Advances in
Science & Research
Open Access Proceedings

An empirical method for estimating probability density functions of gridded daily minimum and maximum temperature

C. Lussana

ARPA Lombardia, Milano, Italy

Correspondence to: C. Lussana (c.lussana@arpalombardia.it)

Abstract. The presented work focuses on the investigation of gridded daily minimum (TN) and maximum (TX) temperature probability density functions (PDFs) with the intent of both characterising a region and detecting extreme values. The empirical PDFs estimation procedure has been realised using the most recent years of gridded temperature analysis fields available at ARPA Lombardia, in Northern Italy. The spatial interpolation is based on an implementation of Optimal Interpolation using observations from a dense surface network of automated weather stations. An effort has been made to identify both the time period and the spatial areas with a stable data density otherwise the elaboration could be influenced by the unsettled station distribution. The PDF used in this study is based on the Gaussian distribution, nevertheless it is designed to have an asymmetrical (skewed) shape in order to enable distinction between warming and cooling events. Once properly defined the occurrence of extreme events, it is possible to straightforwardly deliver to the users the information on a local-scale in a concise way, such as: TX extremely cold/hot or TN extremely cold/hot.

1 Introduction

This paper describes a method for the estimation of daily maximum TX and minimum TN temperature probability density functions (PDFs) from gridded datasets.

The hourly temperature observations have been measured by ARPA Lombardia's meteorological network, a dense surface network of automated weather stations in Northern Italy. The mesonet covers part of the southern alpine ridge and of the Po Plain, where elevation and local advections play important roles; a more detailed description of the network can be found in Lussana et al. (2009a).

As daily temperature extremes are often recorded simultaneously at many stations (Pfahl and Wernli, 2012; Wulfmeyer and Henning-Müller, 2006), it is convenient to use the high-resolution meteorological information provided by the whole surface network through a statistical interpolation procedure.

Lombardy's weather service has implemented an Optimal Interpolation (OI) procedure (Uboldi et al., 2008) to interpolate the sparse hourly temperature observations on a regular grid. The quality of the analysis fields is adequate to resolve the mesoscale features down to the Meso-γ scale (Thunis and Bornstein, 1996). A TX and TN analysis field is computed at each grid point as the daily maximum and minimum (respectively) of the hourly averaged temperature.

This work aims to obtain a reliable spatial description of TX and TN extreme events on a local scale. Extreme temperature events, such as heat waves or cold spells, are of great interest for their impact on human activities, and in a warming climate scenario (Schär et al., 2004) they are expected to become more frequent and intense, but the statistical modelling of climate extremes across space remains a challenging issue (Katz, 2010).

The statistical model chosen for TX and TN PDFs is based on the normal distribution, since TX is often observed to be Gaussian (Katz and Brown, 1992), particularly in summer (Mearns et al., 1984). A stationary climate is implicitly assumed, despite climate change (Toreti and Desiato, 2008; Simolo et al., 2010, 2011), because of the short time period considered in the elaborations. In this preliminary study, to

enable distinction between extreme warm and cold events, the statistical model is designed to be asymmetrical.

The extreme event definition is usually based on percentile thresholds calculated from observed data (Klein Tank et al., 2009), without any explicit assumption regarding the underlying PDF. In this work, due to the short temporal period covered by the initial datasets, any percentile-based threshold could lead to ambiguous results. To overcame these limitations and, hopefully, to infer more general results, the choice has been to include the a-priori knowledge about the expected temperature behaviour through the PDF.

This paper is structured as follows: Sect. 2 addresses the issues of data quality and data coverage (temporal and spatial); Sect. 3 presents the TX and TN PDF models and the related parameter estimation procedure; Sect. 4 discusses the extreme event definition and briefly presents two case studies.

2   Quality of the observations and gridded fields

All new observations of ARPA Lombardia's meteorological network undergo a manual quality control procedure within the network's quality assurance system (Lussana et al., 2012; Ranci and Lussana, 2009), however at present only recent data has been manually verified by meteorologists. To overcame this limitation the OI procedure performs an automated Spatial Consistency Test (SCT), described in Lussana et al. (2010, 2009b). The SCT stops observations likely to be affected by gross measurement errors from entering the OI, and also reduces the effects related to the absence of a data homogenisation procedure because of the filter on large systematic or representativity errors.

The rest of this section concerns the impact of the irregular station distribution on analysis quality. The analysis error's distribution (RMS) at station's location has a median close to $1\,°C$ for all stations but the spread is dependent on station density and is significantly higher for isolated stations. Therefore, though it's reasonable to assume a stable analysis quality for dense stations areas, in areas with poorer station coverage analysis quality is known to be irregular. In the latter case, any long-term statistics is likely to include the nontrivial behaviour of the unstable representativity error (Lussana et al., 2010), and its variability is comparable with the natural variability. It is necessary to reduce the initial dataset in order to use only temporal periods and spatial areas with stationary and adequate station density, thus avoiding ambiguity in the comparisons of the results in different areas or time periods.

A convenient way to quantify station density is through the Integral Data Influence (IDI) parameter described in Uboldi et al. (2008). The IDI field can be intuitively visualised as the analysis obtained by interpolating observations with a constant value of 1 on a constant background field of value 0 (without reference to any particular unit). The IDI analysis combines the effects of irregular station distribution and of OI operational choices, resulting in a field with a straightforward interpretation: dense station areas have IDI values close to 1 while areas with poor station coverage have values close to 0.

In the case of ARPA Lombardia's automated mesonet the bulk of the observations ranges from 2000 to 2011 (with some sparse longer time series). The time series of spatially averaged IDI over Lombardy for this period shows two periods of stationary IDI value: 2000–2002 and 2005–2011. The 2003–2004 period is characterised by a gradual IDI increase corresponding to the mesonet development. Only the latest 7 yr (2005–2011) has been considered in this study.

An IDI-based spatial mask has been computed to include only grid points where the station coverage is sufficiently dense and stationary in time, that is, where the IDI value is mostly close to 1. In Figs. 3 and 5, where the IDI mask is applied, white areas within Lombardy's administrative boundaries represent masked out grid points. The border effects are quite evident, as it's evident the effect of sparse (or variable in time) data coverage in the Alps.

3   Estimating probability density functions

The TX or TN time series at a fixed location for a calendar year shows the noisy weather-related variability overlying the typical bell-shaped main annual cycle. To simplify the presentation the following discussion regards TX, but it can be easily applied to TN.

The TX PDF chosen at any fixed grid point is composed by two normal distributions sharing the same mean value, corresponding to the PDF's location parameter, but with two distinct standard deviations allowing for asymmetry between positive and negative anomalies (i.e. difference between the TX value and the location parameter). The PDF is then completely defined by the three parameters: the location parameter $\langle TX \rangle$, taking into account the main annual cycle, and the two scale parameters σ_+^{TX}, σ_-^{TX} for positive and negative anomalies, respectively, accounting for the daily variability. The PDF $p_{TX}(x)$ can be written as:

$$p_{TX}(x) = \begin{cases} \left[1/\left(\sqrt{2\pi}\sigma_+^{TX}\right)\right]\exp\left[-\frac{1}{2}\left(\frac{x-\langle TX \rangle}{\sigma_+^{TX}}\right)^2\right], & x \geq \langle TX \rangle \\[2mm] \left[1/\left(\sqrt{2\pi}\sigma_-^{TX}\right)\right]\exp\left[-\frac{1}{2}\left(\frac{x-\langle TX \rangle}{\sigma_-^{TX}}\right)^2\right], & x < \langle TX \rangle \end{cases}$$

As a preliminary step in the parameter estimation procedure, each month has been split into three 10-day periods as shown in Fig. 1. The 10-day periods have been used in the selection of the initial statistical datasets for the estimation procedure. It is assumed that each 10-day period is likely to be characterised by similar meteorological conditions. The previous and following 10-day periods show a TX temporal behaviour on longer time scales than the short-term weather variability, but still consistent with the assumptions made in
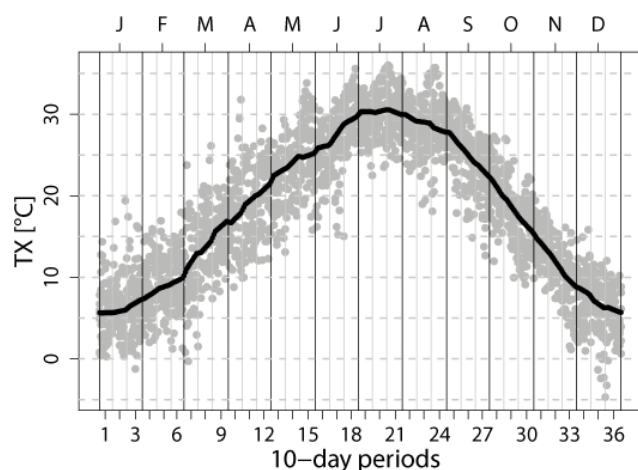
Figure 1. Milan grid point. Location parameter estimates for the calendar year: TX values for the period 2005–2011 in grey; $\langle TX \rangle$ in black.

Figure 2. Milan grid point. Scale parameters estimates for the calendar year: σ_+^{TX}, σ_-^{TX} in black; σ_+^{TN}, σ_-^{TN} in grey; σ_-^{TX} and σ_-^{TN} are reported with opposite sign (negative values).

the estimation procedure, such as the linear trend approximation used for $\langle TX \rangle$.

As a first step, the parameter estimation procedure has been applied independently at every grid point. The location parameter for a fixed calendar day is estimated from an initial dataset of about two hundred TX values (the analysis from the 10-day period containing the calendar day and the two adjacent 10-day periods). Then $\langle TX \rangle$ is obtained by choosing the linear trend which best fits the dataset at the calendar day. The method used for the linear trend estimation is the median of pairwise slope regression (Lanzante, 1996), based on computing the slope between pairs of TX values and considering the median of these values.

The two TX scale parameters for a fixed calendar day are estimated from the same dataset using temperature anomalies instead of TX values. As correlations within the anomaly time series could bias the scale parameters estimates (Wilks, 1995), a resampling procedure has been implemented to create uncorrelated synthetic samples of the three 10-day periods under consideration. The standard deviation is estimated separately for the upper and for the lower part –respect to the median– of each of the synthetic samples and their values are assigned to σ_+^{TX} and σ_-^{TX}, respectively. This procedure is an implementation of the asymmetric biweight standard deviation estimation procedure described in Lanzante (1996). The uncertainties in the σ_+^{TX} and σ_-^{TX} values evaluated from the σ_+^{TX} and σ_-^{TX} distributions obtained form the synthetic samples are about 0.5 °C for both values (the same holds for TN).

The second step in the parameter estimation procedure imposes temporal and spatial consistency by applying a space-time smoothing to the estimates of both location and scale parameters. As the PDF parameters are expected to change slowly in space and in time over the high-resolution gridded domain, for each grid point and fixed calendar day the
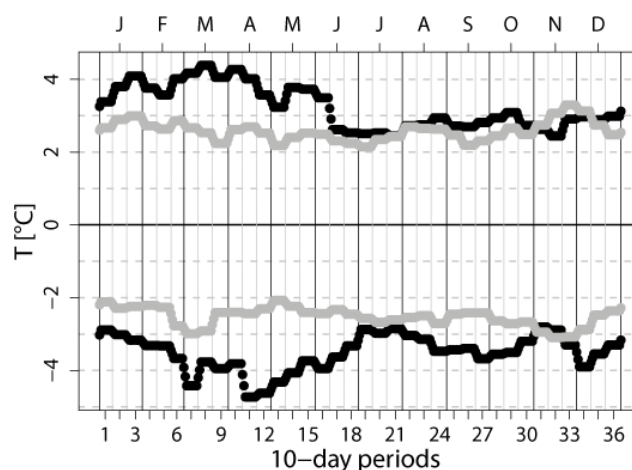
estimated parameters have been replaced with the average of the estimated values in neighbouring grid points and adjacent days.

Overall the parameter estimation procedure is robust and resistant, based upon simple assumptions for the probability model, and rather insensitive to misbehavior of the data.

As an example, the estimation procedure is shown for Milan's grid point. In Fig. 1, grey dots represent the initial TX dataset (each day has 7 values from years 2005 to 2011) and black dots show the rather smoother $\langle TX \rangle$ time series at the same grid point. Figure 2 shows the $\sigma_\pm^{TX}$ and $\sigma_\pm^{TN}$ time series estimated at this grid point. To avoid confusion, scale parameters for negative temperature anomalies are reported with opposite sign (negative values), and TX and TN scale parameters are shown with black and grey dots, respectively. The estimated values are smaller in summer, around 2.5 °C, and higher in winter, around 4 °C. The TN scale parameters are generally lower than the TX parameters.

The maps of seasonally averaged scale parameter, such as those presented in Figs. 3 and 4, often show a significant (i.e. greater than 0.5 °C) difference between σ_+ and σ_-, thus supporting the choice of two distinct scale parameters for positive and negative anomalies. In particular, for TN a significant difference can be found in winter and, for mountainous areas, also in autumn; significant differences for TX have been found in summer, autumn and winter. The spring season is characterised by the highest and more variable values of σ_+ and σ_-. In general, the difference between the two scale parameters is more pronounced for TX than for TN, due to the greater variability associated with daytime weather conditions. Furthermore, σ_- is greater than σ_+ for all seasons except for winter when the situation is the opposite. In winter, σ_+^{TX} is significantly greater than σ_-^{TX} in the Apennines, in the south-western part of the Plain and in the alpine valleys, probably due to the frequent foehn episodes; σ_+^{TN} is
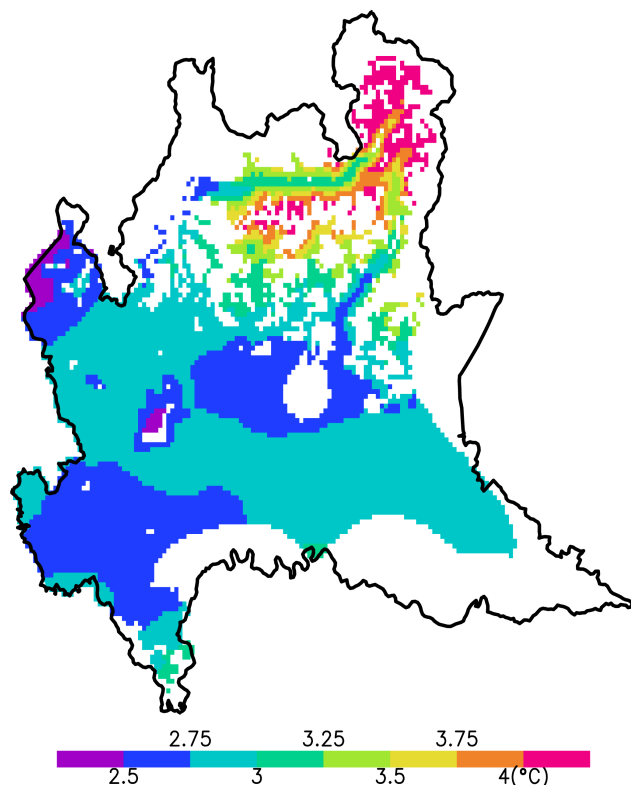
Figure 3. Winter season (December-January-February) averaged σ_-^{TN}. The IDI mask has been applied (see Sect. 2).

Figure 4. Summer season (June-July-August) averaged σ_+^{TX}. The IDI mask has been applied (see Sect. 2).

significantly greater than σ_-^{TN} in the Plain due to frequent and persistent fog. In the other seasons, σ_-^{TX} is increased by the greater variability due to cloudiness, while σ_+^{TX} is associated with the lower temperature variability in clear sky conditions. Figures 3 and 4 show the winter averaged σ_-^{TN} and the summer averaged σ_+^{TX} associated with the occurrence of winter cold spells and summer heat waves, respectively. In general, the study of the seasonal averaged maps reveals the strong influence of the elevation, as expected and showed in Fig. 4; nevertheless it is also possible to identify other distinctive features on a very local scale. In the case of the winter averaged σ_-^{TN}, in Fig. 3, the reduced variability of Milan's urban area is clearly evident in the Western Plain, probably related to the constraining effect of the urban heat island on TN's variability.

4   Extreme events

Extreme events are by definition rare events, and their probability of occurrence is described by the TX and TN PDFs outermost tails. However, due to the limited temporal period available in this study (only 7 yr), it is more likely that extreme events here detected would be in fact moderate extremes (Klein Tank et al., 2009), that typically occurs several times every year.
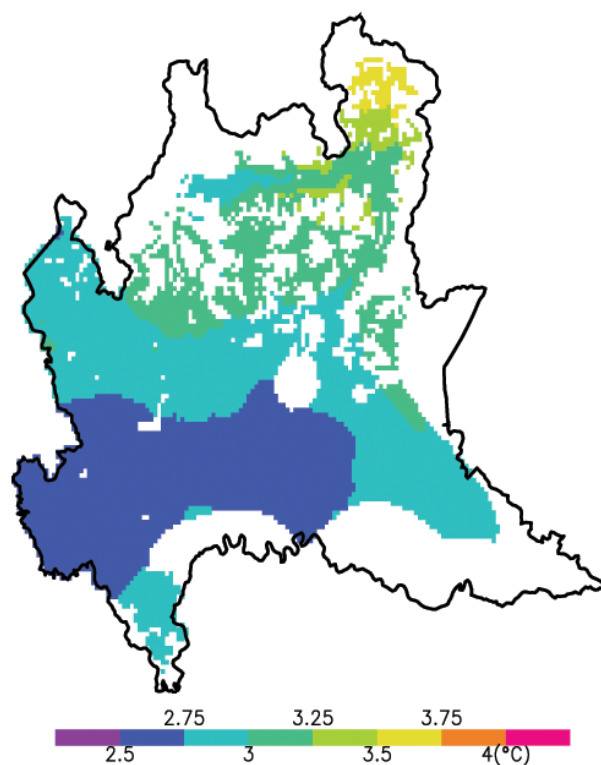
To define an extreme event it is convenient to transform the random variable TX (or TN) to standard form (i.e. a Gaussian having a sample mean of zero and a sample standard deviation of one (Wilks, 1995)). The resulting standardised random variable Z has values z computed as:

$$z = \begin{cases} |x - \langle TX \rangle| / \sigma_+^{TX}, & x \geq \langle TX \rangle \\ |x - \langle TX \rangle| / \sigma_-^{TX}, & x < \langle TX \rangle \end{cases}$$

where x represents a TX realisation. The random variable Z follows a standard Gaussian distribution regardless of the calendar day or the grid point considered, therefore this local-scale definition allows for direct comparison both in space and in time.

The TX and TN events are classified according to their estimated probability of occurrence as: unusual cold $P[Z \leq z] \leq 20\%$ or warm $P[Z \geq z] \leq 20\%$; extreme cold $P[Z \leq z] \leq 5\%$ or warm $P[Z \geq z] \leq 5\%$, whichever more restrictive.

The high sensitivity of extreme event definition to the PDF tails suggests avoiding the use of too small probabilities as thresholds because the Gaussian assumption may be problematic.

As an example, in Fig. 5 the TN analysis map is shown for a winter foehn case resulting in extreme warming in separated areas: two areas close to the conjunction between the Plain and the Alpine region and one area in the South corresponding to the mountainous Apennines region. In this case
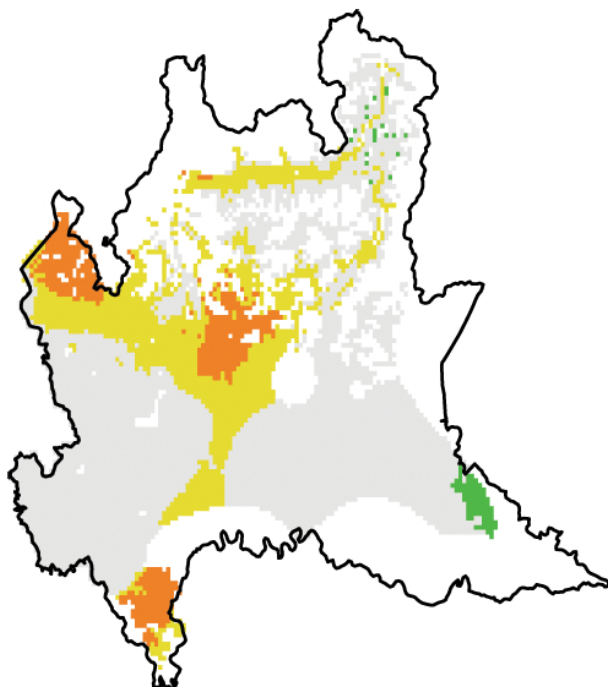
Figure 5. 8 January 2012. TN extreme events classification. Extreme warming: orange. Unusual warming: yellow. Normal conditions: grey. Unusual cooling: green. The IDI mask has been applied (see Sect. 2).

Figure 6. 22 June 2012. TX extreme events classification. Extreme warming: orange. Unusual warming: yellow. Normal conditions: grey. The IDI mask has been applied (see Sect. 2).

the extreme event occurrence is related to local-scale advection, this scheme is able to correctly detect such events and to provide a reasonable description for their spatial extension. Figure 6 presents a summer heat wave case. Respect to the previous foehn case, in this situation the forcing is at the mesoscale, and results in widespread warming in the whole region. Nevertheless, the map can shows gradients in the probability of occurrence, with extreme warming in the central part of the Po Plain while in the upper part of the Plain TX reaches unusual values. In the mountains only the valley floors and only on a local scale show extreme warming.

5 Conclusions

The OI hourly high-resolution temperature analysis produced by ARPA Lombardia's weather service is known to provide valuable information on a local scale (Uboldi et al., 2008). This preliminary study investigates the possibility of exploiting these high-resolution analysis to characterise the daily TX and TN PDFs using an asymmetrical Gaussian-based model.

In studying the daily extremes' statistics it's of crucial importance to use initial temperature analysis fields of adequate and stationary quality. The IDI field can be helpful to locate areas and periods likely to give ambiguous results, which can then be removed from the analysis.
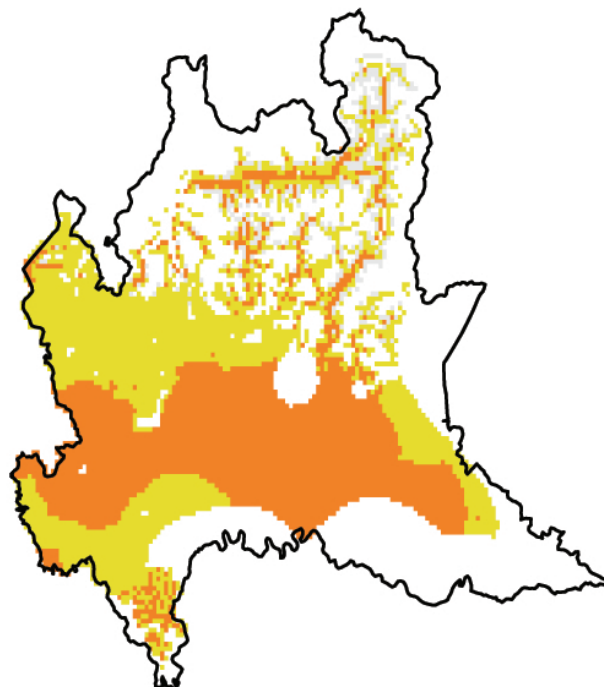
Resistant and robust estimation techniques can be applied to obtain reasonable values for the TX and TN PDF location and scale parameters, also in the case of sparse data. The significant difference in scale parameter value between positive and negative anomalies clearly supports the use of asymmetrical PDFs.

The TX and TN PDFs can be used to detect and evaluate extreme events in the daily TX and TN analysis for each grid point, thus completing the high-resolution spatial information available operationally at Lombardy's weather service.

To improve on this preliminary work the dataset should be enlarged (collecting data from other mesonets in Lombardy and in neighbouring regions), and different choices in statistical model (skew-normal density function, Simolo et al., 2010) and in the parameter estimation procedure (moving-window for the linear trend estimation) should be explored.

References

Katz, R. W.: Statistics of extremes in climate change, Climatic Change, 100, 71–76, 2010.

Katz, R. W. and Brown, B. G.: Extreme events in a changing climate: Variability is more important than averages, Climatic Change, 21, 289–302, 1992.

Klein Tank, A. M. G., Zwiers, F. W., and Zhang, X.: Guidelines on Analysis of Extremes in a Changing Climate in Support of Informed Decisions for Adaptation, Climate data and monitoring, WMO, 2009.

Lanzante, J. R.: Resistant, robust and nonparametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data, Int. J. Climatol., 16, 1197–1226, 1996.

Lussana, C., Salvati, M. R., Pellegrini, U., and Uboldi, F.: Efficient high-resolution 3-D interpolation of meteorological variables for operational use, Adv. Sci. Res., 3, 105–112, doi:10.5194/asr-3-105-2009, 2009a.

Lussana, C., Salvati, M., and Uboldi, F.: Testing the performance of a spatial consistency test for data quality control, in: 9th EMS Annual Meeting, 9th European Conference on Applications of Meteorology (ECAM) Abstracts, held 28 September–2 October 2009 in Toulouse, France, http://meetings.copernicus.org/ems2009/, id. EMS2009-338., Vol. 1, p. 338, 2009b.

Lussana, C., Uboldi, F., and Salvati, M.: A spatial consistency test for surface observations from mesoscale meteorological networks, Q. J. Roy. Meteor. Soc., 136, 1075–1088, 2010.

Lussana, C., Ranci, M., and Uboldi, F.: A Quality-Control-Oriented Database for a Mesoscale Meteorological Observation Network, in: EGU General Assembly Conference Abstracts, Vol. 14, p. 9942, 2012.

Mearns, L., Katz, R., and Schneider, S.: Extreme High-temperature Events: Changes in Their Probabilities with Changes in Mean Temperature, J. Clim. Appl. Meteorol., 23, 1601–1613, 1984.

Pfahl, S. and Wernli, H.: Spatial coherency of extreme weather events in Germany and Switzerland, Int. J. Climatol., 32, 1863–1874, 2012.

Ranci, M. and Lussana, C.: An automated data quality control procedure applied to a mesoscale meteorological network, in: 9th EMS Annual Meeting, 9th European Conference on Applications of Meteorology (ECAM) Abstracts, held 28 September–2 October 2009 in Toulouse, France, http://meetings.copernicus.org/ems2009/, id. EMS2009-335., Vol. 1, p. 335, 2009.

Schär, C., Vidale, P. L., Lüthi, D., Frei, C., Häberli, C., Lininger, M. A., and Appenzeller, C.: The role of increasing temperature variability in European summer heatwaves, Nature, 427, 332–336, 2004.

Simolo, C., Brunetti, M., Maugeri, M., Nanni, T., and Speranza, A.: Understanding climate change-induced variations in daily temperature distributions over Italy, J. Geophys. Res.-Atmos., 115, D22110, doi:10.1029/2010JD014088, 2010.

Simolo, C., Brunetti, M., Maugeri, M., and Nanni, T.: Evolution of extreme temperatures in a warming climate, Geophys. Res. Lett., 38, L16701, doi:10.1029/2011GL048437, 2011.

Thunis, P. and Bornstein, R.: Hierarchy of Mesoscale Flow Assumptions and Equations, J. Atmos. Sci., 53, 380–397, 1996.

Toreti, A. and Desiato, F.: Changes in temperature extremes over Italy in the last 44 years, Int. J. Climatol., 28, 733–745, 2008.

Uboldi, F., Lussana, C., and Salvati, M.: Three-dimensional spatial interpolation of surface meteorological observations from high-resolution local networks, Meteorol. Appl., 15, 331–345, 2008.

Wilks, D. S.: Statistical Methods in Atmospheric Sciences: An Introduction, in: International Geophysics Series, Acad. Press, 1995.

Wulfmeyer, V. and Henning-Müller, I.: The climate station of the University of Hohenheim: analyses of air temperature and precipitation time series since 1878, Int. J. Climatol., 26, 113–138, 2006.