Radboud University

Faculty of Social Sciences

Spatio-temporal analysis and error correction of consumer weather station data

Lisa Tostrams, s4386167

Supervisors:
Prof. dr. T. Heskes (Radboud University)
T. de Ruijter, Msc. (MeteoGroup)

Final version

Nijmegen, January 2017

Abstract

Since there is a need of more temperature observations in urban areas, the high availability of crowd sourced weather data might be of interest for meteorological purposes, such as modelling the higher temperatures in urban areas known as the Urban Heat Island (UHI) effect. However, there are quality issues that arise around Consumer Weather Stations (CWS) that need to be examined. These issues, such as radiation bias and measurement gaps, need to be addressed before the measurements are useful. Here, temperature measurements of around 400 stations in Amsterdam are examined over multiple months. During pre-processing the obvious incorrect measurements are discarded. Then temporal errors are removed using the probabilistic Kalman filter. Station covariance is estimated with Principal Component Analysis. Finally, spatial relationships are examined and corrected for with Kriging to determine validity of local measurements. These methods lead to simple to understand quality measures for individual station quality. The improvement of the data is quantified by the correlation with KNMI measurements and the usability of the data to demonstrate the Urban Heat Island effect in Amsterdam. The proposed methods are successful in removing poor performing stations as well as individual measurements.

Contents

Chapter 1

Introduction

Consumer Weather Stations

Over the past few years, there has been a steady growth in the use of automated weather stations by meteorology hobbyists. Besides being able to monitor your local weather state, these stations can notify users of interesting measurements and can even be integrated with other home automation systems. Many of the most common Consumer Weather Stations (CWS) have the option to upload the measurements to the internet to share and compare measurements with fellow meteorology enthusiasts. Some popular websites are the Netatmo site (www.netatmo.com), Wunderground (www.wunderground.com), WOW-UK (wow.metoffice.gov.uk) and WOW-NL (wow.knmi.nl).

There are currently over 10.000 CWS in the Netherlands regularly uploading weather data, compared to 31 official automated weather stations maintained by the KNMI (The Royal Netherlands Meteorological Institute, see www.knmi.nl/nederland-nu/weer/waarnemingen for a list of stations). In Amsterdam alone there are around 400 CWS of the brand Netatmo (see fig. 1.1). These new measurements could potentially be used in a range of applications, including high spatial resolution weather forecasting. Also, CWS data has been used to model the Urban Heat Island (UHI) effect [Steeneveld et al., 2011]. The UHI effect means that temperatures in urban areas can be expected to be higher than in rural areas, due to daytime heat storage and subsequent heat release after sunset. Steeneveld et al. used CWS data to asses the UHI effect in cities in the Netherlands.

Challenges of Consumer Weather Station data

Before CWS data can be used on a large scale there are some issues regarding the data that need to be addressed. These issues can lead to a high uncertainty as to how well the measurements represent the true state at that location. These challenges include calibration issues of individual sensors leading to bias or drift, design flaws, or sensor quality [Bell, 2014]. Most stations do not have metadata available about the upkeep and location of the modules. Another often occurring problem are software and communication errors, leading to long gaps in measurements. In general, more data is available during the day than during the night. A possible explanation is that some users switch off Wi-Fi at night. The CWS may also be placed in a way that is not optimal for weather monitoring, in the full sun or in an area without ventilation, under umbrellas, or even inside. In the Netatmo manual [NWS01, 2012], Netatmo itself lists a number of reasons why measurements may be missing, including low battery power, inference in signal by walls or wind, the devices being located too close to each other, unpowered inside module, missing Wi-Fi signal, or network change detection.

Meier et al. [Meier et al., 2015] developed a 5-level quality measure for Netatmo stations, based on available station-specific metadata, data availablity and a comparison with UCON atmospheric data. Arguments for different quality levels are listed as Netatmo API and server limits, user-specific operating errors (especially in communication metadata), failure of the wireless network, loss of battery power, user-specific installation error, device being set up indoors, and device not

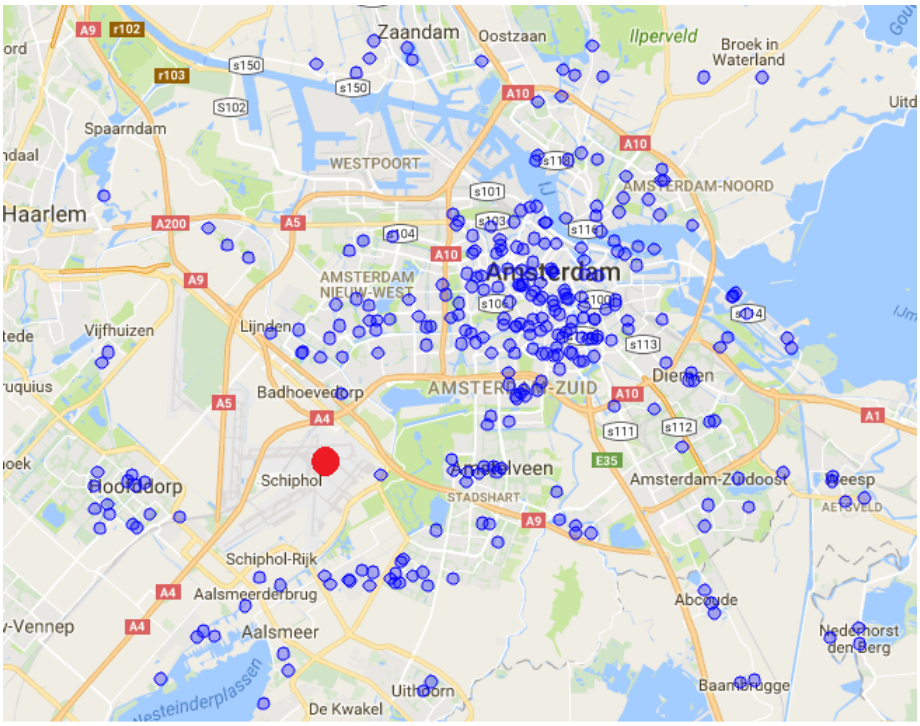


Figure 1.1: Amsterdam, the Netherlands, with locations of 419 Netatmo brand consumer weather stations in November 2016 plotted in blue. The red dot is the location of the nearest KNMI weather station, at Schiphol Airport.

having been set up in the shade. Part of the problems with the stations is attributed to the fact that there are no standard guidelines on how to use the devices, leading to incorrect installment and use of sensors. At the highest quality level, the daily number of available stations varied between 350 and 831, out of 1100 total stations.

Research question

These problems can make CWS data difficult to process. There is research employing meteorological methods to estimate the quality of crowd sourced weather sensor data, Netatmo temperature data, and Netatmo rainfall data [Bell, 2014, Meier et al., 2015, de Vos et al., 2016]. However, there are many other fields of study that deal with processing high uncertainty measurements, such as time series analysis, robotic motion planning and control, and trajectory optimization, among others. These fields of study have developed mathematical and statistical methods for the processing of possibly unreliable measurements. Which leads to the main research question of this thesis: How can we apply existing spatial and temporal analysis to correct for measurement errors and bias in consumer weather station temperature data?

In order to validate the results and quantify improvement, both the improvement in correlation with KNMI measurements and the usability of the data to observe the UHI will be considered.

Chapter 2

Methods

2.1 Data

The data used for analysis came from CWS that uploaded their measurements to the Netatmo website. Over the months of April and November 2016, the air temperature measurements in Amsterdam and its direct surroundings were considered (see figure 2.1). The data on the Netatmo website came from Netatmo branded weather stations [NWS01, 2012]. Besides the outside module, the weather station setup includes an inside device that transfers the data automatically to the website. The popularity of these devices can be attributed to their cost effectiveness, ease of setup, and visualisation of the data via smart phone. The stations use AA batteries for power, and transfer their data over Wi-Fi.

Since it is not known exactly how reliable the CWS data is, and the actual temperatures at the measurement locations are unknown, the measurements from the nearby KNMI station at Schiphol were used as a means of comparison.

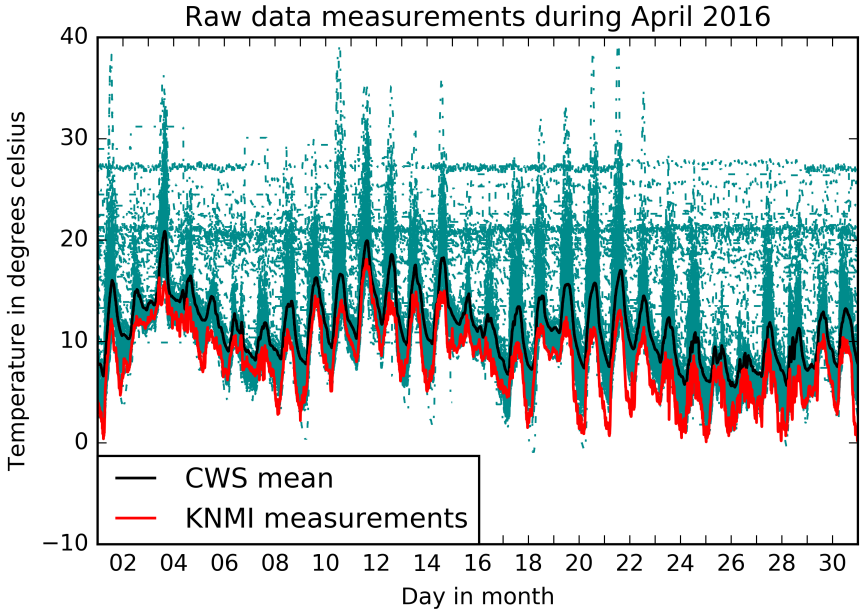


Figure 2.1: Unprocessed temperature measurements of 354 Netatmo stations in Amsterdam, April 2016. Plot includes the mean of CWS measurements and KNMI measurements at Schiphol for comparison.

Netatmo measurements are publicly available, although the API usage limits querying large quantities of data. The Netatmo partner API provides the latest known measurements of world wide locations. Of the 150.000 stations world wide, approximately 10.000 are located in the Netherlands. To collect data, the partner API was polled every 10 minutes during April and November for the measurements of stations located in Amsterdam. These months were chosen to represent different seasonal effects on temperature measurements. Amsterdam was chosen as location for the amount of available stations, and the concurrent research of its urban climate. During data collection, duplicate measurements are removed and possible timestamp shifts are corrected, with the same methodology as in de Vos et al. [de Vos et al., 2016].

To make sure all analysis was done correctly, the most obviously unreliable data is removed during pre-processing. During temporal analysis, the individual station issues are addressed first before collective station behaviour can be analysed. Spatial analysis is done last, since some necessary assumptions could not be done based on the raw and pre-processed data.

2.2 Data Pre-processing

Before analysing the data, some pre-processing was done.

Resampling and imputation

The different stations offered measurements of varying intervals, ranging from a measurement every minute to a few measurements per day. This variation did not only occur between stations, but also within stations, as many stations were affected by gaps in measurements. These varying intervals and gaps made certain computations difficult, and comparisons impossible. The data was resampled at a fixed interval of 10 minutes to synchronize stations. New measurements were computed using the mean of surrounding measurements, the non-numerical values occurring in the data were padded with the nearest numerical value. Stations having over 75% of missing data were considered unreliable and therefore excluded.

Outlier removal

After resampling, the most unlikely measurements ($z - score > 3$, in a normal distribution this includes 0.03% of data) were removed from the data. The data is not normally distributed but has a right skewed distribution, with outliers being extremely high temperature measurements, so generally a z-score would not be considered an appropriate quantifier for outliers. However, considering the nature of the data, a temperature measurement being more than 3 standard deviations away from the mean is very unlikely. Z-score of a measurement x_i from station i at a certain time point t is computed as

$$z_{it} = \frac{x_i - \bar{x}_{*t}}{\sigma_{*t}}$$

with

$$\bar{x}_{*t} = \frac{1}{n}\sum_{i=1}^{n} x_{it}$$

being the mean and

$$\sigma_{*t} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_{*t} - \bar{x}_{*t})^2}$$

being the standard deviation of all n stations at time step t.

Per day, the correlation of each station with measurements from the KNMI was computed. The Pearson product-moment correlation coefficient ρ_i of all 144 measurements x_{i*} of a station i with the KNMI station k measurements x_{k*} is computed as

$$\rho_i = \frac{\text{covariance}(x_{i*}, x_{k*})}{s_{i*}s_{k*}}$$

with s_{i*} and s_{k*} being the standard deviations of the measurements on that day. The covariance of the measurements of two stations i and k, each including $T = 144$ measurements, is

$$\text{covariance}(x_{i*}, x_{k*}) = \frac{1}{T} \sum_{t=1}^{T-1} (x_{it} - \bar{x}_{i*})(x_{kt} - \bar{x}_{k*})$$

with $\bar{x}_{i*}$ and $\bar{x}_{k*}$ being the means of the measurements on that day.

Since a certain type of variation in the data is likely, namely the 'warm during the day, cold during the night' variation or diurnal cycle, a minimal correlation with high quality temperature measurements from the KNMI is likely. To determine a reasonable minimal correlation, measurements of multiple CWS were plotted against KNMI measurements. A visual comparison of 'good' and 'bad' stations (for example, see figure 2.2) determined that a minimal correlation coefficient of 0.3 resulted in the filtering of obviously incorrect data. Since most stations had a high average coefficient, changing the minimum between 0 and 0.6 did not make a significant difference, see figure 2.3.

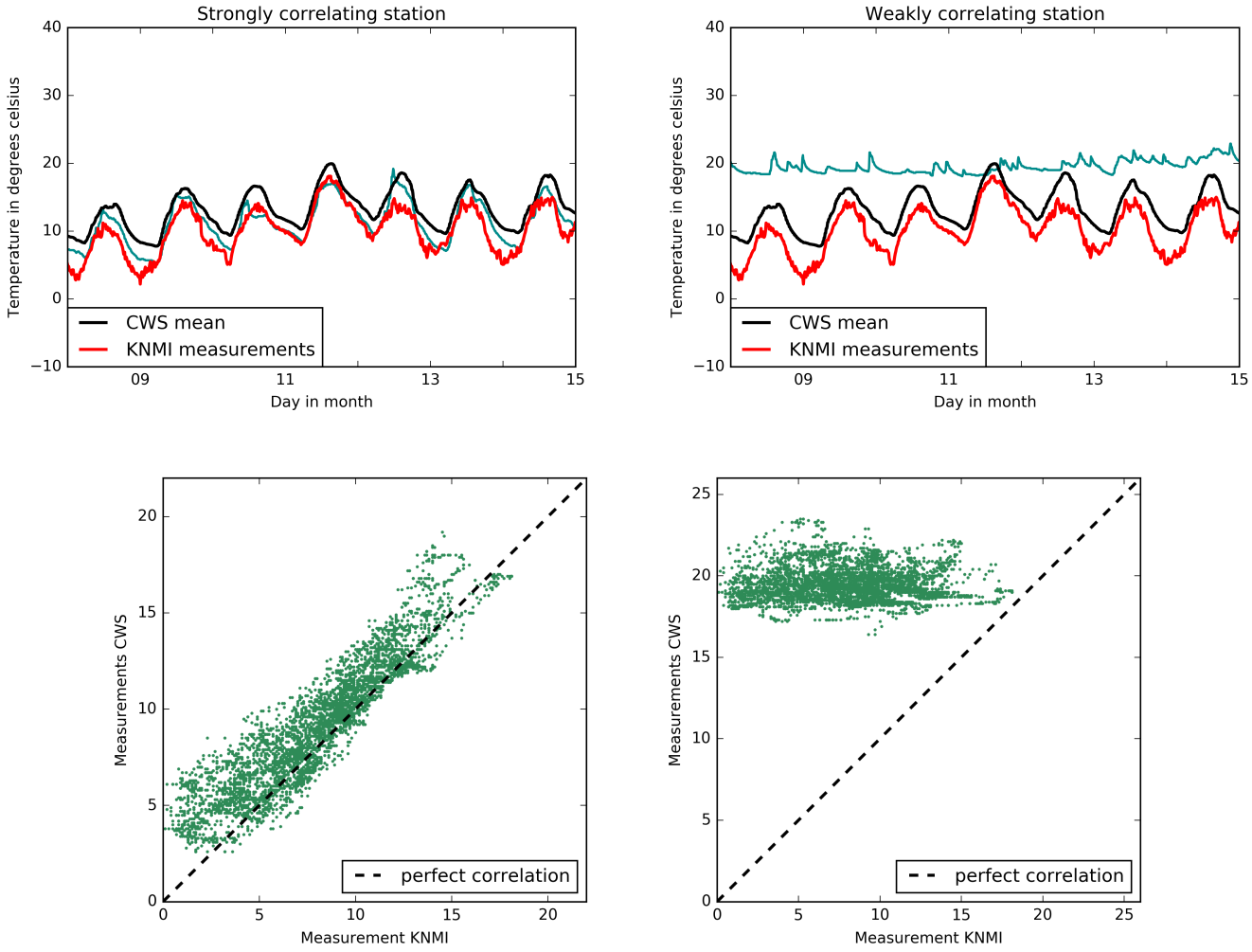


Figure 2.2: Visual comparison of a 'good' (strongly correlating) station vs a 'bad' (weakly correlating) station. Strongly correlating station has an average correlation coefficient of 0.91 per day with KNMI measurements. Weakly correlating station has an overall average coefficient of 0.1, and its measurements are discarded. The weak relationship could be caused by station being placed inside. By plotting the station measurements against the KNMI measurements, the correlation plot shows how the station relates to the KNMI.

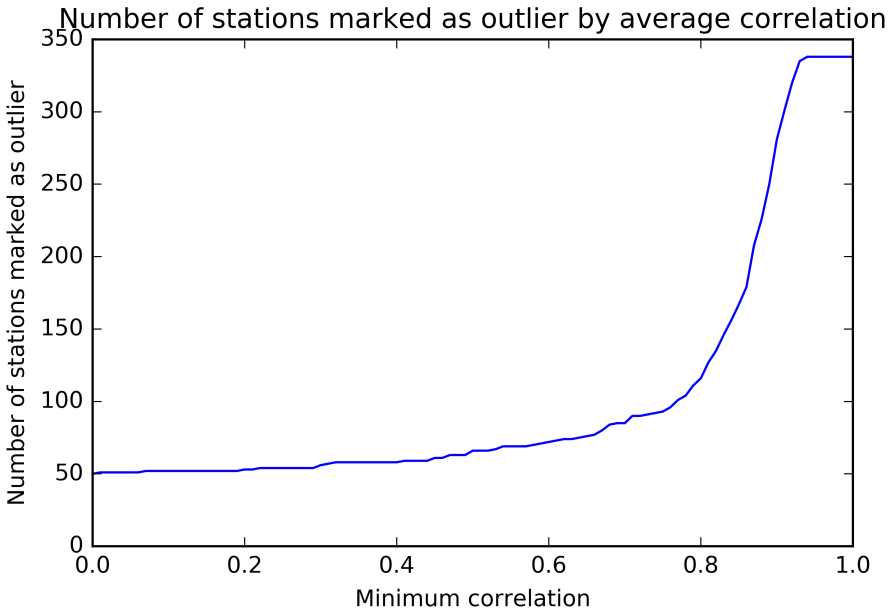


Figure 2.3: Number of stations marked as outlier by their average correlation coefficient.

2.3 Temporal Data Processing

Two methods were used for temporal processing. The Kalman filter, a method for estimating values from measurements, was applied per station and corrects for short term temporal errors. This filter can be applied to reduce measurement noise in temperature sensors [Eom et al., 1996]. The calculated mean square error gives an indication of how affected measurements are by rapid increases of temperature. Principal Component Analysis was used to identify the most important patterns in the data, and the unexplained variation was computed to identify diverging stations. PCA can be used to characterize temperature measurements, as seen in [Benzi et al., 1996].

2.3.1 Individual station variance: Kalman filter

The Kalman filter is a set equations that provides a means to estimate the state of a process, in a way that minimizes the mean of the squared error [Welch and Bishop, 2006]. The Kalman filter addresses the general problem of trying the estimate the true state τ_{it} of a time controlled system from a measurement x_{it} at time step t with

$$x_{it} = \tau_{it} + \epsilon$$

where the variable ϵ has variance R and represents the measurement noise at that time. R may vary with time or may remain constant. In this application, R was expected to vary during the day. The temperature measurements of CWS were assumed to include a non-zero noise component caused by sensor bias [Bell, 2014]. In lack of a precise model, the estimation of this bias was based on the difference with KNMI measurements. This means that application of the filter corrected the CWS measurements to behave similarly to KNMI measurements, especially when it comes to rate of change in temperature. Results of the filter should be interpreted as such.

The noise of the time controlled system is normally distributed with variance Q. Determining the process variance Q is difficult because, in theory, the system that is estimated cannot directly be observed. A more detailed look into the setting of the parameters of the Kalman filter can be found in appendix A.

During the estimation process, only one station at a time was considered, and for convenience the station index i is omitted from equations.

The *a priori* state estimate at step t is defined as $\hat{\tau}_t^-$, and the *a posteriori* state estimate is defined as $\hat{\tau}_t$, at time step t given measurement x_t. The goal of the Kalman filter is to compute the *a posteriori* state estimate $\hat{\tau}_t$ as a linear combination of an *a priori* estimate $\hat{\tau}_t^-$ and a weighted difference between an actual measurement x_t and measurement prediction related to the previous estimate (in this application, simply the previous estimate) $\hat{\tau}_t^-$, given by

$$\hat{\tau}_t = \hat{\tau}_t^- + K_t(x_t - \hat{\tau}_t^-)$$

K_t is chosen to be the *gain factor* that minimizes the *a posteriori* error estimate P_t

$$P_t = (1 - K_t)P_t^-$$

One form of K_t that minimized P_t is given by

$$K_t = \frac{P_t^-}{P_t^- + R_t}$$

The difference $(x_t - \hat{\tau}_t^-)$ is called the measurement *residual*. The residual reflects the difference between the predicted measurement and the actual measurement. When the measurement error R approaches zero, the gain K_t weights the residual more heavily. When the estimate error P_t approaches zero, the gain K_t weights the residual less heavily. After each time and measurement update, the process is repeated with the previous *a posteriori* estimates used to project or predict the new *a priori* estimates.

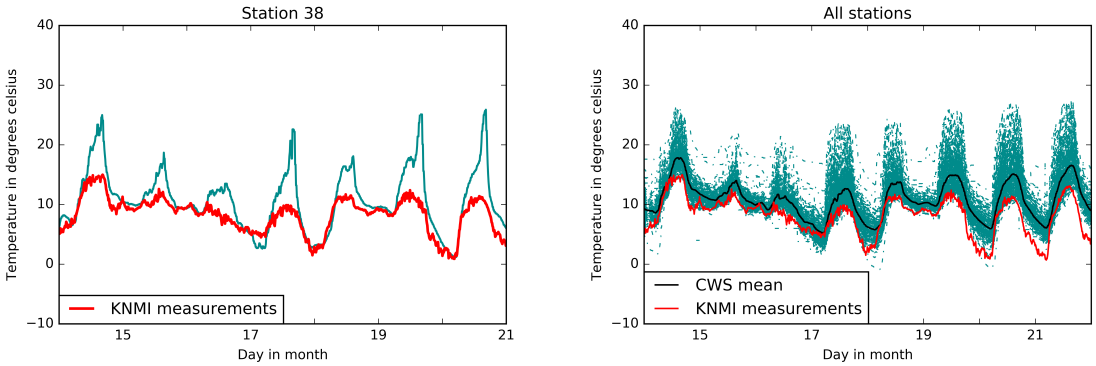


Figure 2.4: Temperature measurements of a few days of a single station that records rapid temperature increases during the day, and all station measurements during that period. Plot includes KNMI measurements at Schiphol for comparison.

The Kalman filter relates the predicted state $\hat{\tau}_t$ to the previous state estimate $\hat{\tau}_t^-$ and the actual measurement x_t. If the expected variance is low, say 0.1 degrees Celsius, but the difference between the previous estimate and the actual measurement was high, say 2.5 degrees Celsius, then the actual measurement becomes more unlikely. In this case, the current state estimate is more strongly related to the previous estimate than to the actual measurement.

CWS seem to measure much higher temperatures during the day than the KNMI station, as seen in figure 2.4. This station records temperature measurements that are likely during most of the night and mornings. In the afternoon the measurements become more unlikely, indicating that the temperature at the station is heating up at a much higher rate than at the KNMI station. This might happen because air temperature measurements are susceptible to solar radiation error [Nakamura and Mahrt, 2005], where the temperature is overestimated due to overheating of components in the sun. The Kalman filter was expected to work especially well here, when the variation from measurement to measurement far exceeds the expected variation rate.

To quantify how much correction the Kalman filter applies to the measurements, the mean squared error (MSE) between measurements and Kalman predictions was computed:

$$MSE = \frac{1}{T}\sum_{t=1}^{T}(\hat{\tau}_t^- - x_t)^2$$

2.3.2 Global station covariance: Principal Component Analysis

Principal component analysis (PCA) is a statistical method that can be used to capture the variability of data in less attributes. PCA uses orthogonal transformation to represent the data in statistical uncorrelated dimensions. From these dimensions, the data can be reconstructed, losing a certain amount of variation in the process [Tan et al., 2013]. PCA tends to identify the strongest patterns in the data. Since patterns caused by unlikely measurements are probably weaker than patterns caused by unlikely measurements, reduction of dimensionality can eliminate some of the measurement noise. When the data is reconstructed, this noise is left behind. The daily patterns of UHI in the temperature measurements were expected to remain constant in most stations, and should not be affected by noise reduction.

The variability of an n by T data set X, containing T measurements of n stations, can be summarized in an T by T covariance matrix C, which has entries c_{pv} defined as:

$$c_{pq} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ip} - \bar{x}_{*p})(x_{iq} - \bar{x}_{*q})$$

where p and q are the p^{th} and q^{th} attribute (column) of the data set X respectively, and represent measurement moments. $\bar{x}_{*p}$ and $\bar{x}_{*q}$ represent the mean of the columns p and q, and thus depict

the average measured temperature at those moments. The entry c_{pv} represents the covariance between two measurement moments.

On the diagonal of the covariance matrix you find the variance in each attribute, or dimension, of the data. The direction of the data with the largest variance is the first Principal Component. The orthogonal direction with the second largest variance is the second Principal Component, and so on. These directions correspond with the eigenvectors V of the covariance matrix C, ordered by their corresponding eigenvalues $s_1, ..., s_T$. The relative size of the eigenvalues also corresponds to the variance explained by each Principal Component.

The data can be represented in these new dimensions by projecting the data onto the first m eigenvectors V_1 to V_m, resulting in n by m projection Z_m:

$$Z_m = X[V_1, .., V_m]^{\mathrm{T}}$$

with $[V_1, .., V_m]^T$ being the transpose of $[V_1, ..., V_m]$. The percentage of variance explained p_m by this projection can be computed from the eigenvalues $s_1, .., s_m$:

$$p_m = \frac{\sum_{l=1}^{m} s_l}{\sum_{l=1}^{T} s_l} \times 100\%$$

The data can be reconstructed from this projection to n by T reconstruction $\hat{X}$:

$$\hat{X} = Z_m[V_1, .., V_m]$$

In this reconstruction, p_m percent of variance is captured.

PCA finds the measurement moments that explain the most variance in the data. The measurements that explain $> 99.99\%$ of the data were calculated, and the remaining unexplained variance was calculated per station. This unexplained variance give an estimation as to how much each stations behaviour relates to the other stations in the dataset.

2.4 Spatial Data Processing

Besides temporal (in-)consistencies, measurements also can also be likely or unlikely based on the homogeneity or nearby measurements. A variogram was calculated to analyse spatial coherence in station measurements. Kriging was applied to estimate the temperature at a station location, based on measurements in the direct vicinity. The difference between the Kriging prediction and the actual measurement was calculated as an estimate of how each stations measurements relates to station measurements nearby. During Kriging, only one measurement moment t is considered at a time.

2.4.1 Spatial variance: Kriging

Kriging is a method of interpolation, originally stemming from geostatistics to predict spatially correlated ore grades in gold mines. The goal of Kriging is to provide best linear unbiased predictions for the value of a function at a given point by computing a weighted average of known values in the neighborhood [Oliver and Webster, 2014]. The estimate is based on assumptions of spatial covariance, for which a function is computed beforehand. Generally, when the distribution of measurements is normal and has an expected mean of zero, Kriging is referred to as 'simple Kriging' or Gaussian Process regression [Rasmussen and Williams, 2016]. When the mean is not expected to be constant, the method is mostly referred to as 'ordinary Kriging'. Since only one time step is considered, the time step notation t is omitted from equations for convenience.

When the mean cannot be assumed to be constant in for all locations, the spatial covariance function is defined as half the variance of the differences. This is called the semivariance, and is a function of distance d.

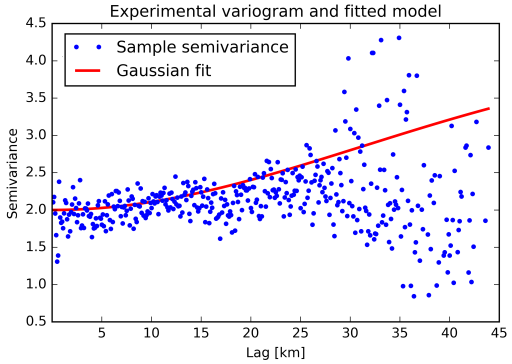


Figure 2.5: The average experimental variogram over the month April, with the Gaussian model fitted. The experimental variogram shows the semivariance of different locations calculated from the actual station measurements. Distances were collected in 100 meter bins, so the semivariance at 5 kilometer is the semivariance of stations located at a distance between 4950 and 5050 meters from each other. The Gaussian model models the continuous semivariance function $\gamma(d)$ by estimating the parameters c_0 (spatially uncorrelated variance, i.e. the variance at zero distance) and c (spatially correlated variance) from the data.

The experimental variogram is the variogram calculated from the data (see figure 2.5), based on estimates of semivariance at each distance d.

$$\text{semivariance}(d) = \frac{1}{2n_d} \sum_{\{i,j\}} (x_i - x_j)^2$$

Here $\{i, j\}$ is the set of pairs of observations x_i, x_j that have $d - 50m < d_{ij} < d + 50m$ with d_{ij} denoting the distance between stations i and j (see figure 2.6 for histogram of distances). n_d depicts the number of pairs of stations at distance d. The next step is to fit a smooth curve on the experimental variogram. Popular models are the exponential, spherical, and Gaussian models. From these three, the Gaussian had the best fit (See appendix B for other models).

The Gaussian model to estimate the semivariance is defined as

$$\gamma(d) = c_0 + c\left(1 - \exp\left(-\frac{d^2}{a^2}\right)\right)$$

The nugget variance c_0 depicts the uncorrelated variance, c is the correlated component of the variation, and a is the effective distance. See figure 2.5 for the fitted Gaussian model on the averaged experimental variogram.

Spatial estimation of the temperature measurement $\hat{x_i}$ at location i is calculated as a linear combination of the η closest observed values $x_{i1}, \cdots, x_{i\eta}$ and corresponding weights for x_i, $w_{i1}, \cdots, w_{i\eta}$:

$$\hat{x_i} = \sum_{j=1}^{\eta} w_{ij} x_{ij}$$

The locations of the η closest stations are denoted as $i1, \cdots, i\eta$ and the distance between two locations i and ij is written as $d_{i,ij}$. The weights for x_i are calculated as follows:

$$\begin{pmatrix} w_{i1} \\ \vdots \\ w_{i\eta} \end{pmatrix} = \begin{pmatrix} \gamma(0) & \cdots & \gamma(d_{i1,i\eta}) \\ \vdots & \ddots & \vdots \\ \gamma(d_{i\eta,i1}) & \cdots & \gamma(0) \end{pmatrix}^{-1} \begin{pmatrix} \gamma(d_{i1,i}) \\ \vdots \\ \gamma(d_{i\eta,i}) \end{pmatrix}$$

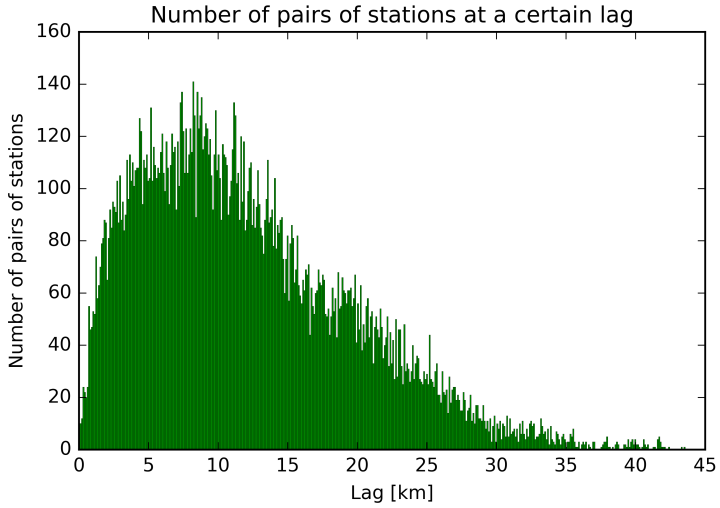


Figure 2.6: Histogram of lag between pairs of stations. For each distance d, the number of pairs of stations at that distance is n_d in the experimental variogram. Distances were collected in 100 meter bins.

The Kriging error variance is given by

$$\sigma^2(x_i) = \sum_{j=1}^{\eta} w_{ij}(\hat{x}_i - x_{ij})^2$$

Since the Kriging prediction is the mean of a normal distributed range of predictions, the Kriging error variation provides the possibility to calculate a 99% confidence interval for the predicted value:

$$\text{conf}_{99} = (\hat{x}_i - 2.58 \times \sigma(x_i), \hat{x}_i + 2.58 \times \sigma(x_i))$$

To keep computations operational, the optimal number of closest observed values η to consider in estimating the average temperatures at each location was determined by comparing the estimated runtime complexity with the mean squared error MSE,

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{x}_i - x_i)^2$$

for $\eta = 1, .., 99$. To estimate runtime complexity, the time to perform all computations was measured in seconds. Both $\eta = 10$ and $\eta = 38$ had the lowest MSE for all η with runtime < 50 seconds (see figure 2.7). Considering the large amount of measurement moments, 4320 per month, $\eta = 10$ was used for all computations. Besides the overal MSE per measurement moment, the average MSE per station over all measurement moments was calculated. This MSE indicates how the measurements of one location relates to the measurements of the η closest stations, and evaluates if the station performs as a spatial outlier.

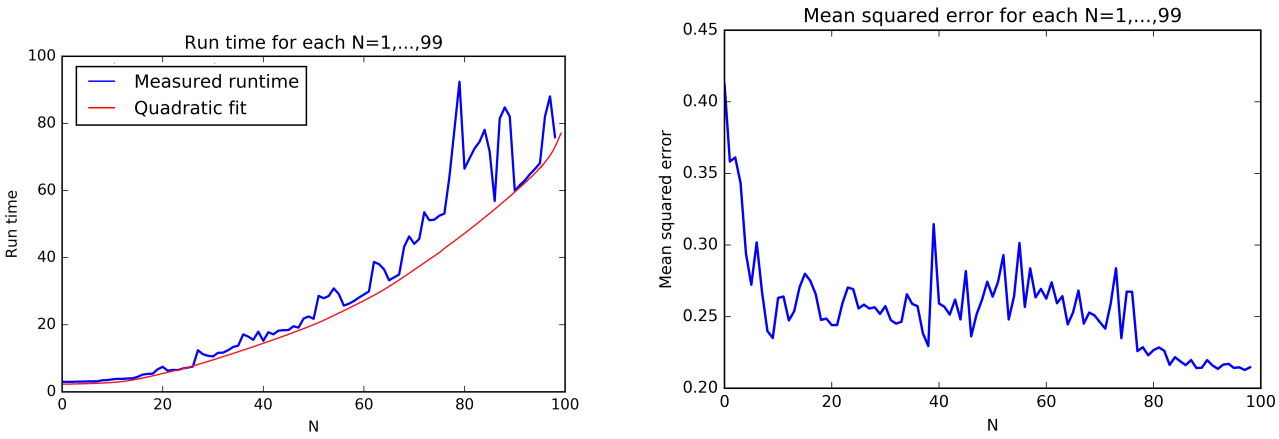


Figure 2.7: Runtime and mean squared error for each number of considered closest observed values. To obtain these values, the computations to estimate the average temperatures at each location were executed with parameter η set to 1,..,99 for all stations, and the MSE between the measurement and prediction of each location was calculated.

Chapter 3

Results

For each applied method described in the previous chapter, a quality evaluation measure was calculated to estimate the reliability of the measurements of individual stations are. By plotting the 10 best or 10 worst performing stations according to each quality evaluation, it is possible to demonstrate the contribution of each method contributes to the overall analysis and quality improvement. The unprocessed measurements are shown in figure 3.1. The analysis started with 354 stations in April and 419 stations in November. In order to asses the improvement, the correlation with KNMI measurements is used as comparison. The unprocessed CWS measurements compared to the KNMI measurements show three major issues:

- A few stations do not show the diurnal cycle.
- Some CWS measure extremely high temperatures during the day.
- Some station measurements do not correlate with KNMI. measurements

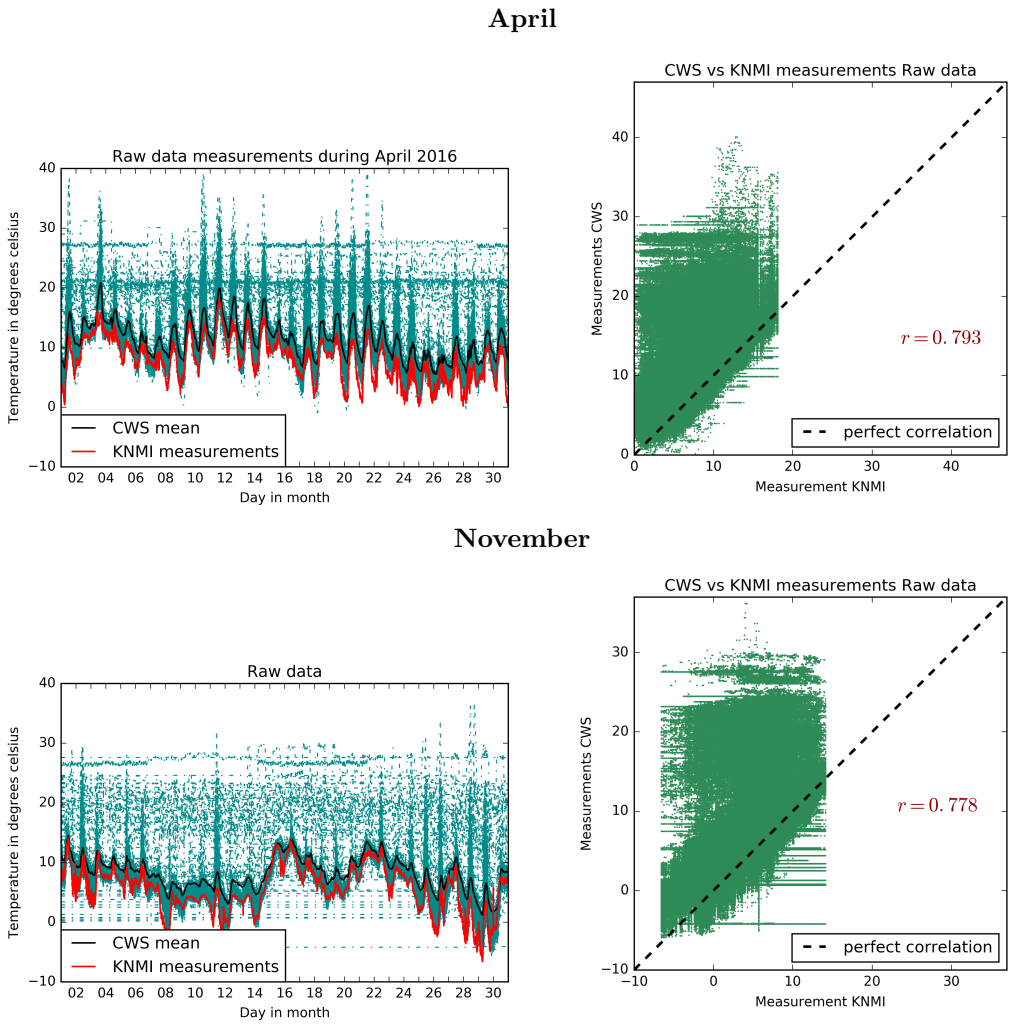


Figure 3.1: Unprocessed temperature measurements during April (top) and November (bottom).

3.1 Pre-processing

After removing stations that missed 75% or more of data, 297 and 330 stations were left to be resampled in April and November, respectively. For all of these stations, the daily correlation coefficient with the KNMI station was computed. By removing the measurements during days with a correlation coefficient lower than 0.3, most of the non-cyclic data is removed (see figure 3.2). For 47 stations in April and 70 stations in November, every day of measurements had a correlation coefficient of below 0.3, and thus these stations were excluded. The measurements that were marked as outlier by their z-score were considered unreliable and thrown out as well. This results in a higher average correlation with the KNMI measurements.

Most stations correlated strongly with the KNMI station, having median coefficients of 0.91 and 0.93 in April and November, respectively (figure 3.3). The stations with a coefficient near 0 do not vary with the KNMI at all. The negative coefficients might be caused by stations having lagged data uploads, reporting measurements from hours ago as if recorded now, resulting in a difference in phases.

Figure 3.4 shows the best and worst performing stations according to their correlation with the KNMI station. This gives an indication of which stations were removed from the dataset during pre-processing.

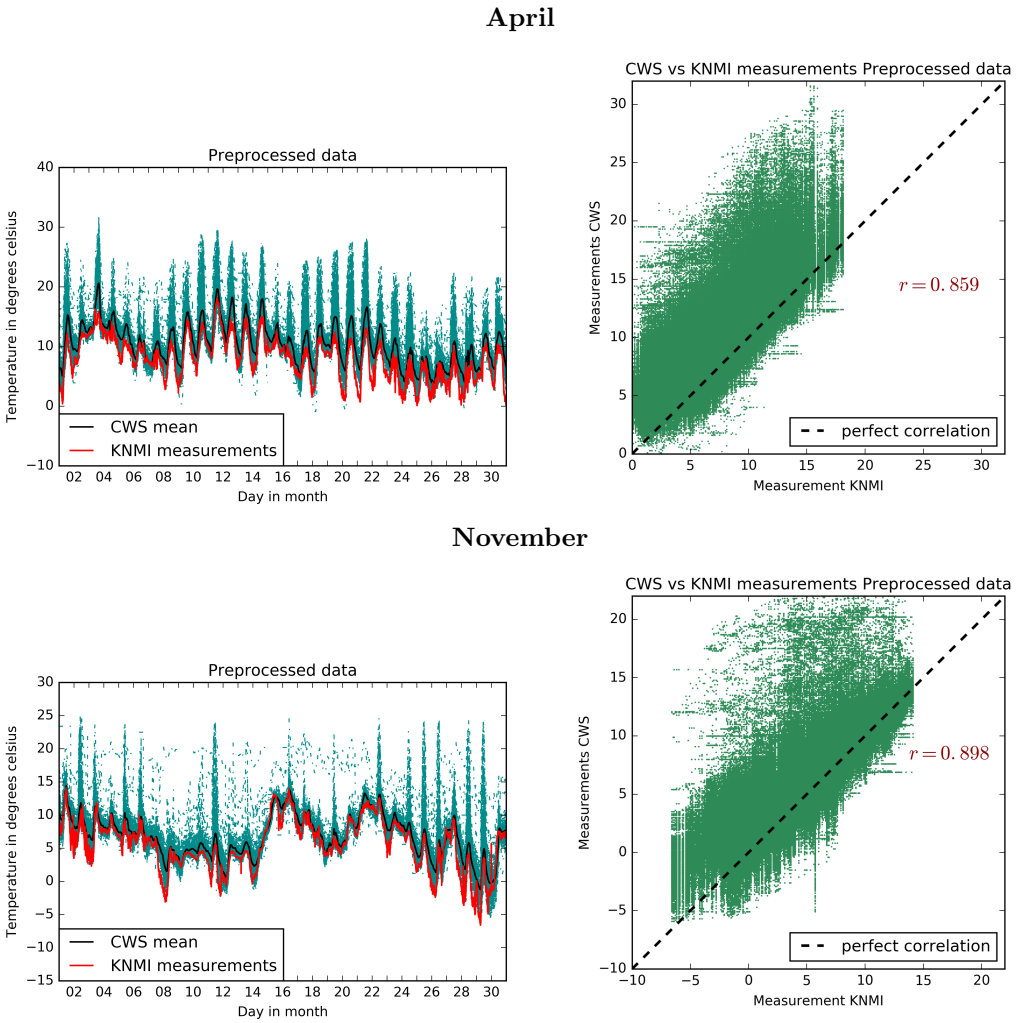


Figure 3.2: Preprocessed temperature measurements during April (top) and November (bottom). Measurements marked as outlier have been removed from the data, either according to z-score or the strength of the correlation with KNMI measurements.

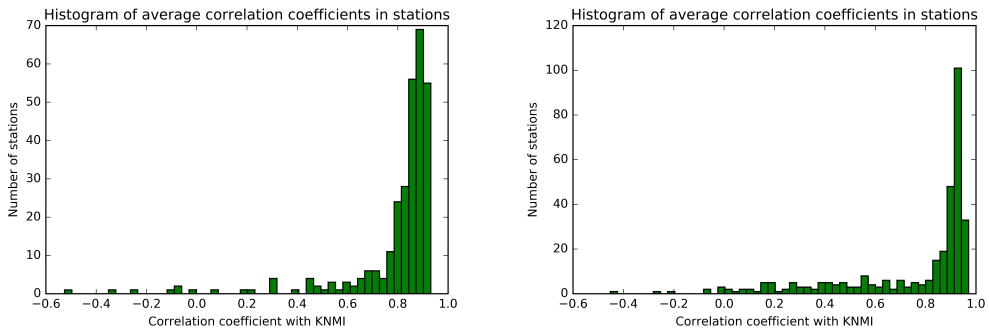


Figure 3.3: Quality evaluation scores of stations during April (left) and November (right), corresponding to the average correlation coefficient with KNMI measurements.

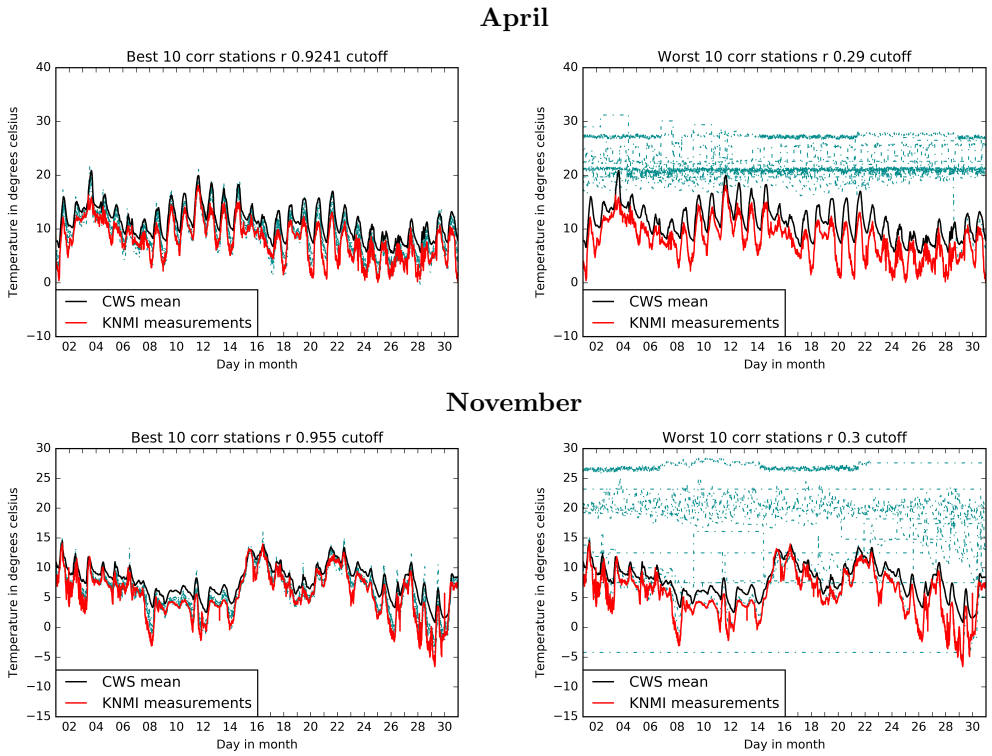


Figure 3.4: Best and worst performing stations during April (top) and November (bottom). The correlation correctly identifies stations that do not show the diurnal cycle in their measurements.

3.2 Temporal processing

3.2.1 Kalman filter

The effect of the Kalman filter on a station that records rapidly increasing temperatures can be seen in figure 3.5. In this example, the filter handles the high temperature increases well. When applied to all stations, the Kalman filter results in a decrease in overestimation of temperature during the day.

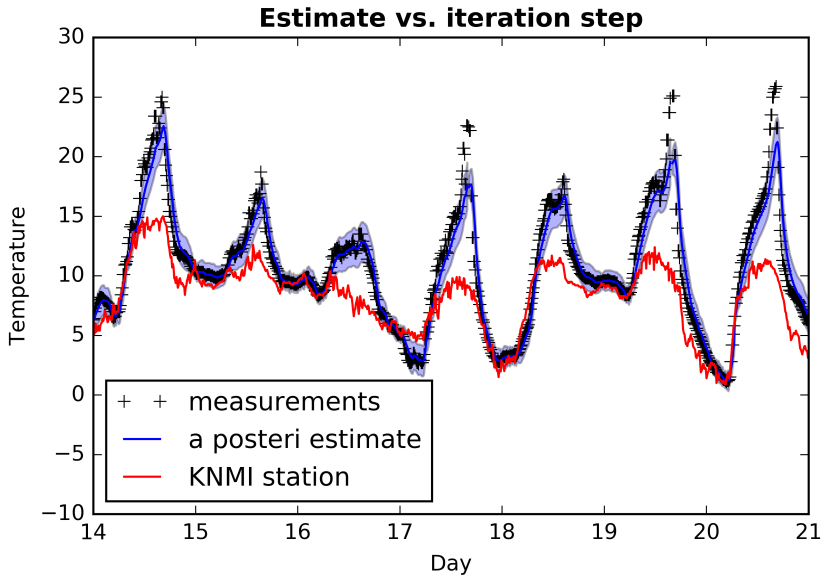


Figure 3.5: Kalman filter effect on station measurements. The blue area represents the 95% confidence interval of the Kalman filter estimate (blue line).

The slight decrease in correlation (see figure 3.6) might be caused by the fact that the Kalman filter mostly influences the measurements when the temperatures are high. This could potentially de-linearize the relationship with the KNMI station slightly.

The difference between the Kalman filter estimate and the original measurement is expressed as the mean squared error (MSE). A high MSE indicates that the Kalman filter generally applied a strong correction to the measurements. In figure 3.7 the MSE of all stations are plotted in a histogram. The MSE scores in November are generally much lower than the MSE scores in April, indicating that either the Kalman filter was dealing with a much higher uncertainty in estimating the measurement error, or that the measurements in November were in less need of correction by the filter than the measurements in April.

Plotting the best and worst performing stations according to the MSE (figure 3.8) shows what type of measurements are corrected most by the filter. As expected, these are measurements from stations that record rapid temperature increments.

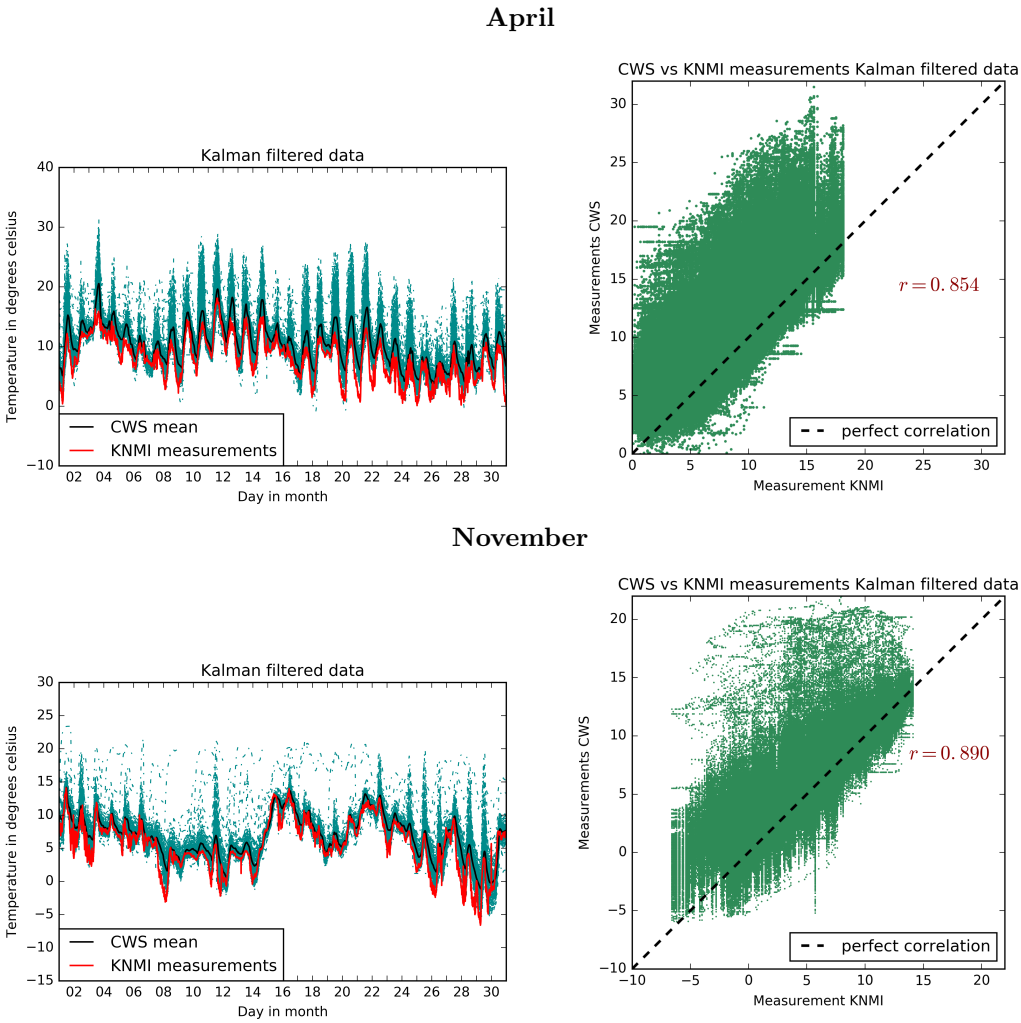


Figure 3.6: Kalman filtered temperature measurements during April (top) and November (bottom). Stations with a high MSE are stations that report high temperature increases during the day.

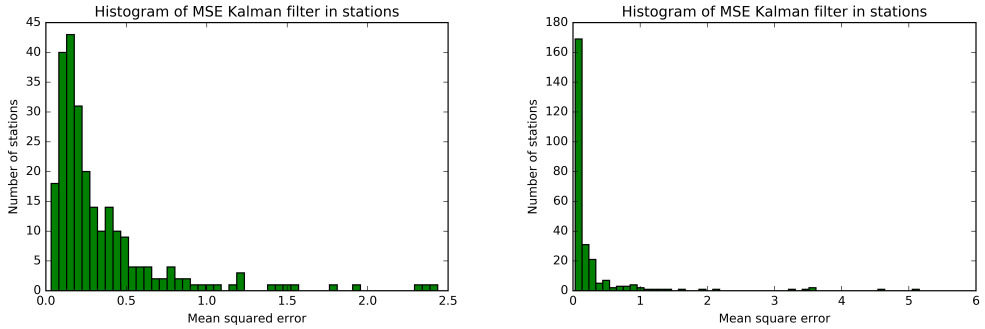


Figure 3.7: Quality evaluation scores of stations based on mean squared error April data (left) and November data (right).

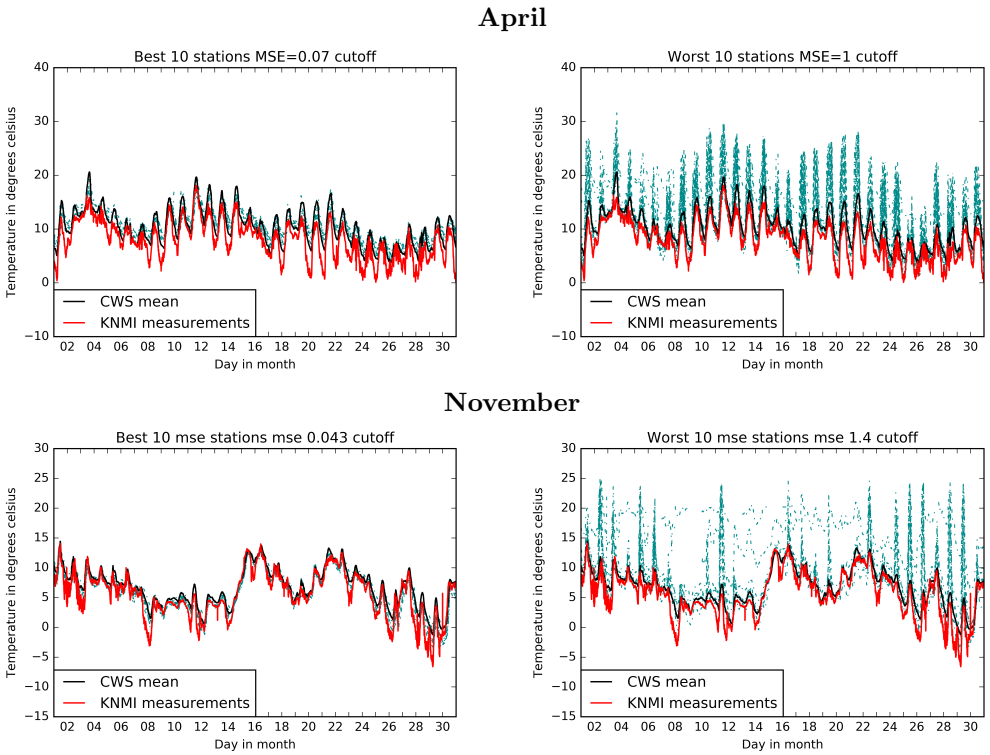


Figure 3.8: Best and worst performing stations in April (top) and November (bottom).

3.2.2 Principal component analysis

During PCA the measurements were projected onto the principal components that explained 99.99% of the variance in the complete dataset. The remaining unexplained variance was computed per station as a measure of how much the measurements diverged from the most important patterns in the data.

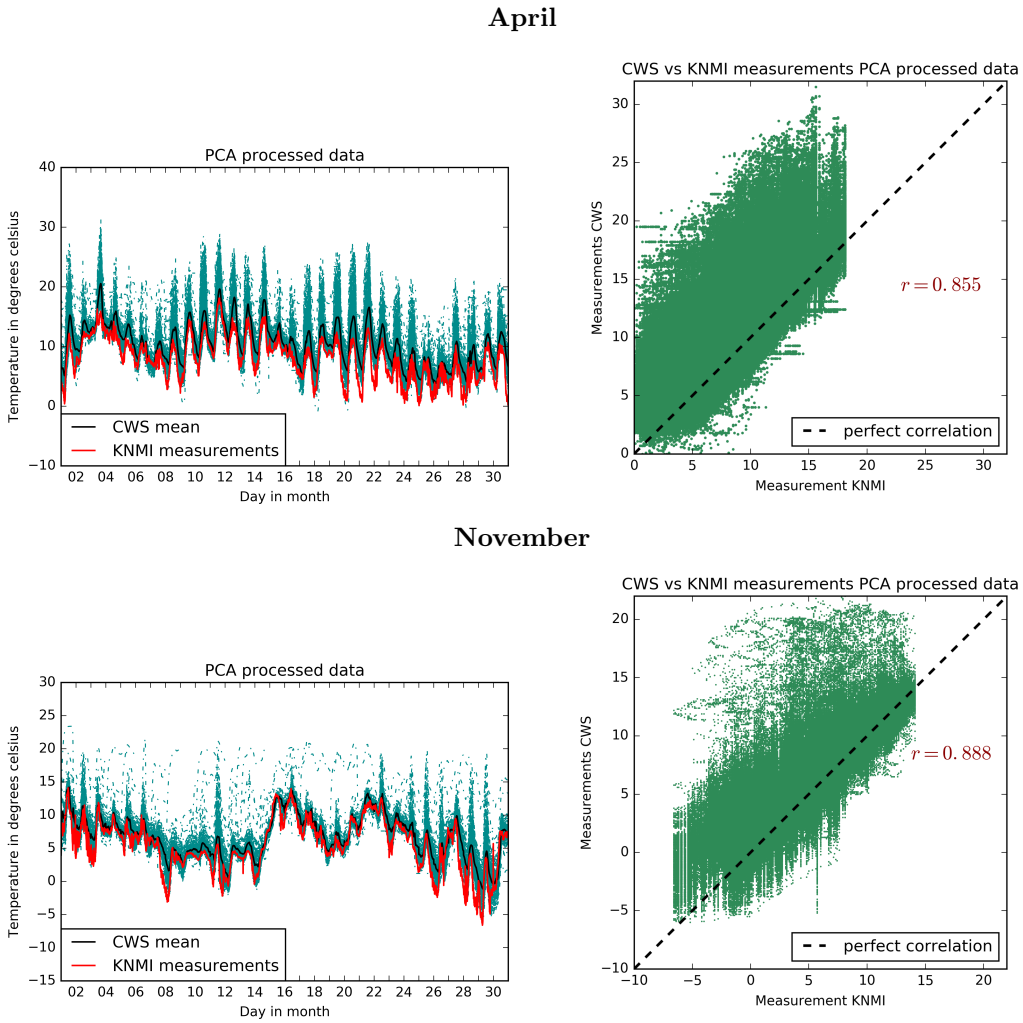


Figure 3.9: PCA processed temperature measurements during April (top) and November (bottom).

Projecting the data onto the first few principal components did not modify the measurements considerably, but computing the unexplained variance per station did identify some seemingly 'noisy' stations, as can be seen in figure 3.10. Contrary to the correlation coefficient with the KNMI, this quality evaluation score does not necessarily identify stations that do not behave like the KNMI station behaves, but it identifies stations that do not behave like the stations in the rest of the city does. The best performing stations do tend to overestimate the temperature compared to the KNMI, but this overestimation is apparently a pattern that can be validated by most stations.

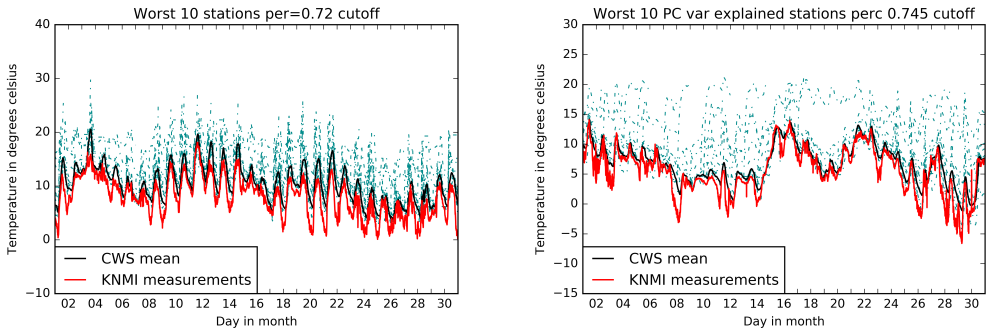


Figure 3.10: Worst performing stations in April (left) and November (right). The PCA identifies 'noisy' stations.

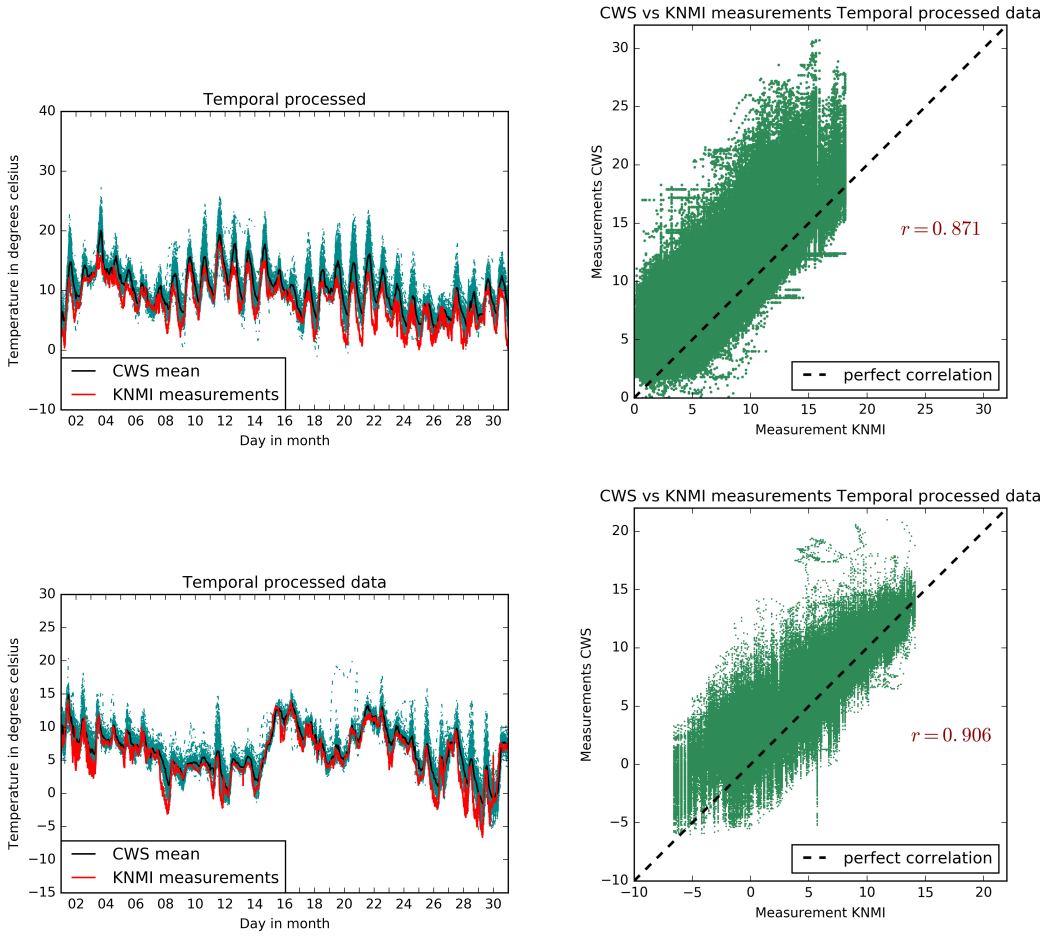


Figure 3.11: Temporally processed data without worst 15% performing stations, according to PCA explained variance during April (top) and November (bottom)

Using the unexplained variance as a quality measure, the 'worst' 15% of the dataset was removed, resulting in the remaining 212 and 221 station measurements depicted in figure 3.11.

3.3 Spatial processing

3.3.1 Kriging

Kriging was performed after temporal processing, since the correlation between station variation and distance was non-existent or even negative before outlier removal and temporal error corrections (see Appendix C). Using the Kriging error variation, a confidence score can be computed for each actual measurement, compared to the Kriging estimate at that location. A confidence score of below 2.58 indicates that the actual measurement falls within the 99% confidence interval of the prediction. To allow the possibility for the urban climate to cause strong spatial variations without those variations being disregarded, a measurement was not marked an outlier until a confidence score of over 6. Outlier measurements were replaced by the Kriging estimate for their locations. Figure 3.12 shows the locations and average measurements of stations before and after Kriging. The model of spatial covariance seemingly allows for warmer temperature measurements towards the center of the city.

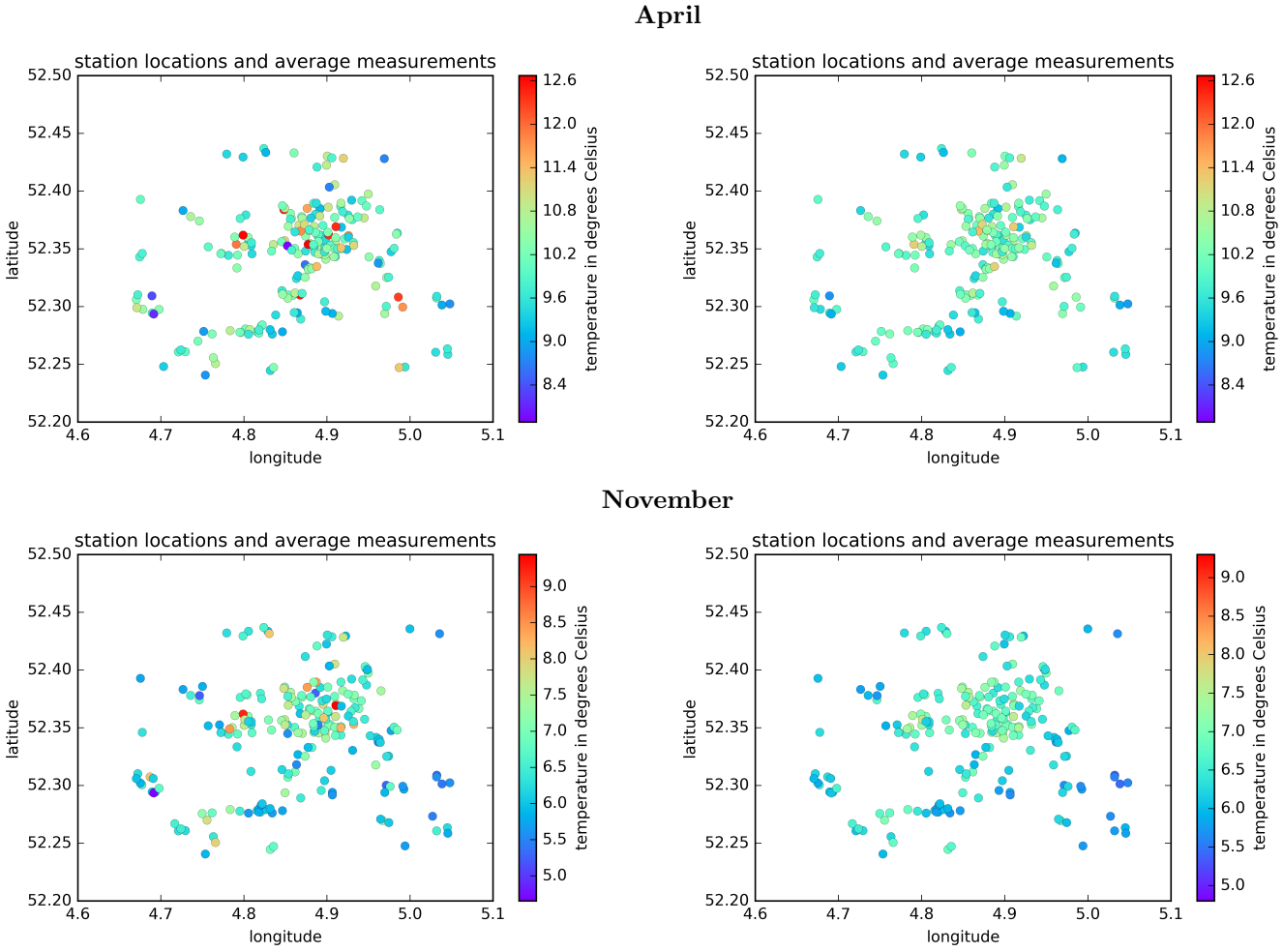


Figure 3.12: Locations and average measurements during April (top) and November (bottom) of stations before and after Kriging correction

Contrary to earlier processing, Kriging identifies measurements that are unlikely based on nearby measurements at the time. The resulting measurements are plotted in figure 3.13.

For each station, the difference between measurements and estimate is expressed in the mean squared error. Again, the measurements during November require fewer corrections than the

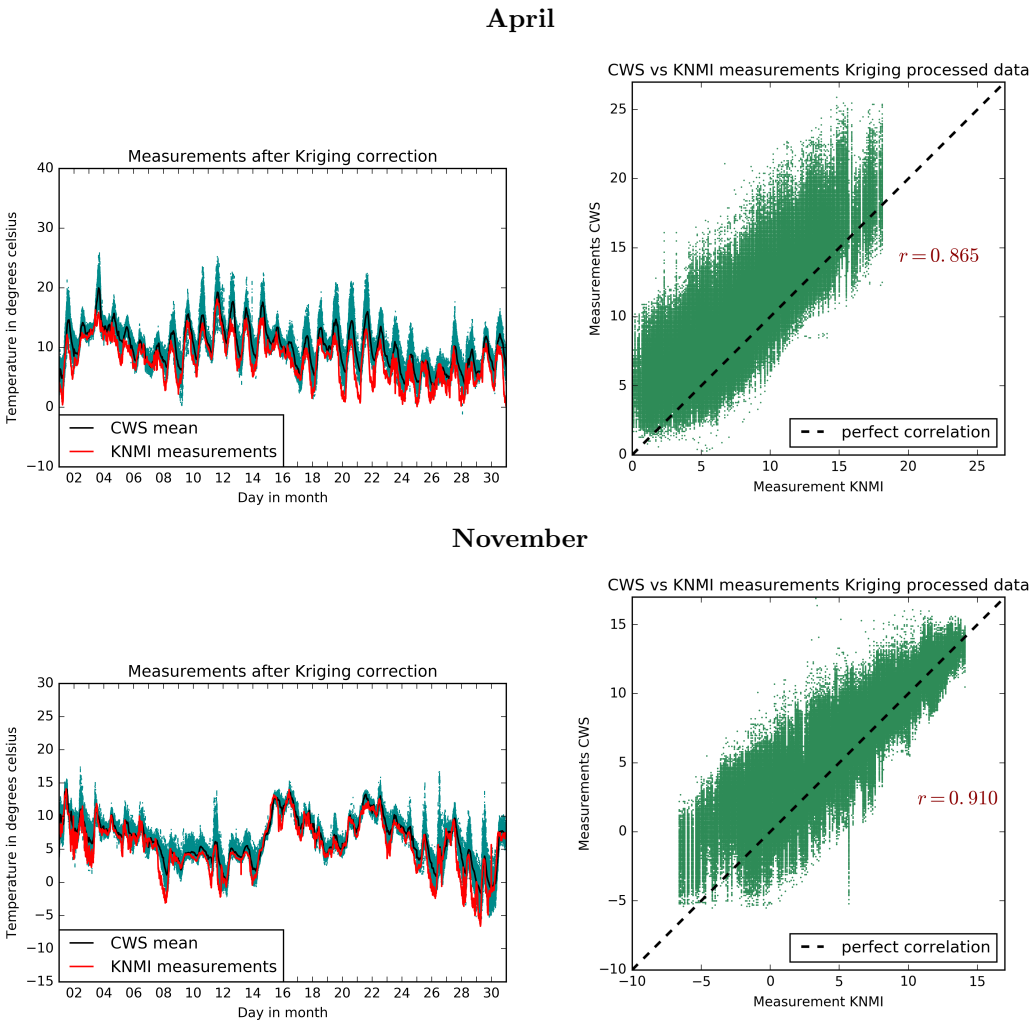


Figure 3.13: Measurements after Kriging correction

measurements during April, resulting in lower MSE scores (figure 3.14). The average MSE of all stations during the day changes between April and November as well, with a generally lower MSE score, a smaller difference between MSE during the day and during the night, and a shift from maximum MSE occurring around 15:00 to occurring around 12:30. It can be concluded that even spatial outliers might be caused by some stations being in the sun while others are not, especially since the higher MSE scores are seemingly strongly influenced by the longer days in April (higher MSE between 07:00 and 19:00) and the shorter days in November (higher MSE between 09:00 and 17:00).

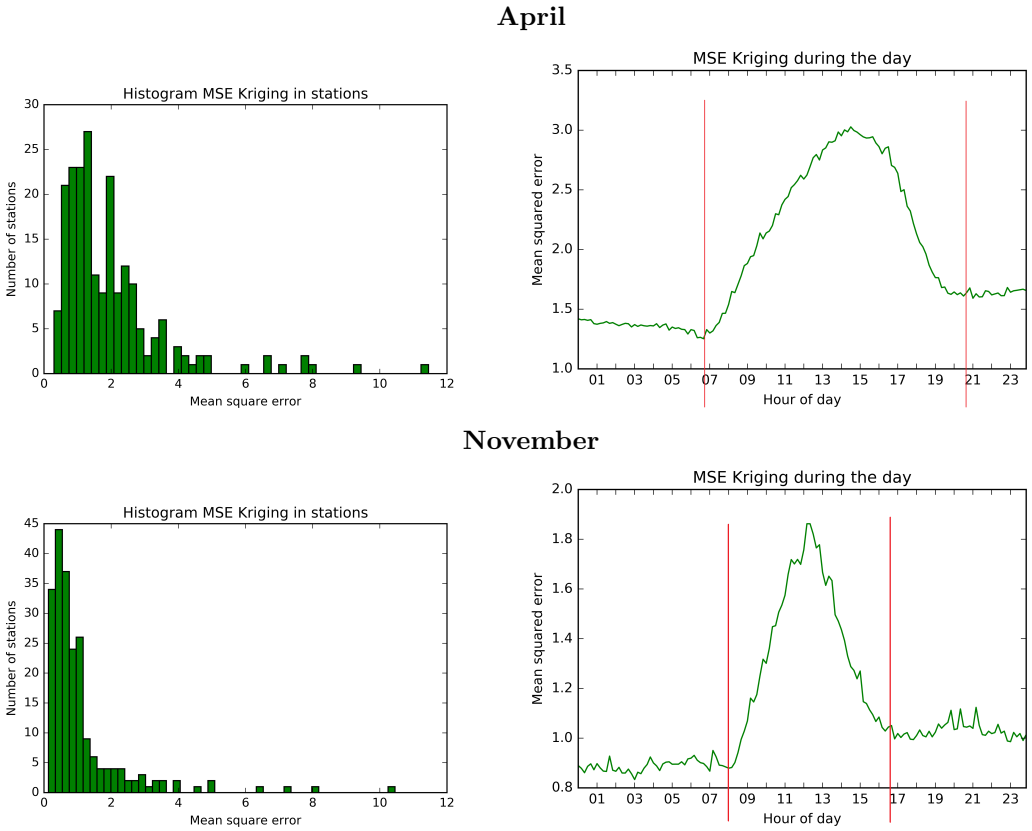


Figure 3.14: Quality evaluation scores of stations based on Kriging mean squared error, and average mean squared error during the day, for April (top) and November (bottom). The red lines depict the sunrise and sunset. On April 15, sunrise was at 06:42 and sunset at 20:40. On November 15, sunrise was at 08:02 and sunset was at 16:49.

Chapter 4

Discussion

4.1 Discussion and Conclusion

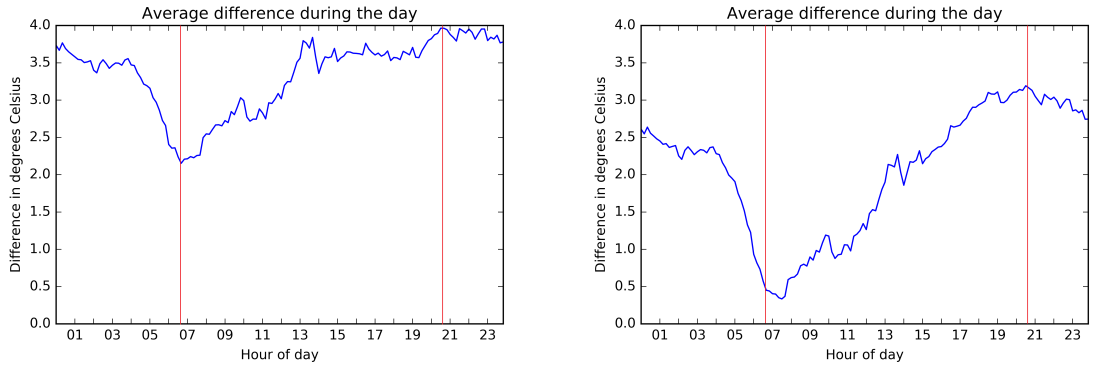
One of the biggest problems of assessing the quality of temperature measurements done by CWS, is that the true temperatures at those locations are unknown. Therefore it can not be said for certain that the processed measurements are closer to the truth. However, it can be shown that the processed measurements allow for a better illustration of the Urban Heat Island (UHI) effect than the unprocessed measurements. UHI are urban areas that experience warmer temperatures than their rural surroundings [Steeneveld et al., 2011].

The UHI effect can be characterized by the difference between the CWS measurements in the city and the KNMI measurements just outside the city. The UHI effect is most apparent in the late afternoon and at night [Steeneveld et al., 2011, Oke, 2006], so differences are expected to be largest at those times. Typically the effect tends to be strongest during daylight, and slightly smaller during the night. The annual mean air temperature in a city can be 1 to 3 degrees Celsius warmer than its surroundings. Figure 4.1 shows the average difference between CWS and KNMI measurements during the day over the months of April and November. Especially right before sunset the UHI effects are strongest compared to the measurements during the morning. After processing, it is more apparent that the difference between KNMI and CWS measurements increase as long as there is daylight, and then decrease slightly during the night, until a sharp decrease occurs in the early morning. Since UHI effect is affected by daylight hours and the sun's intensity [Oke, 1982, Imhoff et al., 2010], the effect is stronger in April than it is in November, with a maximum difference of 3°C and 1.5°C respectively.

4.1.1 Quality evaluation of Consumer Weather Stations

There is a strong diurnal pattern in the temperature bias of CWS measurements compared to KNMI measurements. More seasonal data is needed to validate this effect, but it looks like the pattern is caused by the daylight hours, and it might be concluded that the pattern is related to solar radiation errors. Its reported in [Bell, 2014] that the bias of temperature sensors decreases during the day for some (more expensive) stations, and increases for other (cheaper) stations.

To estimate the improvement of the CWS data quality, the KNMI measurements at Schiphol were used for comparison. However, it is important to note that it cannot be assumed that a higher correlation effectively means higher individual station quality. An intended purpose for the CWS data is to use it to model phenomena that currently cannot be modelled by KNMI measurements from stations near the city alone, such as the urban heat island effect. This means that it is not necessarily beneficial to disregard data that does not demonstrate the same behaviours that the KNMI data does. The average difference with KNMI measurements during the day (see appendix D, figure A.2) shows that the variation between KNMI and CWS data is high during the day and night, except for the early mornings. However, from the average Kriging MSE during the day (figure 3.14) it can be concluded that stations close together show relative high homogeneity



Average differences between CWS and KNMI measurements in April, before (left) and after (right) processing. On April 15th, sunrise was at 06:42 and sunset at 20:40.

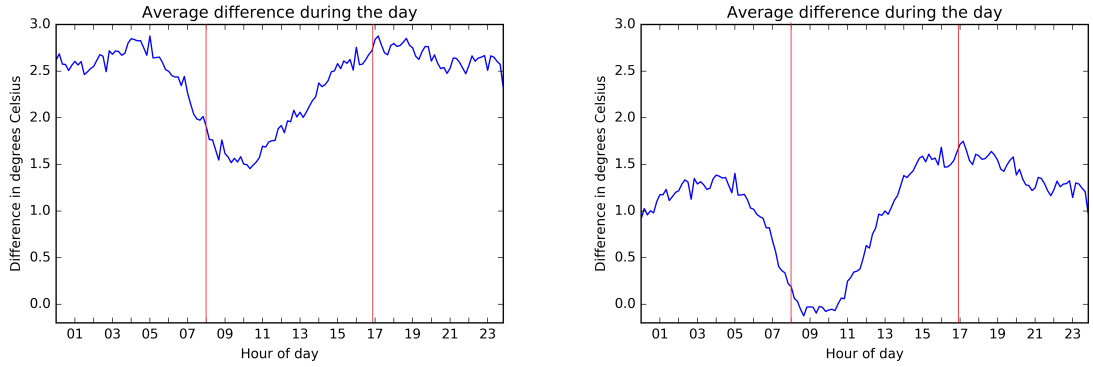


Figure 4.1: Average differences between CWS and KNMI measurements in November, before (left) and after (right) processing. On November 15th, sunrise was at 08:02 and sunset at 16:49.

during the night. So the bias with KNMI measurements at night is similar across stations. Even though the diurnal pattern of temperature bias causes a weakened correlation with the KNMI station, the weaker correlation does not invalidate this pattern.

It is also important to keep in mind that CWS are privately owned stations, set up by different people with different purposes. Considering the vast amount of stations that are apparently set up inside the house, there might be reasons for consumers to use the stations to monitor something different from the weather outside. The CWS measurements are not to be considered 'correct' and 'wrong', but rather optimal for the purpose of urban weather monitoring or suboptimal. Therefore the implied 'quality' that can be concluded from the quality evaluations associated with each method is to be interpreted as more or less optimal for weather monitoring purposes.

Depending on what the data is to be used for, some quality evaluations might be more relevant than others. Stations with a high Kalman MSE might probably be placed in the sun, so short term temporal variations in those stations may be related to clouding. If the data is to be used to model an effect exclusive to inner city climates, PCA unexplained variance might be a better measure to detemine outliers than correlation with KNMI measurements. Stations with high Kriging MSE might be placed in the sun, or might be affected by a local climate effect.

4.1.2 Observations

An interesting effect is the increase of spatial correlation at large distances. Considering the size of Amsterdam, with the distance between the northwest and southeast of the city being approximately 20 kilometers, two stations having a lag larger than $\pm$ 30 kilometer means that

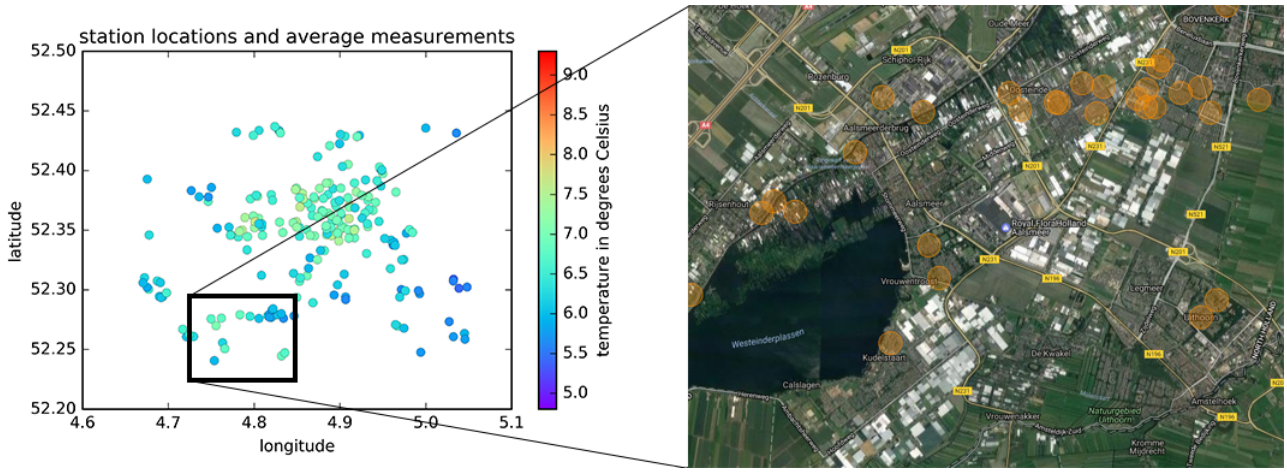


Figure 4.2: Aalsmeer area

they are both on the outskirts of the city, or at least both outside the city center (see Appendix C, figure C.7). Influenced by similar open land climates, the local climates that these stations reside in start to look more like each other again.

A notable exception, as can be seen in the results after Kriging, is the warmer area outside the city center towards the southwest of the map (see figure 4.2). This area is Aalsmeer, a small town that is considered one of the suburbs of Amsterdam. This town is known for its export of flowers, which are grown in greenhouses that cover most of the area. A possible cause for the higher temperatures in Aalsmeer might be the effect of intense agriculture on the local climate [Kalnay and Cai, 2003]. Agriculture tends to increase the minimum temperature and can be consistent with urbanization effects.

4.2 Conclusions

There is an increasing need of a higher spatial resolution in meteorological observations, especially in urban areas [Oke, 1973]. As there is a limited number of official weather stations, the use of observations obtained from CWS might be a solution. However, the quality of CWS data is often unknown. This thesis focused on assessing and improving the quality of CWS data by addressing possible issues such as radiation bias, missing data and unreliable observations. Both temporal and spatial analysis were applied to improve the usability of the data. Besides the increase in correlation with the KNMI station, this improvement now allowed for a possibility to observe the UHI effect.

4.3 Acknowledgements

Many thanks to Tom de Ruijter for daily guidance, the many feedback sessions and data collection. I am very happy I got the chance to intern at MeteoGroup. Thanks to Tom Heskes for the heated discussions about methodology, parameter settings and notations. Last but not least, thanks to the Netatmo users for uploading their data.

Bibliography

S.J. Bell. *Quantifying uncertaintry in Citizen Weather Data*. PhD thesis, Aston University, 2014. 1, 2, 8, 26

R. Benzi, R. Deidda, and M. Marrocu. Characterization of temperature and precipitation fields over sardinia with principal component analysis and singular spectrum analysis. *International journal of climatology*, 17:1231–1262, 1996. 8

L. Chapman, C.L. Muller, D.T. Young, E.L. Warren, C.S.B. Grimmond, X.M. Cai, and E.J.S. Ferranti. The birmingham urban climate laboratory an open meteorological test bed and challenges of the smart city. *American Meteorological Society*, Semptember:1545–1560, 2015.

L. de Vos, H. Leijnse, A. Overeem, and R. Uijlenhoet. The potential of urban rainfall monitoring with crowdsourced automatic weather stations in amsterdam. *Hydrology and Earth System Sciences*, page Manuscript under review, 2016. 2, 4

K.H. Eom, S.L. Lee, Y.S. Kyung, C.W. Lee, M.C. Kim, and K.K. Jung. Improved kalman filter method for measurement noise reduction in multi sensor rfid systems. *International journal of climatology*, 17:1231–1262, 1996. 8

G. Evensen. The ensemble kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53:343–367, 2003.

R. Hernndez, M. Maruri, K Otxoa de Alda, J. Egaa, and S. Gaztelumendi. Quality control procedures at euskalmet data center. *Advances in Science and Research*, 8:129–134, 2004.

M.L. Imhoff, P. Zhang, R.E. Wolfe, and L. Bounoua. Remote sensing of the urban heat island effect across biomes in the continental usa. *Remote Sensing of Environment*, 114(3):504–513, 2010. NASA. 26

E. Kalnay and M. Cai. Impact of urbanization and land-use change on climate. *Nature*, 423: 528–531, 2003. 28

F. Meier, D. Fenner, T. Grassman, B. Jnicke, M. Otto, and D. Scherer. Challenges and benefits from crowdsourced atmospheric data for urban climate research using berlin, germany, as testbed. In *ICUC9 - 9th International Conference on Urban Climate jointly with 12th Symposium on the Urban Environment*. Department of Ecology, Technische Universitt Berlin, Germany,, 2015. 1, 2

R Nakamura and L. Mahrt. Air temperature measurement errors in naturally ventilated radiation shields. *Journal of atmospheric and oceanic technology*, 22(6):1046–1058, 2005. 9

NWS01. *Netatmo User Manual*. Netatmo, version 1 edition, May 2012. http://my.netatmo.com. 1, 3

T.R. Oke. City size and the urban heat island. *Atmospheric Environment*, 7(8):769–779, 1973. 28

T.R. Oke. The energetic basis of the urban heat island. *Quarterly journal of the Royal Meteorological Society*, 108(455):1–24, 1982. 26

T.R. Oke. Initial guidance to obtain representative meteorological observations at urban sites. *World meteorological organization*, Report No. 81:1250, 2006. Retrieved from weather.gladstonefamily.net/UrbanMetOps.pdf. 26

M.A Oliver and R. Webster. A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *CATENA*, 133:56–69, 2014. 10

C.E. Rasmussen and C.K.I. Williams. *Gaussian Processes for Machine Learning*. MIT press, 2016. Retrieved from www.gaussianprocess.org/gpml/chapters/RW.pdf. 10

G.J. Steeneveld, S. Koopmans, B.G. Heusinkveld, L.W.A van Hove, and A.A.M. Holtslag. Quantifying urban heat island effects and human comfort for cities of variable size and urban morphology in the netherlands. *Journal of Geophysical Research*, 116(D20129): doi:10.1029/2011JD015988, 2011. 1, 26

P. Tan, M. Steinbach, and V. Kumar. *Introduction to Data Mining*. Pearson Education Limited, 2013. 9

G Welch and G. Bishop. *An introduction to the Kalman Filter*. Department of Computer Science, University of North Carolina at Chapel Hill, 2006. 8

S. You, K.G. Hubbard, and S. Goddard. Comparison of methods for spatially estimation station temperatures in a quality control system. *International journal of Climatology*, 28:777–787, 2008.

I. Zahumensky. Guidelines on quality control procedures for data from automatic weather stations. *World Meteorological Organization, Commision for instrument and methods of observation, Expert team on surface technology and measurement techniques*, 6:1, 2004.

Appendix A

Kalman Filter parameters

In general, the Kalman filter estimates the state x of a discrete-time controlled processes that is governed by a *linear* difference equation. Since temperature measurements are not considered to be linear, technical term for the application on this dataset is the *extended Kalman filter*. In the extented Kalman Filter, the estimates of noise use the derivative of the process. In principle, the underlying process cannot be observed, but here the assumption is made that the KNMI measurements from the nearby station at Schiphol are representative for the process. Therefore, parameters for expected measurement error and process variation are calculated from KNMI measurement characteristics. For the derivative of the KNMI measurements, see figure A.1.

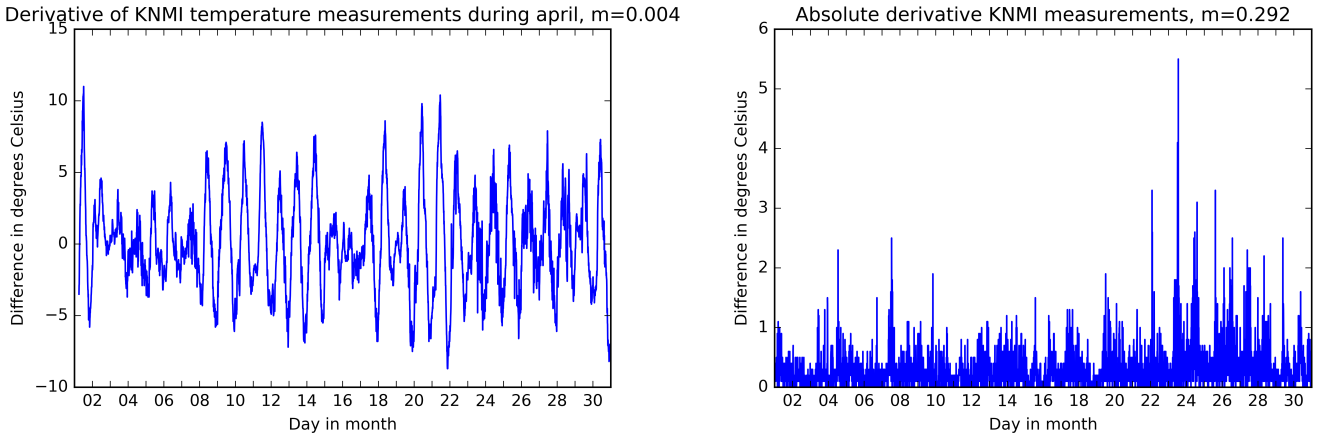


Figure A.1: Derivative of KNMI measurements during April (per hour) and absolute derivative of KNMI measurements (per 10 minutes)

To determine the process noise parameter Q, the absolute derivative of KNMI measurements is considered, see plot A.1. Here the variance from timestep to timestep can be seen. The average m represents the average difference between to concurrent timesteps. This can be seen as the allowed process noise, since some variation between timesteps is expected. In this application, the parameter Q was assumed to be constant and set to the mean m.

The next parameter to be estimated is the magnitude of the measurement error variance and bias, specified by R. Usually R depicts measurement noise variance in the Kalman filter and is a constant and normally small value. In this application it was assumed that the CWS measurements obtain some degree of bias from solar radiation in their temperature sensors. To account for both measurement noise and sensor bias, R was set to vary during the day, and based on the squared difference between KNMI and CWS measurements. Seeing the relationship between difference in measurements and time of day, the expected measurement error can probably be modelled as a

direct function of time. However, due to both the variance of quality of CWS stations (should not 'punish' the well performing station measurements for the behaviour of the badly performing stations) and need for simplicity, the expected measurement error R is simply the difference between the measurement at the CWS station and the measurement at KNMI at each timestep. This reflects both differences in expected measurement error during the day, and difference in quality between CWs stations. However, this R could lead to over fitting of the filter to behaviour of the KNMI station, so results should be interpreted with care.

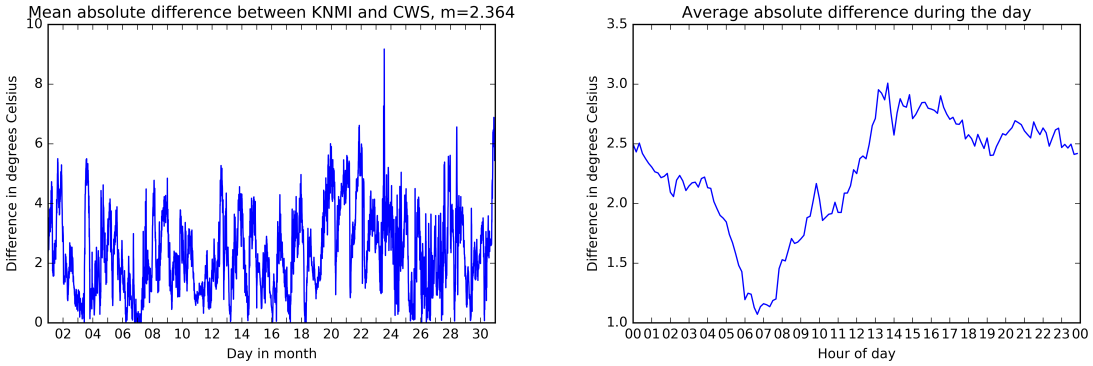


Figure A.2: Average difference between KNMI and CWS measurements at each timestep, and average difference during the day

Appendix B

Variogram models

Common functions for variogram modelling are the spherical, exponential and Gaussian models. In all functions, h is lag in distance, c is the correlated and c_0 is the uncorrelated component of the variance.

The spherical model is defined as

$$\gamma(h) = c_0 + c\left(\frac{3h}{2r} - \frac{1}{2}\left(\frac{h}{r}\right)^3\right)$$

r is the range (maximum distance between stations) of the function. The quantity $c + c_0$ is known as the 'sill'.

The spherical model fitted to the average variogram can be seen in figure B.1. The c_0 is estimated to be low, compared to the other models. This does not seem to reflect the variance at the zero lag. The mean squared error compared to the sample semivariance was 7.030.

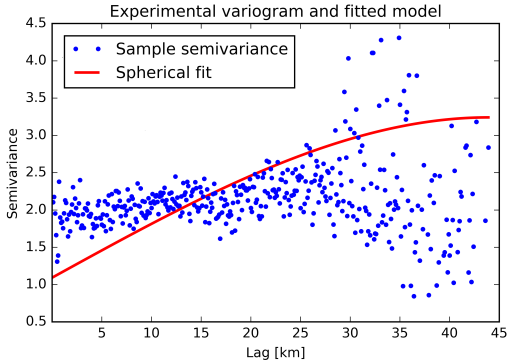


Figure B.1: Average experimental variogram with spherical model fitted.

The exponential model is the function

$$\gamma(h) = c_0 + c\left(1 - exp\left(-\frac{h}{r}\right)\right)$$

The exponential model fitted on the average variogram can be seen in figure B.2. The mean squared error compared to the sample semivariance was 6.229. The increase in variance over the first few kilometers of lag seems to be overestimated.

Since the Gaussian model had the best fit with a mean squared error of 6.075, this model was chosen as variogram model for Kriging (see Methods chapter, figure 2.5).

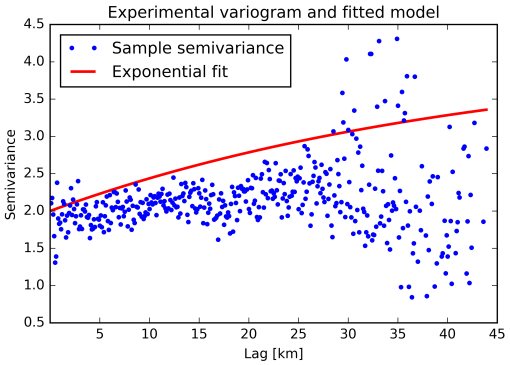


Figure B.2: Average experimental variogram with exponential model fitted.

Appendix C

Validity of assumption of spatial variance correlation

In order to perform Kriging, there has to be an assumption that the data is spatially correlated: correlation between stations is high when stations are close, and low when stations are far apart. In order to determine if this is a valid assumption, the correlogram (correlation between stations as a function of lag) calculated from the samples was considered. The sample correlogram was computed for the raw dataset, the pre-processed dataset and the temporally processed dataset, and visually inspected.

C.1 Spatial correlation in raw data

Figure C.1 shows the average measurements of stations and their locations. Some stations averages are extremely high compared to others. This can be seen in the corresponding correlogram in figure C.2, where correlation is lowest at the zero lag.

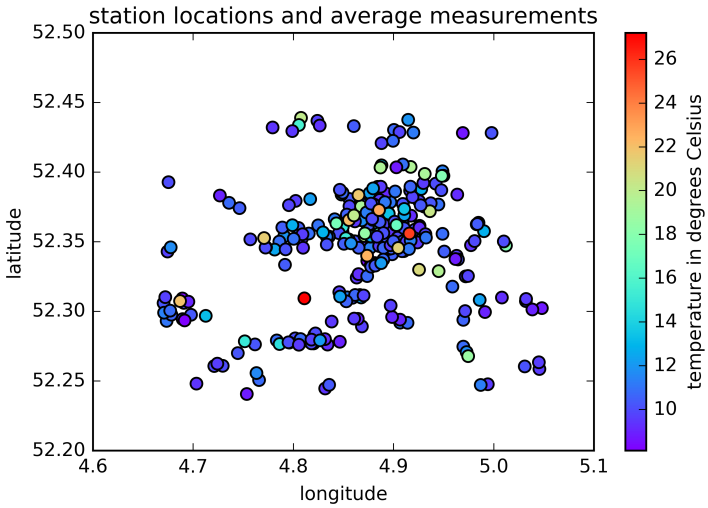


Figure C.1: Location and average measurements of unprocessed stations.

In the correlogram the correlation increases with distance, meaning that the further away a pair of stations is, the more similar they are. This can perhaps be explained by the observation that the stations that report the highest average temperatures seem to be located towards the

center, and thus being close to the rest of the stations. In any case, the assumption of spatial correlation between distance and variance does not hold.

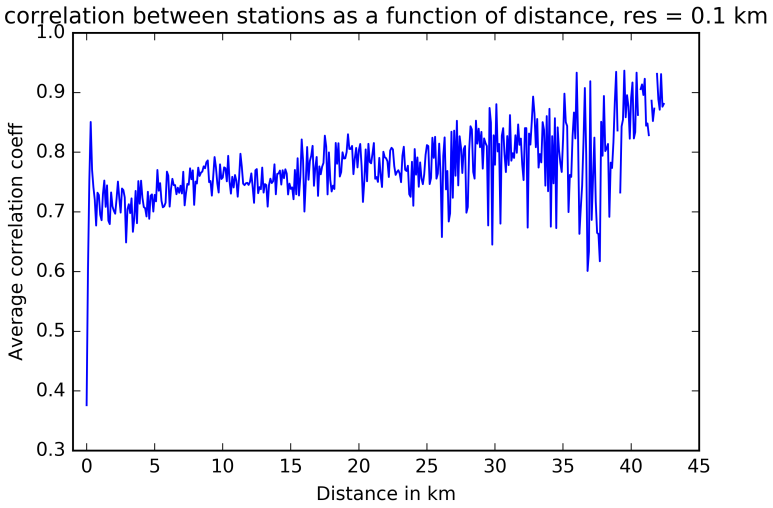


Figure C.2: Correlogram of unprocessed data.

C.2 Spatial correlation in preprocessed data

Figure C.3 shows the average measurements of stations after preprocessing. Besides the one red dot, most stations have similar average measurements. This is reflected in the correlogram in figure C.4, with an increasing correlation at the zero lag.

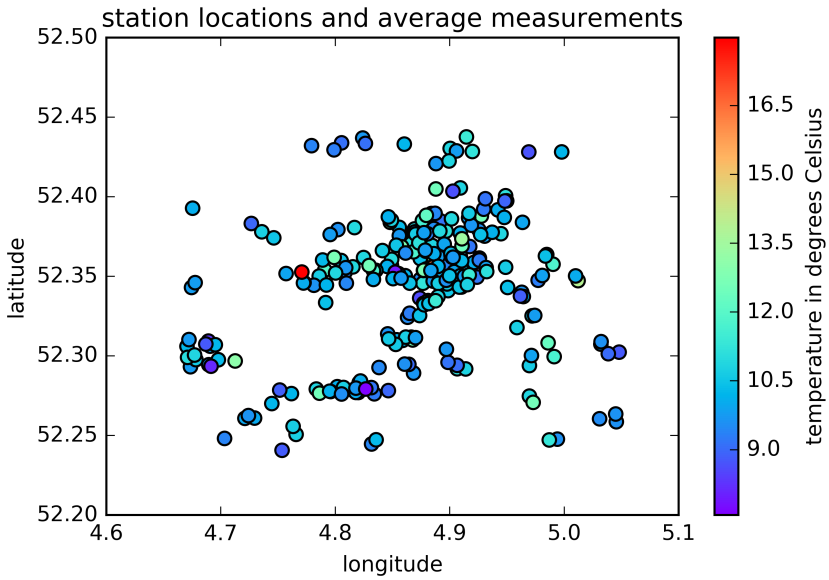


Figure C.3: Location and average measurements of preprocessed stations.

However, there is still no correlation between distance and variance. Only the uncertainty increases with distance, as fewer stations become available at high lags.

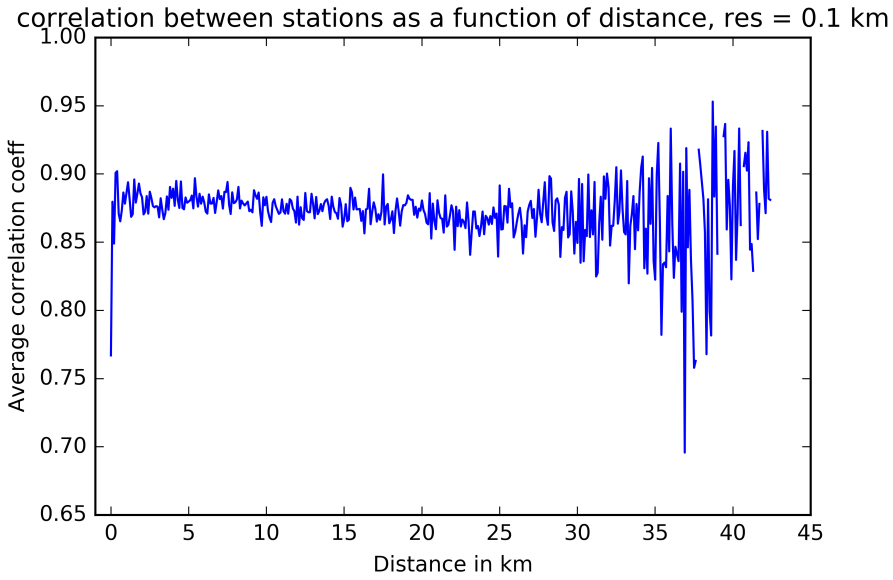


Figure C.4: Correlogram of preprocessed data.

C.3 Temporally processed data

After data has been processed by the Kalman filter and PCA, the resulting averages are as shown in figure C.5. Most stations have very similar averages, wich is reflected by the correlogram with a high correlation at the zero lag.

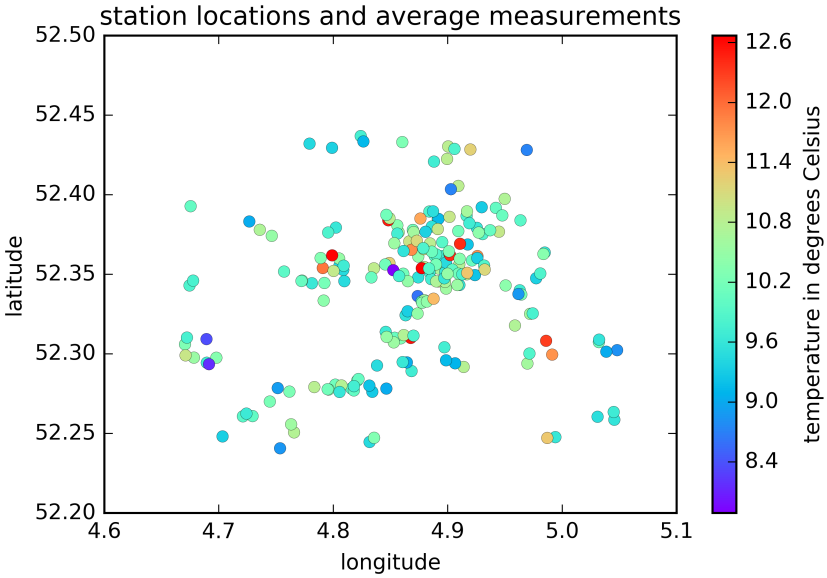


Figure C.5: Location and average measurements of processed stations.

The correlation between distance and variance is positive, with decreasing correlation over distance. There are indications that correlation increases again after 30 kilometer. This is likely due to the increased amount of microscale variations within the city, and the fact that all stations

pairs that are > 30 kilometer apart are in open land, see figure C.7.

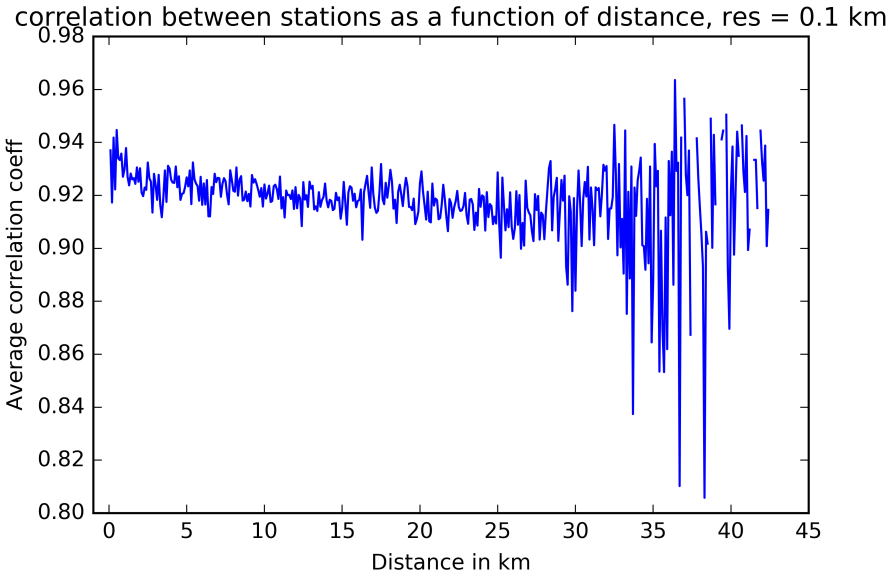


Figure C.6: Correlogram of processed data.

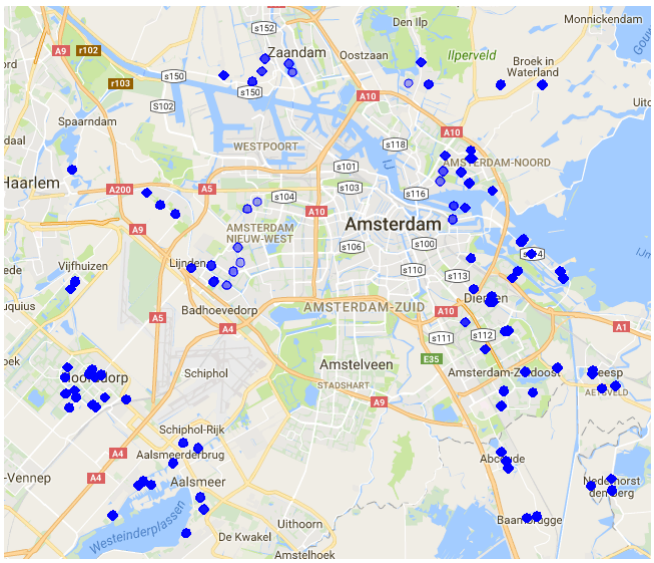


Figure C.7: Stations that are > 30 kilometer removed from another station