g., von Arx et al. 2012).

The overall mean (± 1 SE) RH before and after quality control and data adjustment was 73.00 ± 0.01 and 72.66 ± 0.01 %, respectively. The small change is mainly explained by the balancing positive and negative net effects of detrending. Consequently, the overall mean of absolute changes was greater, in fact reaching 0.98 %, and even increased to 1.96 % when only considering the detrended series. The mean change before and after data adjustment for the 28 weather stations was 0.46 % and ranged from 0.02 to 1.65 %. Beside a random variability, this variation may also reflect specific problems of some stations and/or differential suitability of selected MeteoSwiss reference stations.

In the 80 data series from sensor operation periods with successfully corrected sensor drift, about 19 % of all hourly data were adjusted by no more than ± 0.25 % RH and about half of all data within ± 1.25 % RH limits. A considerable share of 20 % of all hourly data points were adjusted by ± 2.75 % RH and more, and about 10 % by ± 4.75 % RH and more (Fig. 3). The diminishing shares of values requiring substantial adjustment mostly reflect the assumed linear drift starting from correct values, which ultimately results in ever greater deviations towards the end of each sensor operation period. This is illustrated by the comparably large absolute deviation of 4.28 ± 0.29 % RH (mean ± 1

SE) at the end of each sensor operation period when pooling all RH bins of all detrended series. To give an idea of how such differences influence water demand of a plant: when assuming an air temperature of 15 °C and RH of 70 %, a change of 2.75 % RH corresponds to a change in vapor pressure deficit (VPD; a measure of a plant’s evaporation rate) of about 10 %, and a change in RH of 4.75 % would change VPD by about a sixth (Kramer and Boyer 1995). Relatively small differences in RH thus have a disproportionate influence on plant water balance.

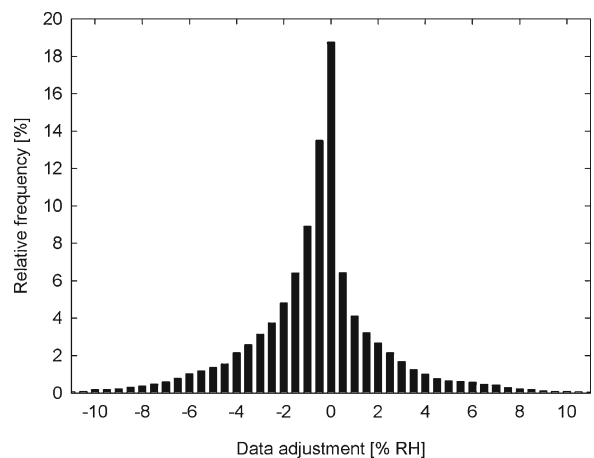


Fig. 3 Frequency distribution of the hourly detrending adjustment amounts for relative humidity bins of 0.5 % when all detrended data series were pooled

Table 1 Distribution of all data sets pooled before and after data adjustments

RH bin (%)	Relative frequency (%)		
	Before	After	Δ
≥ 100.00	2.68	2.20	-0.49
95.00–99.99	7.96	6.96	-1.00
90.00–94.99	10.93	11.41	0.48
80.00–89.99	22.16	22.72	0.56
70.00–79.99	17.02	17.21	0.18
60.00–69.99	13.87	13.91	0.04
50.00–59.99	10.59	10.61	0.02
40.00–49.99	7.78	7.84	0.06
30.00–39.99	4.75	4.83	0.09
20.00–29.99	1.89	1.97	0.08
15.00–19.99	0.26	0.27	0.01
10.00–15.99	0.06	0.07	0.00
≤ 10.00	0.05	0.01	-0.04

Data adjustment did not affect all RH values in equal measure (Table 1). However, except for the values between 90 and 100 %, the mean net effect of positive and negative adjustments was close to zero across the entire RH range (Fig. 4a) and also the upper 95 % and lower 5 % adjustment amounts fluctuated quite symmetrically between ± 2 and ± 4 % RH. The upmost RH values close to 100 % were overall overestimated by up to 1.33 % RH (mean) and 7.29 % RH (upper 95 % adjustment threshold; Fig. 4a) when all data sets were pooled. This result indicates that, while there was no consistent overall drift among sensors in

the RH scale range below 90 %, the used sensors tended to consistently overestimate the largest RH values over time compared to the reference stations. Such systematic trends may not necessarily indicate deficient sensor technology but may have various causes related to the absent ventilation (e.g., vapor condensation and thus overestimated values until complete evaporation) and environmental influences (e.g., bioactivity on sensor filters creating a distinct microclimate, clogged and polluted filters). On the basis of absolute adjustment amounts, the mean change in the lowest RH range was, on average, about 0.5 %, about 1.0 % in mid-range, and 1.33 % at 100 % RH (Fig. 4b). The 95 % limit of absolute adjustment amounts showed a peak of about 5.5 % in mid-range and another peak of about 7 % at 100 % RH (Fig. 4b).

Conclusions

This paper detailed a possible procedure for quality control and data adjustment of high-resolution long-term data series illustrated at the example of hourly RH data. However, the procedure may be used and adapted to any other time series of data for which reference data are available. The procedure described here is particularly suitable when linear drift appears to be the main data inhomogeneity, and when other sources for inhomogeneities such as changes of station location, station surroundings, or instruments have been taken into account. Possible improvements include consideration of nonlinear drift and extension of the employed polynomial functions to interpolate daily drift rates to

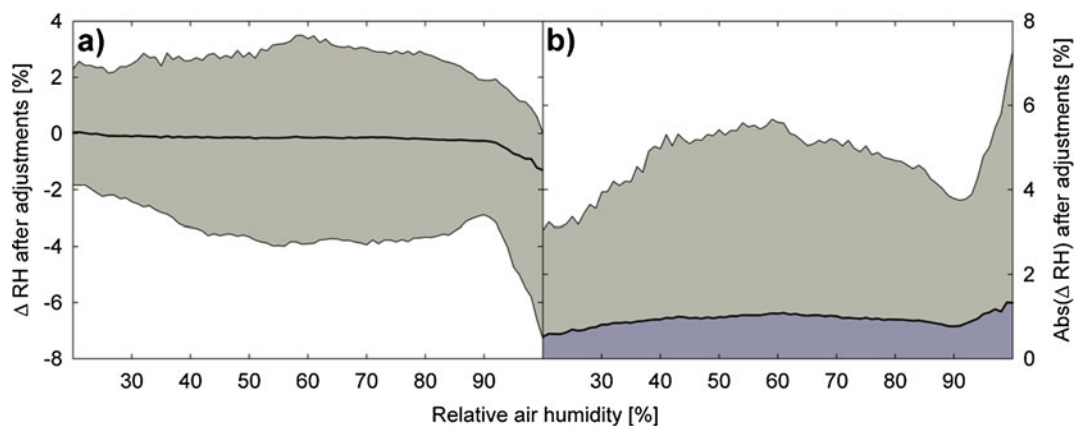


Fig. 4 Overall changes in relative humidity (RH) after data adjustment in dependence on RH values: **a** when considering effective changes and **b** when considering absolute changes. *Thick solid line* mean changes; *shaded area* 5–95 % percentile envelope of changes

more flexible spline functions. This paper also showed that instrumental recordings may significantly deviate from likely correct values due to problems such as sensor drift, after a while. However, such data inhomogeneities can be detected and corrected.

Acknowledgments We thank Gustav Schneiter for meteorological data collection and sharing his expertise of local station situation and Peter Jakob for help with the data base. Michael Begert is appreciated for his valuable advice and input to quality control, data adjustment, and the text. An anonymous reviewer made valuable suggestions to a previous version of this paper. This project was supported by a grant within the program “Forest and Climate Change” from the Swiss Federal Office for the Environment FOEN and Swiss Federal Research Institute for Forest, Snow and Landscape Research WSL.

References

- Aguilar, E., Auer, I., Brunet, M., Peterson, T. C., & Wieringa, J. (2003). *Guidelines on climate metadata and homogenization*. WMO-TD 53. WMO-TD no. 1186, pp. 51
- Begert, M., Schlegel, T., & Kirchhofer, W. (2005). Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000. *International Journal of Climatology*, 25(1), 65–80. doi:10.1002/joc.1118.
- Begert, M., Zenklusen, E., Häberli, C., Appenzeller, C., & Klok, L. (2008). An automated procedure to detect discontinuities; performance assessment and application to a large European climate data set. *Meteorologische Zeitschrift*, 17(5), 663–672.
- Brandsma, T., & Konnen, G. P. (2006). Application of nearest-neighbor resampling for homogenizing temperature records on a daily to sub-daily level. *International Journal of Climatology*, 26(1), 75–89. doi:10.1002/joc.1236.
- Conrad, V., & Pollack, L. W. (1950). *Methods in climatology* (2nd ed.). Cambridge: Harvard University Press.
- Costa, A. C., & Soares, A. (2009). Homogenization of climate data: review and new perspectives using geostatistics. *Mathematical Geosciences*, 41(3), 291–305. doi:10.1007/s11004-008-9203-3.
- Della-Marta, P. M., & Wanner, H. (2006). A method of homogenizing the extremes and mean of daily temperature measurements. *Journal of Climate*, 19(17), 4179–4197. doi:10.1175/jcli3855.1.
- Innes, J. L. (1995). Theoretical and practical criteria for the selection of ecosystem monitoring plots in Swiss forests. *Environmental Monitoring and Assessment*, 36(3), 271–294.
- Jakob, P., Sutter, F., Waldner, P., & Schneiter, G. (2007). Designing a flexible ict framework for processing remote gauging-data information. In J. M. Gómez, M. Sonnenschein, M. Müller, H. Welsch, & C. Rautenstrauch (Eds.), *Technologies in environmental engineering* (pp. 211–220). Berlin: Springer. Environmental Science and Engineering.
- Kramer, J. P., & Boyer, J. S. (1995). *Water relations of plants and soils* (2nd ed.). San Diego: Academic.
- Peterson, T. C., & Easterling, D. R. (1994). Creation of homogeneous composite climatological reference series. *International Journal of Climatology*, 14(6), 671–679. doi:10.1002/joc.3370140606.
- Trewin, B. C., & Trevitt, A. C. F. (1996). The development of composite temperature records. *International Journal of Climatology*, 16(11), 1227–1242. doi:10.1002/(sici)1097-0088(199611)16:11<1227::aid-joc82>3.0.co;2-p.
- von Arx, G., Dobbertin, M., & Rebetez, M. (2012). Spatio-temporal effects of forest canopy on understory microclimate in a long-term experiment in Switzerland. *Agricultural and Forest Meteorology*. doi:10.1016/j.agrformet.2012.07.018.