



ISTITUTO NAZIONALE DI RICERCA METROLOGICA Repository Istituzionale

Metrological evaluation of the effect of the presence of a road on near-surface air temperatures

This is the author's accepted version of the contribution published as:

Original

Metrological evaluation of the effect of the presence of a road on near-surface air temperatures / Coppa, Graziano; Quarello, Annarosa; Steeneveld, Gert-jan; Jandrić, Nebojša; Merlone, Andrea. - In: INTERNATIONAL JOURNAL OF CLIMATOLOGY. - ISSN 0899-8418. - 41:6(2021), pp. 3705-3724. [10.1002/joc.7044]

Availability:

This version is available at: 11696/69030 since: 2021-05-26T17:06:06Z

Publisher:

WILEY

Published

DOI:10.1002/joc.7044

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
WILEY POST PRINT

This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions. This article may not be enhanced, enriched or otherwise transformed into a derivative work, without express permission from Wiley or by statutory rights under applicable legislation. Copyright notices must not be removed, obscured or modified. The article must be linked to Wiley's

(Article begins on next page)

Metrological evaluation of the effect of the presence of a road on near-surface air temperatures.

Graziano Coppa^{1,*}, Annarosa Quarello^{2,3}, Gert-Jan Steeneveld⁴, Nebojsa Jandric⁵,
Andrea Merlone¹

¹Istituto Nazionale di Ricerca Metrologica (INRiM), Strada delle Cacce, 91, Turin (Italy)

²Université Paris Diderot, Paris (France)

³Università di Torino (Italy)

⁴Wageningen University, Meteorology and Air Quality Section, Wageningen (Netherlands)

⁵Institut za Mjeriteljstvo Bosne i Hercegovine (IMBiH), Sarajevo (Bosnia-Herzegovina)

With the purpose of revising World Meteorological Organization's Commission for Instruments and Methods of Observation (WMO/CIMO) Guide #8 on weather stations siting, an experiment to evaluate metrologically the maximum influence of a paved road on 2-m air temperature measurements ("road siting effect") has been designed, installed and run in Italy. It consists of a 100-m long array of seven measurement stations, at increasing distances from a local road, equipped with shielded Pt100 thermometers and ancillary sensors. Data coming from 1 year of observations has been analysed for daily climatological metrics, finding that the road mostly effects minimum temperatures, with average values of $\sim 0.30 \pm 0.18$ °C at a distance of 1 m; then, in order to quantify the instrumental effect on the measurement, data was filtered by applying a Generalized Additive Model, selecting only times when the effect is more intense (during nights, in presence of low winds coming from the road), and the road siting effect has been calculated by modelling the maximum temperature differences by using Extreme Values Analysis. The 1-year return value on 10-min measurements is 1.22 ± 0.30 °C at 1 m from the road, with a gradual decline (~ 0.1 °C/m), while an extrapolation to 100-year return level gives a value of 1.71 ± 0.79 °C with a decline rate of about 0.17 °C/m. This is a first step towards a redefinition of the weather station classification scheme of WMO/CIMO Guide #8, together with building and tree effects experiments which have been run in parallel with the road siting experiment here presented and which will be presented elsewhere.

Keywords: air temperature, uncertainty, siting classification, MeteoMet, WMO, near surface data

1. Introduction

Accurate air temperature measurements are crucial for understanding weather and climate. Forecasts require high-quality observations for model verification (Ebert *et al.*, 2013) and surface observations are also essential for data assimilation in numerical weather prediction models in order

*Corresponding author: g.coppa@inrim.it

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the [Version of Record](#). Please cite this article as doi: [10.1002/joc.7044](https://doi.org/10.1002/joc.7044)

to determine the best initial state (Drusch and Viterbo, 2007). Within the context of climate monitoring, long-term undisturbed temperature records are of high value since they are scarce. Maintaining such long records is challenging both for technical and logistic reasons (Peterson *et al.*, 1998), and understanding the quantitative impact of disturbances in the near vicinity of the observational site is crucial for the interpretation of climate records.

An often-discussed phenomenon related to temperature disturbances is the urban heat island effect. Earlier studies report that for individual stations the influence of urbanization is substantial (Koopmans *et al.*, 2015). For instance, Heusinkveld *et al.*, (2014) report that the weather station at Rotterdam airport overestimates the average nocturnal temperature by 0.7 °C due to a plume of warm air from the built environment and airport runway (NE to SW between 0.7 and 1.6 °C).

There are two recognized ways in which artificial surfaces such as roads affect the local microclimate: first and foremost, asphalt has a greater heat capacity than the natural soil (Mahmood *et al.*, 2014) as well as larger heat conductivity (Madjidi *et al.*, 2013) and lower albedo (Lin *et al.*, 2010; Ibrahim *et al.*, 2018), so it stores more solar energy during the day, leading to different cooling rates during nights (Oke, 1982); the other is due to the reduction of evapotranspiring surfaces (i.e., trees and grass) that raises air temperature (Arnfield, 2003).

Climate monitoring revealed that the currently observed climate trend is asymmetric since the nocturnal temperatures appear to rise faster than the daytime temperatures (Karl *et al.*, 1993; Vose *et al.*, 2005; McNider *et al.*, 2012; Cox *et al.*, 2020). As such, the accurate monitoring of night-time air temperatures is crucial. Moreover, since the daytime boundary layer is characterised with high turbulence intensity, we expect that the small disturbances in the landscape have relatively little impact on the 2-m temperature record. On the contrary, in the nocturnal atmosphere turbulence is relatively weak and other physical processes such as fog, drainage flows, microfronts, and thermal radiation may trigger small-scale temperature differences.

Measurements of Essential Climate Variables for climatological studies, in particular near-surface air temperature measurements, can be affected by local peculiarities of the site (Mahmood *et al.*, 2006; Runnalls and Oke, 2006; Pielke *et al.*, 2007). Since data quality and representativeness are of paramount importance for these purposes, it is necessary to evaluate the effects of the site's features, in terms of the presence of slopes, shades, vegetation, uneven or unrepresentative surface albedo, water bodies and man-made structures, in the measurement themselves. In fact, these

micro-siting external factors can contribute, at a purely instrumental level, to biases that, especially if asymmetrical, could propagate to climate analysis if underestimated.

In 1954 the World Meteorological Organization (WMO), with its Commission for Instruments and Methods of Observation (CI-MO), implemented an operational document – the “WMO/CI-MO Guide on Instruments and Methods #8”, cited in this work henceforth as “the Guide” (WMO, 2012) that adapted an internal procedure created by Météo-France – to provide meteorological stations with a qualitative evaluation of uncertainties associated to their measurements.

Regarding near-surface air temperature, the Guide establishes a 5-class grading for meteorological stations, based on the presence of several possible sources of influence on the sensors themselves, such as 1) topography, 2) surface characteristics, 3) presence of shades, 4) presence and distance of physical obstacles as sources of heat (buildings, concrete surfaces, car parks, roads, trees, water bodies).

The Guide considers stations as unaffected by the obstacles – and therefore “Class 1” if also the other requirements are met – when the minimum distance from any of the aforementioned obstacles is 100 m. Distances shorter than this arbitrary threshold correspond to lower quality classifications: a station is downgraded to “Class 2” if placed between 30 and 100 m from any obstacle; to “Class 3” between 10 and 30 m, to “Class 4” between 3 and 10 m, down to “Class 5” with distance less than 3 m.

In the Guide, some of these classes are given estimates of “additional uncertainty” in temperature measurements. However, there is no clarification on the methodologies adopted to evaluate these uncertainties. The Guide itself states: «the classification is (occasionally) completed with an estimated uncertainty due to siting, which has to be added in the uncertainty budget of the measurement. This estimation originates from bibliographic studies and/or some comparative tests». The uncertainties given in the Guide are stated as 5 °C for Class 5, 2 °C for Class 4 and 1 °C for Class 3 sites (no indications for Classes 2 and 1), without regard for the type of obstacle and influence.

The Guide is currently in the process of being overhauled by WMO, and it is already established that the siting classes will be related to uncertainties in the representativeness of the measurand, not of the instrument. The new classification will explicitly deal with climatology-grade sites, in that user requirements driven by OSCAR (Observing Systems Capability Analysis and Review Tool)

contributing to the definition of the Measurement Quality Classification Scheme will define, among others, a top-level grade (Class A) dedicated to climatological studies.

The purpose of this work, carried out in the framework of project MeteoMet (Merlone *et al.*, 2015, 2018) is to deliver measured results to the WMO expert teams tasked to revising the Guide, by experimentally evaluating with metrological rigour maximum errors and uncertainties in near-surface air temperature measurements in presence of obstacles. This paper focuses on the effect of the presence of a paved road, which in this case means “paved with asphalt”. Effects of nearby trees and buildings have also been studied within MeteoMet with identical experimental setups: results will be presented elsewhere.

Very few literature works deal with the estimation of the influence of paved surfaces on measurements of air temperature. Kumamoto *et al.*, (2013) describe a similar experiment with a much smaller frame in a closed environment; Bogren *et al.*, (2001) focus on the stratification of air temperature at two distances from a road, while Delgado *et al.*, (2007) focus on the effect of a road corridor inside a forest. On the theoretical side, Kinoshita, (2014) proposes a detailed physical model with predictions, using several road sizes.

The paper is organized as follows. Section 2 presents the field experiment and the methods, in terms of data analysis and estimation of the uncertainties involved; Section 3 presents the results of the analysis; Section 4 presents a discussion of the results, along with comparison with literature works; Section 5 shows the conclusions.

2. Methods

Here we present the setup of the experiments for the evaluation of obstacles' influence on temperature measurements. Design and realization have been carried out at the Italian Institute for Metrological Research (INRiM).

Seven measurement stations, installed on masts at 1.5-2 m from the ground, constitute the experimental layout. The masts hosting the instruments have been arranged along a straight line, perpendicular to the influence factor (road, building wall, line of trees). The masts were installed respectively 1 m, 5 m, 10 m, 20 m, 30 m, 50 m and 100 m from the influence factor: this

arrangement has been chosen to evaluate progressively the effect of the distance of the influence factor on instruments, in reference to a WMO class 1 site. Stations will be identified in the remainder of the paper with sequence numbers, starting from the obstacle: station #1 is the closest to the obstacle, while station #7 is the farthest away.

Each measurement station carries specific instruments: Figure 1 shows a diagram of the stations and sensors setup including the distance from the influence factor and the positions of all sensors. Each measurement point features at least one shielded, actively ventilated Pt100 thermometer (station #1 holds two for redundancy). Some of the stations also feature instruments for ancillary measurements: two hygrometers, a pyranometer for the measurement of solar radiation and a sonic anemometer. Throughout the paper, the temperature sensors will be identified by the characters TE followed by a digit representing the station number: for instance, sensor TE5 identifies the temperature sensor hosted by station #5. All connections for the instruments (power lines, data wires etc.) were put in underground conduits.

The experiment to evaluate the maximum effect of the presence of a paved road on 2-m air temperature measurements (henceforth, *road siting effect* – RSE) has been hosted by a private airfield (ICP Srl) in the town of Castelnuovo Don Bosco (45°01'34" N, 7°57'53" E), 40 km southeast of Turin, Italy (Figure 2). The instruments array was placed on a geologically uniform flat field covered by grass, representative of the typical vegetation of the region, kept sufficiently short by regular mowing, never ploughed nor artificially irrigated. Figure 3 shows the actual installation and the position of measurement points with respect to the road.

The field around the sensor array was kept free of any obstacles – except the road, as the influence factor – for a radius of 100 m around each station. In particular, station #7 was placed according to WMO Class 1 and was considered as reference. On the opposite side of the road there were no obstacles as well for a radius of 100 m around station #1. Local road SP16 – the obstacle – was straight at the experiment location, paved with asphalt and 2-laned for a total width of 10 m, lightly trafficked (~2000 vehicles per day). The airfield building in the East is placed more than 100 m away from each sensor: contributions from it are likely to be excluded (Garcia Izquierdo *et al.*, in prep), and would in any case be constant due to a similar distance between the building and each of the sensors. The field shows no slopes for the entire length of the sensors array and around them, while the road is separated from the field by a small ditch (covered during construction) and is 50 cm higher than the terrain: this was taken into account when installing the stations, by levelling the

Accepted Article
heights of sensor #1 (1.5 m) and the others (2 m), while fulfilling the prescription of the Guide in this regard.

The experiment was carried out between September 2015 and November 2016. Subsequently, the layout was modified in order to evaluate instruments' behaviour when subject to the same environmental conditions. All the sensors were therefore relocated in the same spot (chosen to be the station #5 at 30 m, for practical reasons) for 2 months (December 2016 – January 2017). Measurements from this exercise provided a relative “zero” reading between the sensors and contributed to the measurements' uncertainty budget. This will be discussed in more detail in Section 2.2.

The raw data is publicly available online at Zenodo.org, as 230 MB database files in ASCII format, containing 1.2 million 30-s observations in 15 million fields (Coppa *et al.*, 2020).

2.1 Measured parameters

The installation is equipped with eight independent temperature sensors (Pt100), each one mounted in a separate Young 43502 mechanically ventilated radiation shield. Two redundant thermometers were installed on station #1 in order to assure the acquisition of the most valuable information through possible failures. The complete equipment is as follows:

- Temperature sensors (TE_i), connected to a customized Analog-to-Digital Converter MUX DAT-3014. Calibration has been performed at INRiM and is described below.
- A pyranometer (XT5) Model Hukseflux LP02, measuring solar radiation and mounted on station #5. Manufacturer's calibration has been assumed, with a final uncertainty less than 0.9 %.
- The wind sensor (ZT/ST5) is a Gill Wind sonic anemometer mounted 2.5 m above ground on top of station #5. Its uncertainty has been estimated by the manufacturer as ± 1 % at 12 m/s.
- Two hygrometers (UT_i) model ETM-30 by Lombard & Marozzini, mounted inside multi-plate solar radiation shields close to the temperature sensors on stations #3 and #5. Manufacturer's calibration is assumed, with a stated uncertainty of 0.5 %.

All measurements are taken simultaneously every 10 s, while once every 30 s the average of the measurements is calculated and stored into the database; wind measurements are averaged

vectorially. Uncertainties are provided as prescribed by the GUM (Guide to the expression of Uncertainties in Metrology, BIPM and Joint Committee For Guides In Metrology, 2008). Given the non-Gaussian nature of some of the distributions mentioned, a coverage factor $k = 1$ is adopted throughout the paper (coverage factor coincides with σ for Gaussian distributions). The total expanded uncertainty is then provided with coverage factor $k = 2$, representing 95 % significance.

The calibration of Pt100 thermometers (without shield) was carried out at INRIM laboratories, while manufacturers' calibration is assumed for the other sensors. This will not affect the metrological evaluation since they generate only auxiliary data.

Before calibration, the Pt100 thermometers were thermally cycled between $-20\text{ }^{\circ}\text{C}$ and $50\text{ }^{\circ}\text{C}$ in order to evaluate the repeatability of the instruments. The thermometers, inserted in glass tubes, were calibrated in a very stable and homogeneous liquid bath, by comparison with a standard resistance thermometer calibrated at the fixed points of the ITS-90. The thermometers were calibrated at six temperature points: $-20\text{ }^{\circ}\text{C}$, $0\text{ }^{\circ}\text{C}$, $20\text{ }^{\circ}\text{C}$, $30\text{ }^{\circ}\text{C}$, $40\text{ }^{\circ}\text{C}$ and $50\text{ }^{\circ}\text{C}$ to include all possible environmental conditions of the sites. The final calibration uncertainty was evaluated as $0.005\text{ }^{\circ}\text{C}$.

During in-field operation, though, the sensors work in very different environmental conditions: radiation, wind, turbulence, all introduce an uncertainty contribution in the measurements, in addition to the calibration uncertainty. This is the total *measurement uncertainty* that needs to be evaluated and taken into account.

These are parts of the two-step process of this metrological investigation, the first being sensor calibration – which establishes traceability to the SI – while the second is the evaluation of measurement uncertainties – being performed during the co-location phase of the sensors – and their reduction due to the relative analysis performed (Section 2.3).

2.2 Measurement uncertainty evaluation

After the completion of the main field experiment, stations have been rearranged as close as possible in order to evaluate possible differences in thermal behaviour and response under the same environmental conditions (Figure 4).

Data was averaged every 10 min, with wind speed and directions averaged vectorially. Several considerations led to this choice:

- First, 10-min averaged data is much more frequently used for atmospheric science analysis because it reduces turbulence and boundary-layer issues (Lin and Hubbard, 2008; WMO, 2012). In this way, temperature spikes are smoothed and only significant ΔT_i can be examined.
- Second, the averaging made it easier to perform data analysis in terms of manipulation and subsequent filtering.

All differences $\Delta T a_i = T a_i - T a_7$, (being $T a_7$ the temperature recorded by TE7 and $T a_i$ the temperature recorded by other sensors TE*i* during this phase of the experiment) have been computed: for the sake of clarity, only $\Delta T a_1$ is plotted in Figure 5. Distributions of all the $\Delta T a_i$ are shown in Figure 6, along with their values (blue dashed lines). The average differences are peaked at 0 °C, with differences in any case less than 0.06 °C, confirming that all the sensors are measuring the same measurand. These average differences have been subtracted from all the values recorded during the main phase of the experiment, in order to realign the readings of all the instruments.

Figure 5 also shows short-scale $\Delta T a_i$ up to ± 1.5 °C, which are likely caused by turbulence and transient behaviour of the local atmosphere, or occasional shading of the sensor by other elements of the array. As the measurement uncertainty stage was performed to compute the $\Delta T a_i$ averages during the whole 2-month period, these spikes do not alter significantly the results. However, spikes could affect measurements during the main experiment, by altering or masking the RSE: as it will be shown in Section 2.4.2, these spike events were filtered out and discarded from the analysis.

The standard deviations of the $\Delta T a_i$ are here assumed as *measurement uncertainties*. Despite being an oversimplification, this exercise provided an indication of the sensors' relative behaviour in a real-world situation, under uniform environmental conditions. A more thorough method would have required several more months of acquisition, in order to cover a more extended range of temperatures and environmental conditions, and a separate uncertainty evaluation for all the quantities of influence, which was out of the scope of this work. The standard deviation values calculated during this stage will be taken into account during the whole uncertainty budget evaluation, and are shown in Table 1 along with the other identified sources of uncertainty.

2.3 Uncertainty budget

Measurements are accompanied by a metrological investigation that take into account as many aspects of the experiment as possible, ranging from sensor issues to uncertainties due to quantities of influence. Currently, a complete quantitative evaluation of the uncertainty associated to a near surface air temperature measurement by means of contact thermometers is a challenging problem, discussed within both BIPM and WMO working groups and expert teams, with no immediate and comprehensive solution. Many influencing quantities, different typologies of sensors and solar shields, dynamics, calibration methods, siting and environmental conditions generate a multitude of contributions to the overall measurement uncertainty budget. The RSE experiment has been designed to minimize such uncertainty contributions by featuring measurement conditions as uniform, constant and equal as possible for all sensors, as well as relying on relative measurements, thus allowing to cancel some absolute uncertainty contributions. This is part of the two-step process of this metrological investigation mentioned in Section 2.1.

In Table 1, the uncertainty budget for all the temperature measurement points is shown. According to the definitions given by the GUM, uncertainties are listed in the Table by their nature: statistical (type A) and non-statistical (type B). Type A uncertainties taken into account are the measurement one, derived by the exercise described in Section 2.2, and the RSE one, for different return levels (see Section 3.1). Type B uncertainty derives solely from the calibration of the Pt100.

Constant maintenance guaranteed that all measuring points (thermometers and associated shields) were kept in same working conditions. Checks were made to presence of dirt, insects, orientation, fan speed, etc., about weekly all along the experiment. This, together with the measurement uncertainty evaluation stage of the experiment, ensured that the relative differences are effectively related to different air temperatures values.

The following is a list of actual uncertainty sources that, given the relative nature of the measurements performed, have been reduced or eliminated from the total budget:

- **Sensor drift:** Temperature sensors, as well as many other kinds of electrical measurement systems, can change their properties over time. This has been evaluated by repeating calibration at the end of the experiment: no sensor showed a drift larger than 0.01 °C, which is compensated by relative measurements.
- **Shield ageing and solar radiation:** By comparing shields of different age, Lopardo *et al.*, (2014) showed an uncertainty as large as 1.6 °C over a 5-year lifespan due to shield degradation. In the present work, however, all the shields were new, employed for the same

amount of time and exposed to the same homogeneous environmental conditions, causing a similar level of dirtiness. Any potential error induced by shield ageing is shared by all the stations, cancelling out during the calculation of temperature differences.

- **Sensor self-heating:** Many literature works deal with the self-heating evaluation of a platinum resistor in high-quality laboratory measurements (Batagelj *et al.*, 2003; Coppa and Merlone, 2016), but very few are directed towards the same evaluation for meteorological measurements (Pavlasek *et al.*, 2020). The latter, investigating common meteorological temperature sensors, found temperature uncertainties due to self-heating up to 0.2 °C: in our case, all the were sensors identical and powered by the same datalogger, so self-heating do not alter the relative readings among them.
- **Surface features:** Changes in surface reflectivity can affect sensors, which are shielded against direct sun radiation but not against radiation reflected from the ground (Musacchio *et al.*, 2019). Throughout the duration of the experiment, field grass has been kept as short as recommended by the Guide; also, no asymmetric or inhomogeneous variations in surface reflectivity have been detected.
- **Datalogger contribution:** Modern temperature resistors (such as Pt100 used in this study) measure the resistance of a platinum wire run by an electric current, which is then digitalised, converted into temperature and stored in a datalogger. This process introduces other uncertainties, related to the electronics, the resolution and the accuracy of the resistance measurement. In this work, all the thermometers are read using the same datalogger, so the uncertainty introduced is compensated in the relative analysis.

2.4 Data analysis

The 10-min averaged dataset consists in ~48000 records over a 1-year period. Temperature differences $\Delta T_i = T_i - T_7$, analogous to the $\Delta T a_i$ calculated during the previous stage, have been computed: the complete profile of ΔT_1 is shown in Figure 7, while all the distributions are plotted in Figure 8. As it can be seen from the plot, all distributions are skewed towards positive values, with mean values decreasing from ΔT_1 to ΔT_6 , as it can be expected in a case of a distance-dependent warming effect. Mean value of ΔT_5 is higher than ΔT_4 (by ~0.01 °C), possibly due to central processing unit on mast #5 heating up local air.

Data has been processed for other, more common for climatological studies, metrics. In Figure 9, ΔT_i distributions calculated at daily maximum, average and minimum temperatures are plotted. Minima seem to be most affected, with ΔT_1 values up to ~ 0.9 °C and an average value of 0.3 °C, while maxima ΔT_1 show values up to ~ 0.8 °C with an average value of 0.12°C. Maxima and averages seem to be affected only up to ~ 10 m, while effects on minima appear to reach as far as 50 m; uncertainties on maxima make their distributions even compatible with a flat one, i.e. no RSE at all.

These calculations have been performed by using the 10-min dataset, since many recent “first-order” climatological stations are able to acquire data at this interval (WMO, 2018), and taking the values in the period before the full hour. In order to maintain backward compatibility with older stations, which record maxima and minima only and evaluate the average as $(T_{max} + T_{min})/2$, this calculation has also been computed, resulting in slightly higher average ΔT_i and larger standard deviations, as noted e.g. by Bernhardt *et al.*, (2018). Results of both methods, in terms of mean biases and standard deviations, are listed in Table 2.

This straightforward approach, which gives hints on the magnitude of the RSE on metrics relevant for climatological studies, may not tell the whole story. For instance, this analysis does not take into account the RSE on shorter timescales, which may be of interest for other applications. Moreover, climatological metrics lead to an evaluation of a bias, which was out of the scope of the experiment since the beginning. Given that a bias must be used to correct data in other installations, it would need an accurate evaluation of the relationships between it and all the influencing quantities, which would require a more sophisticated experiment, much more time, and several different transects in order to disentangle the effects of road widths, different traffic, local microclimate, etc. As a matter of fact, the uncertainty evaluation on such biases, at the current level of knowledge, would be basically impossible. From a data analysis point of view, this brings in also other effects not strictly related to the RSE, e.g. errors related to fast temperature transients; in fact, each sensor responds with a different time constant to the temperature rise produced by a fast and strong variation in radiation (Kowal *et al.*, 2020). On the other hand, daily averages on a filtered dataset would yield inconsistent results given the non-continuous and non-homogeneous nature of the dataset.

In order to better understand this effect, and help us to focus on the combination of parameters that could maximize the ΔT_i (and ΔT_1 in particular), a statistical semi-parametric GAM (Generalized Additive Model) was developed.

2.4.1 The Generalized Additive Model

Since linearity is not a good assumption on the relationships between ΔT_i and the other relevant variables, a semi-parametric GAM (Hastie and Tibshirani, 1986; Bertaccini *et al.*, 2012) was implemented on ΔT_1 , as an extension of a standard linear model, in order to drop the linearity assumption:

$$E(\Delta T_1) = f_0(\text{season}) + f_1(\text{RAD}, \text{season}) + f_2(\text{WS}, \text{season}) + f_3(\text{WD}, \text{season}) \\ + f_4(\text{RH}, \text{season})$$

Season was included as parametric part, to try and separate relevant effects due to different temperature, and length-of-day effects; solar radiation (RAD), wind speed (WS), wind direction (WD) and relative humidity (RH) were included as smooth thin plate regression splines (Wood, 2003), evaluated simultaneously through Generalized Cross-Validation (Quarello, 2017) using R package “mgcv” (Wood, 2011, 2017). In the model, linear effects are described with estimated coefficient values and non-linear effects are presented graphically in Figure 10 as smooth functions with 95 % confidence band, along with marginal distributions of ancillary measurements. Given the computationally demanding nature of GAM analysis, results here presented only involved ΔT_1 to evaluate the strongest signal, assuming the same behaviour for all other ΔT_i . The analysis gave similar results, but with a weaker signal, also on ΔT_2 . For the same reason data has been treated as independent, so confidence bands in the plot must be taken just as an indication.

Figure 10a shows the relation between solar radiation and the expected values of ΔT_1 . An increase in solar radiation is associated with a decrease in average temperature difference, as already noted by e.g. Bogren *et al.*, (2001) and Leeper *et al.*, (2019). The explanation of this phenomenon is not straightforward: one factor could be that higher solar radiation values may result in more turbulence, which mixes away contrasts among measurement points in the array; another factor can be the different release rates of heat by asphalt and grass, as noted by (Oke, 1982), which warms up air above the road faster. Moreover, heating could be masked during the day by the larger effect due to direct solar radiation on the instruments. While it is true that thermometers are shielded from solar radiation, it is also well known that shields do not totally cancel this influence (Erell *et al.*, 2005).

Wind speed (Figure 10b) appears to be inversely related to the temperature difference, as for example shown also by Bogren *et al.*, (2001). Low wind speeds can cause warm air from the road to

stay close to it for longer times: this can cause the sensors to measure higher temperatures with respect to farthest sensors, which receive relatively cooler air from the grass around them. Moreover, in case of high wind, turbulent mixing of heat to higher atmospheric layers makes its influence on temperature lower at 2 m above ground, where the sensors are placed. This effect is steep up to 1 m/s, then its magnitude decreases more slowly, flattening out beyond 2 m/s.

Wind direction appears to have a positive contribution in average temperature difference, in all seasons (Figure 10c). The maxima shown at 250° are due to the winds coming from the left side of the array. For geometrical reasons, winds coming from this direction encounter a larger asphalted area, air passing over it is heated up more, and this in turn heats up the sensor close to the road more than the farthest ones, increasing ΔT_1 .

A simple wind rose plot (Figure 11) shows that the prevailing winds come from the NW, i.e. from the road. Winds faster than 2 m/s are extremely rare in the site, and they are virtually non-existent from the prevailing direction.

The estimated relative humidity smooth functions, in Figure 10d, show a rather flat behaviour: at values lower than 25 % – not shown in the plot – they tend to decrease but with lower statistical significance due to scarcity of data, and the fact that below that value the hygrometer readings tend to be less reliable.

2.4.2 Data filtering

The model described in the previous Section was successful in narrowing down the huge amount of data. It suggests that the RSE should be largest during nocturnal ($Rad \rightarrow 0 \text{ W/m}^2$) periods of low-speed winds coming from the road. Relative humidity appears to marginally affect the expected ΔT_1 and was not considered for the analysis.

The 10-min-averaged dataset was therefore filtered for zero solar radiation, wind speed below 0.5 m/s, on a 120°-wide angle centred on the road.

An additional filtering was performed on the 10-min dataset, in order to remove residual fast transients in temperature like those introduced by some föhn wind events, which are not so rare in the region (Cassardo *et al.*, 2007). This was performed by smoothing the time series with a running mean and removing all the instances where the local derivative of the temperature function was

greater (in absolute value) than a certain value (in this case 0.05 °C was chosen: this depends heavily on the smoothing performed earlier). Finally, only two or more consecutive observations were kept in the final dataset, in order to leave out residual spikes and retain observations at least 20 min long. Thus, the filtered working dataset consisted in ~1300 records (2.7 % of the 10-min averaged dataset), whose distributions are represented in Figure 12 and in Table 3 as main statistics summary.

3. Results

In order to associate confidence intervals to such rare events as maxima of ΔT_i , it is necessary to model the distribution of a sizeable number of them. For this reason, in order to perform a consistent and robust statistical analysis on such rare events, an Extreme Value Analysis (EVA, Fisher and Tippett, 1928; Gumbel, 1958) was applied.

3.1 Extreme Values Analysis

According to the EVA, observations lying above a sufficiently high threshold h can be modelled after a Generalized Pareto Distribution (GPD), whose cumulative distribution function assumes this analytical form:

$$G_{\xi,\sigma}(x) = \begin{cases} 1 - \left(1 + \frac{\xi x}{\sigma}\right)^{-1/\xi} & \xi \neq 0 \\ 1 - e^{-x/\sigma} & \xi = 0 \end{cases}$$

where σ is called “scale parameter” and ξ is called “shape parameter”.

This “Peaks Over Threshold” (POT) analysis allows for fitting the extreme data, calculating the probabilities of an excess occurring within a given period of time (“return period”) and associate a confidence interval to the extreme values (Goda, 1988). One of the most critical steps of this analysis is the choice of the threshold h . Several diagnostics have been developed to help during this process, like the mean residual life plot and the threshold stability plot (Coles, 2001), but simpler rules of thumb are also used, like choosing the p largest out of the total n observations: commonly used are $p = \sqrt{n}$, $p = n^{2/3} / \log(\log(n))$ and the 90th percentile (Scarrott and MacDonald, 2012). In fact, when choosing a threshold with one of the available methods, the user must take into account the

number of exceedances left for the analysis: if it is too large, too few values will be left to model the distribution tail correctly as the variance is likely to be large due to only very extreme observations remaining. On the other hand, choosing a threshold too low, too many exceedances will remain giving a high bias (Coles, 2001).

After the choice of thresholds and a declustering process, a GPD can be fitted to the remaining excesses, providing values for the parameters ξ and σ (Table 4). The usual way to visualize the results of the fitting is via the return level plots (Figure 13). In the plots, each dot represents a measured excess over the threshold, and the data is fitted with GPD's (red lines). The 68 % confidence intervals, represented as grey bands, are calculated via the delta method (Doob, 1935). Return period is strictly related to the probability of each return level (longer period corresponds to lower probability). All these calculations are performed using R package "extRemes" (Gilleland and Katz, 2016), selecting "GP" (Generalized Pareto) as the fit type.

Figure 14 shows the values of the ΔT_i maxima as calculated by the EVA, for three different return periods (1, 10 and 100 years), with error bars representing the 68 % uncertainty of the GPD fit at each distance. These intervals are then used as an uncertainty component in the overall budget shown in Table 1, which reports the complete budget for each of the three return periods, while the RSE values for the three return periods are tabulated in Table 5.

As expected, all the return values show a decreasing trend from the road. The 1-year return level goes from 1.25 °C at 1 m from the road, decreases sharply to 1 °C at 5 m then more slowly until 0.6 °C at 50 m. The rate of temperature decrease is roughly 0.1 °C every 10 m. Uncertainties are of the order of ~0.1 °C.

The 10- and 100-year return values show very similar trends, but uncertainties grow substantially, up to ~0.4 °C at 100-year return level. Both return values show a more linear behaviour: the 10-year level extrapolation shows a maximum ΔT_1 of 1.5 °C, going down to 1.3 °C at 20 m, then reaching 0.75 °C at 50 m; maximum difference values for the 100-year return level are around 1.7 °C then again down to 0.85 °C at 50 m.

4. Discussion

While this experiment was not intended to provide a final answer to the evaluation of RSE, given that larger, more or less trafficked, differently oriented roads in different environments could yield different results, we think that a local 2-laned road with light traffic, at mid-latitudes, may represent a sizeable fraction of the roads that can be found in the area, thus providing a reasonable representativeness for the chosen site. Moreover, this kind of roads is the most expected to be found near WMO stations, since those are often located in relatively rural regions. In fact, within the local large area presented by the Guide as representative of a site (100-1000 km²), the majority of roads are very similar to the one featured in this experiment: a little more elevated than the surrounding land, with a ditch nearby, both to prevent accumulation of water and to drain it properly. In this sense, these features can be regarded as representative and concur in generating the RSE.

Calibration of air temperature sensors is still an open issue in temperature metrology. There is currently no universally agreed-upon procedure for calibration of air temperature sensors directly in air so, in order to maintain traceability to the SI, a liquid bath calibration was necessary. The necessity to devise and code a metrological procedure for calibration in air is in fact included in the BIPM Consultative Committee on Thermometry, as a key objective of the roadmap up to 2027. There is currently a EURAMET project called ATM –led by INRiM – to research into the best calibration method in air, including radiative and convective analysis. The implication is that calibration of air sensors in liquid considerably underestimates uncertainties: for this reason, the co-location measurements mentioned in Section 2.2 were necessary in order to provide more realistic uncertainties, while retaining traceability to SI through conventional liquid bath calibration.

In terms of data analysis, in this work GAM was successfully used to isolate the largest effect on ΔT_i by each of the quantities of influence: while in principle a similar approach could be used to explore the interactions between these predictors, we think that such an analysis would go outside the scope of this work. Indeed, an experiment to study the interaction between predictors would be interesting, but would have to feature different instruments and most probably a different location: analyse the interaction of wind and radiation in a place where wind is almost always absent would have very little significance.

The other statistical technique employed in this work, EVA, has proven able to provide reliable estimates of extreme events probabilities, such as the maximum temperature differences (within

given time frames) measured between one reference point and any other point affected by a heat source.

As already mentioned in Section 3.1, the choice of thresholds is a rather delicate one. However, by trying with different values at very small steps, we noted that choosing a threshold too large suddenly yields return level values with extremely large uncertainties, so that these mistakes are easy to spot; on the other hand, choosing a threshold too low changes the overall fit very little, usually by flattening (for shorter distances from the road) or steepening (for longer distances) the weighted fits in Figure 14.

We note from the return level plots (Figure 13) that they do not represent the maximum possible ΔT_i : they just give an indication of them in a given period of time. However, the shape parameter ξ determines the boundedness of GPD: a value $\xi < 0$ leads to a bounded right tail, that is $x \in [0, -\sigma/\xi] + h$ (Shevchenko, 2011). As a matter of fact, in our case all fits returned negative ξ values, so it is theoretically possible to compute asymptotic fits for all ΔT_i maxima. The exercise seems to have little value though, given that all the confidence intervals of ξ include zero, resulting in unbounded asymptotic uncertainties.

At the beginning of the experiment there was no guarantee that a maximum distance of 100 m would be sufficient for the evaluation of RSE. It was chosen because it is the current reference distance adopted by the Guide: adding measurement stations farther away would have made the task of finding the site even harder, with an unclear outcome in usefulness. In order to evaluate the possibility of RSE even on Class-1 compliant sites, trends reported in Figure 14 have been extrapolated to 100 m by means of a weighted linear fit: in all cases, the trends are compatible, within the uncertainties, to $\Delta T_7 = 0$ °C, meaning no RSE.

4.1 Comparison with literature

As mentioned in the introduction, very few literature works dealt with RSE on temperature measurements, at least with the kind of metrological depth required for the characterization of measurements for climatological studies. Some works on effects of advected heat on temperature measurements do exist in literature (Bassett *et al.*, 2016, 2017), but they tend to refer to larger

scales and to compare urban and rural settings, rather than study the phenomenon per se in comparison to an unbiased setting.

Delgado *et al.*, (2007), among others, studied a similar but different issue: the *road edge effect*, i.e. the effect of the presence of a narrow corridor due to a road on a pre-existent wooded area. This coalesces very different issues – the effects of roads and trees on air temperature measurements– which in our works were kept separate, as well as others like different solar illumination and the wind canyon effect. Given that the cited work employed low-quality instruments like unshielded thermocouples and a measurement time of only 1 h, a direct comparison with our findings seems difficult. The authors find that the road edge effect vanishes merely few meters within the forest, which is likely due to the constant shadow of the trees, which prevents accumulation and subsequent gradual release of heat by the soil. Voldan (2016, private communication) found that an expanse of trees acts as a thermal capacitor, capable of effectively levelling the thermal gradients due to external factors.

Bogren *et al.*, (2001) measured temperatures at different heights above a road and few meters in the surroundings and found that the largest difference between them was around 1.1 °C at 2.5 m of height, a value compatible to our findings, considering that the nature of “surroundings” and the distance are never explicitly mentioned.

A more directly relatable work was performed by Kumamoto *et al.*, (2013). Their reference thermometer is placed 10 m from the road, so the measured differences are lower, the sensors were placed at different heights (0.5, 1.5 and 2.5 m), and all the comparisons are performed with respect to the temperature of the road itself. In their paper, they both used short (10 s) and long timescales (10 min), focusing more on the interaction with the ancillary measurements than the evaluation of the largest event. Moreover, no uncertainty evaluation is performed. For these reasons, only a cautious comparison can be made between their differences at 10 m and the difference between ΔT_1 and ΔT_3 calculated by EVA in our work: findings are similar and of the order of 0.25 °C.

It is also useful to compare the results of the present work with theoretical predictions made by Kinoshita, (2014) for ΔT_{pmax} , defined as the maximum value of daily ΔT_{max} . The simulation computed featuring a 10-m-wide road shows a similar ΔT profile to the one shown in Figure 14a, with a similar value of 1.1 °C for ΔT_1 , a steeper descent up to a distance of 20 m and lower ΔT values thereafter. This discrepancy can be due to a number of assumptions made during the simulation –

for instance a fixed temperature difference between the road surface and the grass, which of course depends on the nature of the soil and the kind of asphalt used, and the simulation being based on cloudless days only – that were, beside a simplification, also tailored for the atmospheric and geographic particularities of the Japanese environment and may not represent the actual features of a different site. Considering all these aspects, the results of Kinoshita seem comparable.

Results can be compared to the uncertainties stated in the Guide for classes 3 through 5. The Guide provides much larger errors: compare the 5 °C values for class 5 to the 1.25 °C value found in this work; 2 °C from the Guide for class 4 to 1 °C in this work. It must be noted that the current classification mixes up many different contributions to each class – presence of obstacles, terrain with a slope or untrimmed grass, shades etc. It is likely that combinations of different issues and obstacles for a given station can raise the uncertainty on the air temperature measurement up to values closer to the ones given in the Guide, at least for some of the current classes.

5. Conclusions

In the framework of project MeteoMet 2, an experiment dedicated to the evaluation of near surface air temperature uncertainty in measurements due to the presence of a road has been presented. An array of seven measurement points, at increasing distances from a road, has been built and the 8 Pt100 thermometers, 2 hygrometers, 1 radiation meter and 1 anemometer have acquired data every 30 s for 1 year. At the end of the main experiment, a rearrangement of the whole array in just one measurement point has been performed to evaluate the behaviour of the temperature sensors in the same environmental conditions and evaluate measurement uncertainties.

A Generalized Additive Model (GAM) was developed to explore the relationships between the quantities of influence and the temperature differences between stations.; the model results indicated a way to effectively reduce the amount of data to analyse, focusing only on data recorded during nights, with low-speed winds coming from the road, when the effect was found to be maximum.

An Extreme Values Analysis (EVA) has then been carried out. By fitting the extreme ΔT_i over a threshold, return levels – and their uncertainties – at different return periods are obtained. The analysis revealed a rate of error decrease of ~ 0.1 °C every 10 m for a return period of 1 year, rising

to 0.17 °C for a return period of 100 years, which can be easily parametrized and taken into account in the general frame of completing the evaluation of measurement uncertainties in near-surface temperatures. Mean temperature differences are evaluated as less than 0.2 °C in the vicinity of the road.

By providing a metrological evaluation of uncertainties, these results show how the current classification could be improved, together with other effects evaluated within project MeteoMet 2 – which will be presented elsewhere. These results suggest that particular sources of influence to air temperature measurements can and must be investigated, in order to provide a more robust evaluation of the uncertainties. Similar experiments can be set up to evaluate the influence of other parameters cited in the Guide – bodies of water, terrain slopes – and even different properties of already investigated ones – roads of different widths, or with less/more intense traffic, in different climates. All these findings could be then parametrized and the future Guide could provide users with a much finer evaluation of thermal uncertainties, with the goal of providing much higher quality data for all meteorological and climatological studies. The main goal of this work was to provide a better evaluation of the maximum uncertainties related to the presence of a road on temperature measurements. The idea is to associate to any station, and therefore to any time series – either for meteorological or climatological studies – a metadata with quantified uncertainties on how reliable these measurements are, so that this information can be added to the models that are used for weather forecasts or the analysis of climatic trends.

Acknowledgements

This work has been developed within the frame of the EMPIR (European Metrology Programme for Innovation and Research), jointly funded by the EMPIR participating countries within EURAMET and the European Union. GJS acknowledges funding from the Netherlands Organization for Scientific Research (NWO) VIDI Grant 'The Windy City' (file no. 864.14.007).

We wish to thank the personnel of ICP Srl at Castelnuovo don Bosco airfield for their kind hospitality and effective logistical support for the set-up and operation of the experiment. We also acknowledge Francesca Romana Pennechi and Chiara Musacchio of INRiM for useful discussions and field work.

References

- Arnfield AJ. 2003. Two decades of urban climate research: A review of turbulence, exchanges of energy and water, and the urban heat island. *International Journal of Climatology*, 23(1): 1–26. <https://doi.org/10.1002/joc.859>.
- Bassett R, Cai X, Chapman L, Heaviside C, Thornes JE. 2017. The Effects of Heat Advection on UK Weather and Climate Observations in the Vicinity of Small Urbanized Areas. *Boundary-Layer Meteorology*. Springer Netherlands, 165(1): 181–196. <https://doi.org/10.1007/s10546-017-0263-0>.
- Bassett R, Cai X, Chapman L, Heaviside C, Thornes JE, Muller CL, Young DT, Warren EL. 2016. Observations of urban heat island advection from a high-density monitoring network. *Quarterly Journal of the Royal Meteorological Society*, 142(699): 2434–2441. <https://doi.org/10.1002/qj.2836>.
- Batagelj V, Bojkovski J, Drnovšek J, Pušnik I. 2003. Influence of SPRT Self-Heating on Measurement Uncertainty in Fixed Point Calibration and Calibration by Comparison. *AIP Conference Proceedings*. AIP Publishing, 315–320.
- Bernhardt J, Carleton AM, LaMagna C. 2018. A comparison of daily temperature-averaging methods: Spatial variability and recent change for the CONUS. *Journal of Climate*, 31(3): 979–996. <https://doi.org/10.1175/JCLI-D-17-0089.1>.
- Bertaccini P, Dukic V, Ignaccolo R. 2012. Modeling the short-term effect of traffic and meteorology on air pollution in Turin with Generalized Additive Models. *Advances in Meteorology*, 2012: 16. <https://doi.org/10.1155/2012/609328>.
- BIPM, Joint Committee For Guides In Metrology. 2008. *Evaluation of measurement data — Guide to the expression of uncertainty in measurement*. International Organization for Standardization Geneva ISBN.
- Bogren J, Gustavsson T, Karlsson M. 2001. Temperature differences in the air layer close to a road surface. *Meteorological Applications*, 8(4): 385–395. <https://doi.org/10.1017/S1350482701004017>.
- Cassardo C, Fratianni S, Cremonini R. 2007. A study on the foehn in Piedmont, Italy. *29TH INTERNATIONAL CONFERENCE ON ALPINE METEOROLOGY (ICAM)*. Meteo France, 389–392.
- Coles S. 2001. *An Introduction to Statistical Modeling of Extreme Values*. Springer. Springer London: London.
- Coppa G, Merlone A. 2016. A study on ITS-90 type 3 non-uniqueness between freezing points of Al and Ag. *Measurement: Journal of the International Measurement Confederation*. Elsevier Ltd, 89: 109–113. <https://doi.org/10.1016/j.measurement.2016.04.014>.
- Coppa G, Musacchio C, Merlone A. 2020. Road Siting Experiment. Zenodo, <https://doi.org/10.5281/ZENODO.4300299>.
- Cox DTC, Maclean IMD, Gardner AS, Gaston KJ. 2020. Global variation in diurnal asymmetry in temperature, cloud cover, specific humidity and precipitation and its association with leaf area index. *Global Change Biology*, 26(12): 7099–7111. <https://doi.org/10.1111/gcb.15336>.
- Delgado JD, Arroyo NL, Arévalo JR, Fernández-Palacios JM. 2007. Edge effects of roads on temperature, light, canopy cover, and canopy height in laurel and pine forests (Tenerife, Canary Islands). *Landscape and Urban Planning*, 81(4): 328–340. <https://doi.org/10.1016/j.landurbplan.2007.01.005>.
- Doob JL. 1935. The Limiting Distributions of Certain Statistics. *The Annals of Mathematical Statistics*, 6(3): 160–169. <https://doi.org/10.1214/aoms/1177732594>.

Drusch M, Viterbo P. 2007. Assimilation of screen-level variables in ECMWF's integrated forecast system: A study on the impact on the forecast quality and analyzed soil moisture. *Monthly Weather Review*, 135(2): 300–314. <https://doi.org/10.1175/MWR3309.1>.

Ebert E, Wilson L, Weigel A, Mittermaier M, Nurmi P, Gill P, Göber M, Joslyn S, Brown B, Fowler T, Watkins A. 2013. Progress and challenges in forecast verification. *Meteorological Applications*. John Wiley and Sons Ltd, 20(2): 130–139. <https://doi.org/10.1002/met.1392>.

Erell E, Leal V, Maldonado E. 2005. Measurement of air temperature in the presence of a large radiant flux: An assessment of passively ventilated thermometer screens. *Boundary-Layer Meteorology*. Springer, 114(1): 205–231. <https://doi.org/10.1007/s10546-004-8946-8>.

Fisher RA, Tippett LHC. 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*. Cambridge University Press, 24(2): 180–190. <https://doi.org/10.1017/S0305004100015681>.

Gilleland E, Katz RW. 2016. ExtRemes 2.0: An extreme value analysis package in R. *Journal of Statistical Software*, 72(8). <https://doi.org/10.18637/jss.v072.i08>.

Goda Y. 1988. ON THE METHODOLOGY OF SELECTING DESIGN WAVE HEIGHT. *Coastal Engineering Proceedings*. Publ by ASCE, 1(21)(67): 899–913. <https://doi.org/10.9753/icce.v21.67>.

Gumbel EJ. 1958. *Statistics of Extremes*. Columbia. Columbia University Press.

Hastie T, Tibshirani R. 1986. Generalized Additive Models. *Statistical Science*. Institute of Mathematical Statistics, 1(3): 297–310. <https://doi.org/10.1214/ss/1177013604>.

Heusinkveld BG, Steeneveld GJ, Van Hove LWA, Jacobs CMJ, Holtslag AAM. 2014. Spatial variability of the rotterdam urban heat island as influenced by urban land use. *Journal of Geophysical Research*. Wiley-Blackwell, 119(2): 677–692. <https://doi.org/10.1002/2012JD019399>.

Ibrahim SH, Ibrahim NIA, Wahid J, Goh NA, Koesmeri DRA, Nawi MNM. 2018. The Impact of Road Pavement on Urban Heat Island (UHI) Phenomenon. *International Journal of Technology*, 9(8): 1597–1608. <https://doi.org/10.14716/ijtech.v9i8.2755>.

Karl TR, Knight RW, Gallo KP, Peterson TC, Jones PD, Kukla G, Plummer N, Razuvayev V, Lindsey J, Charlson RJ. 1993. A New Perspective on Recent Global Warming: Asymmetric Trends of Daily Maximum and Minimum Temperature. *Bulletin of the American Meteorological Society*, 74(6): 1007–1023. [https://doi.org/10.1175/1520-0477\(1993\)074<1007:ANPORG>2.0.CO;2](https://doi.org/10.1175/1520-0477(1993)074<1007:ANPORG>2.0.CO;2).

Kinoshita N. 2014. An Evaluation Method of the Effect of Observation Environment on Air Temperature Measurement. *Boundary-Layer Meteorology*, 152(1): 91–105. <https://doi.org/10.1007/s10546-014-9918-2>.

Koopmans S, Theeuwes NE, Steeneveld GJ, Holtslag AAM. 2015. Modelling the influence of urbanization on the 20th century temperature record of weather station De Bilt (The Netherlands). *International Journal of Climatology*, 35(8): 1732–1748. <https://doi.org/10.1002/joc.4087>.

Kowal A, Merlone A, Sawiński T. 2020. Long-term stability of meteorological temperature sensors. *Meteorological Applications*, 27(1). <https://doi.org/10.1002/met.1795>.

Kumamoto M, Otsuka M, Sakai T, Hamagami T, Kawamura H, Aoshima T, Fujibe F. 2013. Field Experiment on the Effects of a Nearby Asphalt Road on Temperature Measurement. *SOLA*, 9(1): 56–59. <https://doi.org/10.2151/sola.2013-013>.

Leeper RD, Kochendorfer J, Henderson TA, Palecki MA. 2019. Impacts of Small-Scale Urban Encroachment on Air Temperature Observations. *Journal of Applied Meteorology and Climatology*, 58(6): 1369–1380. <https://doi.org/10.1175/JAMC-D-19-0002.1>.

Lin T, Matzarakis A, Hwang R, Huang Y. 2010. Effect of pavements albedo on long-term outdoor thermal

comfort. *Proceedings of the 7 th Conference on Biometeorology*, 497–503.

Lin X, Hubbard KG. 2008. What are daily maximum and minimum temperatures in observed climatology? *International Journal of Climatology*. John Wiley & Sons, Ltd, 28(3): 283–294. <https://doi.org/10.1002/joc.1536>.

Lopardo G, Bertiglia F, Curci S, Roggero G, Merlone A. 2014. Comparative analysis of the influence of solar radiation screen ageing on temperature measurements by means of weather stations. *International Journal of Climatology*, 34(4): 1297–1310. <https://doi.org/10.1002/joc.3765>.

Madjidi F, Bidokhti AA, Ghader S, Mansori N. 2013. Cooling and Warming Effects of a Grass Covered Area and Adjacent Asphalt Area in a Hot Day. *International Journal of Occupational Hygiene*, 5(2): 56–63.

Mahmood R, Foster SA, Logan D. 2006. The GeoProfile metadata, exposure of instruments, and measurement bias in climatic record revisited. *International Journal of Climatology*, 26(8): 1091–1124. <https://doi.org/10.1002/joc.1298>.

Mahmood R, Pielke RA, Hubbard KG, Niyogi D, Dirmeyer PA, Mcalpine C, Carleton AM, Hale R, Gameda S, Beltrán-Przekurat A, Baker B, McNider R, Legates DR, Shepherd M, Du J, Blanken PD, Frauenfeld OW, Nair US, Fall S. 2014. Land cover changes and their biogeophysical effects on climate. *International Journal of Climatology*, 34(4): 929–953. <https://doi.org/10.1002/joc.3736>.

McNider RT, Steeneveld GJ, Holtslag AAM, Pielke RA, Mackaro S, Pour-Biazar A, Walters J, Nair U, Christy J. 2012. Response and sensitivity of the nocturnal boundary layer over land to added longwave radiative forcing. *Journal of Geophysical Research: Atmospheres*, 117(D14): 106. <https://doi.org/10.1029/2012jd017578>.

Merlone A, Lopardo G, Sanna F, Bell S, Benyon R, Bergerud RA, Bertiglia F, Bojkovski J, Böse N, Brunet M, Cappella A, Coppa G, del Campo D, Dobre M, Drnovsek J, Ebert V, Emardson R, Fernicola V, Flakiewicz K, Gardiner T, Garcia-Izquierdo C, Georgin E, Gilabert A, Grykalowska A, Grudniewicz E, Heinonen M, Holmsten M, Hudoklin D, Johansson J, Kajastie H, Kaykisizli H, Klason P, Kňazovická L, Lakka A, Kowal A, Müller H, Musacchio C, Nwaboh J, Pavlasek P, Piccato A, Pitre L, de Podesta M, Rasmussen MK, Sairanen H, Smorgon D, Sparasci F, Strnad R, Szmyrka-Grzebyk A, Underwood R. 2015. The MeteoMet project - metrology for meteorology: Challenges and results. *Meteorological Applications*, 22(January): 820–829. <https://doi.org/10.1002/met.1528>.

Merlone A, Sanna F, Beges G, Bell S, Beltramino G, Bojkovski J, Brunet M, del Campo D, Castrillo A, Chiodo N, Colli M, Coppa G, Cuccaro R, Dobre M, Drnovsek J, Ebert V, Fernicola V, Garcia-Benadí A, Garcia-Izquierdo C, Gardiner T, Georgin E, Gonzalez A, Groselj D, Heinonen M, Hernandez S, Högström R, Hudoklin D, Kalemci M, Kowal A, Lanza L, Miao P, Musacchio C, Nielsen J, Nogueras-Cervera M, Oguz Aytekin S, Pavlasek P, de Podesta M, Rasmussen MK, Del-Río-Fernández J, Rosso L, Sairanen H, Salminen J, Sestan D, Šindelářová L, Smorgon D, Sparasci F, Strnad R, Underwood R, Uytun A, Voldan M. 2018. The MeteoMet2 project—highlights and results. *Measurement Science and Technology*, 29(2): 025802. <https://doi.org/10.1088/1361-6501/aa99fc>.

Musacchio C, Coppa G, Merlone A. 2019. An experimental method for evaluation of the snow albedo effect on near-surface air temperature measurements. *Meteorological Applications*, 26(1): 161–170. <https://doi.org/10.1002/met.1756>.

Oke TR. 1982. The energetic basis of the urban heat island. *Quarterly Journal of the Royal Meteorological Society*, 108(455): 1–24. <https://doi.org/10.1002/qj.49710845502>.

Pavlasek P, Merlone A, Sanna F, Coppa G, Izquierdo CG, Palencar J, Duris S. 2020. Determination of automatic weather station self-heating originating from accompanying electronics. *Meteorological Applications*, 27(1). <https://doi.org/10.1002/met.1844>.

Peterson TC, Easterling DR, Karl TR, Groisman P, Nicholls N, Plummer N, Torok S, Auer I, Boehm R, Gullet D, Vincent L, Heino R, Tuomenvirta H, Mestre O, Szentimrey T, Salinger J, Førland EJ, Hanssen-Bauer I,

Alexandersson H, Jones P, Parker D. 1998. Homogeneity adjustments of in situ atmospheric climate data: A review. *International Journal of Climatology*, 18(13): 1493–1517. [https://doi.org/10.1002/\(SICI\)1097-0088\(19981115\)18:13<1493::AID-JOC329>3.0.CO;2-T](https://doi.org/10.1002/(SICI)1097-0088(19981115)18:13<1493::AID-JOC329>3.0.CO;2-T).

Pielke RA, Davey CA, Niyogi D, Fall S, Steinweg-Woods J, Hubbard K, Lin X, Cai M, Kim YK, Li H, Nielsen-Gammon J, Gallo K, Hale R, Mahmood R, Foster S, McNider RT, Blanken P. 2007. Unresolved issues with the assessment of multidecadal global land surface temperature trends. *Journal of Geophysical Research Atmospheres*. Blackwell Publishing Ltd, 112(24): D24S08. <https://doi.org/10.1029/2006JD008229>.

Quarello A. 2017. Experimental evaluation of temperature uncertainty components due to siting condition with respect to WMO classification. Università di Torino.

Runnalls KE, Oke TR. 2006. A technique to detect microclimatic inhomogeneities in historical records of screen-level air temperature. *Journal of Climate*. American Meteorological Society, 19(6): 959–978. <https://doi.org/10.1175/JCLI3663.1>.

Scarrott C, MacDonald A. 2012. A review of extreme value threshold estimation and uncertainty quantification. *Revstat Statistical Journal*, 10(1): 33–60.

Shevchenko P V. 2011. *Modelling operational risk using bayesian inference*. Springer. Springer Berlin Heidelberg: Berlin, Heidelberg.

Vose RS, Easterling DR, Gleason B. 2005. Maximum and minimum temperature trends for the globe: An update through 2004. *Geophysical Research Letters*, 32(23): L23822. <https://doi.org/10.1029/2005GL024379>.

WMO. 2012. *WMO/CIMO #8 Guide to meteorological instruments and methods of observation*. .

WMO. 2018. *Guide to climatological practices, second edition*. .

Wood SN. 2003. Thin plate regression splines. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 65(1): 95–114. <https://doi.org/10.1111/1467-9868.00374>.

Wood SN. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 73(1): 3–36. <https://doi.org/10.1111/j.1467-9868.2010.00749.x>.

Wood SN. 2017. *Generalized additive models: An introduction with R, second edition*. CRC. Chapman and Hall/CRC.

Figure 1. Layout of the road experiment. For each measurement point, at the specified distance from the obstacle, position, number and nature of the instruments installed are also indicated.

92x92mm (300 x 300 DPI)

Figure 2. Google Earth snapshot (© Google, 2017) of the experimental site. Visible in the picture are the paved road, the experiment array, the grassy airstrip and the main airfield building. In order to maintain the prescribed distance to the building in the East and the wooded area in the West, the array could not be positioned perpendicular to the road: an angle of $\sim 75^\circ$ was chosen instead. For this reason, the nominal distances of the sensors to the road are to be reduced by $1 - \sin(75^\circ) \cong 3.4\%$.

409x244mm (96 x 96 DPI)

Figure 3. All the measurement stations in their definitive configuration. The picture was taken right after the end of the installation works: the barren land along the line of the array, due to earthworks, was naturally restored to its grassy status after few weeks, and data acquired during this recovery phase was excluded from the analysis.

Figure 4. The sensors' configuration during the measurements uncertainties evaluation stage.

Figure 5. Temperature differences recorded during the measurement uncertainty evaluation stage of the experiment. For clarity, only ΔT_1 has been plotted.

319x179mm (600 x 600 DPI)

Figure 6. Distributions of ΔT_{a_i} during co-location of the measurement stations. Blue vertical dashed lines represent distributions' means.

319x179mm (600 x 600 DPI)

Figure 7. Temperature differences recorded during the main experiment. Again, for clarity, only ΔT_1 has been plotted.

319x179mm (600 x 600 DPI)

Figure 8. Distributions of ΔT_i recorded during the main experiment. Blue vertical dashed lines represent distributions' means.

319x179mm (600 x 600 DPI)

Figure 9. Distributions of ΔT_i calculated at daily maximum, average and minimum temperatures.

319x179mm (300 x 300 DPI)

Figure 10. Graphic representation of GAM results. The four quadrants represent the influence of radiation, wind speed, humidity and wind direction respectively (clockwise, from top left). Humidity values lower than 25 %, not reliable as per the manufacturer's specs, are cropped out the plot. On top of each panel, marginal density plots of the ancillary data are represented.

319x179mm (600 x 600 DPI)

Figure 11. A windrose plot of the site. The red line represents the measurement stations' alignment.

319x179mm (600 x 600 DPI)

Figure 12. Distributions of ΔT_i from filtered data, and their means.

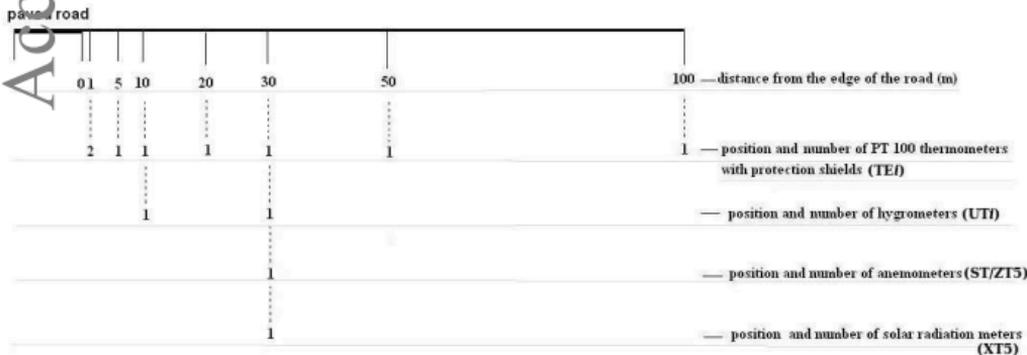
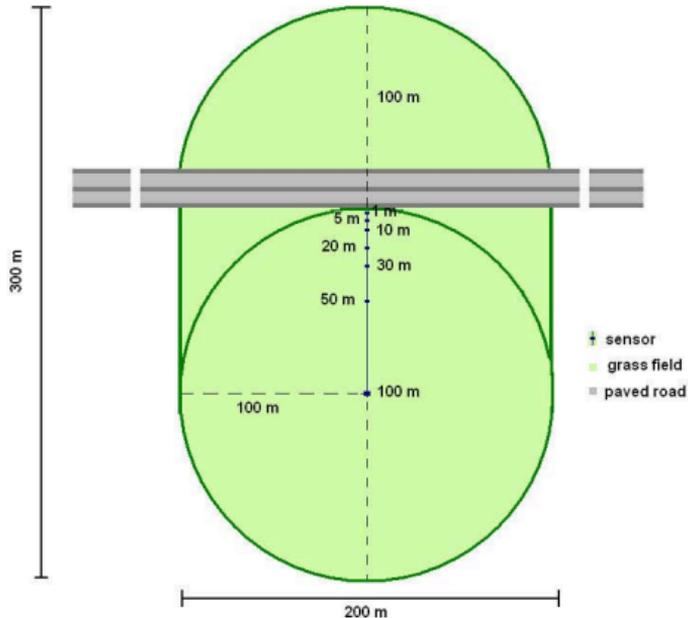
319x179mm (600 x 600 DPI)

Figure 13. Return level plots of ΔT_i excesses over threshold. The red solid lines are the Generalized Pareto distribution fits, while the shaded bands represent the 68 % confidence levels.

319x179mm (600 x 600 DPI)

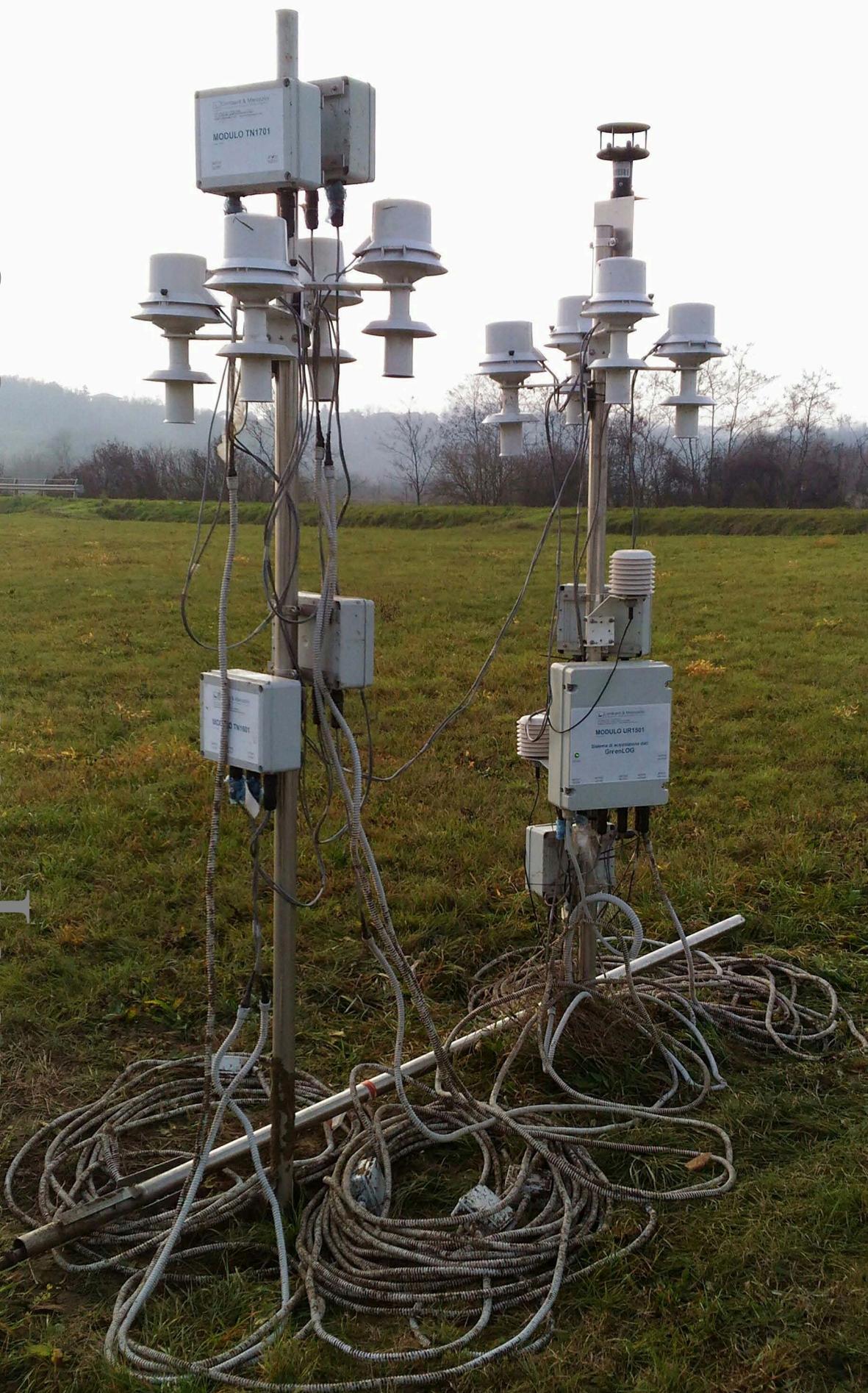
Figure 14. Plot of ΔT_i maxima as calculated by the Extreme Values Analysis for 1-year, 10-year and 100-year return periods. Points represent the return level values, at different distances from the road, of the Generalized Pareto fits, with error bars representing the 68 % uncertainty of the GPD fit. The dot-dashed lines represent a weighted linear fit, extrapolated to 100 m, with 68 % confidence bands shown in grey.

319x179mm (600 x 600 DPI)

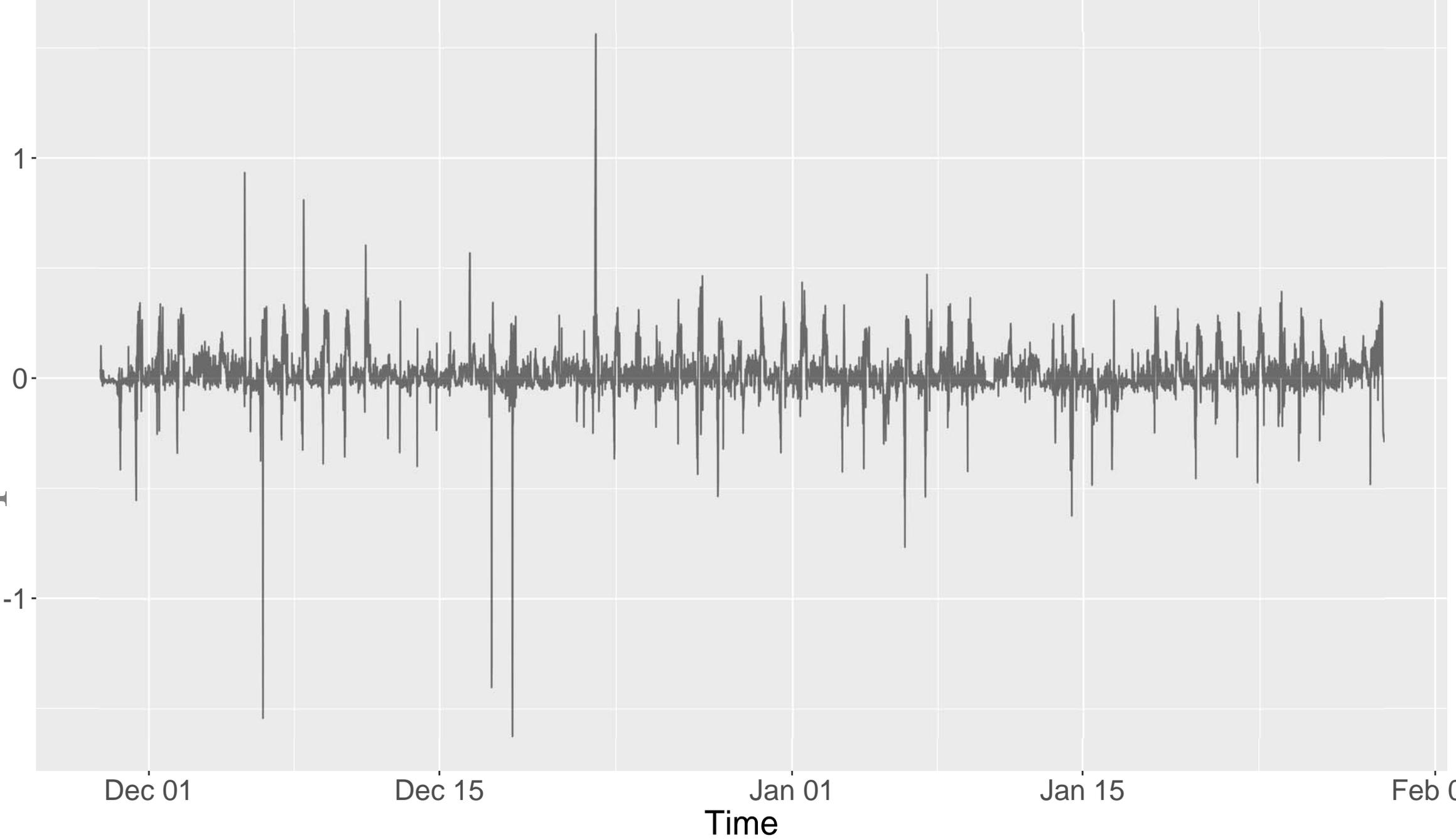


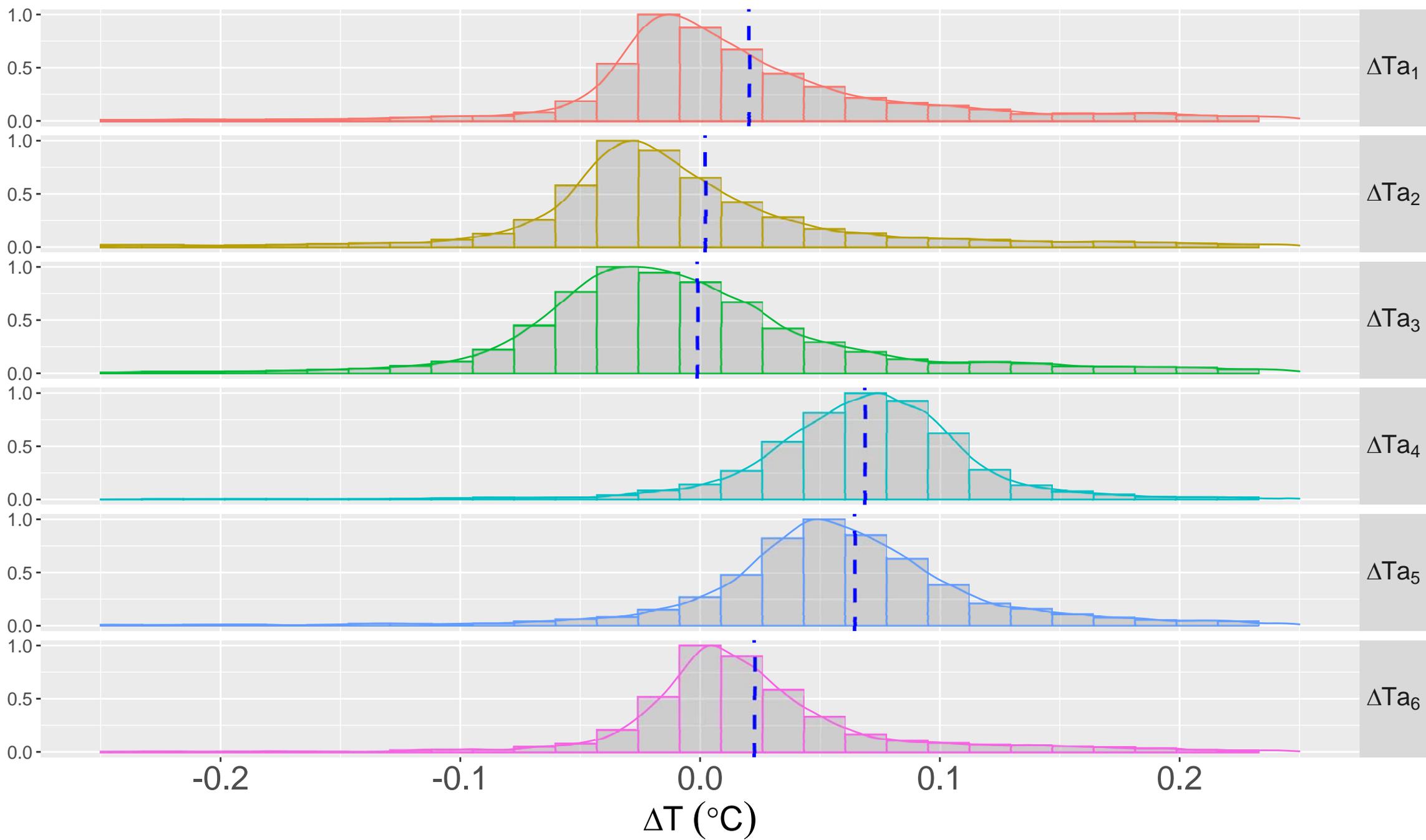




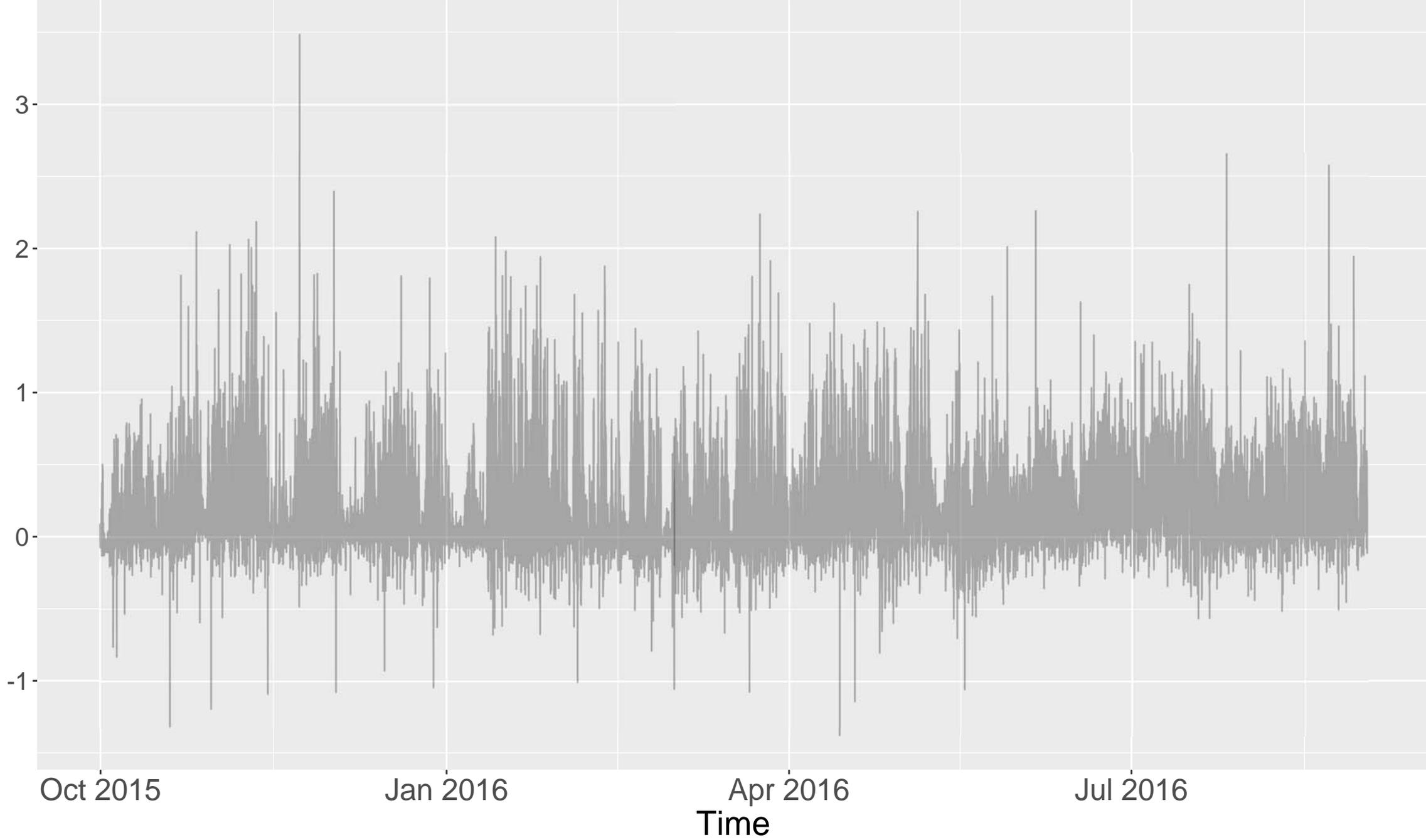


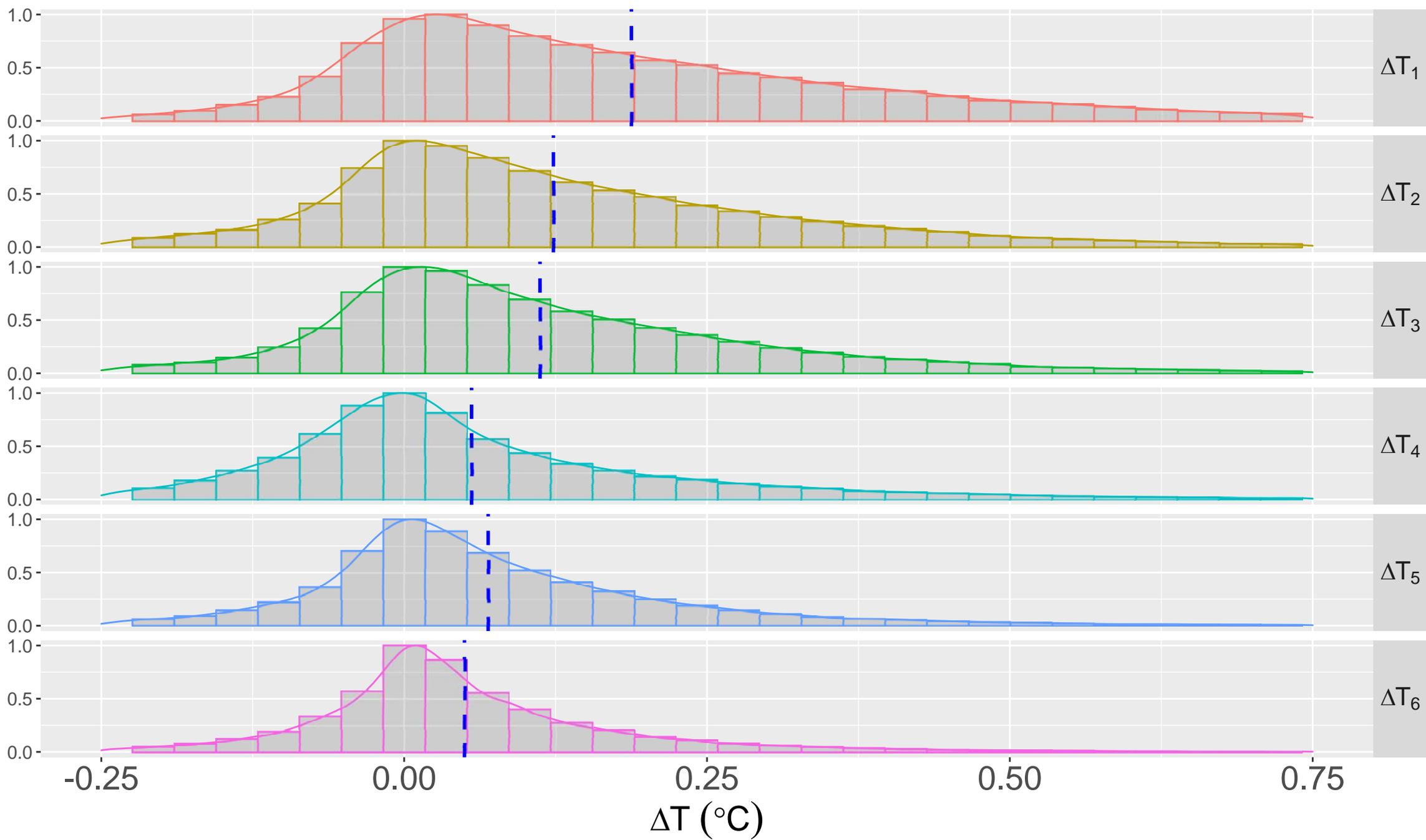
Accepted Article

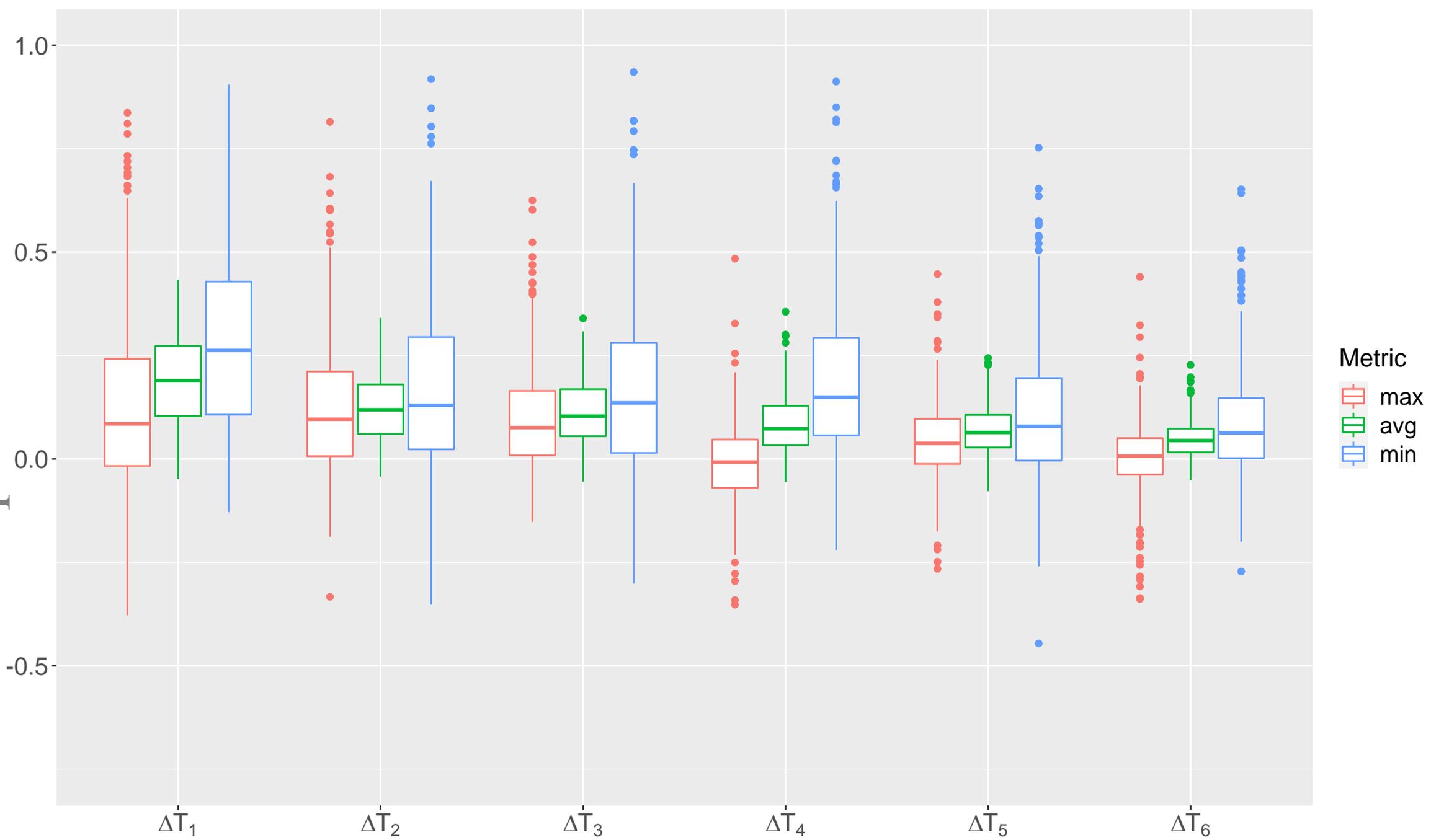


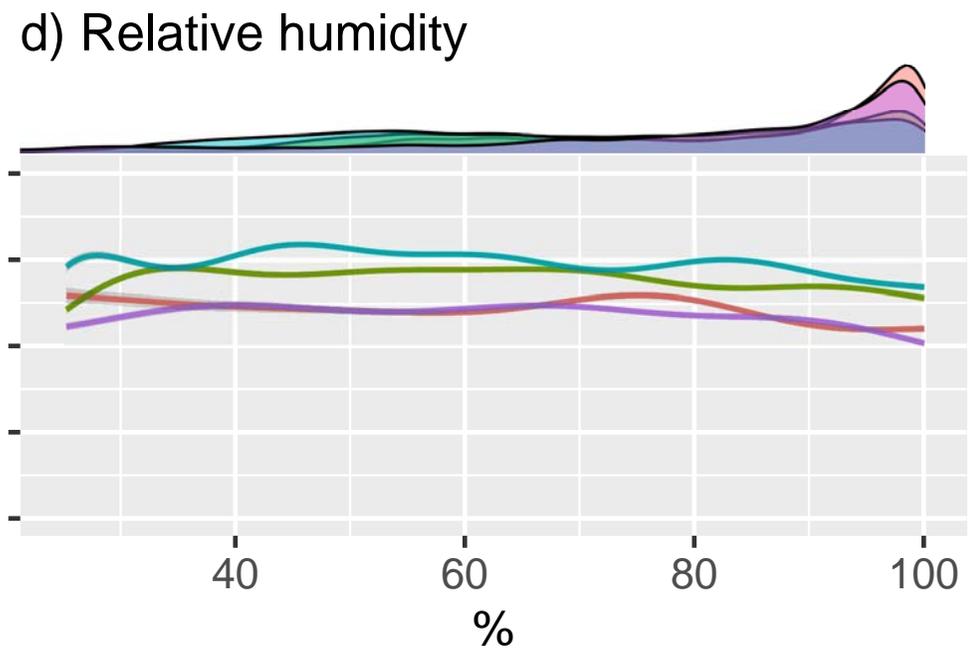
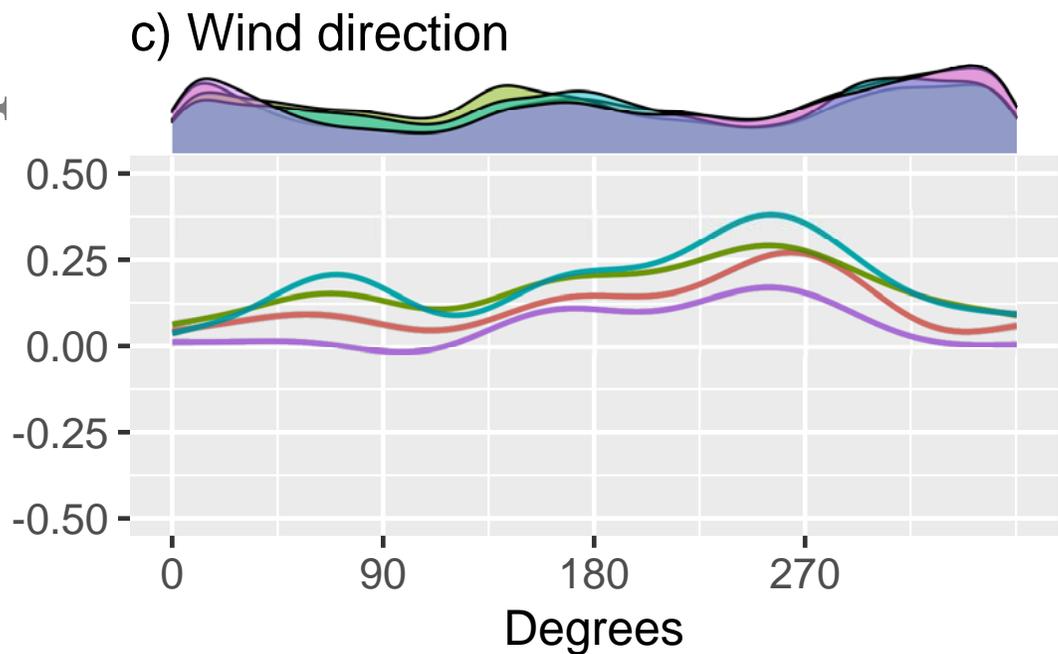
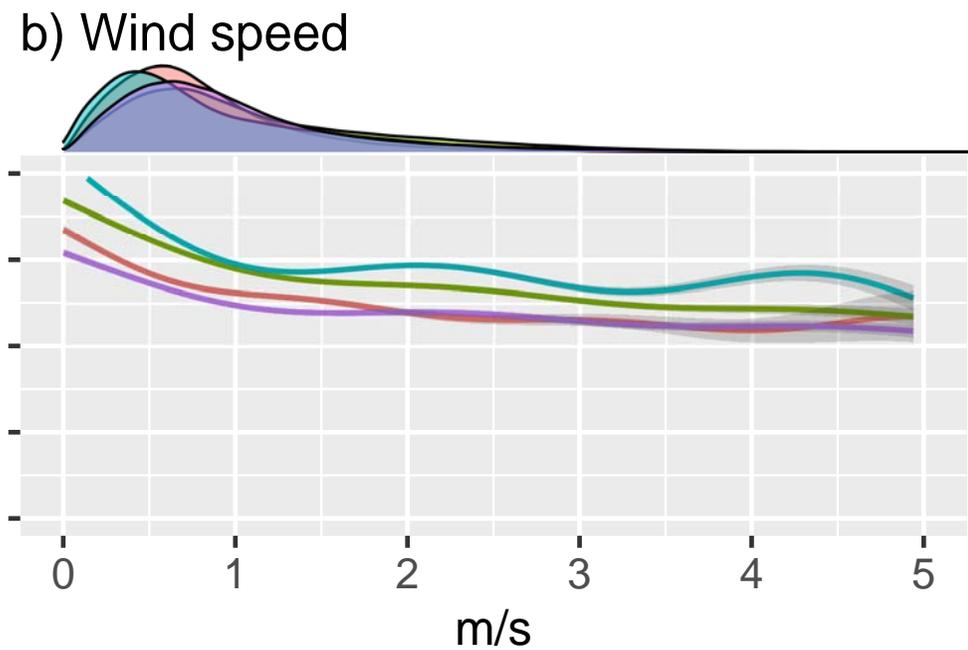
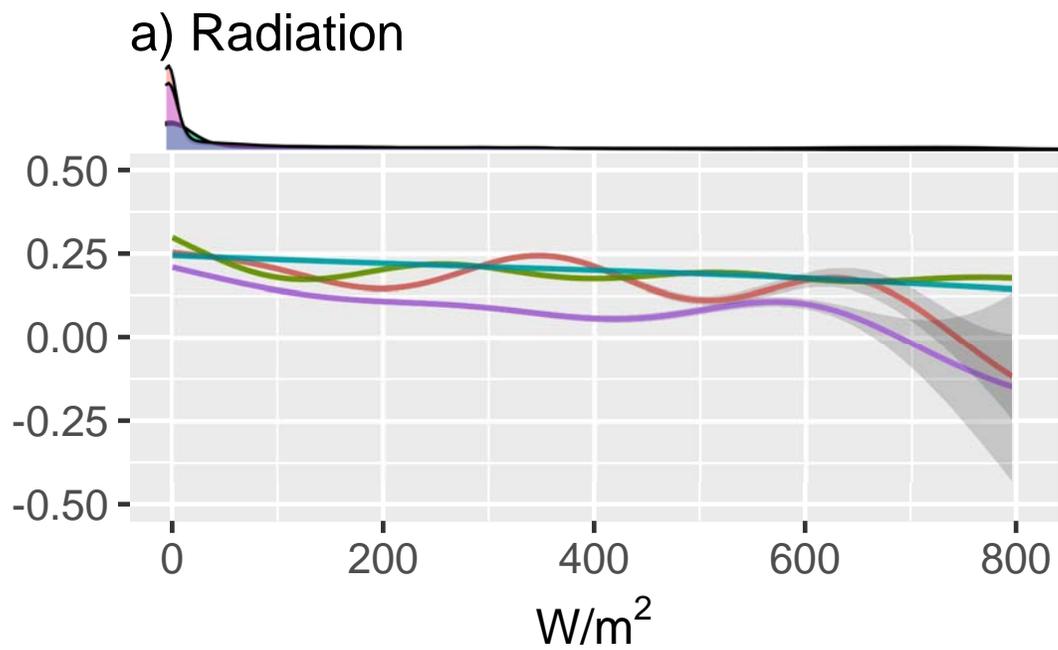


Accepted Article



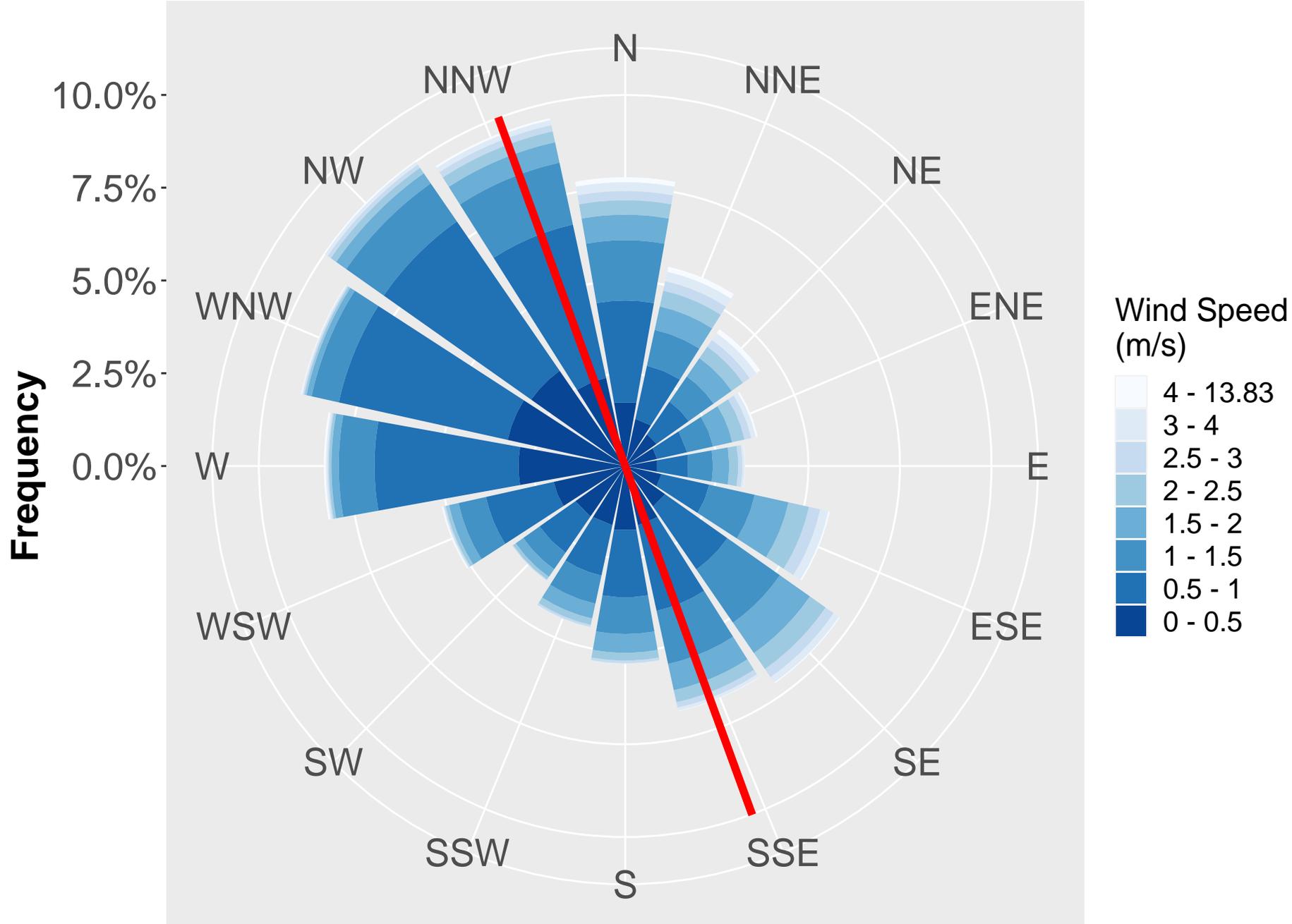


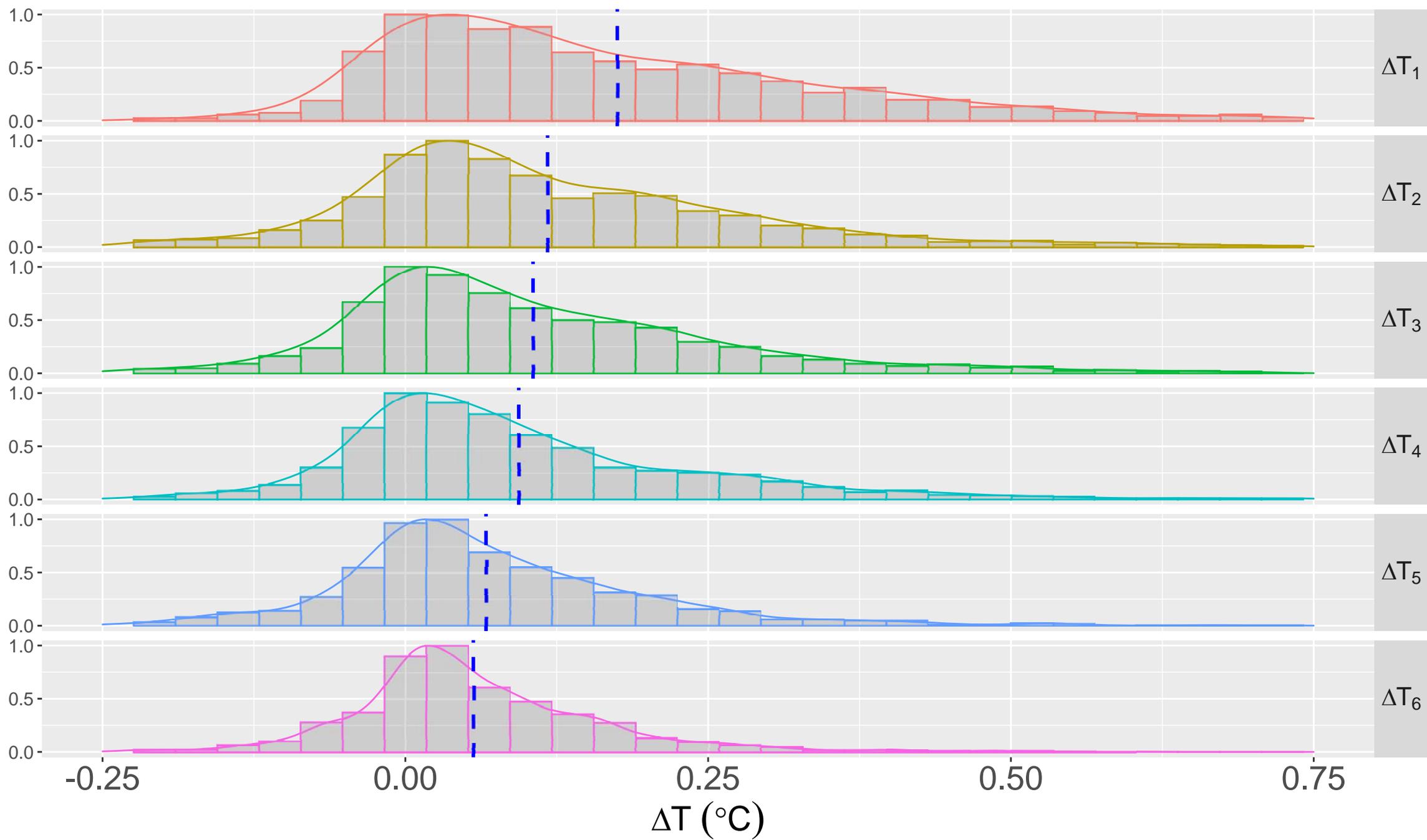


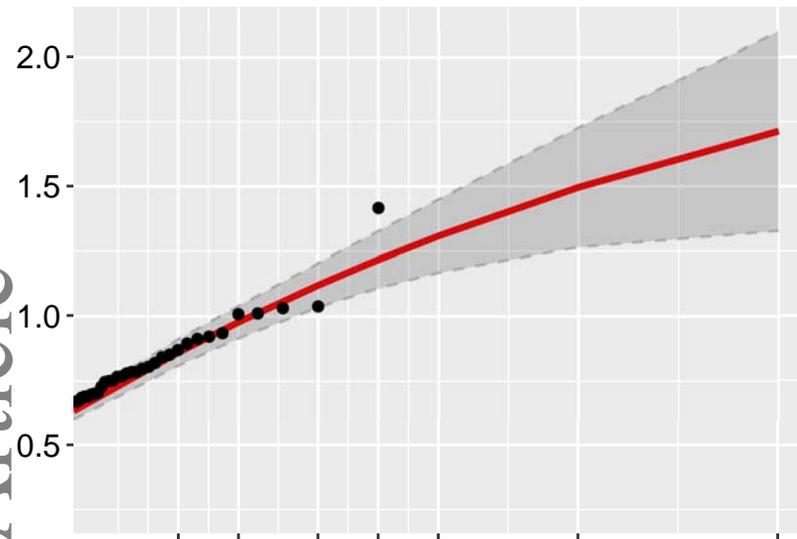
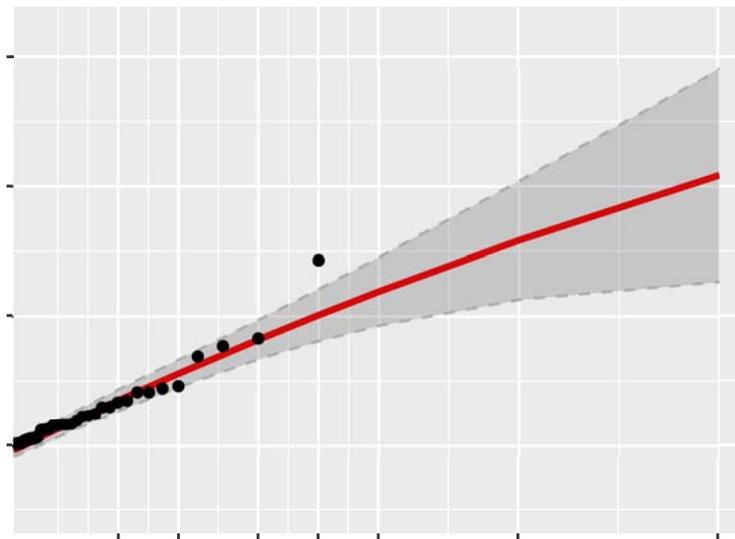
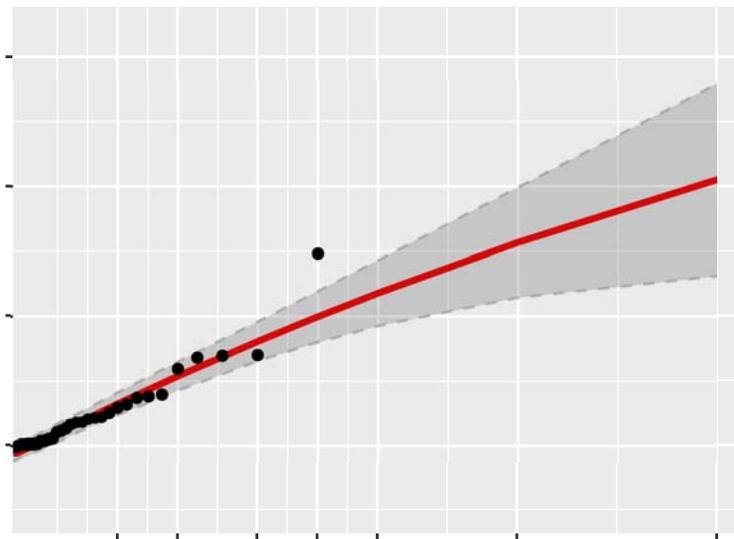
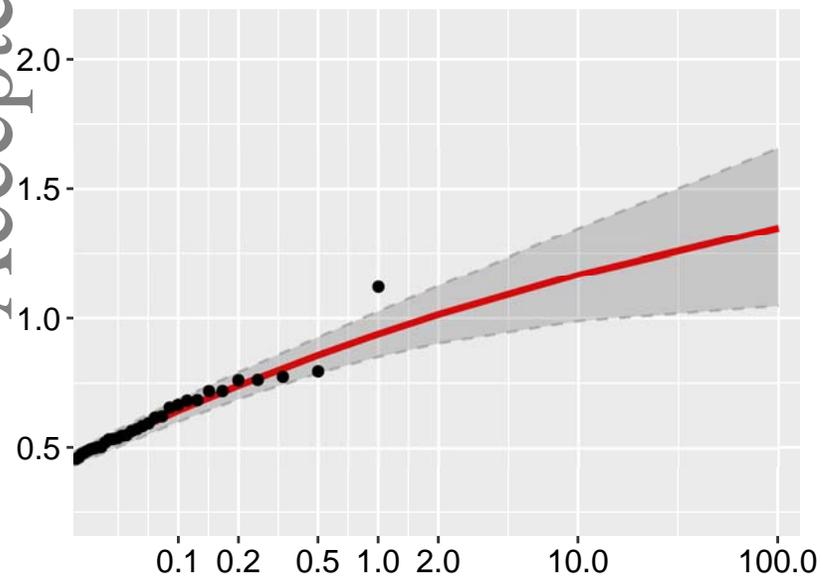
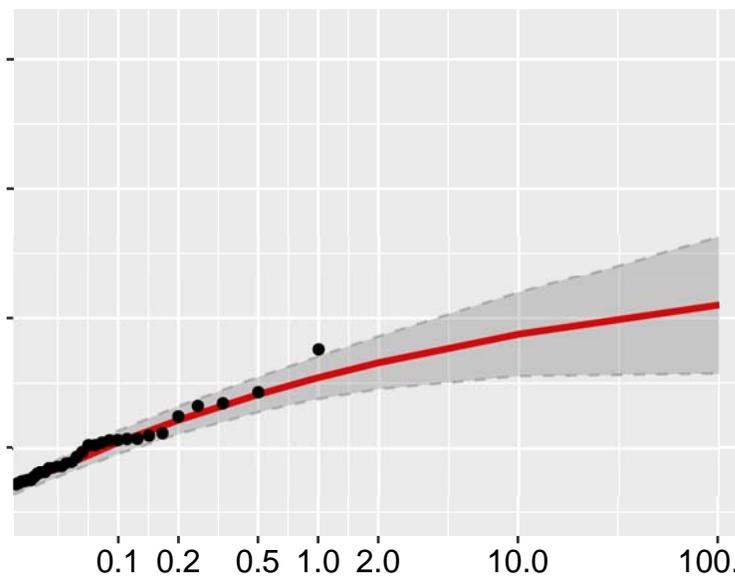
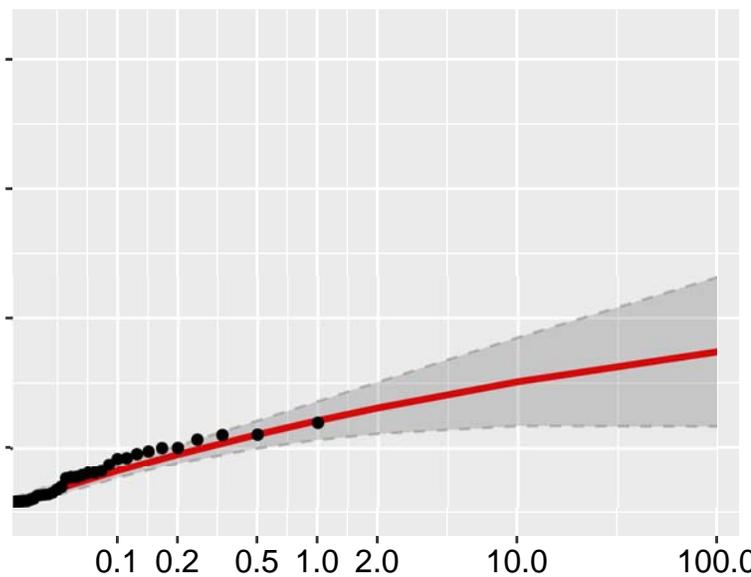


season

- autumn
- spring
- summer
- winter

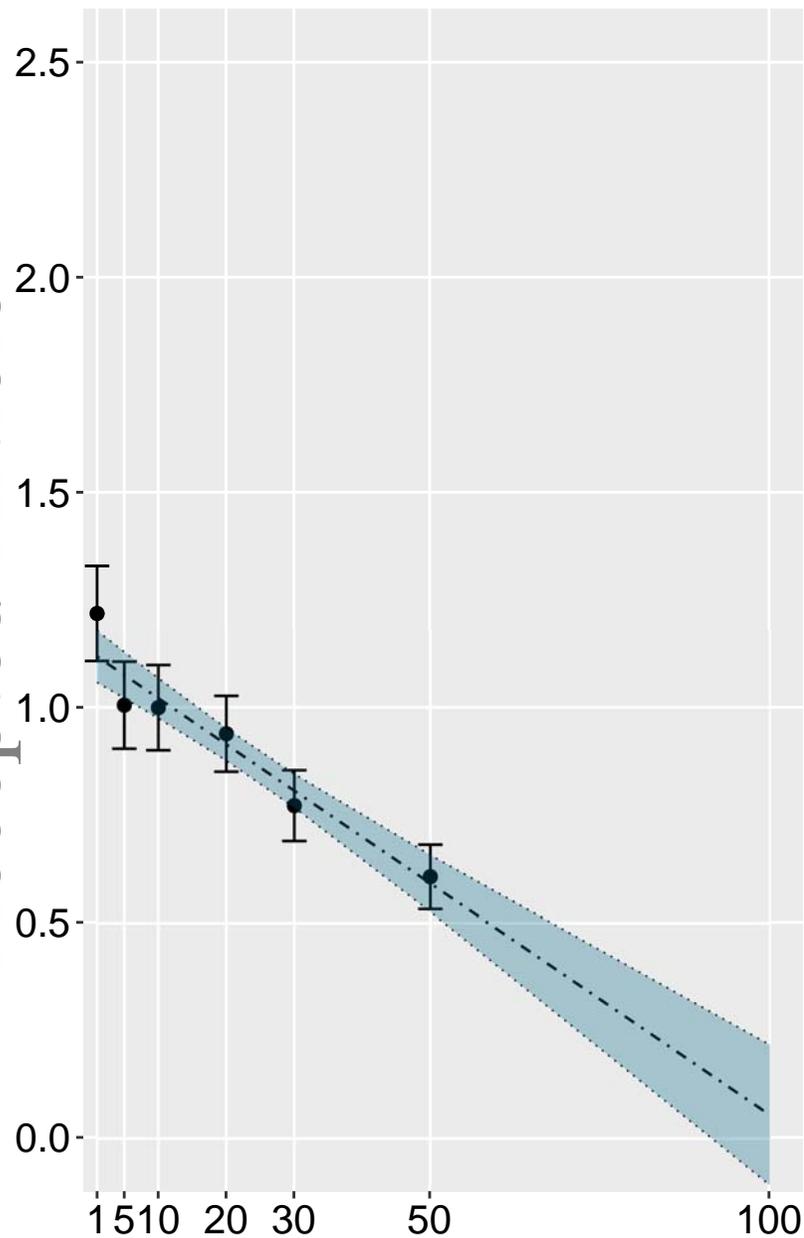




a) ΔT_1 b) ΔT_2 c) ΔT_3 d) ΔT_4 e) ΔT_5 f) ΔT_6 

Return Period (Years)

a) 1-year return level



- 1 *Table 1. Uncertainty budget, listing all contributions to total uncertainty considered during the study, with normal (k=1) and*
 2 *extended (k=2) coverage. The calculations are truncated to the second digit so discrepancies may arise.*

	1 m	5 m	10 m	20 m	30 m	50 m
Type A						
Measurement	0.10	0.11	0.10	0.06	0.09	0.08
Road siting effect (1 year)	0.11	0.10	0.10	0.09	0.08	0.08
Road siting effect (10 years)	0.23	0.23	0.21	0.18	0.16	0.17
Road siting effect (100 years)	0.38	0.41	0.37	0.30	0.26	0.29

Type B						
Calibration	0.005	0.005	0.005	0.005	0.005	0.005
Total (1-year) <i>(k=1)</i>	0.15	0.15	0.14	0.11	0.12	0.11
Extended (1-year) <i>(k=2)</i>	0.30	0.30	0.28	0.22	0.24	0.22
Total (10-year) <i>(k=1)</i>	0.25	0.26	0.23	0.19	0.18	0.19
Extended (10-year) <i>(k=2)</i>	0.50	0.51	0.46	0.38	0.37	0.38
Total (100-year) <i>(k=1)</i>	0.39	0.42	0.38	0.31	0.27	0.30
Extended (100-year) <i>(k=2)</i>	0.79	0.85	0.77	0.61	0.55	0.60

6 Table 2. Statistical indices of ΔT_1 computed at daily maximum, average and minimum temperatures.

	From maxima and minima only						From hourly data					
	Max		Avg		Min		Max		Avg		Min	
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
ΔT_1	0.16	0.23	0.23	0.16	0.31	0.22	0.12	0.21	0.19	0.10	0.29	0.24
ΔT_2	0.16	0.18	0.16	0.14	0.16	0.18	0.12	0.16	0.12	0.08	0.17	0.20
ΔT_3	0.11	0.14	0.14	0.12	0.16	0.18	0.10	0.14	0.11	0.08	0.17	0.20
ΔT_4	-0.03	0.10	0.06	0.09	0.14	0.17	-0.01	0.11	0.08	0.07	0.18	0.18
ΔT_5	0.05	0.10	0.07	0.09	0.09	0.15	0.04	0.12	0.07	0.06	0.10	0.17
ΔT_6	0.01	0.09	0.04	0.08	0.07	0.13	0.00	0.11	0.05	0.05	0.09	0.13

9 *Table 3. Summary statistics of the filtered working dataset of Figure 13.*

	ΔT_1	ΔT_2	ΔT_3	ΔT_4	ΔT_5	ΔT_6
Minimum	-0.212	-0.377	-0.332	-0.256	-0.381	-0.414
1st quartile	0.030	0.013	-0.001	0.028	-0.007	0.000
Median	0.123	0.082	0.071	0.091	0.046	0.041
Mean	0.175	0.117	0.105	0.124	0.067	0.056
3rd quartile	0.273	0.200	0.185	0.187	0.129	0.105
Maximum	1.416	1.216	1.241	1.120	0.880	0.598

10

11

12 Table 4. Results of GPD fits, in term of scale (s) and shape (ξ) parameters, along with their uncertainties.

	Threshold value (h)	Scale parameter (σ)	Uncertainty on σ (u_σ)	Shape parameter (ξ)	Uncertainty on ξ (u_ξ)
ΔT_1	0.50	0.214	0.036	-0.108	0.114
ΔT_2	0.40	0.170	0.030	-0.059	0.117
ΔT_3	0.38	0.178	0.031	-0.070	0.108
ΔT_4	0.38	0.171	0.030	-0.104	0.113
ΔT_5	0.35	0.161	0.037	-0.173	0.159
ΔT_6	0.20	0.119	0.023	-0.113	0.156

	1 year		10 year		100 year	
	value	<i>u</i>	value	<i>u</i>	value	<i>u</i>
ΔT_1	1.22	0.11	1.49	0.23	1.71	0.38
ΔT_2	1.00	0.10	1.29	0.23	1.54	0.41
ΔT_3	0.99	0.10	1.28	0.21	1.52	0.37
ΔT_4	0.93	0.09	1.17	0.18	1.35	0.30
ΔT_5	0.77	0.08	0.93	0.16	1.05	0.26
ΔT_6	0.60	0.08	0.75	0.17	0.87	0.29