Correcting Citizen-Science Air Temperature Measurements across the Netherlands for Shortwave Radiation Bias

Richard C. Cornes[1,2*], Marieke Dirksen[1], and Raymond Sluiter[1]

[1] *Royal Netherlands Meteorological Institute (KNMI), De Bilt, Netherlands*
[2] *Current affilation: National Oceanography Centre, Southampton, UK*
[*] *Corresponding author: richard.cornes@noc.ac.uk*

Abstract

Citizen-science thermometer measurements have the potential to provide information about surface air temperature fields on scales smaller than is typically quantified by the official monitoring network. As such National Meteorological Services are becoming increasingly interested in these measurements as a possible source of data for use in weather monitoring or forecasting. However, in order for the data to be used, biases in the data need to be assessed. The most important source of bias is the potential over-heating of the thermometer due to inadequate shielding or exposure. Previous research has indicated that information about the nature of the instrument and its exposure is important for correcting this bias. However, in the majority of cases this information is not available for amateur stations. In this paper we develop a statistical correction for shortwave radiation-bias in the air temperature data recorded at 159 Weather Observations Website (WOW) stations across the Netherlands during 2015 – 2016. Generalized Additive Mixed Modelling (GAMM) is used to quantify and correct for shortwave radiation bias in the hourly measurements, using a background temperature field generated from the official 34 automatic weather stations along with satellite-derived shortwave radiation estimates. It is demonstrated that the corrected WOW data add local detail to the hourly temperature field, which may provide a useful source of data to supplement official measurements.

Keywords: amateur observations; crowd sourced; Generalized Additive Model; Urban Heat Island; private weather stations

Running head: Citizen Science Temperature Measurements

Funder: KNMI "Meerjaren Strategisch Onderzoek (MSO)" project grant

1 Introduction

In 2011 the UK Met Office launched the Weather Observations Website (WOW) in association with the Royal Meteorological Society (`https://wow.metoffice.gov.uk`). The project aims to provide a web application through which individuals can upload their weather observations, recorded manually or via automatic weather stations, and access data recorded by other observers. The repository contains both near-real time observations and historic observations from the WOW network. Although a relatively new venture, the WOW project builds on earlier initiatives such as Weather Underground (www.wunderground.com), which has provided web access to measurements recorded using private weather stations since 1993, as well as projects that pre-date the internet age, notably the Climate Observers Link (COL) (Brugge, 2010). The WOW network has grown substantially since its inception, with data now being received from over 2000 sites worldwide, and associate projects have been developed in the Netherlands (WOW-NL, `https://wow.knmi.nl`) and Belgium (WOW-BE, `https://wow.meteo.be`) by the Royal Netherlands Meteorological Institute (KNMI) and the Royal Meteorological Institute of Belgium (KMI), respectively.

A particular advantage of the WOW and Weather Underground initiatives is the ability to capture observations automatically from private weather stations. A range of variables are recorded in the databases, including air temperature, wind speed and rainfall, and with data uploaded as regularly as every ten minutes they constitute a large source of temporally high-resolution readings (Bell et al., 2013). Given the wealth of observations in the WOW repository, there has been a great deal of interest in the potential use of these data by National Meteorological Services for weather monitoring and forecasting (Krennert et al., 2018). In contrast to the locations of World Meteorological Organization (WMO) approved official weather stations, which aim to provide measurements that are representative over a wide area, amateur stations are mostly sited in populated areas, and as such may provide important local weather information, particularly in relation to urban environments (Wolters and Brandsma, 2012; Bell et al., 2013; Muller et al., 2013; Chapman et al., 2017; Fenner et al., 2017; Meier et al., 2017; de Vos et al., 2017; Napoly et al., 2018).

In the Netherlands, there has been considerable research interest in the use of non-standard meteorological measurements to supplement the network of official measurements. These have ranged from weather stations attached to lampposts (Ronda et al., 2017) to the amateur measurements that are contained in the Weather Underground database (Steeneveld et al., 2011; Wolters and Brandsma, 2012). Despite the different types of instruments used, all of the studies have sought to use the measurements to analyse urban meteorology, and especially to improve knowledge about the Urban Heat Island (UHI) effect (Oke, 1982; Chapman et al., 2017; Steeneveld et al., 2011; Theeuwes et al., 2016; Thorsson et al., 2014; Lindberg and Grimmond, 2011). This is a particularly important field of research given the increasing proportion of urbanization in the country (DESA, 2017) along with the projected increase in heatwaves in future decades (van den Hurk et al., 2006; Haines et al., 2006) — the effects of which are potentially amplified by the UHI (Heusinkveld et al., 2010, 2014; Li and Bou-Zeid, 2013; Li et al., 2015; Zhao et al., 2018), and are associated with an increase in thermal discomfort (Molenaar et al., 2016).

The thermometer observations contained in the WOW repository can potentially provide useful information about urban temperature but in order to fully utilise this information the data must be corrected for potential biases. The siting of the instruments as well as the type of instruments used can introduce biases that exceed the manufacturer-stated tolerances. This was demonstrated in the year-long test of several commonly used amateur meteorological stations alongside the official UK Met Office measurements at the Winterbourne meteorological enclosure in Edgbaston, Birmingham by Bell (2014) and Bell et al. (2015). The air temperature measurements displayed a marked bias as a result of inadequate radiation shielding — a feature that has also been noted by Jenkins (2014) and Meier et al. (2017)— although the severity of this bias was dependent on the type of instrument used. Using this information, a statistical approach was developed by Bell (2014) to correct for radiation bias, along with the likely bias introduced from poor instrument calibration.

In this paper we analyse the hourly, near-surface air temperature data recorded at WOW sites across the Netherlands during the 24-month period from January 2015 until December 2016. These data have not previously been assessed for radiation bias in a systematic way. Taking the findings of the previous studies described above as a starting point, we derive a station-by-station correction for the WOW data using a statistical model that takes into account the background temperature field (derived from the official temperature measurements), and an estimate of local direct shortwave radiation obtained from satellite data. We demonstrate that the corrected WOW data add local detail to the hourly temperature field, which may provide a useful source of data to supplement official measurements.

2 Data and Methods

2.1 The nature of the WOW temperature data

The WOW air temperature data for stations situated in the Netherlands were obtained from the UK Met Office data repository, along with the latitude/longitude coordinates of the station and the time of observation. The number of WOW stations covering the Netherlands marks a substantial increase over the number of official sites. During the period 2015–2016 a total of 318 stations supplied data to the WOW database, compared to the 34 official weather stations (Figure 1). However, not all of the WOW station series are complete for this period and stations were only used if they supplied (interpolated) hourly values that were 80 % complete for each month. This resulted in a sample of 159 stations for use in this analysis. It should be noted that the official KNMI automatic weather station (AWS) data are now included in the WOW database, and when referring to the WOW stations we have excluded the AWS stations.

The WOW stations are generally situated in urban environments, often in people's gardens or on school premises. The instruments typically used are relatively low-cost and are manufactured, for example, by Davis Instruments or Oregon Scientific (Bell et al., 2015), although the instruments from other manufacturers may be used as there is no stipulation on instrument-type for the entry of data to the WOW. In addition, while there is the ability for observers to supply meta-data in a standardized format, this is not mandatory; Bell (2014) estimated that 15 % of observers omit this information, and clear information about the hardware used was available in only around 60 % of stations. The coordinates of the station are mandatory information, although the station altitude is not required; Bell (2014) estimated that for the UK this altitude information was supplied by fewer than 75 % of observers.

In this analysis, all of the 159 WOW stations were analysed. However, four sample stations were selected for closer examination. These stations represent the range of instruments and exposure of the network across the Netherlands; the properties of these stations are summarized in Table 1. Site A is a high-quality instrument situated in the De Bilt official WMO meteorological enclosure close to the official temperature measurements. Site B is a typical instrument sited in a sheltered suburban garden. Site C uses the same instrument as Site B, but is located in a low-cut grass field with an open exposure, where the nearest building is 10m from the weather station. Site D has a standard residential exposure and is situated in the east of the country; the station uses the relatively low-cost AcuRite 5 in 1 instrument.

The WOW observations are recorded on varying time scales. In this analysis we are only interested in hourly values and in order to simplify the analysis, the observation times were rounded to the nearest 10 minute and these values were then interpolated to regular hours by fitting a cubic smoothing spline to each station series. Gaps of longer than two hours (in the 10-minute values) without an observation were marked as missing.

2.2 The correction model

Generalized Additive Modelling (GAM) was used as the basis for the correction of the WOW data. GAMs are an extension of Generalized Linear Modelling (GLM), which themselves are a more flexible version of ordinary-least-squares regression, and allow a model to be fitted to a dependent variable that is not necessarily from a Gaussian distribution (Hastie and Tibshirani, 1990). GAMs extend GLMs by allowing the use of one or more unknown (smooth) functions, and may also include linear coefficients. Generalized Additive Mixed Modelling (GAMM) is a further development that allows random effects and correlation structures to be accommodated in the model (Wood, 2006).

Using the findings of Bell (2014) as a basis we modelled the WOW station data (T_{wow} with outcome at time i) at each site under a semi-parametric scheme (Hastie and Tibshirani, 1990) using shortwave radiation (Rad) and background temperature (T_{bg}) terms as

$$T_{wow,i} = \beta_0 + \beta_1(Rad_i) + f(T_{bg,i}) + \epsilon_i, \qquad \epsilon_i = \phi_i \epsilon_{i-2} + v_i,$$

where T_{bg} is formed from an interpolation of the hourly temperature measurements recorded at the official KNMI weather stations, and Rad is formed from a local estimate of incoming solar radiation derived from satellite retrievals. β_0 represents an intercept term, and ϵ is random error assumed to be identically and independently distributed (i.i.d). Rad values represent hourly averages over the hour leading up to the temperature observation. This corresponds to the findings of Bell (2014) who demonstrated a 1-2 hour lagged response of the WOW measurements to shortwave radiation at the Winterbourne test site. In contrast, T_{bg} represent estimates of concurrent temperature measurements. The local derivation of Rad and T_{bg} is described below

The smooth function f can be derived in a number of ways. Since we are dealing with a univariate function the piece-wise cubic polynomial spline is an obvious choice as it is relatively quick to converge. However, better model fitting was achieved by using a Thin-Plate Regression Spline (TPRS) (Wood, 2003). TPRS are a more general form of cubic splines, and in practice with the data used here produced splines that were slightly smoother and which were more physically plausible.

Since the WOW observations analysed in this paper are comprised of hourly observations, temporal autocorrelation in the observations needs to be taken into account in the models in order to satisfy the i.i.d assumption. Neglecting temporal autocorrelation may lead to inaccurate parameter estimation and poor uncertainty estimates of the model terms (Wood, 2006). We used a lag-2 autoregressive model in the GAMM, where the autoregressive coefficient (ϕ) is estimated as part of the model fitting. The autogressive function was nested in each month of data in order to speed-up the calculation. A variety of lag intervals were tested and the lag-2 autocorrelation model effectively counteracted temporal autocorrelation in ϵ, up to a lag of 10 hours. A low-level of autocorrelation at around lag-24 remained in ϵ for many stations. This appears to represent a local diurnal cycle not quantified in the model covariates, and is a likely feature of the local temperature data.

In this model the form of each of the parameters is chosen via a backfitting algorithm that iteratively selects an optimal fitting of each function. The fitting of the smooth function f represents a balance between over- and under-fitting of the function, i.e. between a spline that is too smooth and one that is too "wiggly". A penalization is imposed to the function f to avoid over-fitting of the spline. An optimal fitting of the function in these models is obtained through the calculation of a score that measures the degree to which the predictive error is minimised. In the WOW-correction models used in this paper Marginal Likelihood (ML) scores are used.

The fitting of the smoothed terms in the models used in this paper rely on the prior setting of an upper limit (k) on the effective degrees of freedom (EDF) of the smoothing terms

(Wood, 2003). This allows for a more efficient way of fitting the smoothing functions through an eigendecomposition, but necessitates the subjective selection of k. As stressed by Wood (2006) however, while this selection of k is subjective, the actual selection of the EDF uses the pre-selected optimization procedure (Marginal Likelihood in this case), up to a limit of $k-1$. Values of $k = 30$ were chosen for the models used in this paper following application of the heuristic tests recommended by Wood (2006).

This model assumes that $T_{wow,i}$ can be modelled as a non-linear response to $T_{bg,i}$ (via f) plus a linear response to Rad_i (via the coefficient β_1). By making this assumption we are able to ensure that the response to radiation scales from a partial intercept at zero. We also tested whether a simple model where the response to T_{bg} was also linear, i.e. in the form of a GLM. A significant difference was observed in these models in terms of the explained variance and we therefore opted to retain the use of the smoothing function f under the GAMM scheme. The non-linearity in f represents the local distortion to the background temperature field that is assumed to arise from the temperature environment of the WOW station.

Prior to model-fitting, the T_{wow} values were quality-controlled against the respective T_{bg} values: $|T_{wow} - T_{bg}| > 8°C$ values were removed, and identical values of T_{wow} for more than four consecutive hours were excluded. The Urban Heat Island in cities across the Netherlands has been estimated by Steeneveld et al. (2011) to be of the order of 6 °C during calm, fair weather and hence the threshold of 8 °C does not preclude the capturing of these types of features in the data.

Since the GAM(M)s are additive in nature, the partial effects of each of the covariates can be assessed individually. Specifically, the contribution of the shortwave radiation term $\beta_1(Rad)$ may be extracted from the WOW temperature data as

$$T^{'}_{wow,i} = T_{wow,i} - \beta_1(Rad_i)$$

to produce the corrected WOW values ($T^{'}_{wow,i}$). In this way the shortwave radiation effect, as modelled by the partial regression coefficient β_1, is removed from T_{wow}, and the corrected temperature values are obtained from the (non-linear) relationship to T_{bg} (estimated from f) plus any residual effect contained in ϵ. Although the correction is zero at nighttime, the models were applied to the data over the full 24-hour period to increase the sample size for fitting of the function $f(T_{bg})$. Since the correction is applied through $\beta_1(Rad)$, this correction should be viewed as a parameterization of the shortwave radiation effect at a given WOW station, under the assumption that any shortwave radiation effect detected in the data is an artificial bias that results from inadequate shielding or siting of the instrument.

2.3 The background temperature field (T_{bg})

The background temperature values (T_{bg}) were calculated by fitting a GAM to the temperature measurements recorded at the 34 official KNMI AWS sites (Figure 1), which was then interpolated to the WOW station locations. In this case a tensive spline was used to model the three-dimensional interaction of temperature across space (using longitude and latitude coordinates) and time:

$$T_{bg,i} = \beta_0 + f(lon_i, lat_i, time_i) + \epsilon_i.$$

This space-time model is preferable to more common spatial interpolation techniques such as kriging — which typically use only spatial coordinates and construct separate models for each time-step — as the sample size for model fitting is increased considerably. This is important as there are only 34 AWS stations across the Netherlands, and this represents a small sample size for fitting the interpolating model. The tensive spline has the advantage of being insensitive to the units of measurement of the covariates (Wood, 2006), (degrees in the case of longitude and latitude, and hours in the case of time). The spline is formed using the joint interaction of longitude/latitude and time, where both components take a thin-plate spline basis. As such an anisotropic relationship over space is modelled. The temperature lapse rate is best captured by the longitude/latitude coordinates, since the altitude is relatively constant across the Netherlands and for these purposes only becomes significant across the south of the country. Tests were carried out using altitude as an additional covariate. However, the fitted function did not have a plausible physical interpretation, which likely resulted from the masking of the lapse-rate by hourly-scale temperature variations. Furthermore, altitude values are not supplied in the WOW-station metadata and would need to be estimated via a Digital Elevation Model adding to the uncertainty in lapse-rate estimation. Three other environmental parameters (coastal proximity, slope and aspect) were also tested in this model but were found to be insignificant, and are most likely also accommodated by the joint-interaction of the latitude and longitude components.

It is impractical to fit this background temperature model to all of the data points simultaneously, since the model would take a considerable length of time and significant computing resources to converge. Therefore the model was fitted on a day-by-day basis. To limit the occurrence of edge effects, an overlap of six hours was used and hence for a given day data from 6pm UTC of the previous day to 6AM UTC of the following day were used to produce a moving-window of overlapping models.

2.4 The local shortwave radiation data (Rad)

Estimates of global solar radiation (the sum of direct and diffuse radiation) at the WOW sites are derived from the MSG-CPP dataset (www.msgcpp.knmi.nl). These values are calculated from the Meteosat SEVIRI imagery using the KNMI Cloud Physical Properties (CPP) algorithm (Roebeling et al., 2006). In the visible range of the spectrum this dataset provides 15-minute retrievals at a ca. 1km resolution. At each WOW station the surface downwelling shortwave (SDS) radiation values (in the unit of W/m^2) were calculated as the average of the nearest 3x3 pixels after Greuell et al. (2013).

The CPP algorithm calculates the SDS values using cloud retrievals and satellite-derived reflectances. Deneke et al. (2008) evaluated the shortwave radiation data obtained from the MSG-CPP dataset over the Netherlands through a comparison against pyranometer measurements recorded at the official KNMI stations. They found that surface irradiance values were comparable to ground-based instruments in the summer, although during the winter the accuracy was lower as a result of the low sun elevation in combination with the large satellite viewing angle across the Netherlands. Hence the data are considered to be suitable for the correction of the WOW data, which are generally only affected by shortwave radiation bias at higher radiation values, as indicated by the linear scaling against the radiation values from a zero intercept. Night-time values (missing in the MSG-CPP visible spectrum data) were set to zero.

3 Results and Discussion

3.1 The background temperature and radiation values

Reliable estimations of the background temperature (T_{bg}) and shortwave radiation values (Rad) at each of the WOW sites are essential for successfully modelling, and ultimately correcting, the WOW temperature measurements as any deficiency would potentially produce skewed model residuals that would violate the i.i.d assumption of the model. To provide an indication of the reliability of the background temperature model a leave-one-out cross validation exercise was conducted, using the observations from the official 34 official AWS (see Figure 1). This exercise consisted of removing one AWS station at a time and interpolating to that candidate station over all time steps in the years 2015–16 using the data from the remaining stations. This cross-validation was repeated for each AWS in turn. The error of T_{bg} relative to the official measurements was then calculated, with the assumption being that a similar degree of error relative to T_{bg} would also be applicable to the WOW stations. The results indicate that the interpolation produces a broadly unbiased estimate of the local temperature (Figure 2). Root-mean squared error (RMSE) values for most stations are in the range 0.4–0.8 °C, although there is a degree of variation during different seasons, with higher RMSE values at certain stations during spring, summer and autumn. The largest of these values (RMSE>1.5 °C) occur at the border regions, and likely result from the relatively large distances from the nearest stations. Improvements could be made by incorporating data from neighbouring countries.

The degree of error indicated in this cross-validation exercise compares favourably with the results obtained by Bell (2014). That study similarly constructed background temperature fields for the correction of UK-based WOW temperature readings. However, the method used to interpolate the official measurements in the present study is greatly simplified compared to the fields constructed by Bell (2014), which included many more covariates. The atmospheric environment of the Netherlands does not require the range of covariates that are necessary for an interpolation of temperature across the UK, but also the tensive spline used here provides a more optimal fit of the data, using the few covariates that are employed. Notably, we do not include short-range weather forecast data as a covariate in the model, as was the case with Bell (2014), since a potential application of these data is assimilation in numerical weather models, and including the corrected WOW data could confound the determination of local temperature information afforded by the WOW data. The GAMs are, however, flexible enough to allow incorporation of such variables as additional model terms if required in future extensions of the method.

The reliability of the local shortwave radiation estimates (Rad) were assessed by evaluating the MSG-CPP values against surface shortwave radiation values recorded at each of the official weather stations. The MSG-CPP radiation estimates are strongly linearly related with station-based shortwave radiation measurements during the year 2015–2016 (Figure 3), and have a small bias of -2.29 W/m^2 and a RMSE value of 36.93 W/m^2. These values are in accordance with the findings of Deneke et al. (2008).

3.2 Evaluation of the GAMM for Station B

A GAMM was developed for each of the selected 159 WOW stations, following the method described in Section 2, using data for the years 2015 and 2016. To provide an example of the nature of the statistical models and the radiation correction, the results for station B (see Table 1) are evaluated in this section.

The statistical model for Station B explains 98 % of the variation in the hourly WOW temperature data (Figure 4 a), and the residuals from the model are normally distributed with a standard deviation of 0.9 °C (Figure 4 b). This indicates that the WOW temperature measurements can be successfully modelled using the background temperature and shortwave radiation parameters since a deficiency in the model, for example through an important missing parameter, would produce a skewed distribution in the residuals. The incorporation of the autocorrelation factor into the model has significantly reduced the degree of temporal autocorrelation in the model's residual. This is demonstrated in Figure 4 c, where residual autocorrelation up to a lag of 41 hours from the lag-2 model are compared against those derived where no temporal autocorrelation is assumed. The residual autocorrelation is reduced at all lag intervals, particularly at lag-1. This plot also displays a moderate degree of autocorrelation at around a lag of 24 hours, which indicates that the model is failing to capture a diurnal cycle in the WOW temperature values. This may be a result of a inadequate representation of heating from solar radiation, or may be a true feature of the local temperature field that represents a diurnal urban temperature cycle.

In the model for this station both of the parameters [$f(T_{bg})$ and $\beta_1(Rad)$] are highly significant predictors (p<0.001). A strong relationship with T_{bg} is to be expected, however the importance of Rad as a predictor is used as an indication of significant shortwave radiation bias in the WOW readings for this station. In Figures 4 d and e the partial model terms are plotted. It should be notated that the model-fitting is applied to values of y expressed as deviations from the mean of y and hence in those figures the temperature response values are relative to the intercept (β_0). The function $f(T_{bg})$ with 13 degrees of freedom shows a non-linear relationship to the background temperature, with the largest deviations from linearity occurring at the lowest temperature values. The radiation coefficient term ($\beta_1 = 2.09e - 03$) scales from zero to a value of 1.67 °C (± 0.09 °C, 95 % confidence interval) at $Rad = 800W/m^2$..

Since the statistical models used here are additive in nature, the original time series of WOW values can be decomposed into each of the model terms. The results from the time series decomposition for Station B are demonstrated in Figure 5 for the year 2016. It should be noted when evaluating this figure that the sum of the background temperature component, shortwave radiation component, residual and the model intercept, equal the raw WOW values. The annual and diurnal cycles in the shortwave radiation component are readily apparent from this figure.

The magnitude of the estimated shortwave radiation bias in Station B is much larger than that measured using a similar instrument (Davis *Vantage Pro 2*) by Bell (2014) in the field-test at the Winterbourne meteorological enclosure; that instrument showed no appreciable radiation bias. However, the instrument used in that experiment was equipped with fan-assisted aspiration, whereas the instrument at used at Station B is naturally ventilated; the lack of assisted ventilation therefore appears to have a detectable effect in these results and leads to significant over-heating under moderate-to-high levels of insolation (>500 W/m^2).

3.3 Evaluation of the Radiation Bias in all Test Stations

In Figure 6, tile plots are produced for the four Test Stations (Table 1). These plots show the average bias per month and per hour of the day (c.f similar plots by (Bell et al., 2015) for the Winterbourne test site). Stations A and C have the lowest estimated shortwave radiation bias, up to average values of 0.6–0.7 °C in the summer months at around noon. Since Site A uses a high-quality instrument and is sited in the official De Bilt Meteorological enclosure, a relatively low bias would be expected. The results from Site D show the largest bias, with average noon biases reaching values above 3 °C during high-summer.

Sites B and C both use a passively-ventilated Davis Vantage Pro2 instrument. The main difference in these sites is the exposure of the instruments: the instrument at Site C is sited in an open field that consists of short grass, whereas Site B is in a more enclosed residential setting. These results therefore suggest that the increased air-flow that would be expected at Site C enhances the passive ventilation and reduces the shortwave-radiation bias in the temperature readings. This effect could possibly be incorporated into the statistical models through the incorporation of wind speed as an additional covariate. However, while many stations record wind speed, it is not ubiquitous across the network and the values are highly susceptible to local wind-flow distortion. Nonetheless, incorporation of this parameter in the models could be considered in future updates to the method.

Since Station A is located in the De Bilt meteorological enclosure the validity of the radiation correction can be assessed through comparing the WOW data against the official AWS temperature measurements. In Figure 7 we have plotted the root-mean squared error of the raw and corrected WOW data relative to the AWS data for at each hour of the day for four selected months. The correction to the data has reduced the error to values consistent with those observed during the night, when no corrections are applied to the data. This background level of error is the same order of magnitude as the precision of the measurements that has been estimated by the instrument manufacturer to be between ± 0.3–0.4 °C. Variations around this range in the corrected readings in Figure 7 likely result from sampling bias, arising from the relatively small sample sizes used here and because the errors are not taken under laboratory conditions.

3.4 Correcting Shortwave Radiation Bias across the WOW network

To examine the T_{bg} and Rad responses across the network, we have plotted the partial model terms for all of the selected 159 WOW stations in Figure 8, in a similar manner to Figures 4 d and e above; the terms for the four test stations are also indicated. The departure from linearity of the T_{bg} terms is most evident at the lower temperature values, and is apparent in many of the stations. This non-linearity is likely a result of the urban distortion to the background temperature field, beyond the shortwave radiation effect that is captured by the term $\beta_1(Rad)$.

The radiation coefficient (β_1) is significant in all of the station models. However, the magnitude of bias estimated using this coefficient varies greatly across the network (Figure 8 b). Several stations have radiation bias that is below 0.3 °C across the range of radiation values, and this is below the usual manufacturer-stated precision that is typically of the order of ± 0.3 °C. Other stations have large estimated bias that are exceed 2 °C at $Rad = 800$ W/m^2. It is likely that the range of β_1 coefficients across the network results from the different instruments employed. This cannot be verified as the meta-data for all stations are not easy to acquire (c.f. Bell (2014) who used 'web-scraping' to obtain information for certain UK WOW stations), however the results from the four test stations — for which we do know the type of instruments used and their general situation — suggest that this is the case. Test stations A and C are at the lower range of this spread, whereas Station D is the station that is most affected by shortwave radiation bias by this estimation. However, as suggested above, it would seem that the degree of shortwave radiation effect is a combination of the nature of the instrument and the exposure/situation of the instrument. In addition to the enhanced ventilation in the Davis VP2 instruments, these siting-effects likely also include a myriad of factors such as the sky-view factor, the local land-use or local boundary conditions. An evaluation of a network of sensors such as that provided Netatmo (e.g. Napoly et al. (2018)) would be useful in this respect, since the confounding effect of different instrument types would be removed.

Several previous studies have indicated a non-linear response of citizen-science temperature observations to shortwave radiation bias (Jenkins, 2014), particularly for those instruments that were particularly vulnerable to shortwave radiation bias (Bell, 2014). A linear function was used in the GAMMs developed in the present study in order to ensure a scaling of the radiation component from a partial zero intercept, but this is likely to be a simplification. The analysis by Bell (2014) indicated that a quadratic function was most suitable for capturing the radiation effect in the stations most affected by such biases, and hence the linear function used here is likely to be a conservative estimate of shortwave radiation bias.

One of the main applications of the WOW data is for the examination of the UHI effect, which is generally not captured by the official network of weather stations. A risk with the correcting of the WOW station temperature data is that true UHI is mistaken for shortwave radiation bias. To examine this we have calculated the diurnal cycle of temperature for each season as averages from the official AWS data and both the raw and corrected WOW temperature measurements (Figure 9). The results indicate a typical feature found in the diurnal temperature cycle of urban areas relative to the background rural temperature (represented here by the AWS data) (Oke et al., 2017). Temperatures during the night are generally warmer in urban areas but around dawn this difference reduces. This remains the case until the afternoon when the difference increases again. The corrected WOW values in Figure 9 clearly show this diurnal variation. In contrast, the raw WOW values show an augmented diurnal cycle with elevated temperatures between around 9–14UTC, particularly during the spring and summer seasons. These results indicate that while the largest shortwave radiation bias is removed from the data, the urban-related diurnal cycle is retained in the corrected WOW data.

3.5 Mapping the Temperature Data

To assess the value of the corrected WOW temperatures in examining the temperature field across the Netherlands, we have mapped the AWS temperature values and the AWS together with the corrected WOW data maps (AWS+WOW). This has been done for the hottest and coldest events in the Netherlands during the 2015–2016 period: the 1300UTC readings from 2 July 2015, and the 0300UTC readings from 19 January 2016 (Figure 10). By comparing the AWS and AWS+WOW maps the extra detail added by the WOW data can be examined. It should be noted, however, that while the AWS are relatively evenly spatially distributed the WOW are concentrated in urban areas and hence the AWS+WOW interpolation will be biased towards the urban areas.

The July 2015 event was connected with a southerly airflow resulting from a high-pressure system centred over Scandinavia, and occurred during a week that saw very high temperatures recorded across northwest Europe. The spatial pattern of temperature observed from the AWS and AWS+WOW stations is broadly similar, with a gradient of temperatures evident across the country from 36 °C in the south to less than 24 °C at the coast. However, the map calculated using the AWS+WOW data indicates local detail not seen in the maps generated from the AWS data alone. In particular, several urban heat and cool islands (Oke et al., 2017) are apparent in the data, which correspond to urban centres and parkland respectively.

The January 2016 event was associated with the development of a ridge of high-pressure that had built from the east over the previous two days. As with the July 2015 maps, the spatial gradient is broadly consistent in the maps produced from the AWS and AWS+WOW data, although additional spatial features are apparent in the maps using the AWS+WOW data. Temperatures are on the whole are warmer in the AWS+WOW map, which likely reflects the urban-bias in the location of the WOW stations. A relatively large difference also occurs in the south-western extremity of the country. The nearest official AWS at the location is at Vlissingen, which although located on land is significantly affected by oceanic conditions. The use of the corrected WOW data result in a cooler interpolated temperature for that region, which further indicates the warm bias at the Vlissingen site during this cold winter event.

A factor that is not taken in correction of the WOW temperature data is the spatial-scale lengths that are represented by the WOW data. The AWS locations are chosen so that the data are representative of conditions over a relatively wide area, although as discussed above in the case of certain stations such as Vlissingen this ideal may not always be reached. Similarly, the WOW stations should be representative of the general conditions that are experienced in the vicinity of the station, albeit in this case often from an urban environment. In practice, however, the stations are sited at the convenience of the observer and hence the readings will represent a very small area (Bell, 2014), possibly on the scale of a few metres. This need not be the case, however, as indicated in the case of sample station C (see Table 1), which appear to represent well the conditions surrounding the observation-site on account of the instrument being located in an open situation, which is typical of the local environment. Nonetheless, in order to make full use of the (corrected) WOW data for examining temperature fields an evaluation of the likely spatial scale represented by each of the station would need to be conducted. This would entail an examination of the surrounding environment of the station, through the use of digital terrain information. This would also require precise station coordinates, which are not always provided. However, representativity error could be reduced when using the WOW data (in combination with the official AWS data) for country-scale mapping through a spatial smoothing procedure. The degree of smoothing used in Figure 10, for example, is determined by the nugget variance of the variograms, which is estimated automatically. Hence the maps calculated from the AWS+WOW data depict a spatial-scale that is larger than the station-scale but smaller than the AWS maps. However, these maps are not able to detect all features of urban temperature, given the large intra-urban variability in air temperature.

3.6 Aspects to Consider for Operational use of the Corrected Data

The WOW data are uploaded by users in near-realtime, with often only a delay of minutes before the data are made available on the WOW website. Similarly, both the satellite data and the AWS values are available in near-realtime. The WOW data and this correction method are therefore of potential use in operational observing or forecasting procedures by National Meteorological Services. The question therefore arises: how can the GAMM approach to correcting the WOW temperature data developed in this paper be used in such operational situations?

A major limitation to the use of the WOW data is the short duration of many records. In this analysis only around half of the total stations that supplied data to the WOW repository during the period 2015–2016 were used, since criteria were placed on the minimum number of missing values in a series (see Section 2.1). While this ensures that the statistical models are comparable between stations, it severely limits the pool of available stations. A different approach was taken by Bell (2014). His analysis used an algorithm that gradually adjusted the uncertainty of the bias as new data were added. Such an approach could be used in the GAMM models presented here through use of the standard error metrics that can be calculated for the smoothing splines (see for example Figure 4 e). These uncertainty estimates would be expected to reflect sampling density, and could be used to indicate uncertainty through a resampling scheme, through which draws from the posterior distribution of the spline could be taken; these values could be used to produce a range of plausible corrections relative to the spline uncertainty.

The question also arises in an operational setting about the speed of computation. The GAMM takes considerable time to converge (several hours for a given station). However, in an operational context a given model could be saved for each station and the corrections applied as new data are received. The background temperature values are quicker to derive since an autocorrelation is not embedded in the model. Similarly the satellite-derived shortwave radiation estimates are quick to produce, since their derivation only relies on a pixel-overlay across the station network.

4 Conclusions

A correction for shortwave radiation-bias has been calculated for the hourly temperature measurements taken at 159 WOW sites across the Netherlands. The corrections were derived on a station-by-station basis for data recorded over the 2015–2016 period, with the aim of retaining the local-scale information contained within the data, whilst removing the bias resulting from inadequate shortwave radiation shielding. Although derived for the WOW network across the Netherlands, the technique developed in this paper could equally be applied to other stations in the WOW network — which has the highest density across the UK, the Netherlands and Belgium – and to other platforms such as Weather Underground or Netatmo, which offer a wider area of investigation.

The correction is calculated by fitting a Generalized Additive Mixed Model (GAMM) to each station series, using satellite-derived shortwave radiation estimates and a local estimate of the background temperature as model covariates. By decomposing the WOW temperature series into components relating to each of the covariates, the effects of the shortwave radiation component can be extracted from the data. The GAMM-approach offers an extremely flexible way of modelling and hence correcting the WOW data, through the ability to model both linear and non-linear responses and to incorporate a temporal autocorrelation component.

We have focused on shortwave radiation-bias in this paper, as this the most significant limitation that is expected in readings from relatively low-cost weather stations, and which through necessity are generally not sited in optimal locations. The shortwave radiation biases estimated through the models developed in this paper are broadly comparable in magnitude to those measured or estimated in previous analyses. However, these models are able to correct for the shortwave radiation bias in the absence of meta-data about the nature of the instruments or their exposure – information that is typically missing or incomplete in databases of citizen-science observations. The correction relies on the assumption that any direct relationship between global shortwave radiation and temperature values beyond the background (rural) temperature temperatures are due to shortwave radiation-bias. Under this assumption shortwave radiation biases are detected in all of the WOW stations but the magnitude of bias varies considerably across the network. This appears to be related to a combination of the nature of the instruments and their exposure. Further analysis is required on the spatial-scale that is depicted in the corrected WOW data. Nonetheless, the data potentially allow a more detailed picture to be developed about temperature variability at a scale smaller than can be depicted by the official network of instruments, and particularly with regard to urban effects on the temperature field.

Acknowledgements

We are grateful to the UK Met Office for maintaining the WOW network (`https://wow.metoffice.gov.uk.nl`), and to all of the observers who supply data to the network. The authors would like to thank Jake Brown and Katharine O'Boyle (Met Office) for extracting the WOW data, Reinout Boers (KNMI) for constructive criticism of the manuscript and Remie van Bommel for providing details about Test Station C. This research was funded by a KNMI "Meerjaren Strategisch Onderzoek (MSO)" project grant. The models developed in this paper were produced using the *mgcv* package for R (Wood, 2018).

References

Bell, S. (2014). *Quantifying Uncertainty in Citizen Weather Data*. PhD thesis, Aston University.

Bell, S., Cornford, D., and Bastin, L. (2013). The state of automated amateur weather observations. *Weather*, 68(2):36–41.

Bell, S., Cornford, D., and Bastin, L. (2015). How good are citizen weather stations? Addressing a biased opinion. *Weather*, 70(3):75–84.

Brugge, R. (2010). Forty years of the Climatological Observers Link. *Weather*, 65(5):139–143.

Chapman, L., Bell, C., and Bell, S. (2017). Can the crowdsourcing data paradigm take atmospheric science to a new level? A case study of the urban heat island of London quantified using Netatmo weather stations. *Int. J. Climatol.*, 37(9):3597–3605.

de Vos, L., Leijnse, H., Overeem, A., and Uijlenhoet, R. (2017). The potential of urban rainfall monitoring with crowdsourced automatic weather stations in amsterdam. *Hydrology and Earth System Sciences*, 21(2):765–777.

Deneke, H., Feijt, A., and Roebeling, R. (2008). Estimating surface solar irradiance from METEOSAT SEVIRI-derived cloud properties. *Remote Sens. Environ.*, 112(6):3131–3141.

DESA, U. (2017). World population prospects. United Nations Department of Economic and Social Affairs (UN DESA) Population Division, New York, NY, USA, revision edition.

Fenner, D., Meier, F., Bechtel, B., Otto, M., and Scherer, D. (2017). Urban Meteorology Intra and inter 'local climate zone' variability of air temperature as observed by crowdsourced citizen weather stations in Berlin, Germany. 26(5):525–547.

Greuell, W., Meirink, J. F., and Wang, P. (2013). Retrieval and validation of global, direct, and diffuse irradiance derived from SEVIRI satellite observations. *J. Geophys. Res. Atmos.*, 118(5):2340–2361.

Haines, A., Kovats, R., Campbell-Lendrum, D., and Corvalan, C. (2006). Climate change and human health: Impacts, vulnerability and public health. *Public Health*, 120(7):585–596.

Hastie, T. and Tibshirani, R. (1990). *Generalized additive models*. Chapman & Hall/CRC.

Heusinkveld, B. G., Hove van, L., Jacobs, C. M. J., Steeneveld, G. J., Elbers, J. A., Moors, E. J., and Holtslag, A. A. M. (2010). Use of a mobile platform for assessing urban heat stress in Rotterdam. In *Proc. 7th Conf. Biometeorol.*, pages 433–438, Freiburg, Germany.

Heusinkveld, B. G., Steeneveld, G. J., van Hove, L. W. A., Jacobs, C. M. J., and Holtslag, A. A. M. (2014). Spatial variability of the Rotterdam urban heat island as influenced by urban land use. *J. Geophys. Res. Atmos.*, 119(2):677–692.

Jenkins, G. (2014). A comparison between two types of widely used weather stations. *Weather*, 69(4):105–110.

Krennert, T., Pistotnik, G., Kaltenberger, R., and Csekits, C. (2018). Crowdsourcing of weather observations at national meteorological and hydrological services in Europe. *Adv. Sci. Res*, 15:71–76.

Li, D. and Bou-Zeid, E. (2013). Synergistic interactions between urban heat islands and heat waves: the impact in cities is larger than the sum of its parts*. *Journal of Applied Meteorology and Climatology*, 52(9):2051–2064.

Li, D., Sun, T., Liu, M., Yang, L., Wang, L., and Gao, Z. (2015). Contrasting responses of urban and rural surface energy budgets to heat waves explain synergies between urban heat islands and heat waves. *Environmental Research Letters*, 10(5):054009.

Lindberg, F. and Grimmond, C. S. B. (2011). The influence of vegetation and building morphology on shadow patterns and mean radiant temperatures in urban areas: Model development and evaluation. *Theoretical and Applied Climatology*, 105(3-4):311–323.

Meier, F., Fenner, D., Grassmann, T., Otto, M., and Scherer, D. (2017). Crowdsourcing air temperature from citizen weather stations for urban climate research. *Urban Climate*, 19(nil):170–191.

Molenaar, R. E., Heusinkveld, B. G., and Steeneveld, G. J. (2016). Projection of rural and urban human thermal comfort in The Netherlands for 2050. *Int. J. Climatol.*, 36(4):1708–1723.

Muller, C. L., Chapman, L., Grimmond, C. S., Young, D. T., and Cai, X. (2013). Sensors and the city: A review of urban meteorological networks. *Int. J. Climatol.*, 33(7):1585–1600.

Napoly, A., Grassmann, T., Meier, F., and Fenner, D. (2018). Development and application of a statistically-based quality control for crowdsourced air temperature data. *Frontiers in Earth Science*, 6(nil):nil.

Oke, T. R. (1982). The energetic basis of the urban heat island. *Quarterly Journal of the Royal Meteorological Society*, 108(455):1–24.

Oke, T. R., Mills, G., Christen, A., and Voogt, J. A. (2017). *Urban Climates*. Cambridge University Press, Cambridge.

Roebeling, R. A., Feijt, A. J., and Stammes, P. (2006). Cloud property retrievals for climate monitoring: Implications of differences between Spinning Enhanced Visible and Infrared Imager (SEVIRI) on METEOSAT-8 and Advanced Very High Resolution Radiometer (AVHRR) on NOAA-17. *J. Geophys. Res.*, 111(D20):D20210.

Ronda, R. J., Steeneveld, G. J., Heusinkveld, B. G., Attema, J. J., and Holtslag, A. A. (2017). Urban finescale forecasting reveals weather conditions with unprecedented detail. *Bull. Am. Meteorol. Soc.*, 98(12):2675–2688.

Steeneveld, G. J., Koopmans, S., Heusinkveld, B. G., Van Hove, L. W., and Holtslag, A. A. (2011). Quantifying urban heat island effects and human comfort for cities of variable size and urban morphology in the Netherlands. *J. Geophys. Res. Atmos.*, 116(20):1–14.

Theeuwes, N. E., Steeneveld, G.-J., Ronda, R. J., and Holtslag, A. A. M. (2016). A diagnostic equation for the daily maximum urban heat island effect for cities in northwestern europe. *International Journal of Climatology*, 37(1):443–454.

Thorsson, S., Rocklöv, J., Konarska, J., Lindberg, F., Holmer, B., Dousset, B., and Rayner, D. (2014). Mean radiant temperature - a predictor of heat related mortality. *Urban Climate*, 10(nil):332–345.

van den Hurk, B., Klein Tank, A., Lenderink, G., van Ulden, A., van Oldenborgh, G. J., K. C., van den Brink, H., Keller, F., Bessembinder, J., Burgers, G., Komen, G., Hazeleger, W., and Drijfhout, S. (2006). KNMI Climate Change Scenarios 2006 for the Netherlands. KNMI, De Bilt.

WMO (2014). Guide to meteorological instruments and methods of observation.

Wolters, D. and Brandsma, T. (2012). Estimating the Urban Heat Island in Residential Areas in the Netherlands Using Observations by Weather Amateurs. *J. Appl. Meteorol. Climatol.*, 51(4):711–721.

Wood, S. (2018). *mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation*.

Wood, S. N. (2003). Thin plate regression splines. *J. R. Stat. Soc. Ser. B (Statistical Methodol.*, 65(1):95–114.

Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC.

Wood, S. N., Li, Z., Shaddick, G., and Augustin, N. H. (2017). Generalized additive models for gigadata: Modeling the u.k. black smoke network daily data. *Journal of the American Statistical Association*, 112(519):1199–1210.

Zhao, L., Oppenheimer, M., Zhu, Q., Baldwin, J. W., Ebi, K. L., Bou-Zeid, E., Guan, K., and Liu, X. (2018). Interactions between urban heat islands and heat waves. *Environmental Research Letters*, 13(3):034003.

Tables

Table 1: Properties of the four sample stations. The Urban Climate Zone (UCZ), Temperature and Exposure are subjective assessments of the site and are supplied by the respective site operators. The UCZ is as defined by the WMO (2014) [1] 5 - Medium development; 6 - Mixed use with large buildings in open landscape; 7 - Semi-rural development. [2] A - Standard instruments in Stevenson Screen; B - Standard instruments in Stevenson Screen or manufacturer supplied AWS shortwave radiation screen; C - Standard instruments in Stevenson Screen or manufacturer supplied AWS shortwave radiation screen, site exposure 1 or less. [3] Exposure: I - Sheltered exposure; II - Restricted exposure; III - Standard exposure; V - Very open exposure.

	Longitude (°E)	Latitude (°N)	Instrument	Urban Climate Zone[1]	Instrumentation[2]	Exposure[3]
A	5.176	52.098	Vaisala WXT520	6	B	II
B	5.253	52.079	Davis VP2	5	C	I
C	5.294	51.396	Davis VP2	7	A	V
D	6.687	52.348	AcuRite 5-in-1	5	B	III

641 # Figures

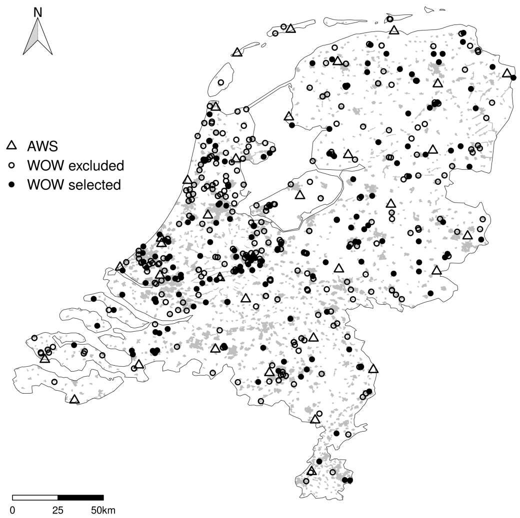

Figure 1: The locations of the official KNMI AWS and the WOW stations across the Netherlands. The WOW stations are categorized into those stations used in the study (WOW selected), and those that contained too many missing values for the models to be applied (WOW excluded). Urban areas defined using the Coordination of Information on the Environment (CORINE) land-cover dataset (CLC2018) are indicated by shading.

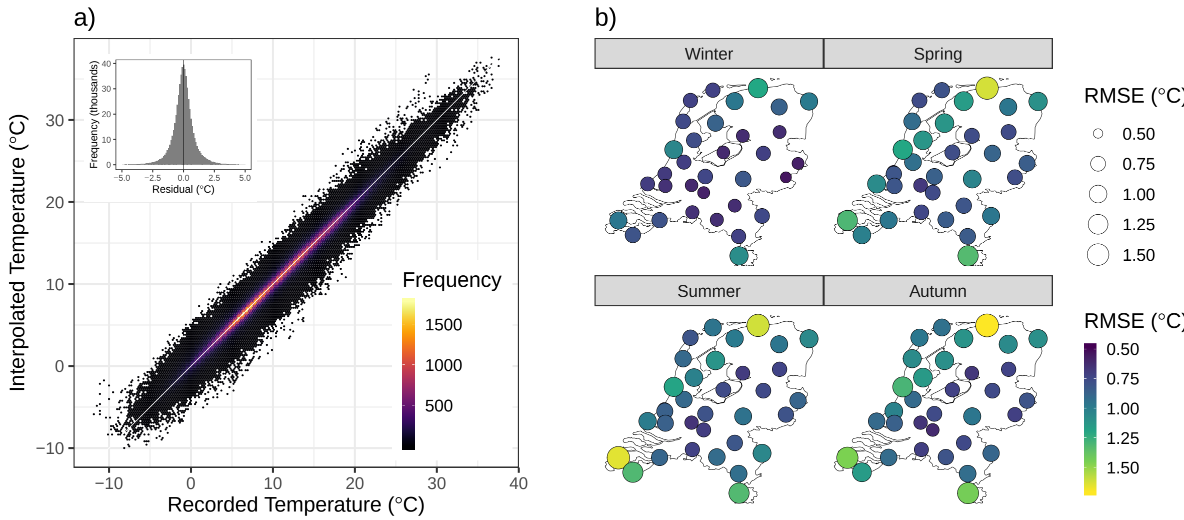

Figure 2: Results of the leave-one-out cross-validation of the KNMI temperature observations. a) shows a scatter plot of the interpolated values relative to the recorded temperature values, and b) shows the root-mean square error between the interpolated and recorded values. The inset in a) shows a histogram of the interpolated *minus* the recorded values. The seasons take the conventional meteorological definition (Winter as Dec-Feb, Spring as Mar-May etc.).

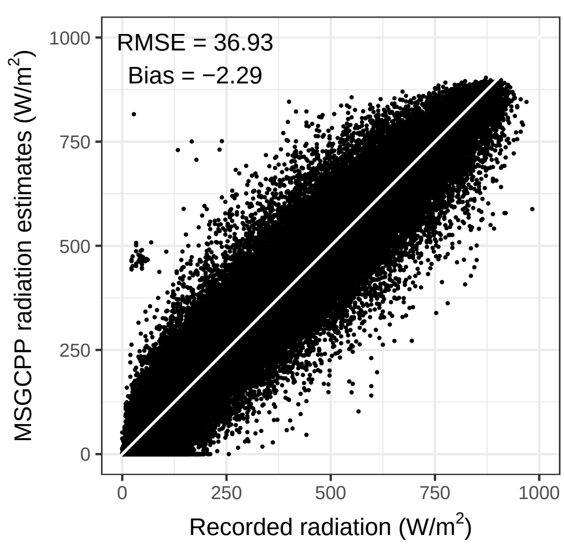

Figure 3: Hourly-mean comparison of the MSG-CPP direct shortwave radiation estimates relative to recorded shortwave radiation at the official KNMI stations over the period 2015–2016.

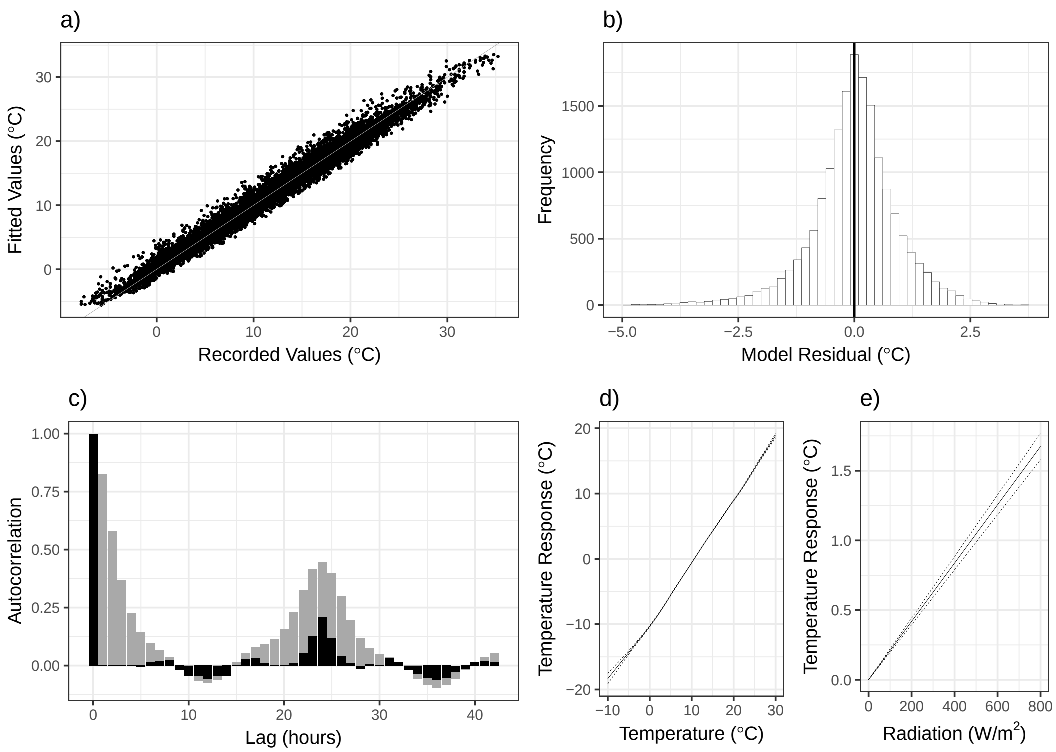

Figure 4: Plots of several model attributes for station B: a scatter plot of fitted values against predictand values (a); a histogram of the model residuals, with the vertical line indicating the mean value (b); autocorrelation function (ACF) for a GAMM with assumed lag-2 autocorrelation (black) and assuming independent residuals (grey) after Wood et al. (2017) (c); and the temperature (d) and shortwave radiation (e) partial terms from the model. In d) and e) the standard errors of the partial model terms are also indicated (dashed lines).

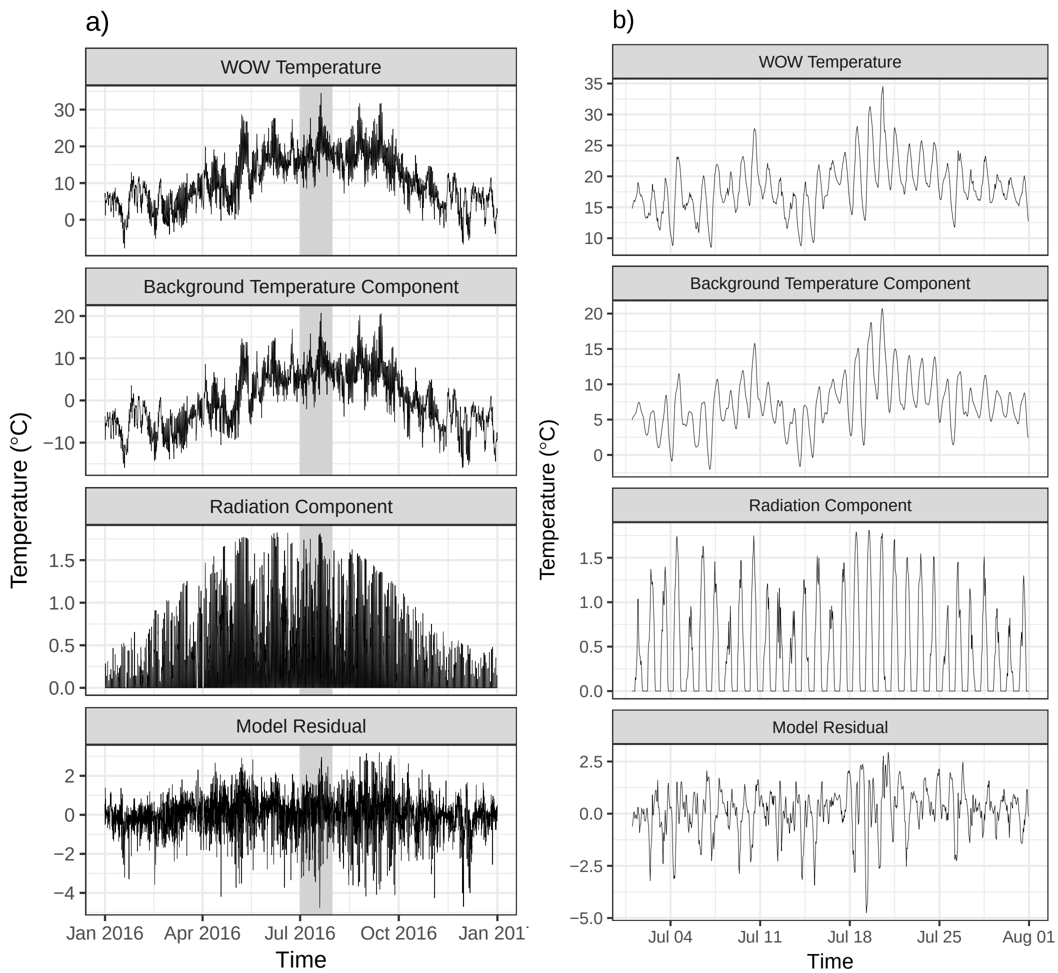

Figure 5: Time series plots showing the original WOW temperature data at test station B, the background temperature and shortwave radiation terms, and the model residual for the year 2016 (a) and for July 2016 (b). The shaded region in (a) indicates the period covered in (b). Note the slightly different y-axis scaling between (a) and (b)

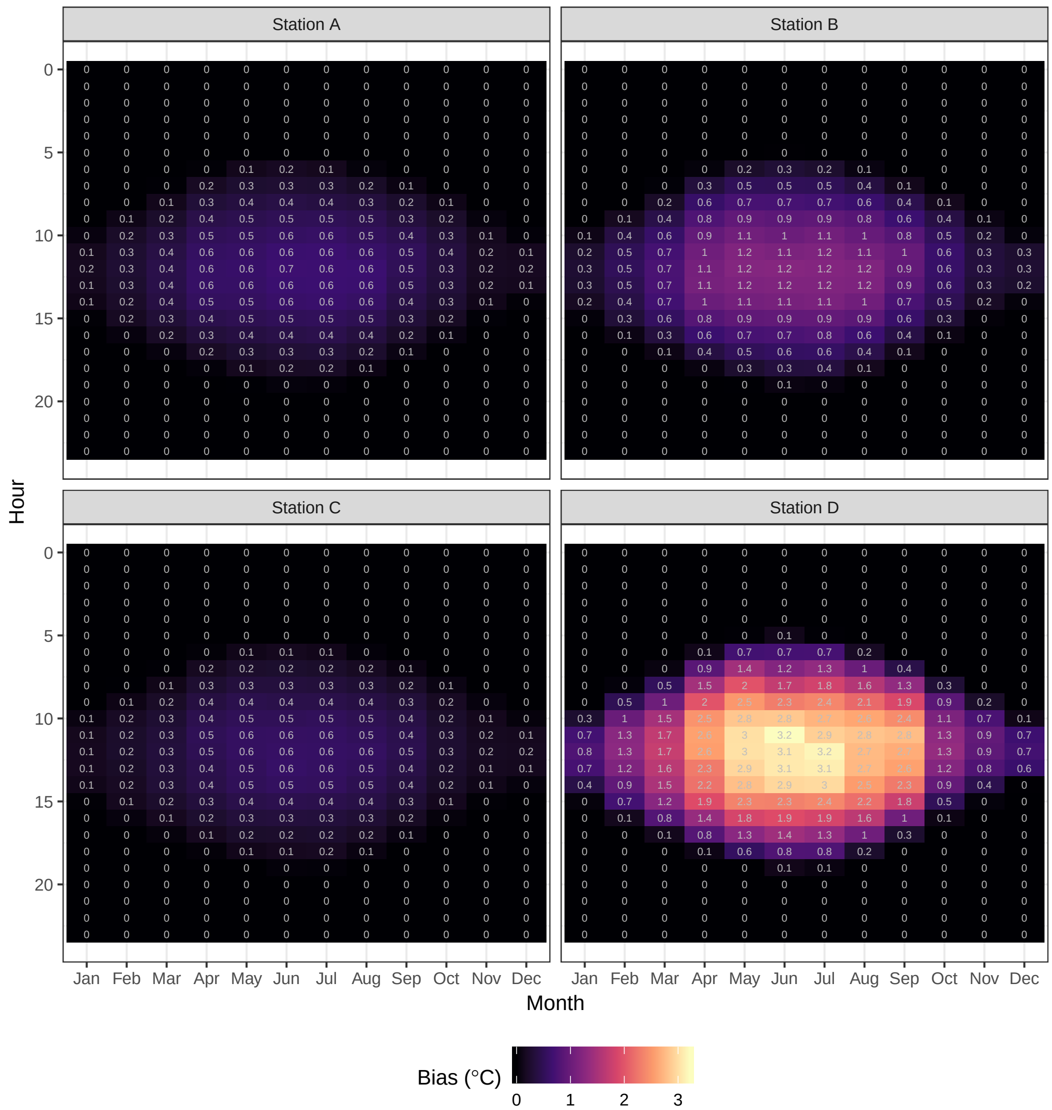

Figure 6: The estimated shortwave radiation bias ($T'_{wow,i} - T_{wow}$) for the four WOW test stations over the period 2015–2016. The plots indicate the average bias per month (x-axis) against the hour of the day (UTC, y-axis).

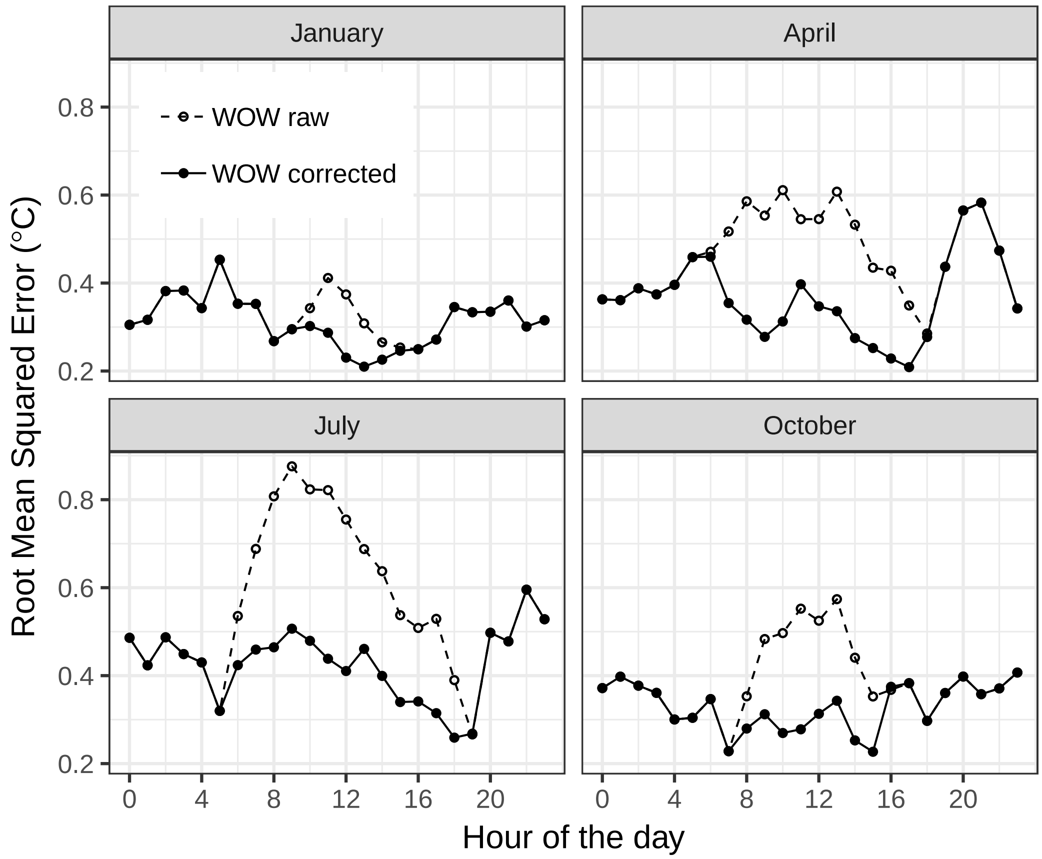

Figure 7: Root-mean squared error values of the raw and corrected WOW data at the De Bilt meteorological enclose relative to the official AWS values over the period 2015–2016. Values are calculated for each hour of the day in the months indicated.

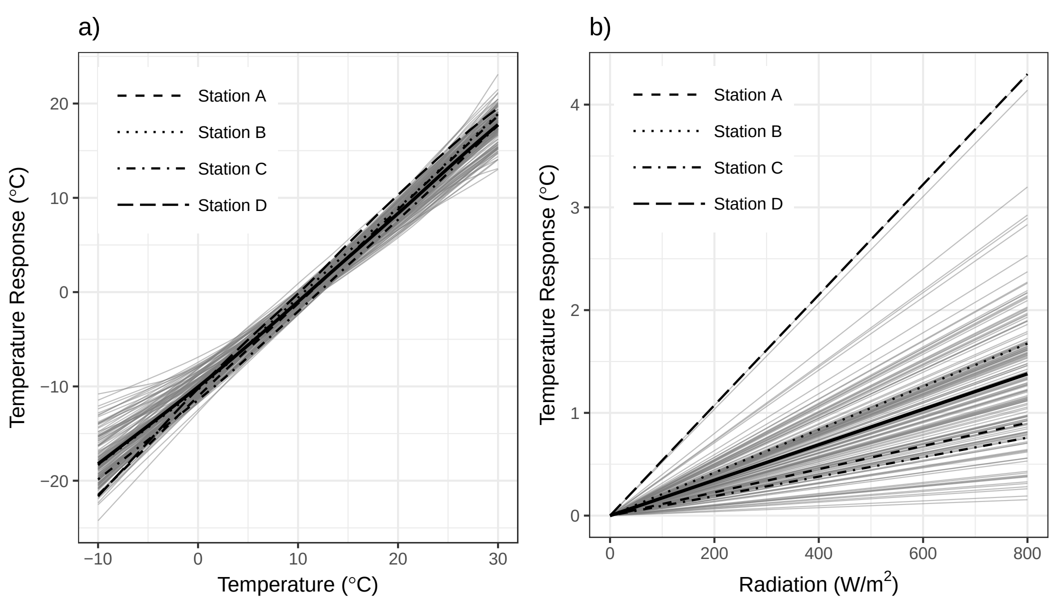

Figure 8: Plots showing the background temperature (a) and shortwave radiation (b) partial model terms calculated from the GAMMs for each WOW station (grey lines). The continuous black line indicates the mean across the station models, and the values for the four test stations are highlighted. The models are fitted using data from the full 24-month period.

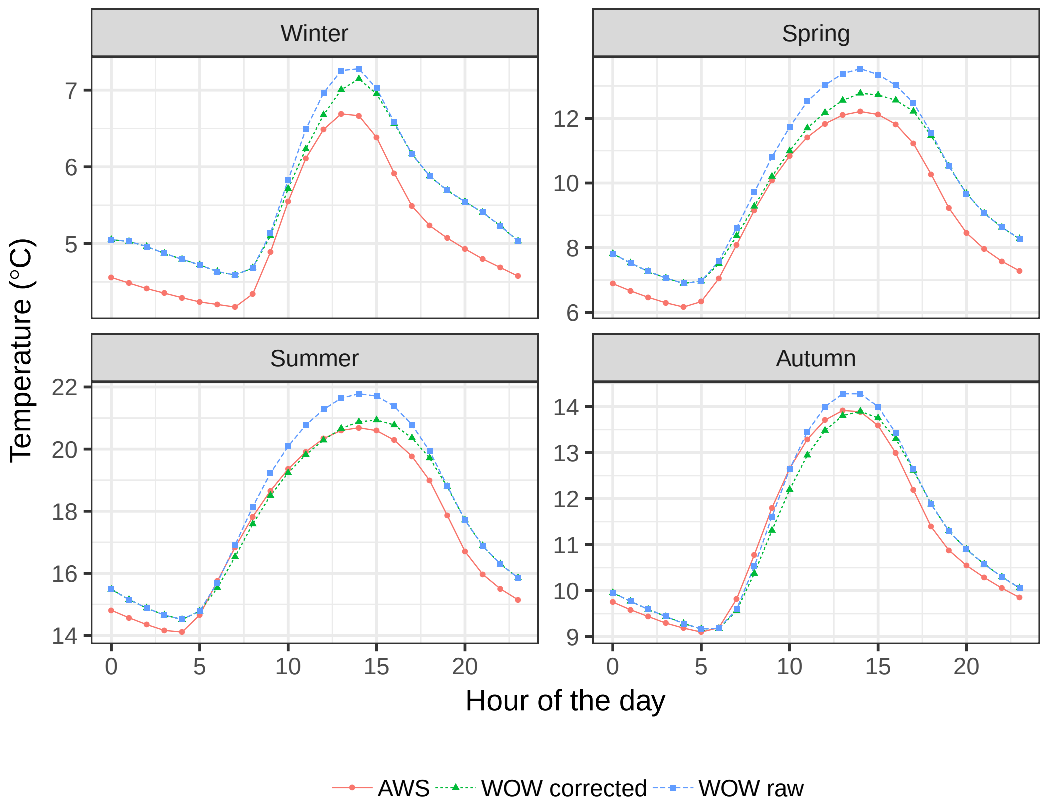

Figure 9: Average temperature values per hour of the day (UTC) over the period 2015–2016 for each each season calculated as the mean from the KNMI AWS weather stations and the raw/corrected WOW stations. Note the different y-axis in each of the panels. The seasons take the conventional meteorological definition (Winter as Dec-Feb, Spring as Mar-May etc.)

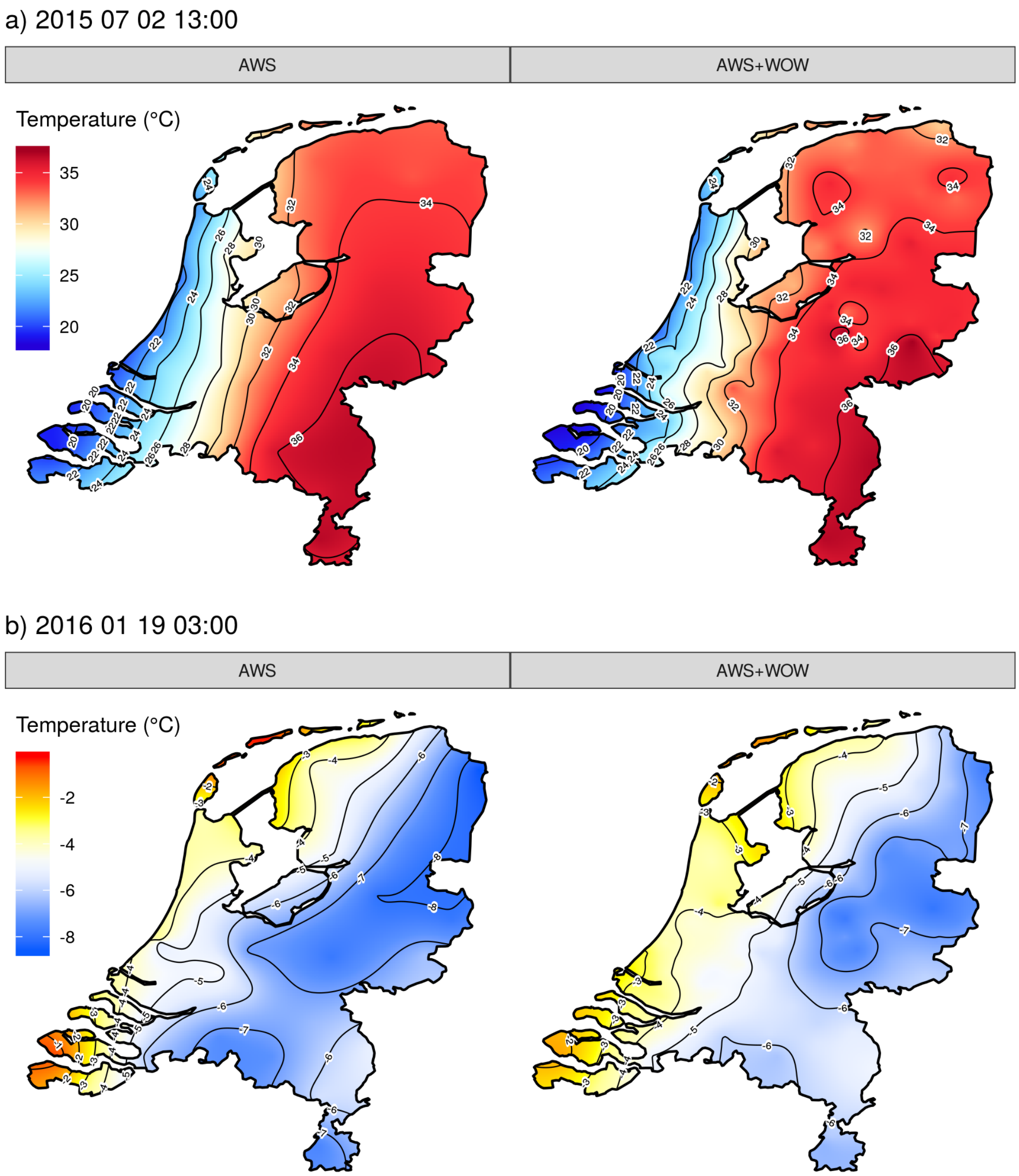

Figure 10: Maps of interpolated temperature for a) 1300 UTC on 2nd July 2015 and b) 0300 UTC on 19th January 2016. The maps on the left (AWS) are produced using only the official KNMI station readings, and the maps on the right (AWS+WOW) use both the corrected WOW temperature measurements and the AWS data. The maps are produced using ordinary kriging, with a separate variogram fitted to each map. Note the different contour spacing in a) (2 °C) compared to b) (1 °C).