

UTRECHT UNIVERSITY

DEPARTMENT OF MATHEMATICS

RESEARCH PROJECT: MASTER'S THESIS (WISM106)

Quality Control and Verification of Citizen Science Wind Observations

Author:
Jieyu CHEN
student ID: 6350178
Mathematical Sciences

Daily supervisors:
dr. Kirien WHAN (*KNMI*)
dr. Kate SAUNDERS (*TUD*)
Project supervisor:
prof. dr. ir. Jason FRANK (*UU*)
Second reader:
dr. Sjoerd DIRKSEN (*UU*)



Utrecht University



Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Milieu

August 13, 2020

Abstract

Wind observations collected by citizen weather stations (CWS) are valuable for forecasting wind and issuing warnings, yet their quality is not guaranteed. Few people have worked on the quality of such wind data so far. In this thesis, we develop methods to improve the quality of wind data collected by CWS. The methods are applied to filter suspect observations and correct systematic biases, processes that are known as quality control and bias correction. We focus on the wind speed observations recorded by citizen weather stations in the province of Utrecht, the Netherlands, and our data is provided by the third-party platform, WOW-NL¹. WOW-NL allows users to upload and view their weather observations in real-time. This thesis consists of four parts: (1) pre-processing the raw data; (2) performing standard quality control that checks the internal consistency, the plausible range, and the temporal consistency of observations; (3) correcting the bias by empirical quantile mapping to reduce the errors mainly caused by low sensor heights; and (4) implementing spatial quality control that compares observations from neighboring stations, where the Earth mover's distance is introduced to select neighbors. More than one-third of low-quality stations are excluded from the study before the standard quality control, based on a lack of data completeness. About three-quarters of the remaining citizen wind speed observations pass all quality control tests. We compare the citizen science data with official data, and use statistical indicators such as Kolmogorov–Smirnov statistic and root mean square error to quantify the improvements in data quality after each step. Our results show that the bias correction substantially reduces the errors after the standard quality control, and the spatial quality control further improves the data such that it is comparable with official data. This study demonstrates that the citizen science wind data match well with official data after quality control and bias correction. This means that this data can potentially be used in many applications like analyzing localized extreme wind events that require denser observation networks than currently provided by the official stations.

¹<https://wow.knmi.nl/>

Contents

1	Introduction	3
1.1	Motivation	3
1.2	Research background	4
1.3	Structure of thesis	5
2	Data and Pre-processing	6
2.1	Data	6
2.2	Pre-processing WOW data	8
2.3	Wind climatology	9
2.4	Data completeness test	10
3	Standard Quality Control	12
3.1	Standard quality control methods	12
3.1.1	Internal check	12
3.1.2	Range check	13
3.1.3	Temporal consistency check	14
3.1.4	Summarizing standard checks	18
3.2	Statistical methods to compare WOW and KNMI	19
3.2.1	Pearson correlation	20
3.2.2	Kolmogorov–Smirnov statistic	21
3.3	Standard quality control results	22
3.4	Discussion and future works	26
3.4.1	Improving parameter selection	26
3.4.2	Limitations of current temporal check	27
4	Bias Correction	29
4.1	General explanation	29
4.2	Bias correction methods	30
4.2.1	Linear scaling	33
4.2.2	Nonlinear scaling	35
4.2.3	Distribution mapping	35
4.2.4	Empirical quantile mapping	36
4.3	Statistical indicator to evaluate bias correction	37
4.3.1	Root mean square error	37
4.4	Bias correction results	38
4.5	Discussion for future works	39

5	Spatial Quality Control	42
5.1	Selecting reference stations	42
5.1.1	Pre-selecting reference stations	43
5.1.2	Determining reference stations by Wasserstein distance	45
5.1.3	Selection procedure of reference stations	48
5.2	Spatial quality control methods	49
5.2.1	Inverse distance weighting	50
5.2.2	Spatial weighted regression test	52
5.3	Spatial quality control results	55
5.4	Discussion on other feasible directions	56
6	Conclusion and Outlook	59
6.1	The overall quality control procedures	59
6.2	Outlook for further studies	61
A	List of WOW stations	63
B	Results of each WOW station	65

Chapter 1

Introduction

1.1 Motivation

The wind is an important meteorological component for regular weather forecasts, as well as being a significant criterion to issue warnings. Wind forecasts and warnings affect human's daily life, for instance, it is hard and dangerous to cycle or sail during heavy wind, and a storm may affect air travels and transports of freight in trucks. Moreover, winds provide power sources for industries, and hence the prediction of wind matters.

To identify extreme wind events and give skillful warnings, meteorologists take special interests on localized and small-scale wind behaviors. However, the Netherlands has only 34 official weather stations that meet WMO [2014] (World Meteorological Organization) standards to record wind. The limited size of the observing network is likely due to the high costs of construction and maintenance. This means that the official wind observation network is sparsely distributed. The low observation resolution is a major source of uncertainty in forecasting, as it provides insufficient initial conditions that could cause large fluctuations in computing.

The increased proliferation of personal recording devices such as citizen weather stations (CWS) provides wind observations at many more locations, which could potentially increase the spatial density of the network. CWS are the automatic weather stations placed by citizen scientists around their own houses, like a roof or garden, and share the observation data online. However, the quality of such devices has to be taken into account because they are manufactured at a lower standard compared to official ones, and they are not placed according to WMO standards [Droste et al., 2019]. Therefore, the quality of observations from CWS could be a critical issue, and it is unknown whether these CWS can be used to complement the official wind observation network of high-quality data.

This research aims to assess the quality of CWS observation networks. We review existing approaches from the literature on determining suspicious observations in temperature and precipitation. Then we discuss whether these approaches are appropriate for wind observations and use statistical techniques to make improvements, calibrations, and evaluations.

1.2 Research background

The wind is a directional weather variable, and mathematically a two-dimensional vector field, as it blows from one place to another, unlike temperature and precipitation. Thus it is not sufficient to take only one quantity for indicating wind. Generally, wind observations consist of three components, (i) wind speed, (ii) wind direction, and (iii) wind gust. (i) Wind speed indicates the average velocity of wind during a fixed time. (ii) Wind direction shows the dominant direction from where the wind blows during a fixed time. (iii) Wind gust presents the maximum instantaneous wind speed within a certain period, which gives a view of how heavy the wind can achieve.

The term ‘quality control’ in data science represents a procedure intended to ensure that the concerned data set meets the requirements of some specific standards. The purpose of quality control in weather data is to automatically identify suspect or inaccurate observations, preferably in real-time. In our research, we mainly focus on wind speed, and our mission is to construct quality control procedures for citizen science wind speed observations to detect suspect records from raw data and to correct systematic biases.

A lot of researchers like DeGaetano [1997], Zahumenskỳ [2004], Jiménez et al. [2010], Estévez et al. [2011] have made contributions in quality controls of wind observations. In an early example of quality control for wind data, DeGaetano [1997] generated an automatic quality control routine based on hourly wind speed and direction data, which focused on the excessive variability and inordinately constant observations for both wind speeds and wind directions. After that, Jiménez et al. [2010] proposed a comprehensive quality assurance for surface wind speed and direction observations; they mainly considered three parts, manipulation errors, limits consistency, and temporal consistency. Zahumenskỳ [2004] and Estévez et al. [2011] discussed automatic quality assurance for multiple weather observations, including temperature, precipitation, and wind. Zahumenskỳ [2004] checked the plausible range, time consistency, and internal consistency for both wind speeds and wind directions, while Estévez et al. [2011] only applied these tests on wind speeds. Most of the quality control steps they have discussed are based on similar ideas, which are also summarized in the guide by WMO [2014] (World Meteorological Organization). We call such steps the standard quality control methods in our study.

These previous studies did not discuss the quality control in a spatial aspect that considers neighboring observations. Fiebrich et al. [2010] summarized the quality assurance tests used by the Oklahoma Mesonet, where the spatial checks for wind observations are included. Nevertheless, the spatial quality control specifically for wind observations has rarely been discussed. Researchers like Estévez et al. [2011] and Hubbard et al. [2005] generated spatial quality control tests, but only for temperature data. Therefore, we need to discuss the ideas for temperature spatial checks and adapt them to wind speeds.

We study the crowdsourced CWS wind observations provided from Weather Observations Website (Netherlands, WOW-NL [2020]), a third-party platform on which the CWS owners can share, view, compare current and past weather observations. The term ‘crowdsourcing’ introduced by Howe [2006] is a model to collect data from a distributed network of individuals, based on the internet. In meteorological studies, it is usually regarded as accessing data of numerous non-official public devices through online platforms [Muller et al., 2015].

The use of crowdsourced data in meteorology has become more prominent in recent years.

Bell et al. [2013] discussed the crowdsourced weather observations in humidity, temperature, and pressure. Bell et al. [2015] investigated the quality of crowdsourced data for temperature, rainfall, humidity, and dew point observations. They developed field studies and concluded that the CWS data could contain significant biases; therefore, any application of such crowdsourced data should require a quality control system that can both remove errors and correct biases.

The growing number of CWS motivates some researchers to study the quality control for crowdsourced automatic weather stations. Some researchers discussed the quality controls for CWS observations, and they have shown that there is much potential for crowdsourced data in temperature observations [Meier et al., 2017] [Napoly et al., 2018], precipitation observations [De Vos et al., 2017], and surface wind observations in an urban setting [Droste et al., 2019]. They followed some of the traditional quality control approaches, and it was found that those methods are also well-suited for high-resolution weather observations.

1.3 Structure of thesis

This research focuses on the quality control procedures of crowdsourced wind speed observations from third-party platform WOW-NL. We divide the whole process into three parts. After introducing the raw crowdsourced wind data with some necessary pre-processing steps in Chapter 2, we outline and discuss the standard quality control methods in Chapter 3, which are traditionally applied for single automatic weather stations. Next, in Chapter 4, we provide bias correction approaches to calibrate the systematic instrument errors for individual stations. After correcting the bias, the wind observations from each station could be compared spatially in the network. Therefore in Chapter 5, we generate the advanced spatial quality control methods that compare neighboring observations. Finally, in Chapter 6, we conclude our results and discuss the outlook for future studies.

Chapter 2

Data and Pre-processing

In this chapter, we first introduce the crowdsourced data set and then carry out essential pre-processing steps. Subsequently, we perform a preliminary analysis of the data to give a general view from the climatological perspective. At last, we select CWS that have enough valid wind observations for the quality control study.

2.1 Data

The crowdsourced data set used in this study is provided by Weather Observations Website (Netherlands, WOW-NL [2020]), a third-party platform on which the personal weather station owners can share, view, compare current and past weather observations. The data set contains weather records by 93 CWS (citizen weather stations) located in the Utrecht province of the Netherlands. We refer to these CWS as WOW stations throughout the thesis. All observations of WOW stations are reported in Greenwich Mean Time (GMT). The time range of the data is the three years from ‘2016-01-01 00:00:00 GMT’ to ‘2018-12-31 23:59:59 GMT’.

The WOW data set includes both metadata information and observations of multiple weather parameters. The metadata information mainly provides the site ID, latitude, longitude, and height (above sea level) of each station. There are 12 weather parameters in the data set, which are listed below, with corresponding units. We only consider the first three ones related to the wind for study.

- | | |
|-----------------------------------------|------------------------------------------|
| 1. wind speed (meter per second) | 7. mean sea level pressure (hectopascal) |
| 2. wind gust (meter per second) | 8. relative humidity (percentage) |
| 3. wind direction (degree) | 9. rainfall amount (millimetre) |
| 4. wind speed (knot) | 10. rainfall rate (millimetre per hour) |
| 5. wind gust (knot) | 11. dew point temperature (Celsius) |
| 6. air pressure (hectopascal) | 12. dry bulb temperature (Celsius) |

To construct the quality control, we need accurate official data for a reference. Our reference data are provided by the Royal Netherlands Meteorological Institute (Dutch: Koninklijk Nederlands Meteorologisch Instituut, KNMI [2020]), which records official observations, provides forecasts and warnings. KNMI has an observation network of 48 official weather measuring stations spread over the Netherlands and the North Sea. We select 12 inland official sites located around the province of Utrecht and use their wind observations as reference data, where the coastal stations are excluded because they behave differently compared with inland sites [Gatey and Miller, 2007]. We refer to these 12 official stations as KNMI stations throughout the thesis. The locations of KNMI and

WOW stations are drawn on the map in Figure 2.1. Corresponding to WOW data, we choose KNMI surface wind observations from ‘2016-01-01 00:00:00 GMT’ to ‘2019-01-01 00:00:00 GMT’ as reference KNMI data. We choose ‘2019-01-01 00:00:00 GMT’ rather than ‘2018-12-31 23:59:59 GMT’ because KNMI stations do not have records at the latter one.

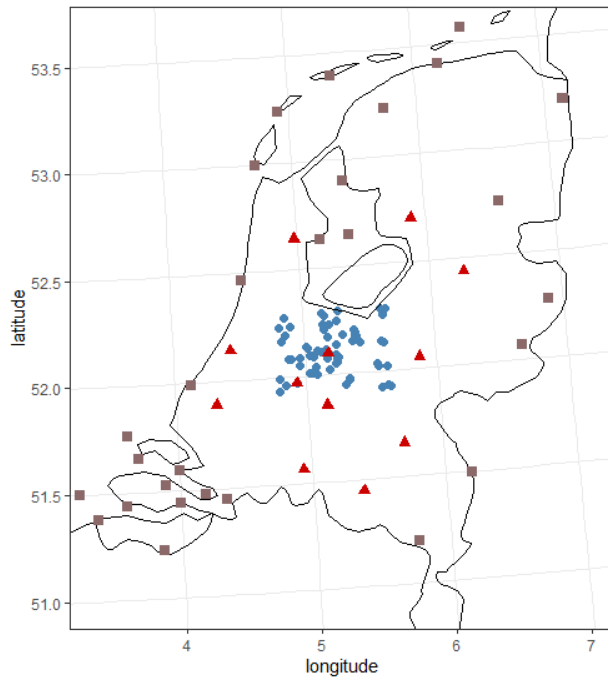


Figure 2.1: A map showing the locations of WOW (blue points) and KNMI (red triangles) stations. The grey squares represent other KNMI stations that are either too far away for consideration or are coastal sites.

The WMO sets the standards for all meteorological observations. They recommend that the wind sensor should be placed at 10 meters height, the measurement site is flat and unobstructed by trees or buildings. For forecasting purpose, WMO [2014] suggests using observations recorded every 10 minutes. The standard measuring method for wind speed, direction, and gust are listed below, according to WMO [2014].

1. Wind speed is computed by the average velocity during the last ten minutes;
2. Wind direction is given by the average direction over the past ten minutes;
3. Wind gust is chosen as the maximum three-second mean speed in the last ten minutes.

For wind sensors of official weather stations, they usually record instantaneous read-outs of wind with a high sampling frequency. Wind speeds and directions read the average of these instant records within the last ten minutes, while wind gusts compute means every three seconds and choose the highest one during the past ten minutes.

Most KNMI stations meet the WMO standards. All stations in our reference KNMI data, except for ‘De Bilt (260_W_a)’, are placed at the recommended 10 meters height. A correction is applied to ‘De Bilt’ to adjust for obstructions in the surrounding location. We make a necessary assumption that the KNMI data are correct and measuring the real wind at 10 meters height.

The WOW stations do not follow the WMO standards, and they have various reporting date and time sequences due to different manufacturing standards. For instance, some stations record

observations every 5 minutes, while some other record observations every 15 minutes. We checked the CWS of the brand Netatmo, and find that their wind observations are averaged in the last 5 minutes [Netatmo Helpcenter, 2020]. We also investigated the CWS of the brand Davis. Their anemometers measure the average wind over a 2.25-second interval, which is a much lower sampling frequency than official ones, and they provide the average and high wind for two-minute and ten-minute intervals [Dann and Prodata Weather Systems, 2020]. In our data set, most WOW stations do not follow the standard minutes-seconds time sequence that is used by KNMI data, for example, (MM:SS) ‘00:00, 10:00, 20:00, 30:00, 40:00, 50:00’, which leads to a problem on matching the WOW observations directly with the KNMI observations. We deal with this issue in the following section.

2.2 Pre-processing WOW data

Our approach for matching WOW data to KNMI reporting times is to look at the recording times of a WOW station and select the time closest to the standard KNMI time sequence. We assume that WOW observations at the chosen times indicate wind at the corresponding KNMI time points. The following is a simple example. Suppose a WOW station records wind at time points (HH:MM:SS)

$$\begin{array}{ccccccc} 00:01:00, & 00:06:00, & 00:11:00, & 00:16:00, & 00:21:00, & 00:26:00, & 00:31:00, \\ \hline \text{value 1,} & \text{value 2,} & \text{value 3,} & \text{value 4,} & \text{value 5,} & \text{value 6,} & \text{value 7,} \end{array}$$

while the standard time sequence from KNMI stations is

$$00:00:00, 00:10:00, 00:20:00, 00:30:00.$$

Therefore, we choose the WOW time points that are closest to these standard time points and generate a new time sequence

$$\begin{array}{cccc} 00:01:00, & 00:11:00, & 00:21:00, & 00:31:00, \\ \hline 00:00:00, & 00:10:00, & 00:20:00, & 00:30:00, \\ \hline \text{value 1,} & \text{value 3,} & \text{value 5,} & \text{value 7,} \end{array}$$

where we take WOW observations at the new time points as the observations at the standard time sequence. The assumption is made to ignore the actual differences of observations between WOW and standard time points. In this way, we get a new WOW data set with the identical recording time sequence as KNMI stations. However, for wind gusts, this approach might throw away larger wind gust records during the ten-minute interval and leads to an underestimate of actual gusts. We further discuss the limitations of this pre-processing step in the following.

Limitations of the matching method

This approach to match WOW data is not perfect as it omits a lot of data which might be useful. We use a simple example to show the possible consequences by the deficiencies of the method. Recall the previous example, ‘value 2’, ‘value 4’, and ‘value 6’ are excluded from the data set. Suppose these are the wind gust observations, ‘value 2’ is larger than ‘value 1’. Then the maximum gust of wind during the ten-minute should be ‘value 2’, which is removed by pre-processing. Thus the pre-processed wind gusts would be lower than actual ones. We also consider another case, where ‘value 1’ is implausibly large due to the sensor error and it fails quality control tests, while ‘value 2’ and ‘value 4’ are normal records. Then the wind gust observation at time (HH:MM:SS) ‘00:10:00’ would be empty because ‘value 3’ fails the quality control and is removed. Therefore, we have no idea of

what wind is like at that time if ‘value 2’ and ‘value 4’ are not considered, which is a loss in the data.

This leads to a trade-off between considering all WOW observations with more complex running programs, and removing unnecessary observations for quicker implementations. In our research that focuses on the quality control methods of wind speed observations by CWS, we think it is not worth considering all WOW data if it takes much longer processing time and more complicated computations for implementation. Therefore, we assume that the chosen WOW observations well represent wind at relating KNMI time points in this study. However, more sufficient wind observations should be considered in future works.

2.3 Wind climatology

We investigate the reference KNMI data to find seasonal differences for wind in the Netherlands. We split KNMI wind speed observations first by seasons and then by years, and compare their statistical distributions.

As shown in Figure 2.2, there are seasonal variations in wind speed, with the extended winter season (Nov-Apr) experiencing higher average wind speeds, compared to the extended summer season (May-Oct). Interannual variations are also evident, but they are smaller than seasonal changes in this 3-year data set.

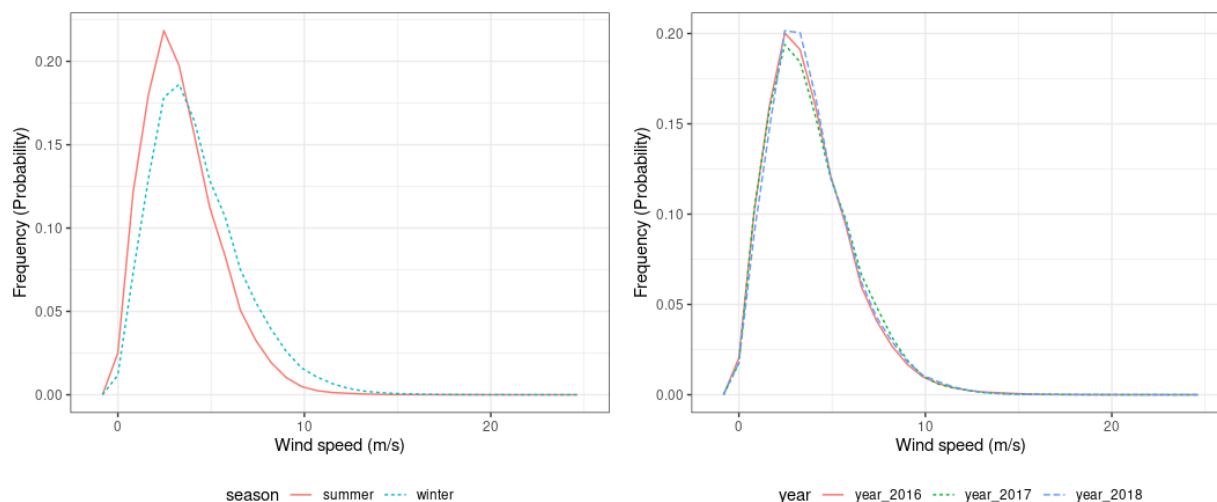


Figure 2.2: Frequency plots of KNMI wind speed in different seasons (left) and in each year (right).

We only consider stations with at least one year of data available, as we want to compare the statistical distributions of all wind speed observations from a station. If a station has observations only in winter while the other station has observations only in summer, it is hard to tell whether the differences between their distributions of wind speed result from seasonal variations or different locations. Therefore, a weather station should record wind speed in at least one year to be analyzed. We remove WOW stations that do not have enough observations, which is the completeness test for the WOW data.

2.4 Data completeness test

The completeness test checks whether a station in the WOW data set has enough available observations. This is a necessary step before quality control since a station with few records or many repeated records is meaningless for further analysis. This data completeness check consists of two parts, (a) *the ‘null’ test* and (b) *the ‘duplicate’ test*. If a station cannot pass either test, it will be removed from the original data set and not considered in future studies.

(a) *The ‘null’ test* examines what percentage of a station’s wind speed/direction/gust observations does not have any valid values. A station fails the ‘null’ test if more than two-third of the observations are missing, and we label the station as ‘incomplete’. Here we use two-third because our wind observations are in the three years 2016-2018, and we only consider stations with more than one-year data.

(b) *The ‘duplicate’ test* considers how many repeated observations in a station. This test is necessary because, in some cases, a station may have enough available sensor records, but most of them hold the same value and keep unchanged. A station fails the ‘duplicate’ test if more than 95% of the records hold a constant value, and we label the station as ‘broken’, speculating that the station gets stuck in some technical problem.

Figure 2.3 provides an example station that passes the ‘null’ test but fails the ‘duplicate’ test. The station has more than one-third available wind speed records in the three-year period of our raw data set. However, most of the records are zero, and only a few of them have positive values. This type of station is unrealistic compared to official stations, and so they are removed from further analysis.

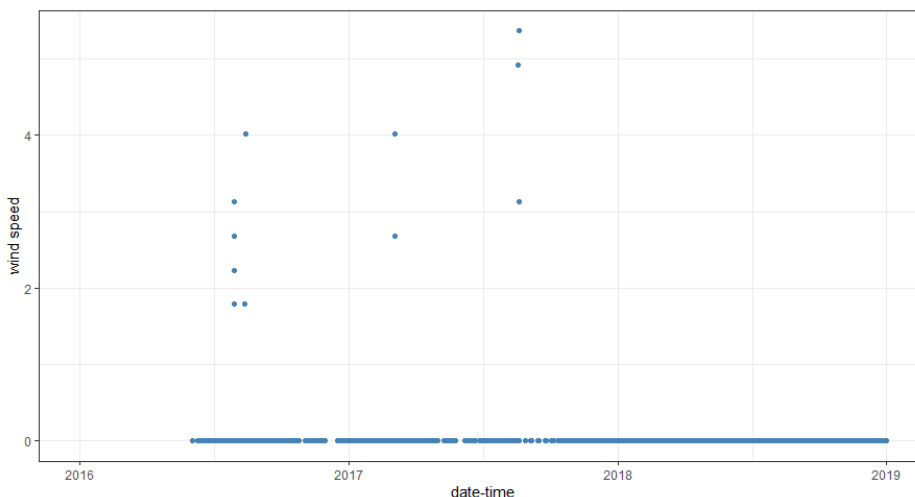


Figure 2.3: An example station (WOW #45) that fails the ‘duplicate’ test.

There are 65 WOW stations in the raw WOW data set that record wind speeds, after implementing the null and duplicate tests, only 39 WOW stations are left as shown in Figure 2.4. The result indicates that more than one-third of CWS located in the province of Utrecht do not have enough observations for quality control study. However, we can see from the upper part of Figure

2.4 that there are multiple newly activated CWS. This means we can expect a more densely distributed WOW observation network in the near future. Nevertheless, these 39 stations are situated with much more spatial density than official sites, as there is only one KNMI station ‘De Bilt’ (260-W_a) in the region. The wind observations of the 39 WOW stations generate a new WOW data set used for the standard quality control checks in the next chapter.

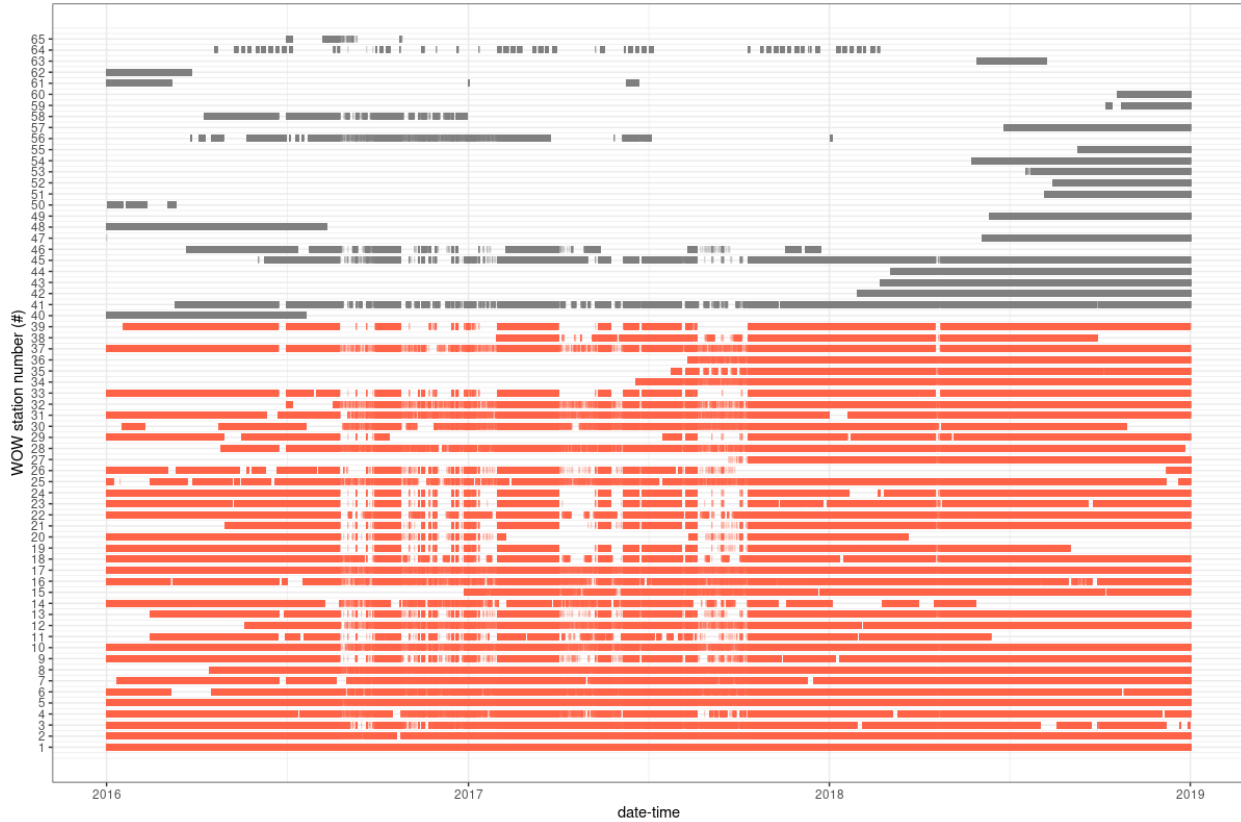


Figure 2.4: The active period of each WOW station. The 39 stations with red blocks pass the data completeness test.

Chapter 3

Standard Quality Control

In the first part of this chapter, we introduce the standard quality control methods and make adjustments so that we can apply these methods to the WOW data set. Then in the next section, we discuss several statistic indicators to evaluate the performances of quality control. Finally, we show and discuss the results of standard quality control on the WOW data set.

3.1 Standard quality control methods

Wind speed and direction are quantities that change rapidly, sometimes with diurnal dependence [Taylor and Loescher, 2013]. Moreover, Droste et al. [2019] noted that urban wind speed and direction are especially hard to quantify due to the strong turbulent nature of wind. Therefore, we must be much more careful when dealing with wind quality control for crowdsourced CWS observations than other weather components. We summarize and adapt the standard quality control methods into three steps; (1) the internal check, (2) the range check, and (3) the temporal check. These steps can be applied for real-time wind observations by single CWS, and we explain each of them in the following sections.

To introduce quality control checks clearly, we need to make use of math notations. Wind observations consist of three components, wind speed, wind direction, and wind gust, and we let s, d, g to represent each of them respectively. We use \mathcal{T} to denote the standard recording time sequence with ten-minute intervals. For any time point $t \in \mathcal{T}$, s_t, d_t, g_t represent the observations at time t . For a certain station, all its wind speed observations over time can be treated as a time series at the time sequence \mathcal{T} , represented as $\{s_t\}_{t \in \mathcal{T}}$. If this station does not have valid wind speed records at some time points, the corresponding positions on the time series show missing values. Wind gust time series $\{g_t\}_{t \in \mathcal{T}}$ and wind direction time series $\{d_t\}_{t \in \mathcal{T}}$ are similar to the settings of $\{s_t\}_{t \in \mathcal{T}}$.

3.1.1 Internal check

The internal check verifies whether a wind observation is reasonable or consistently recorded. According to the internal consistency checks from WMO [2014], Zahumenskỳ [2004], Fiebrich et al. [2010], the following situations are abnormal in an official station that meets all standards proposed by WMO:

- (i) Wind speed is zero, while wind direction keeps changing;

(ii) Wind speed is larger than wind gust.

The first one does not apply to our data set, because the WOW stations have various measuring standards and sensor qualities, and their wind direction records have different behaviors when there is nearly no wind speed. We checked what wind direction observations look like under such circumstances. There are mainly three kinds of wind direction recordings by WOW stations when the synchronous wind speed is recorded zero. One is logging no data for the wind direction observation; another is recording a zero wind direction; the last one is getting a wind direction value identical to its previous record. We cannot find a universal standard to flag observations in such cases, in the meanwhile, we do not want to throw away wind speed records that could be trusted. Therefore, the first abnormal check is not suitable for our study, and we will skip them.

The second internal test states that wind speed should be less than or equal to wind gust for all observations; otherwise, both wind speed and gust records are suspect. However, for CWS with different manufacturers, we have no guarantee that the wind gusts are measured correctly following WMO standards. We have less trust in wind gusts than wind speeds, so we do not want to discard wind speed observations on the basis of wind gusts.

To apply the second test on the WOW data, we flag both wind speed and gust observations if they meet the condition that wind gust is lower than wind speed. Flagging an observation means that this observation from a certain station at a particular time point of the time series fails the test. For a WOW station, its wind speed s_t and wind gust g_t at time point t of the time series $\{s_t\}_{t \in \mathcal{T}}$ and $\{g_t\}_{t \in \mathcal{T}}$ pass the test if

$$s_t \leq g_t.$$

The flagged wind gust records are removed from the WOW data after this test, while the flagged wind speed records remain as we discussed earlier. These wind speed observations are not rejected and are further validated during other quality control checks whether they should be wrong records or not.

3.1.2 Range check

The range check is used for detecting implausible values, for instance, observations higher than world records, or abnormal negative records. The main purpose of this check is to filter climatologically impossible observations. Filtering means to remove observations that do not pass the test. To achieve this, we need to set up reasonable thresholds for wind speed, gust, and direction. We select the thresholds by looking at historical KNMI weather data in the Netherlands. Such bounds may vary in different months, due to seasonal variations in wind speed (see Chapter 2.3).

The wind direction must lie between 0 and 360 degrees; this is an unconditional threshold for direction angles at all times. Moreover, wind speed and wind gust must be non-negative without any doubt. Thus the remaining question is to determine the upper bound of wind speed and wind gust.

Range check approach 1:

The standard approach to determine the upper bound for wind speed and gust is to select the historical highest records as a universal threshold [Estévez et al., 2011]. We use ‘SPD_MAX’ and ‘GUST_MAX’ to stand for the highest wind speed and wind gust records, respectively. For a WOW station, the conditions that its wind observations at time t pass the range check are listed below:

- Wind direction should lie between 0° and 360° ; this means d_t passes the test if

$$d_t \in [0, 360].$$

Here we consider both 0 and 360 because of the rounding errors and different recording standards by various brands of CWS.

- Wind speed should be non-negative and smaller than SPD_MAX m/s, this means s_t passes the test if

$$s_t \in [0, \text{SPD_MAX}].$$

- Wind gust should be non-negative and smaller than GUST_MAX m/s, this means g_t passes the test if

$$g_t \in [0, \text{GUST_MAX}].$$

We look at KNMI historical wind speed and gust observations during the past 120 years in the Netherlands¹, and take the highest records 35.0 (m/s) for SPD_MAX and 64.0 (m/s) for GUST_MAX. In this way, we assume that the WOW wind speeds in the data set could never exceed the historical highest record.

Range check approach 2:

However, we find that the first approach is not perfect because it will miss flagging abnormal high wind speeds in the calm period of a year. According to Chapter 2.3, the wind has seasonal variations but much lower annual changes, thus we can use monthly thresholds to flag values that are abnormally high for the particular season. Here we make assumptions that the seasonality behaviors of KNMI and WOW wind observations are similar, and WOW wind speeds are less than the highest historical record of KNMI each month. We use ‘SPD_MAX_M’ and ‘GUST_MAX_M’ to represent the upper bounds for wind speed and wind gust in different month ‘M’. The conditions that wind observations pass the plausible range check are similar to those in Approach 1, and the only change is that wind speed and wind gust have a different threshold in each month.

We take the highest KNMI historical wind speed and wind gust in each month from the same website in Approach 1, these maximum records are listed in Table 3.1. We look at wind speeds and gusts in our reference KNMI data (in the three years 2016-2018, by the selected 12 KNMI stations) for comparison. For observations in each year, we choose monthly maximum wind speed and wind gust, then compare them with the upper bounds ‘SPD_MAX_M’ and ‘GUST_MAX_M’. We can see from Figure 3.1 that those maximum values vary in different months, and all of them are far below the upper bounds for plausible range check. We make assumptions that the WOW wind speeds are similar to KNMI wind speeds in the three years. This implies that our thresholds are suitable, even if some WOW stations record higher wind speed or gust than KNMI stations, they still pass this check unless the records are abnormally large.

3.1.3 Temporal consistency check

The temporal check considers a plausible rate of change for observations. On the one hand, the variations of wind speed/gust should not be too sharp, which means a huge jump in wind speed/gust

¹Maximum average wind speed and maximum wind gust data from KNMI: <http://projects.knmi.nl/klimatologie/daggegevens/selectie.cgi>.

Month	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
SPD_MAX_M	33.4	29.8	29.8	29.8	26.2	23.1	25.0	26.8	35.0	31.0	30.9	32.4
GUST_MAX_M	48.0	64.0	44.8	51.0	37.0	36.0	34.0	36.0	37.0	42.0	48.0	44.0

Table 3.1: The upper bounds ‘SPD_MAX_M’ (m/s) and ‘GUST_MAX_M’ (m/s) for wind speed and wind gust used in plausible range check.

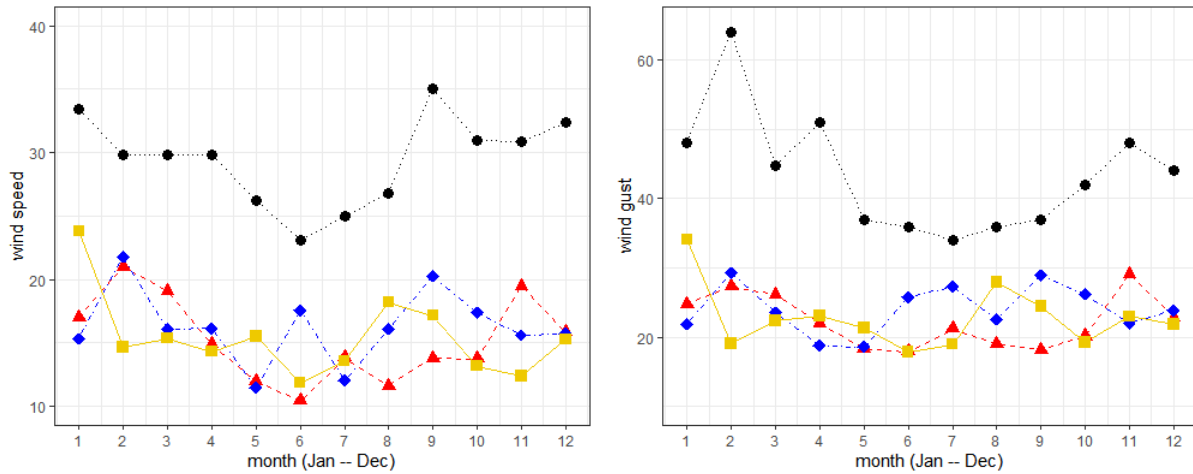


Figure 3.1: The upper bounds (black points) for wind speed (left) and wind gust (right) in different months, and the highest KNMI wind speed (left) and wind gust (right) observations for each month in the year 2016 (red triangles), 2017 (blue diamonds), and 2018 (yellow squares).

records is questionable, and this is sometimes called the step test. On the other hand, wind speed/gust should not remain the same, or with slight changes in a long period, this is sometimes called the persistence test [Fiebrich et al., 2010]. Figure 3.2 shows simple examples for both step and persistence tests.

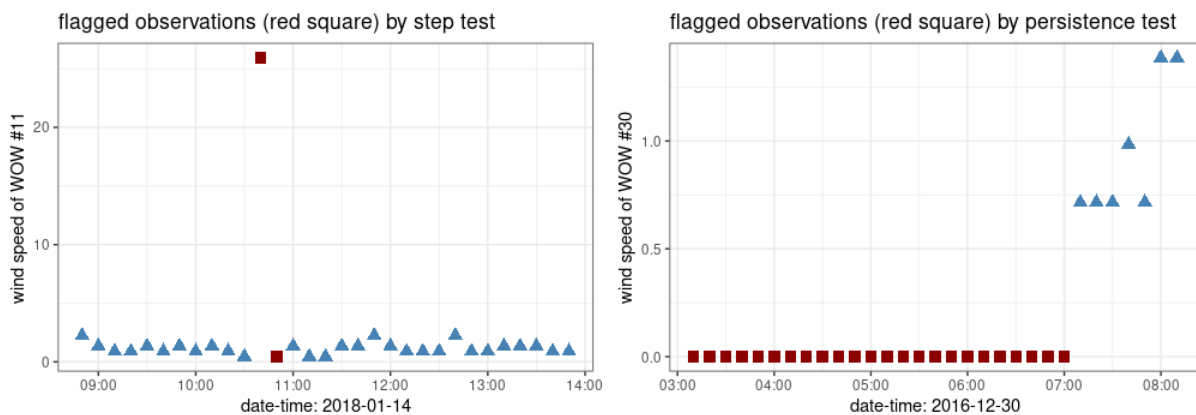


Figure 3.2: Examples of step test (left) and persistence test (right), red squares are the flagged wind speed observations.

Step test:

This step test aims to identify excessive wind speed or gust variability. We do not consider wind direction in this test since directions are more changeable than speeds, and the variation of wind direction can range from 0 to 360 within 10 minutes [Shafer et al., 2000]. The idea for this test is to check the variability of wind speed or gust with time, and decide whether the rate of change is greater than a maximum allowed value. The conditions to pass the step test for wind speed and gust records are as follows:

- Wind speed does not change more than STEP_SPD m/s in STEP_SPD_T minutes; this means s_t passes the test if

$$\max_{\tau \in [t - \text{STEP_SPD_T}, t] \cap \mathcal{T}} \{s_\tau\} - \min_{\tau \in [t - \text{STEP_SPD_T}, t] \cap \mathcal{T}} \{s_\tau\} \leq \text{STEP_SPD}.$$

- Wind gust does not change more than STEP_GUST m/s in STEP_GUST_T minutes; this means g_t passes the test if

$$\max_{\tau \in [t - \text{STEP_GUST_T}, t] \cap \mathcal{T}} \{g_\tau\} - \min_{\tau \in [t - \text{STEP_GUST_T}, t] \cap \mathcal{T}} \{g_\tau\} \leq \text{STEP_GUST}.$$

Determination of parameters

1. *STEP_SPD_T* and *STEP_GUST_T*: In this step test, we want to detect large jumps in the wind observation sequence. Therefore, the time intervals STEP_SPD_T and STEP_GUST_T should be small enough for considering sharp increases or decreases. The standard recording times in our data set have 10-minute intervals, and naturally, we set both STEP_SPD_T and STEP_GUST_T as 10 (min).
2. *STEP_SPD* and *STEP_GUST*: We use the reference KNMI data (in the three years 2016-2018) to determine the thresholds. Firstly, we calculate the difference between a wind speed/gust record and its previous wind speed/gust record for all observations in the KNMI data set. Then for each KNMI station, we find the maximum fluctuations of wind speed/gust from our calculation results, which are shown in Figure 3.3. We can see that the maximum variations in wind speeds are usually smaller than that in wind gusts. We choose the largest one among those maximum variations in wind speed and wind gust as the value of STEP_SPD and STEP_GUST , respectively. Therefore, STEP_SPD is 15.51 (m/s) and STEP_GUST is 27.41 (m/s) in our step test.

Persistence test:

The persistence test aims to detect abnormally low changes in wind observations with time, that may be caused by a stuck wind sensor. The idea of this test is to check whether a CWS records nearly repeated values over a long period of time. We set up the minimum required variabilities for wind speed, gust, and direction. The conditions for passing the persistence test are listed below:

- Wind speed should change more than PERSIST_SPD m/s in PERSIST_SPD_T minutes; this means s_t passes the test if

$$\max_{\tau \in [t - \text{PERSIST_SPD_T}, t] \cap \mathcal{T}} \{s_\tau\} - \min_{\tau \in [t - \text{PERSIST_SPD_T}, t] \cap \mathcal{T}} \{s_\tau\} > \text{PERSIST_SPD}.$$

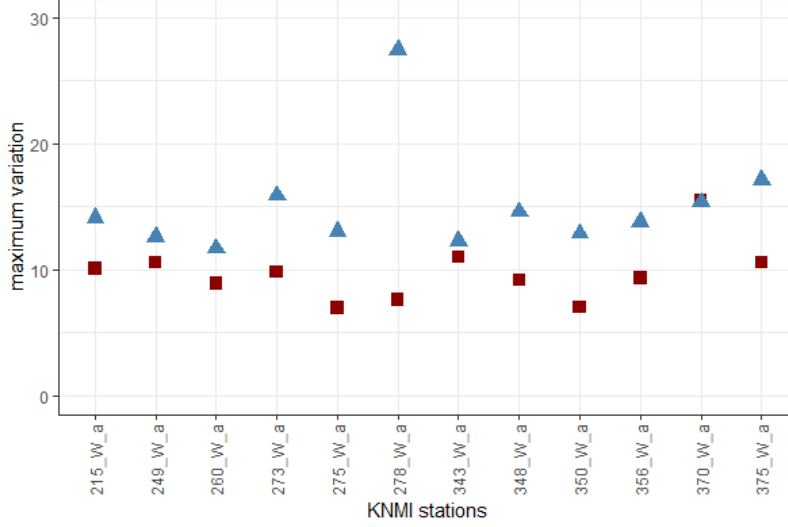


Figure 3.3: Maximum variations of wind speed (red square points) and wind gust (blue triangle points) from each station in the reference KNMI data.

- Wind gust should change more than `PERSIST_GUST` m/s in `PERSIST_GUST_T` minutes; this means g_t passes the test if

$$\max_{\tau \in [t - \text{PERSIST_GUST_T}, t] \cap \mathcal{T}} \{g_\tau\} - \min_{\tau \in [t - \text{PERSIST_GUST_T}, t] \cap \mathcal{T}} \{g_\tau\} > \text{PERSIST_GUST}.$$

The calculation for wind direction variations: Here we take special treatment to calculate wind direction fluctuations with time because it is not proper to compute the arithmetic difference directly. Consider a simple example with two wind direction observations 355° and 5° . The absolute value of their arithmetic difference is 350° . However, if we look at the two directions in the plane, the actual angle fluctuation between them should be 10° , while the arithmetic difference 350° does not indicate the real variation. Therefore, we come up with an adding step to solve this problem, as shown below.

- Wind direction should change more than `PERSIST_DIR` degree in `PERSIST_DIR_T` minutes, this means d_t passes the test if

$$\text{let } \Delta d_t = \max_{\tau \in [t - \text{PERSIST_DIR_T}, t] \cap \mathcal{T}} \{d_\tau\} - \min_{\tau \in [t - \text{PERSIST_DIR_T}, t] \cap \mathcal{T}} \{d_\tau\},$$

and then $\min\{\Delta d_t, 360 - \Delta d_t\} > \text{PERSIST_DIR}.$

Determination of parameters

1. *PERSIST_SPD* and *PERSIST_GUST*: In the persistence test, we aim to find abnormally unchangeable speed/gust records in WOW observations. Therefore, `PERSIST_SPD` should be small enough to indicate almost constant wind speed. By looking at wind observations in the WOW data set, we find that both wind speed and gust are recorded with the unit knots² and rounded to one decimal. Since we use the unit meter per second for speed in our study, we set `PERSIST_SPD` and `PERSIST_GUST` as 0.05 (m/s), which is rounded from 0.1 knots.

²1 knot = 1 nautical mile per hour = 1.852 km/h \approx 0.514 m/s.

2. *PERSIST_DIR*: This is similar to what we did for *PERSIST_SPD*. WOW wind direction is recorded with the unit degree and rounded to an integer. Therefore, we take 1 (degree) as the value of *PERSIST_DIR*.
3. *PERSIST_SPD_T*: By following the earlier step in determining *STEP_SPD*, we get a new fluctuation data of KNMI wind speed with time. Next, we check which ones of KNMI wind speed variations are less than 0.05 (m/s), and find the time interval length of those continuous low variations. In this way, we get a distribution of durations that keep near-constant wind speed by KNMI stations. Figure 3.4 shows a cumulative density function of this distribution, where we add a 0.99-probability horizon line. We find the smallest integer that is larger than the 99th-percentile is 4, which means more than 99 percent of the periods when KNMI keeps near-constant wind speeds are less than 40 minutes. Therefore, we take 40 (min) as the threshold *PERSIST_SPD_T*.
4. *PERSIST_GUST_T*: This has the same idea as the approach to determine *PERSIST_SPD_T*. By checking the fluctuation data of KNMI wind gust with time, we see that more than 99 percent of the periods when KNMI keeps near-constant wind gust are less than 40 minutes. Therefore, we take 40 (min) as the threshold *PERSIST_GUST_T*.
5. *PERSIST_DIR_T*: This has the same idea as the approach to determine *PERSIST_SPD_T*. By checking the fluctuation data of KNMI wind direction with time, we see that among the periods when KNMI keeps near-unchangeable wind direction, more than 99 percent of them are less than 90 minutes. Therefore, we take 90 (min) as the threshold *PERSIST_DIR_T*.

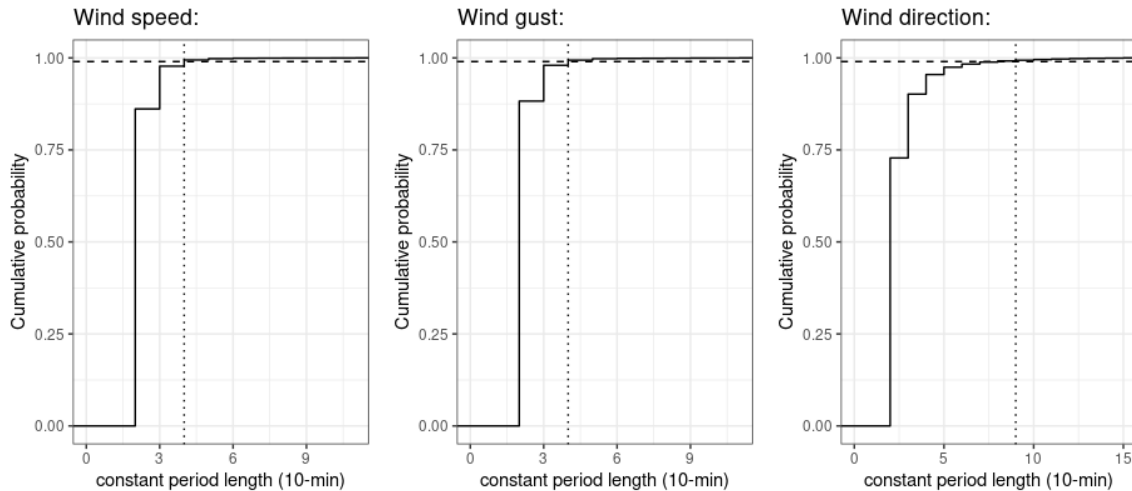


Figure 3.4: The cumulative density distribution of ‘constant’ wind speed/gust/direction time periods, the x label shows the ‘constant’ period equals how many 10-minute intervals.

3.1.4 Summarizing standard checks

We sum up all the previous tests, and Table 3.2 shows the proper order of those standard quality control checks. We define different flags for the failed observations in each test; in this way, we can know straightforward whether a record passes a particular test. Moreover, we set ‘null’ flags for missing values of WOW wind observations. If a wind observation does not have valid preceding

records in the time series, we flag it as ‘isolated’ because it has no previous observations to compare for the temporal checks.

For those isolated observations, we preserve their values after applying the quality control filter. The quality control filter means to remove the value of observations that do not pass the tests from the time series of a single WOW station. The length of such time series does not change, but the number of missing values increases. We are not confident about the quality of isolated observations since they cannot be examined by temporal checks. Therefore, by flagging them as ‘isolated’, we may recognize manually afterward, and let potential users decide whether to keep these observations or not.

Standard quality control procedure			
variant	test	conditions to pass	failed flag
Internal check			
wind speed, wind gust	internal consistency test	wind speed less or equal to wind gust	IN
<i>Comments:</i>	<i>wind speed records with flag IN should not be filtered</i>		
Range check			
wind speed	plausible range test	between 0 and 35.0 m/s	RS
wind gust	plausible range test	between 0 and 64.0 m/s	RG
wind direction	plausible range test	between 0° and 360°	RD
<i>Comments:</i>	<i>wind speed and wind gust have monthly ranges, see Table 3.1. Here we use the universal thresholds 35.0 (m/s) and 64.0 (m/s).</i>		
Temporal check			
wind speed	step test	changes no more than 15.51 m/s in 10 minutes	TS1
wind gust	step test	changes no more than 27.41 m/s in 10 minutes	TG1
wind speed	persistence test	changes at least 0.05 m/s in 40 minutes	TS2
wind gust	persistence test	changes at least 0.05 m/s in 40 minutes	TG2
wind direction	persistence test	changes at least 1 degree in 90 minutes	TD

Table 3.2: Summarized standard quality control steps for WOW wind observations.

3.2 Statistical methods to compare WOW and KNMI

To measure the wind data quality of a WOW station and quantify the improvements after implementing quality control, we need to compare its observations with a reference KNMI station. This involves analyzing the similarity of wind observations between a WOW station and a corresponding KNMI station. In statistics, we have multiple methods to compare two sample distributions. We

introduce the ones that are used here and discuss the results in the next section.

For the rest of the study, we only focus on wind speed observations. Wind speed is the most crucial component of wind observations, and the behavior of wind speed is more consistent in space than wind gust and direction. As mentioned in Chapter 3.1, for crowdsourced data, we have more confidence in wind speed data than in direction and gust data due to the non-standard placement and fabrication of CWS. Thus in this part, the comparisons between WOW and KNMI are applied for wind speed observations only.

Let $X = \{x_t\}_{t \in \mathcal{T}}$ and $Y = \{y_t\}_{t \in \mathcal{T}}$ be two time series of wind speed by a WOW station and a KNMI station, which we want to compare. Let $\mathcal{T}_X \subset \mathcal{T}$ be the subset of time sequence where X has no missing values, and the same to $\mathcal{T}_Y \subset \mathcal{T}$. To make comparisons between any two stations, we consider wind speed observations when both of them have valid data, which is sometimes called pairwise completeness. For the two time series X and Y , we use $\mathcal{T}' = \mathcal{T}_X \cap \mathcal{T}_Y$ to represent the time points where both two have valid values. The comparisons are performed on the two new time series $X = \{x_t\}_{t \in \mathcal{T}'}$ and $Y = \{y_t\}_{t \in \mathcal{T}'}$ with the same length.

3.2.1 Pearson correlation

A common approach to measuring the relationship between two sequences of variables is to check the linear relationship. For a WOW station and its nearest KNMI station, we draw a scatterplot of the paired time series $(X, Y) = \{(x_t, y_t)\}_{t \in \mathcal{T}'}$ in Figure 3.5, and it shows a linear trend between WOW and KNMI observations. We make assumptions that WOW stations have linear relations with KNMI stations on wind speed. This is climatologically reasonable because when a KNMI station records a high wind, its nearby WOW sites are likely to detect a high wind as well.

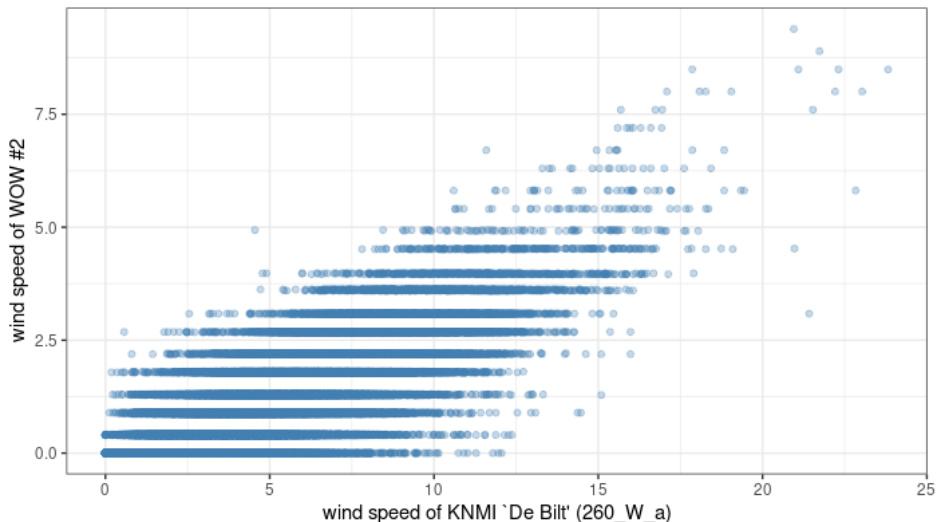


Figure 3.5: A scatterplot that compares simultaneous wind speeds between a WOW station and *relating* KNMI station, before quality control.

We use Pearson correlation coefficient to quantify the linear association between WOW and KNMI wind speeds. For a wind speed time series $X = \{x_t\}_{t \in \mathcal{T}'}$, the mean and standard deviation

of X are computed by

$$\mu_X = \mathbb{E}X = \frac{1}{|\mathcal{T}'|} \sum_{t \in \mathcal{T}'} x_t, \quad \sigma_X = \sqrt{\mathbb{E}[(X - \mu_X)^2]} = \sqrt{\frac{1}{|\mathcal{T}'|} \sum_{t \in \mathcal{T}'} (x_t - \mu_X)^2}.$$

μ_Y, σ_Y for time series Y are the same. Then the covariance between X and Y is defined by

$$\text{Cov}(X, Y) = \mathbb{E}\left((X - \mu_X)(Y - \mu_Y)\right) = \frac{1}{|\mathcal{T}'|} \sum_{t \in \mathcal{T}'} (x_t - \mu_X)(y_t - \mu_Y).$$

And the correlation between X and Y is defined by [Wasserman, 2013]

$$\rho_{XY} = \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\sum_{t \in \mathcal{T}'} (x_t - \mu_X)(y_t - \mu_Y)}{\sqrt{\sum_{t \in \mathcal{T}'} (x_t - \mu_X)^2} \sqrt{\sum_{t \in \mathcal{T}'} (y_t - \mu_Y)^2}}. \quad (3.1)$$

This coefficient is also known as the Pearson product-moment correlation coefficient.

The correlation ρ_{XY} ranges from -1 to 1 , and it measures how strong the linear relationship is between X and Y . If $Y = aX + b$ for some constants a, b , then $\rho_{XY} = 1$ when $a > 0$ and $\rho_{XY} = -1$ when $a < 0$. If X and Y are independent, then $\text{Cov}(X, Y) = \rho_{XY} = 0$ [Wasserman, 2013].

As shown in Figure 3.5, the WOW wind speeds and the KNMI wind speeds do not strictly follow a linear model, and they are not independent either. Thus the Pearson correlation between WOW and KNMI wind speeds should be a positive real number less than 1.

Spearman correlation: Sometimes a linear model does not fit the relations well, such as an exponential function relation. The Spearman's rank correlation can deal with such issue by measuring the statistical dependence between the rankings of two random variables. Here the rankings for a time series are the indices of the corresponding permutation of the time series in increasing order. The Spearman correlation coefficient indicates how well the relationship could be described using a monotonic function [NIST/SEMATECH, 2020], which is calculated by the Pearson correlation between the rankings of two random variables.

Taking the exponential function relation as a trivial example, let $\{1, 2, 3, 4\}$ and $\{2, 4, 8, 16\}$ be the numerical values of two time series, then the corresponding rankings are $\{1, 2, 3, 4\}$ and $\{1, 2, 3, 4\}$. We can see that the two rankings follow a linear model, but the numerical values do not. In this case, the Spearman correlation is 1, while the Pearson correlation is lower.

We mainly analyze the Pearson correlations between WOW and KNMI wind speeds, and take the Spearman correlations as a supplement, since we are more interested in the linear relations between WOW and KNMI. Moreover, Pearson correlations are more sensitive to outliers than Spearman correlations, where an outlier is an observation that deviates more than some threshold with other observations [Grubbs, 1969]. The Spearman correlation has less sensitivity in outliers because it diminishes the effect of abnormally large values by assigning them with rankings.

3.2.2 Kolmogorov–Smirnov statistic

The Kolmogorov–Smirnov (K-S) statistic measures the discrepancy between two cumulative probability distributions of samples. It is the test statistic used in the Kolmogorov–Smirnov test, a type

of hypothesis test in statistics that assess whether a list of samples fit a distribution or whether two lists of samples come from the same distribution.

For a time series $X = \{x_t\}_{t \in \mathcal{T}'}$, its empirical cumulative distribution function of X is given by [Rui Castro, 2020a]

$$\hat{F}_X(p) = \frac{1}{|\mathcal{T}'|} \sum_{t \in \mathcal{T}'} \mathbb{1}\{x_t < p\}, \quad \forall p \in \mathbb{R}, \quad \mathbb{1}\{x_t < p\} = \begin{cases} 1, & x_t < p, \\ 0, & x_t \geq p, \end{cases}$$

where $\mathbb{1}$ is the indicator function. Then the K-S statistic for the two-sample case is defined by [Rui Castro, 2020b]

$$D_{XY} = \sup_{p \in \mathbb{R}} |\hat{F}_X(p) - \hat{F}_Y(p)|. \quad (3.2)$$

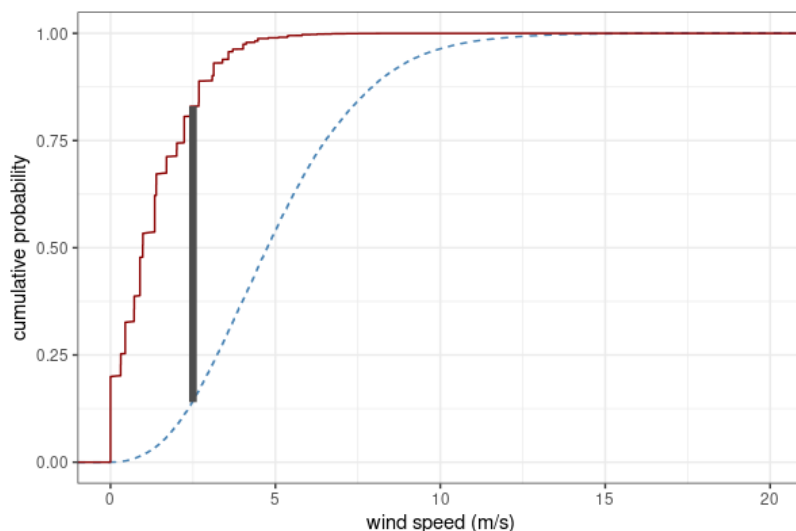


Figure 3.6: The cumulative probability distributions of a WOW station (red line) and a *relating* KNMI station (blue dashed line). The grey bar indicates the K-S statistic between them.

This K-S statistic indicates the largest vertical distance between two empirical cumulative probability distributions, as shown in Figure 3.6. Generally, if the K-S statistic is not large, this implies that the two cumulative probability distributions are close to each other, then we can conclude that the two distributions are similar. The K-S statistic is sensitive to any changes in the sample data [NIST/SEMATECH, 2020], because the changes in sample data lead to the changes in cumulative probability distribution and then affect the vertical distances. For a WOW station and a *relating* KNMI station, we use the K-S statistic to measure how well the WOW wind speeds fit the KNMI wind speed distribution.

3.3 Standard quality control results

We apply the standard quality control methods on the WOW data set. The implementation follows the order in Table 3.2, where the data used in each step is inherited from the filtered data after its previous step. Table 3.3 shows the percentage of WOW observations that fail each standard quality control check. Note that the optional internal check is implemented on WOW raw data set, but we do not filter the failed observations as been mentioned in Chapter 3.1.1. We can see from the table that more than 80% of the valid data from the 39 WOW stations meet the requirements of standard quality control, which is a promising result for CWS. Moreover, most of the filtered

observations are flagged in the temporal persistence test, which means the WOW stations are likely to get stuck in wind sensors.

Standard quality control	wind speed	wind gust	wind direction
Internal check (<i>optional</i>)	0.61%	0.61%	
Range check	0.0002%	0.003%	0.0006%
Temporal step test	0.0008%	0.001%	
Temporal persistence test	14.82%	8.88%	9.99%
Overall pass percent	85.09%	90.20%	89.96%

Table 3.3: Failed percentages in each standard quality control check for WOW data.

The WOW data set after filtered by standard quality control is an important intermediate outcome. For each WOW station, we compute the Pearson correlation between a KNMI station and this WOW station and find which KNMI station has the highest Pearson correlation after standard quality control, as shown in Figure 3.7. For each WOW station, we refer to the KNMI station that has the highest wind speed Pearson correlation with the WOW as its *relating* KNMI station, for the rest of this thesis. In this way, we compare a WOW station with the KNMI that has the greatest linear relationship. We also see from Figure 3.7 that wind directions have a lower correlation with the *relating* KNMI at most times, and even negative sometimes. This suggests that the quality of wind direction observations varies a lot in different WOW stations, some are good enough for comparison, while some have no useful information since the correlations are too low.

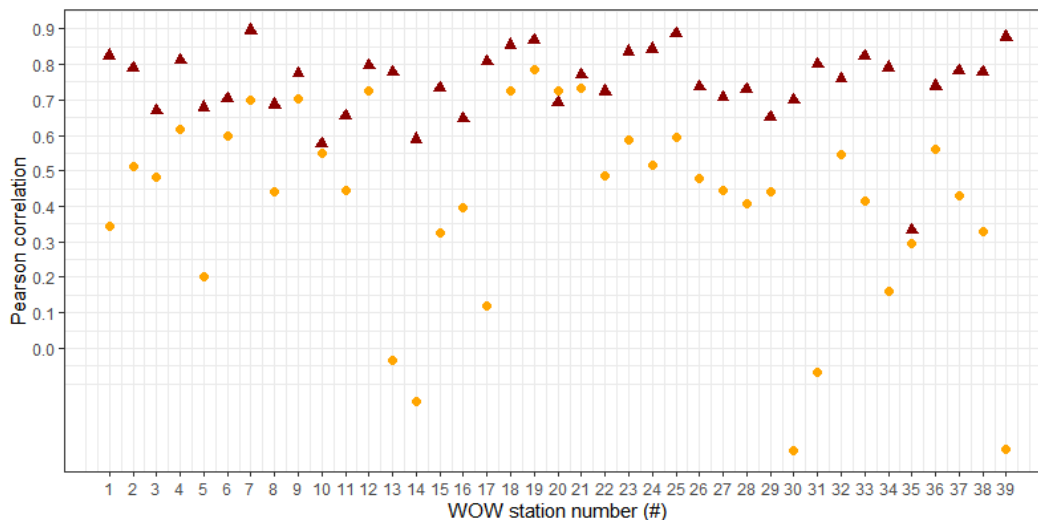


Figure 3.7: The Pearson correlation of wind speed (red triangles), and wind direction (orange points) between a WOW station and its *relating* KNMI station (the KNMI that has the highest wind speed Pearson correlation with the WOW).

We also compute both the Pearson correlation and Spearman correlation between each pair of WOW stations or each pair of KNMI stations as functions of geographical distance, as shown in Figure 3.8 and Figure 3.9. We find that WOW stations do not have a decay behavior with distance, while the KNMI stations show a trend of reducing correlations with larger distances. The reasons

might be that WOW stations have a low quality of data, and so the noise amplitude of correlations covers the trend that decays with distance. The WOW stations are densely located, which means the climatological differences between any two are not distinct. This motivates us that the nearest station might not be a good match for a WOW station. We need to compare a WOW station with others (both WOW and KNMI) throughout the thesis; these figures suggest us to choose similar stations by statistical relationship rather than geographical distances.

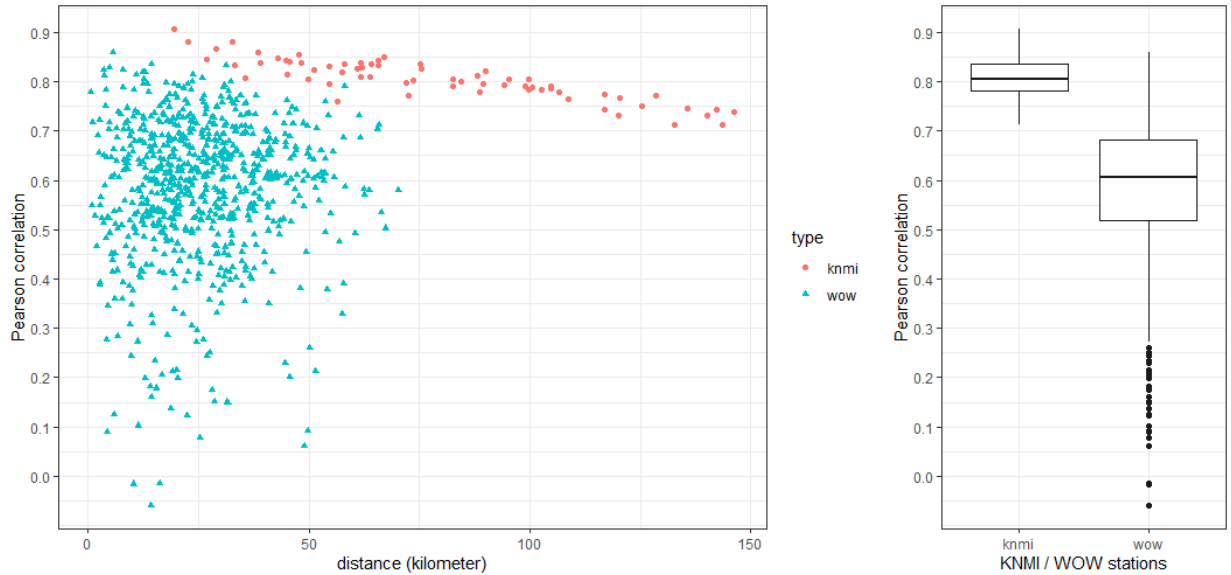


Figure 3.8: Pearson correlation of wind speed distributions between each pair of WOW (blue triangles) or each pair of KNMI (red points) stations.

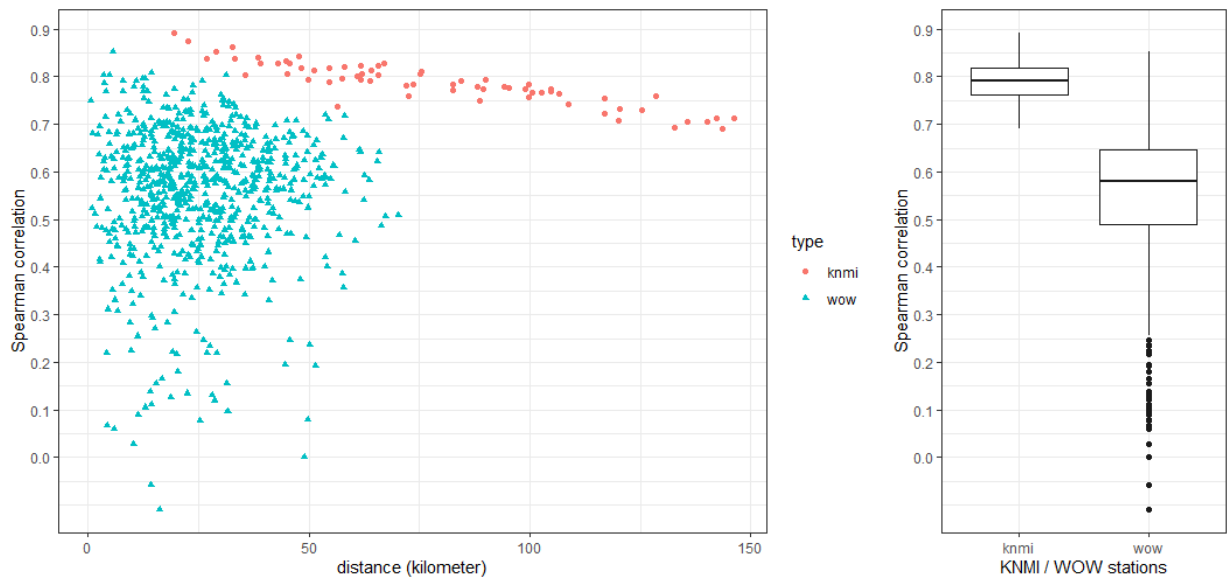


Figure 3.9: Spearman correlation of wind speed distributions between each pair of WOW (blue triangles) or each pair of KNMI (red points) stations.

We compare wind speeds by a WOW station and a *relating* KNMI station for both before and after standard quality control, and figure out how well the standard quality control methods improve the relation between WOW and KNMI data. Figure 3.10 shows the changes in K-S statistics after implementing the standard quality control on WOW data. We can see that the distributions of filtered WOW wind speeds are closer with the distributions of *relating* KNMI wind speeds.

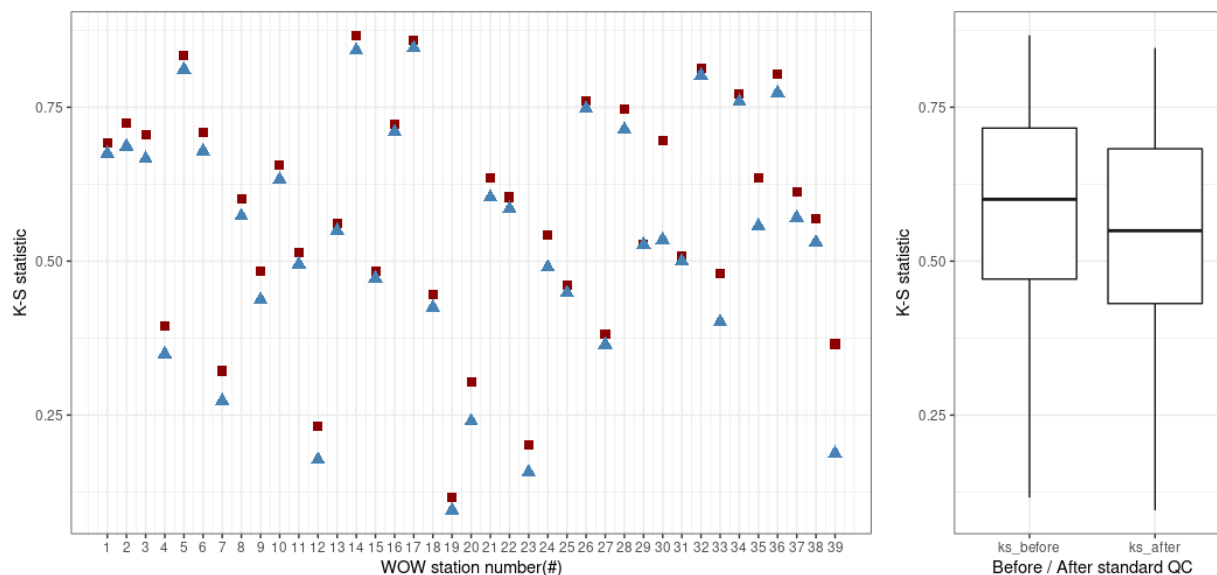


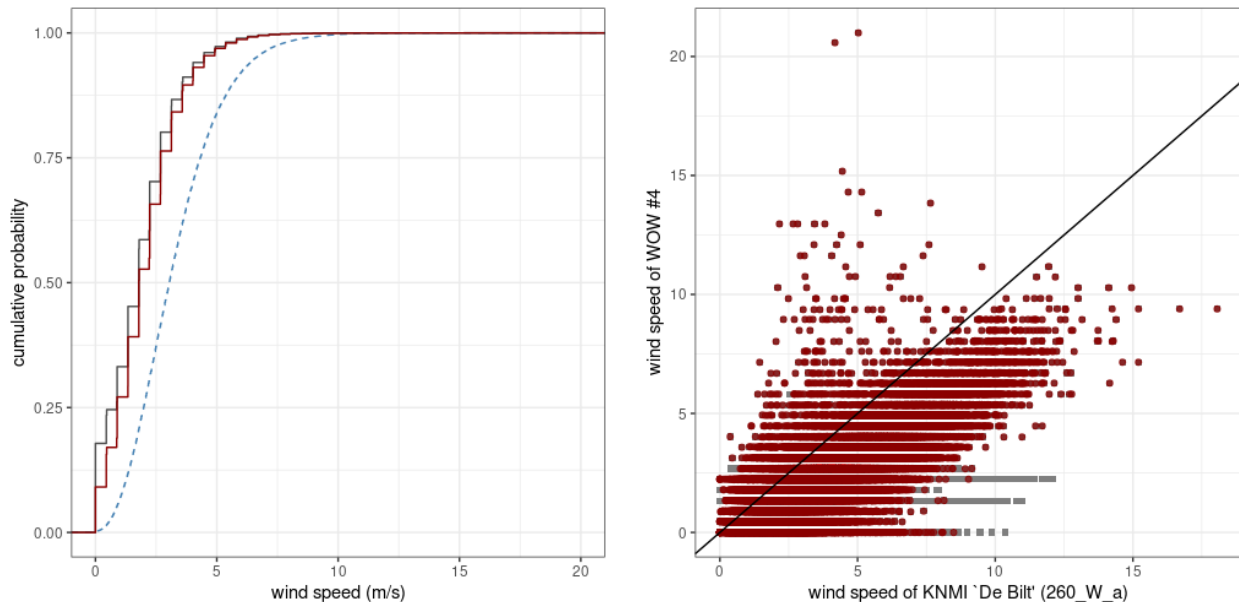
Figure 3.10: Kolmogorov–Smirnov statistic of wind speed distributions between a WOW and a *relating* KNMI, before (red squares) and after (blue triangles) standard quality control.

From the failed percentages table, we have an overview of how many WOW observations could be trusted for describing wind. The figures for statistical indicators show that we improved the quality of WOW data. Next, we take an in-depth look at the specific observations at a representative station WOW #4. We match the wind speed data of this example station with its *relating* KNMI station. Figure 3.11a compares the cumulative probability distributions of wind speed, while Figure 3.11b provides another way to look at WOW wind speed behavior associated with KNMI. We can see from those two graphs that the WOW station records a lot more zero wind speed than the KNMI station, and that the standard quality control filtered many, but not all of them.

This example WOW station has a website to share weather observations and information ³, on which the owner wrote that his anemometer is placed on the house with a 14-meter height. We checked the location of this WOW station and found that it is in the urban area with some trees and buildings nearby. These obstructions might be the reason why most of the wind speed observations from this WOW station are lower than its *relating* KNMI station. Such situations happen a lot for WOW stations, and so most WOW stations record lower wind speeds than the real ones. Although there are a couple of anemometer issues that could lead to low-speed readings, a more likely reason why CWS record lower wind speeds is that the user fails to appreciate the importance of anemometer height and exposure in recording higher wind speeds [Dann and Prodata Weather Systems, 2020]. Therefore, we need to calibrate these wind speed observations to a larger value so that they indicate real wind correctly. This calibration step is known as bias correction,

³<http://weerstation-dorestad.nl/>

and we talk about them in the next chapter.



(a) Cumulative probability distributions for wind speed by the WOW #4 station before (grey line) and after (red line), and the *relating* KNMI station (blue dashed line).

(b) A scatterplot that draws simultaneous WOW #4 and nearest KNMI wind speed records, grey points represent observations filtered after standard quality control.

Figure 3.11: Comparison of wind speed between the WOW and KNMI.

3.4 Discussion and future works

In the last section of this chapter, we discuss the limitations in our current standard quality control, and provide some possible directions that could be tested in future studies.

3.4.1 Improving parameter selection

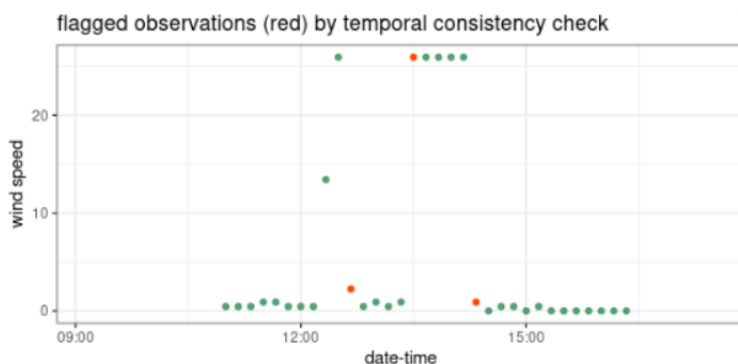
In the temporal checks, the parameters are chosen by the reference KNMI data. However, that data only has wind observations by 12 KNMI stations in the three years 2016-2018. It might be insufficient for summarizing wind behaviors in the region. Therefore, we may consider a more extended period of historical wind observations with more stations, for instance, the historical wind observations of all KNMI stations in the Netherlands over the past 50 years. The parameters selected based on the new large data set should be more representative and convincing.

Another idea to improve the parameter selection is to choose parameters of the temporal checks in different seasons, as we did in the range check. We have discussed in Chapter 2 that the wind has seasonal variations, and thus it could be a proper way to determine parameters for winter or summer seasons based on the distinct wind behaviors.

3.4.2 Limitations of current temporal check

A desired requirement of our quality assurance procedure is that the steps can be implemented in real-time WOW observations. To fulfill this goal, we only consider past observations for comparison when constructing the temporal checks. Unfortunately, this operation causes some problems in flagging suspect observations. Thus in the last section of the standard quality control methods, we discuss the limitations of our temporal checks and try to find a feasible solution.

We find a representative example of the shortage in our temporal checks, as shown in Figure 3.12. The red points are the flagged observations by step test, while the green ones are those who pass temporal checks. We can see that only three wind speed records failed the step test since they have a large gap with their former records. However, when we look at those observations manually, it is reasonable that the five continuous high wind speed records should all be flagged, while the lower two records are not necessary to be flagged. There are a few such cases in our data set, and it is important that we flag those observations correctly.



clusters in the fluctuation data set of WOW wind observations. In this way, those continuous constant observations set up a type of clusters since their variations are all small enough. While the ordinary wind observations set up a different kind of clusters as their variations lie in reasonable thresholds. Then the sharply changed wind observations form the isolated points between clusters. By constructing proper conditions, we can correctly flag all the suspect observations.

Chapter 4

Bias Correction

Many WOW stations measure wind speeds lower than actual wind observed by KNMI stations, especially for urban sites, as discussed in Chapter 3.3. This is mainly because of the lower placing height and surrounding obstructions of wind sensors. In this research, we want to improve the data quality of WOW wind observation networks and combine them with the sparse KNMI observation network. We have assumed that the KNMI network indicates true wind. If the gaps between WOW observed and actual wind speeds are huge, the combined network could not indicate wind speeds at the same level. Therefore, we need to estimate and eliminate the error between observational and true speed values. This step is known as bias correction in meteorological studies.

Our bias correction step is constructed after finishing standard quality control, and thus the WOW data used in this chapter is the filtered data after standard quality control checks in Chapter 3. In this chapter, we would first introduce the ideas and methods for bias correction. Given these methods have been largely developed for temperature and precipitation. We discuss their suitability for wind data and adapt these methods to be suitable for CWS that may contain different types of observation errors. Finally, we apply the appropriate methods to the WOW wind speed data and use several statistical indicators to measure the improvements in errors compared to KNMI stations.

4.1 General explanation

In statistics, the term bias comes from point estimation. Point estimation refers to providing a single ‘best guess’ of some quantity of interest, and the bias is the difference between the expected value of ‘best guess’ and the actual value. [Wasserman, 2013]. For a wind speed time series $X = \{x_t\}_{t \in \mathcal{T}}$ of a WOW station, our quantity of interest is the corresponding true wind speed time series $Y = \{y_t\}_{t \in \mathcal{T}}$. For a certain true wind speed y_t at time t , we treat the simultaneous WOW observation x_t as the estimate of y_t . The bias of this estimate x_t is then given by

$$\text{bias}(x_t) = x_t - y_t.$$

We assume the *relating* KNMI wind speeds could indicate true wind at the location of the WOW station, and then Y is the wind speed time series from the *relating* KNMI. We aim to reduce the bias of WOW wind speeds in this chapter, which means we need to reduce the gap between x_t and y_t for any $t \in \mathcal{T}$.

Figure 4.1 compares the mean wind speed values of different WOW with their *relating* KNMI stations, namely μ_X and μ_Y . We can see from this figure that most WOW stations hold a relatively large bias (more than 1 meter per second) on average compared to their *relating* KNMI station. This implies that the errors between WOW observed and true wind speed are substantial, and so the bias correction step is essential in the quality assurance procedure for WOW data.

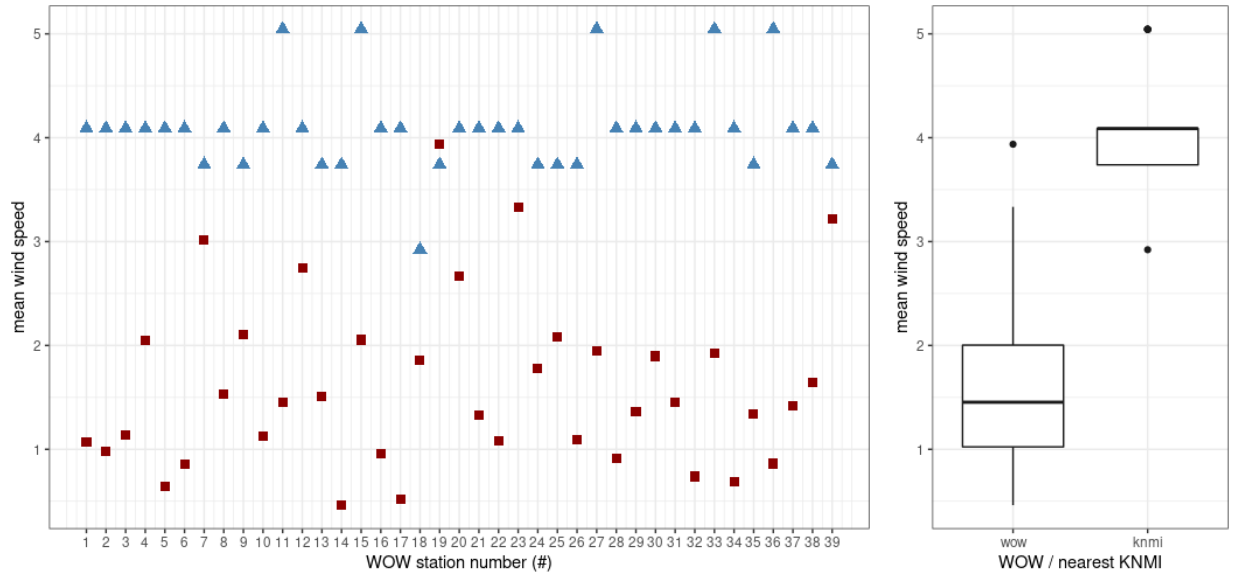


Figure 4.1: Differences between mean WOW (red squares) and KNMI (blue triangles) wind speeds

To study how such bias is reflected in the wind speed observation sequence, we choose a representative WOW station that is the same in Chapter 3.3. Figure 4.2 gives the scatterplot of simultaneous wind speed between WOW (*y*-axis) against its *relating* KNMI (*x*-axis), as well as the corresponding two-dimensional joint frequency plot. We can see from both graphs that a lot of points lie below the diagonal, which means most WOW wind speed records are lower than their corresponding KNMI observations. Moreover, WOW stations record a lot of zero wind speeds, even after the temporal checks in Chapter 3.1. In contrast, the true distribution for wind speed, as observed at KNMI stations, does not have zero-inflated observations. These two features may explain why there is a gap in mean wind speed for WOW versus KNMI.

In the bias correction step, we want to calibrate WOW wind speed such that most points in the scatterplot distribute around the diagonal. We can then check if the distribution of observations at WOW stations is comparable to that of KNMI stations.

4.2 Bias correction methods

The bias correction is achieved through building a mathematical model from input WOW and KNMI data to correct WOW wind speeds. We split the input data into two parts, the training data and the test data. Then we use the training data to fit the model and apply the test data for evaluation, as summarized in Figure 4.3.

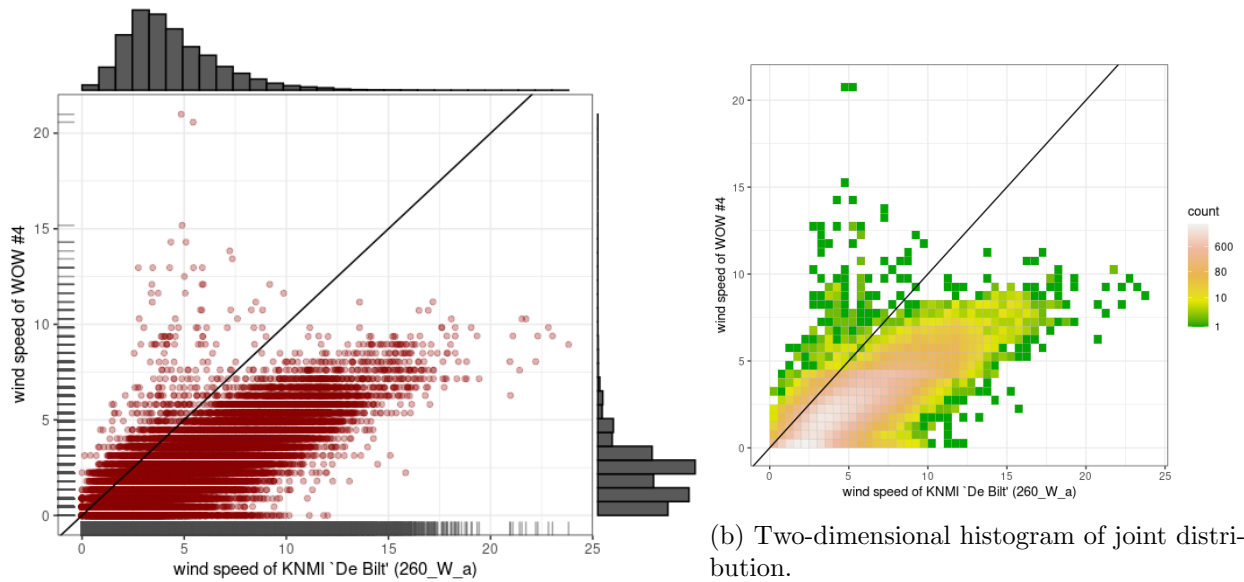


Figure 4.2: Synchronous wind speed observations between WOW (y -axis) and its *relating* KNMI (x -axis) station, with the diagonal line ($y = x$).

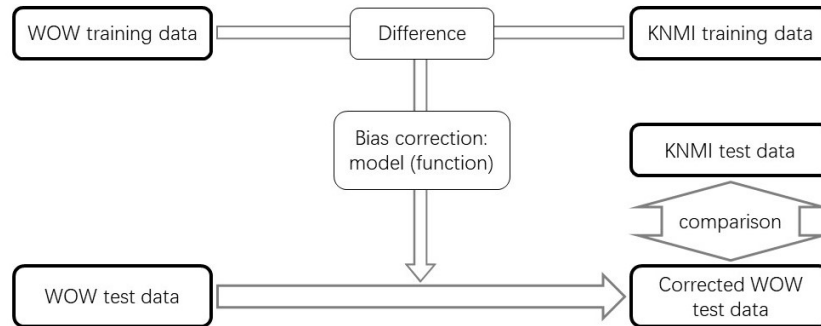


Figure 4.3: A flow chart that explains the general scheme of bias correction methods.

Generally, the training and test data should be representative of the input data. Since we have wind speed observations in the three years 2016-2018, and wind speeds have low annually variations, we naturally set historical observations in 2016 and 2017 as the training data, and set observations in 2018 as the test data.

Before introducing bias correction methods, an essential assumption should be made. We assume that the differences in simultaneous wind speeds between WOW and KNMI are similar in both training and test data. The meteorological understanding of this assumption is that the deviations between WOW observed and true wind speeds are consistent over time. This assumption is suitable for CWS, because the main cause of the bias in WOW data is the systematic errors and placements of wind sensors, as we outlined previously. Systematic errors are dependent on instrument settings, not time, and so our data satisfy the assumption.

Bias correction methods are widely implemented in meteorological studies, not only for correcting current observations, but also for calibrating weather or climate model output. Many researchers have made contributions in this area; they tried multiple mathematical models to fit the errors on different weather variables. Table 4.1 lists some of their previous work. We find that most bias correction research has focused on temperature and precipitation, and that the linear scaling method (discussed in Chapter 4.2.1) is the most widely used.

Bias correction methods		parametric methods			nonparametric methods
Study	Variable	Linear scaling	Nonlinear scaling	Distribution mapping	Empirical quantile mapping
[Terink et al., 2008]	temperature	Yes	Yes	No	No
[Hawkins et al., 2013]	temperature, precipitation	Yes	No	No	No
[Ho et al., 2012]	temperature, precipitation	Yes	No	No	No
[Teutschbein and Seibert, 2012]	temperature, precipitation	Yes	Yes	Yes	No
[Navarro-Racines and Tarapues, 2015]	temperature, precipitation	Yes	No	No	Yes
[Akhter et al., 2017]	temperature, precipitation	Yes	No	No	Yes
[Fang et al., 2015]	temperature, precipitation	Yes	Yes	Yes	Yes
[Luo et al., 2018]	temperature, precipitation	Yes	Yes	Yes	Yes
[Gudmundsson et al., 2012]	precipitation	Yes	Yes	No	Yes
[Droste et al., 2019]	wind speed	Yes	No	No	No

Table 4.1: Commonly used bias correction methods in meteorological and climatological studies.

In the following sections, we start from the commonly used linear scaling method and then introduce other methods in Table 4.1. We analyze the performance of each method from a mathematical point of view, and finally figure out the most suitable method to do bias correction for the WOW data.

These methods can be described as either parametric or nonparametric. In statistics, mathematical models are divided by parametric and nonparametric models. A parametric model expressed by

$$\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta \subset \mathbb{R}^d\}, \quad d < \infty$$

is a model that can be described by a finite set of d parameters θ , and the dimension of the model (number of parameters) is fixed a priori regardless of the data [Rui Castro, 2020a]. We can treat the model itself as a collection of distributions or functions. If we know these values of parameters, then we can get the specific distribution or function. We say a model is nonparametric if it cannot be described by finite parameters.

Recall that in Chapter 3.3, we use paired time series of simultaneous WOW and KNMI wind speeds during the same pairwise complete period for comparison, denoted by $X = \{x_t\}_{t \in \mathcal{T}'}$ and $Y = \{y_t\}_{t \in \mathcal{T}'}$. We follow the same notations in bias correction, where x_t is a variable representing WOW wind speeds that need to be calibrated, and variable y_t indicates the *relating* KNMI wind speeds that are treated as the true values for corrected WOW wind speeds.

Our goal for bias correction is to construct a proper model that fits the relation between wind speeds x_t of a WOW station and the *relating* KNMI observations y_t , for all $t \in \mathcal{T}'$, so that we can apply this model on real-time WOW wind speed observations that do not have simultaneous KNMI records to refer.

4.2.1 Linear scaling

The linear scaling method is the most widely used approach for bias correction. Many researchers applied linear models on calibrating temperature or precipitation data, because the linear transformation provides a direct way to fit one distribution to the other. For wind speed in our data, we can also try linear transformations to fit a WOW distribution to a *relating* KNMI distribution, as investigated in Chapter 3 that there are some linear relations between WOW and KNMI wind speeds. Figure 4.4 compares the wind speed distribution between a WOW and its *relating* KNMI station. The linear transformation corrects the WOW distribution by shifting or scaling.

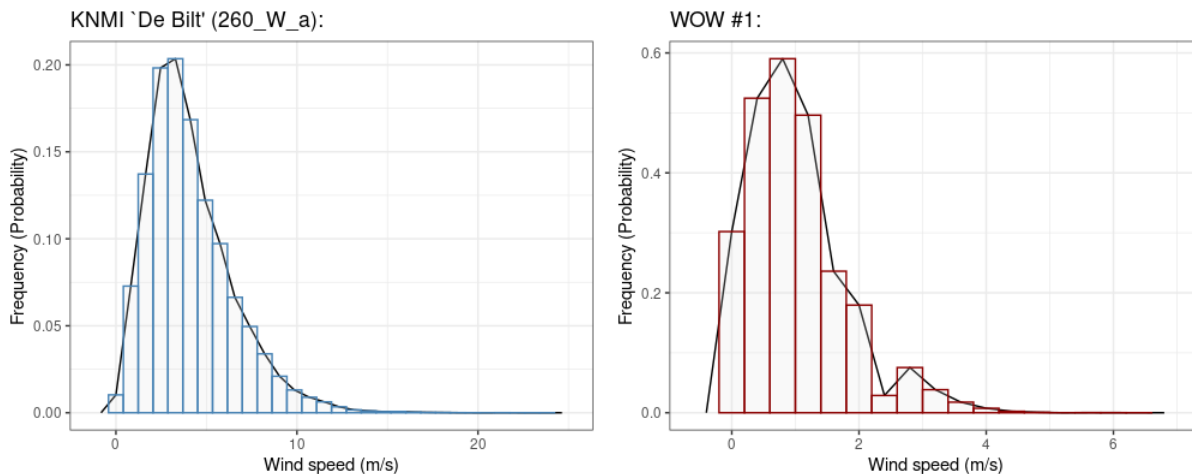


Figure 4.4: The histogram (frequency) plots of a KNMI (left) station and a WOW (right) station.

The linear model fits the relation between WOW observed wind speed x_t and the true wind speed y_t by a linear equation expressed as

$$y_t = \alpha \cdot x_t + \beta. \quad (4.1)$$

This requires the estimation of parameters α and β from the training data. We need to find the proper estimated parameters for each WOW station individually, since the errors between observational and true wind speeds are distinct from station to station. A direct way to find the two parameters is to do linear regression and get the corresponding parameters. However, for bias correction, we want to correct the climate distributions as a whole, so that it is comparable

with the true one. This means to calibrate the mean and variance of the distribution, while linear regression parameters are not good estimates for the two. There are mainly three ways to estimate the parameters α and β , listed as follows.

- (i) (simple addition) If there is a constant deviation between original and reference data, then we can estimate the deviation by the difference between means for the two. The parameters α and β in (4.1) are

$$\alpha = 1, \quad \beta = \mu_Y - \mu_X, \quad (4.2)$$

where μ_Y and μ_X are the mean values of time series Y and X respectively.

Teutschbein and Seibert [2012] applied this estimation on temperature based on the monthly mean difference. However, this approach does not apply to wind speed because the bias is not constant, as we can see from Figure 4.4 that shifting can not fit the two distributions. It works well for temperatures that fit a normal distribution, but is not appropriate for wind speeds which are not normally distributed.

- (ii) (simple multiplication) If there is a constant ratio relation between original and reference data, then we can estimate the ratio by the quotient of means for the two. The parameters α and β in (4.1) are

$$\alpha = \frac{\mu_Y}{\mu_X}, \quad \beta = 0. \quad (4.3)$$

Fang et al. [2015] applied this estimation on precipitation data. We find this approach is not suitable for wind speed, as can be seen from the example in Figure 4.2, if all the WOW wind speeds are multiplied by a constant larger than 1, some of the corrected wind speed would be too high compared with KNMI's.

Droste et al. [2019] are the only authors investigating the bias on crowdsourced CWS wind speed observations, and they tried linear scaling methods to correct CWS wind speeds by simply multiplying a linear regression coefficient. However, as we discussed before, this might not be a good estimate for the parameter. His results showed that the bias correction did correct a lot of low wind speed observations to a larger value, but it also calibrated some observations to abnormally large ones. This motivates us to seek more suitable approaches for wind speed bias correction.

- (iii) (variance scaling) If the differences between original data and its mean hold a constant ratio relation with the differences between reference data and its mean, then we can estimate the ratio by the quotient of standard deviations for the two. The parameters α and β in (4.1) are

$$\alpha = \frac{\sigma_Y}{\sigma_X}, \quad \beta = (\mu_Y - \alpha \cdot \mu_X), \quad (4.4)$$

where σ_Y and σ_X are the standard deviations of Y and X respectively. Then the linear model (4.1) becomes

$$y_t - \mu_Y = \frac{\sigma_Y}{\sigma_X} \cdot (x_t - \mu_X).$$

In this way, we correct both the mean and variance of the distribution. Luo et al. [2018] applied this estimation on temperature data. The idea is first to shift the two distributions so that they both have zero mean, and then compute the variance ratio between two zero-centered distributions.

However, this approach is not suitable for wind speeds that have strictly positive, skewed distributions. If we apply this approach to WOW wind speeds, some of the corrected speed values would be negative, which is not the desired outcome for bias correction.

4.2.2 Nonlinear scaling

The insufficiency of linear scaling methods prompts us to seek new approaches for wind speed bias correction. A traditional approach for calibrating precipitation data is to try parametric nonlinear models instead of linear models. Previous researchers applied the power transformation for correcting precipitations, as precipitations tend to have more near-zero observations than temperatures and are strictly nonnegative. Our WOW wind speeds also have a lot of zero and near-zero values, and so we investigate this nonlinear scaling method and discuss how well it suits our wind speed data.

The power transformation is expressed in the following equation,

$$y_t = \alpha \cdot x_t^\beta. \quad (4.5)$$

Teutschbein and Seibert [2012] applied the nonlinear scaling method using power transformation on bias correction of precipitation data, as the previous linear scaling method (ii) did not correct the differences in the variance for precipitation distributions. They consider a nonlinear correlation in an exponential form that is used to adjust the variance specifically. Their estimation for parameters α and β is given by

$$\alpha = \frac{\mu_Y}{\mu_X}, \quad \beta = \arg \min_{\beta \geq 0} \left\{ \left| \frac{\sigma_Y}{\mu_Y} - \frac{\sigma_{X^\beta}}{\mu_{X^\beta}} \right| \right\},$$

where β commonly takes value of $\frac{1}{3}$ in practice. We take this nonlinear scaling method into consideration because for some WOW stations, their Spearman's correlation is higher than Pearson's correlation, which implies that these WOW stations are more likely to hold a nonlinear relation with their *relating* KNMI stations, as we have discussed in Chapter 3.4. However, the properties of power transformation indicate it tends to get abnormally large values for corrected wind speeds, especially for high wind speeds in the original data. Therefore, this nonlinear scaling method does not suit the WOW wind speed, and hence we do not apply this approach.

4.2.3 Distribution mapping

Given limitations of linear scaling, in this section, we explore an alternate approach, the mapping methods. Mapping methods are based on the statistical distribution of training data, and the aim is to find a proper map between two distributions. Unlike scaling methods that take universal transformation on all data, mapping methods correct the data according to their quantiles, which is the percentage location in the cumulative distribution. The Pearson correlation does not change by linear scaling, and the Spearman's correlation does not change by nonlinear scaling, but both of them would change after applying the mapping methods on data.

We first try a parametric distribution mapping method. The idea is to correct the distribution function of the WOW data to agree with the distribution function of the KNMI data. This approach involves applying proper statistical distributions to fit the training data of each station, and assuming these distributions from any station are of the same type. Teutschbein and Seibert [2012] discussed the distribution mapping methods on both temperature and precipitation data. They applied the Gamma distribution on precipitation data and the Normal distribution on temperature data.

By looking at the histograms of wind speed from a representative KNMI station and a WOW station in the training data (Figure 4.4), we see that the low wind speed distribution presents a

mountain peak shape while the high wind speed distribution presents a rapidly shrinking tail. This shows a similar behavior with the Weibull distribution. Droste et al. [2019] also applied the Weibull distribution to fit both official and crowdsourced CWS wind speeds. Therefore, we consider the Weibull distribution in the distribution mapping method by assuming that the wind speeds of all WOW and KNMI stations follow Weibull distributions.

The Weibull cumulative distribution function group can be represented by [Papoulis and Pillai, 2002]

$$F(p; k, \lambda) = 1 - \exp \{ - (p/\lambda)^k \},$$

where $k > 0$ is called the shape parameter and $\lambda > 0$ is called the scale parameter of the distribution. We fit the two parameters for each station in the training data by applying the maximum likelihood. Let functions $F_X(\cdot)$ and $F_Y(\cdot)$ denote the fitted Weibull distributions of X and Y , respectively. Then the inverse function of F_Y is the corresponding quantile function F_Y^{-1} , that maps a probability to a certain value. The model used to correct WOW wind speed can be written as

$$y_t = F_Y^{-1}(F_X(x_t)). \quad (4.6)$$

For a given WOW wind speed, we first find its cumulative probability according to the fitted Weibull model $F_X(\cdot)$, and then get the corrected wind speed at this cumulative probability level, in the way of applying KNMI fitted Weibull quantile function to map the value.

However, this mapping method does not work well for WOW stations in the training data, because the wind speed distributions of most WOW stations do not fit the Weibull distributions exactly. A significant problem is that some WOW stations record a lot of zero wind speed, as we discussed earlier in this chapter. The large number of zero wind speed observations means that the Weibull distribution does not fit WOW stations well. Droste et al. [2019] tried a mixture of two Weibull distributions on fitting CWS wind speed data, and he solved the zero-inflated records problem by throwing away all the low wind speed observations. We want to keep the low wind speed records, and some WOW stations do not seem to follow the Weibull distribution even after excluding the zero wind speeds. Therefore, this distribution mapping method only performs well on correcting wind speed data from a high-quality WOW station that has a similar shape in wind speed distribution with official ones.

4.2.4 Empirical quantile mapping

The parametric methods discussed so far are not well-suited for all WOW wind speed data, and thus we seek nonparametric methods that do not rely on parameters for a better solution. We expect the nonparametric methods to perform better than parametric methods, since they are generated by the data itself, which may avoid the problem that the estimation of parameters may not be suitable for some observation values.

The most commonly used nonparametric method for fitting two distributions is to use the empirical quantile function, which is known as the empirical quantile mapping method. The general idea of empirical quantile mapping is similar to distribution mapping. The only difference is that here we would not try a certain statistical distribution to approximate the training data, but use the empirical cumulative distribution function.

For a time series $X = \{x_t\}_{t \in \mathcal{T}'}$, the empirical distribution function \hat{F}_X is a step function defined by

$$\hat{F}_X(p) = \frac{1}{|\mathcal{T}'|} \sum_{t \in \mathcal{T}'} \mathbb{1}\{x_t < p\},$$

where $\mathbb{1}$ is the indicator function. When $|\mathcal{T}'|$ goes to infinity, the empirical distribution function shows the exact distribution of the data, that avoid the distributional assumptions. Therefore, we would expect a more accurate calibration result for WOW wind speed compared to the distribution mapping, especially for some WOW stations that do not follow the Weibull distribution.

Let $\hat{F}_X(\cdot)$ and $\hat{F}_Y(\cdot)$ represent the empirical distribution functions of observed and true wind speed separately, then this nonparametric bias correction model can be expressed by

$$y_t = \hat{F}_Y^{-1}(\hat{F}_X(x_t)). \quad (4.7)$$

By following the analysis of all the previous methods for bias correction, we conclude that the empirical quantile mapping method is the most suitable one for correcting our WOW data. Therefore, we implement empirical quantile mapping to do the bias correction and discuss the performance results in Section 4.4.

4.3 Statistical indicator to evaluate bias correction

In order to assess how well the bias correction performs for wind speed, and quantify the improvement after implementing bias correction, we need multiple statistical indicators to evaluate the bias correction methods. Previous researchers have used multiple indicators for evaluating bias correction. Teutschbein and Seibert [2012], Fang et al. [2015], Luo et al. [2018] applied multiple scores including the mean, standard deviation, coefficient of determination (R^2), and mean absolute error (MAE). Akhter et al. [2017] tried statistics like index of agreement (d -index) and Nash-Sutcliffe efficiency (NSE). The main idea of using those statistical indicators for evaluation is to (a) quantify the reduction in errors between two observation sequences, and to (b) compare the similarity of two observation distributions after bias correction.

We use the root mean square error (RMSE) to measure the reduction of bias in WOW wind speeds, and apply the Kolmogorov–Smirnov (K-S) statistic to indicate the improvements in distribution similarities. The K-S statistic is already discussed in Equation (3.2) of Chapter 3.2.2, and we introduce the RMSE in the following section.

4.3.1 Root mean square error

The root mean square error (RMSE) is a measure of differences between predicted/estimated values and true values. We apply the RMSE to quantify the average error of wind speeds between a WOW station and a *relating* KNMI station. We assume the KNMI data indicates real wind speeds, and the WOW data represents the synchronous estimates, then the RMSE measures how large the WOW records deviate from the real wind speeds.

The RMSE computes the square root of the arithmetic mean of the squared errors. For two time series $X = \{x_t\}_{t \in \mathcal{T}'}$ and $Y = \{y_t\}_{t \in \mathcal{T}'}$, the RMSE between X and Y is defined by

$$\text{RMSE}(X, Y) = \sqrt{\mathbb{E}[(X - Y)^2]} = \sqrt{\frac{1}{|\mathcal{T}'|} \sum_{t \in \mathcal{T}'} (x_t - y_t)^2}. \quad (4.8)$$

From Equation (4.8) we can see that RMSE aggregates the errors in each time points of two sequences, and it is standardized for the length of sequences. The RMSE is always a nonnegative value. A lower RMSE value indicates smaller differences between X and Y . In our case, low RMSE between WOW and KNMI wind speeds means the WOW wind speeds are well suited for indicating true wind.

K-S statistic compares WOW and KNMI wind speeds from the overall distribution perspective, while RMSE compares the differences between each pair of simultaneous observations. We compute RMSE and K-S statistic to measure the performances of bias correction, so that we can have a view of comparison between WOW and KNMI wind speeds from both integrated and particular aspects.

4.4 Bias correction results

In this section, we implement the empirical quantile mapping method for bias correction, which has been selected as the most appropriate one for correcting WOW wind speed. For each individual WOW station, we first train the model from the WOW observed and KNMI true training data in the two years 2016-2017, and then apply the model to correct the WOW observed test data in the year 2018. After that, we get three wind speed sequences for comparison, the WOW observed wind speeds, the WOW corrected wind speeds, and the corresponding KNMI real wind speeds, in the year 2018.

We use the statistics introduced in the previous section to measure the differences for evaluation. (i) The differences between WOW observed and KNMI true wind speeds are taken as the initial bias of WOW observations. (ii) The differences between WOW corrected and KNMI true wind speeds are also calculated to measure the new bias of WOW after empirical quantile mapping. (iii) The differences between WOW observed and WOW corrected wind speeds are computed to quantify the improvements by bias correction.

We show the values of RMSE (Figure 4.5) and the K-S statistics (Figure 4.6) between WOW and KNMI both before and after bias correction. The two figures show that the empirical quantile mapping method reduces the RMSE and K-S statistic to a much lower level. The bias correction with empirical quantile mapping calibrates large differences to a much lower and similar level, which implied that the corrected WOW wind speed could indicate true wind at the same height with KNMI. This is essential for spatial quality control in the next chapter that compares neighboring observations.

We also look at the cumulative distribution of the example WOW station in Figure 4.2 to investigate the performances of empirical quantile mapping. From the scatterplot in Figure 4.7, we find that the majority of the pairs of simultaneous WOW and KNMI wind speed are located around the diagonal after bias correction. This implies that a lot of WOW wind speeds are corrected to the true wind speeds. As shown in Figure 4.7, the cumulative density distribution of the WOW wind speeds has been fitted so that it has a similar shape and scaling with the KNMI wind speeds. From such results, we can conclude that the WOW wind speeds after bias correction are comparable with KNMI true wind speeds in the way that they indicate wind on similar levels.

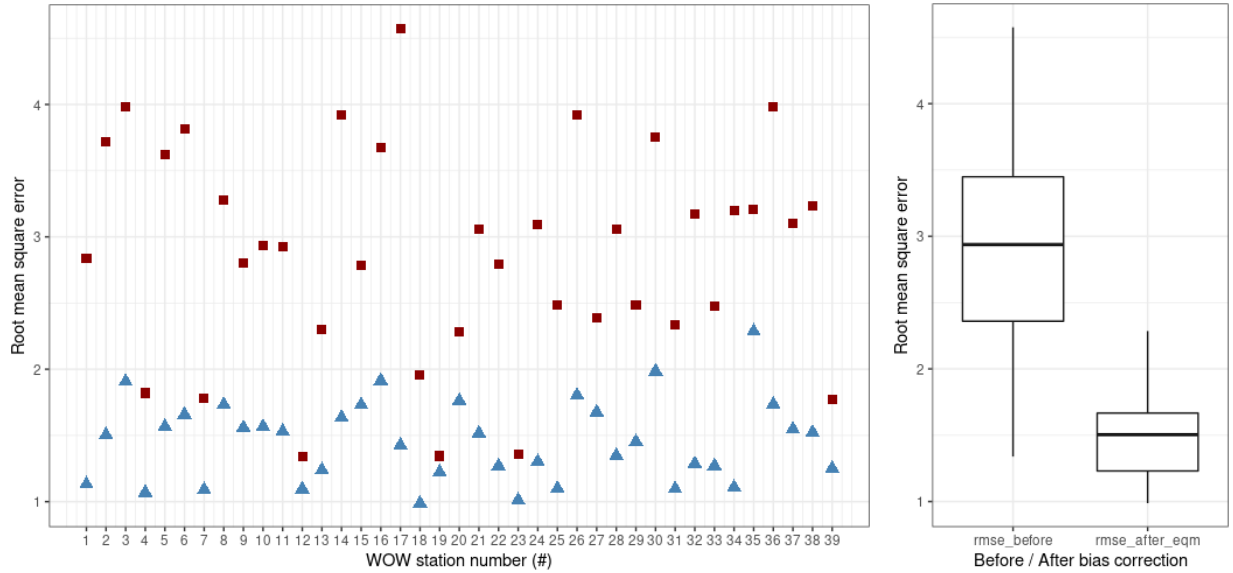


Figure 4.5: Root mean square errors of wind speeds between a WOW and a *relating* KNMI, before (red squares) and after (blue triangles) bias correction.

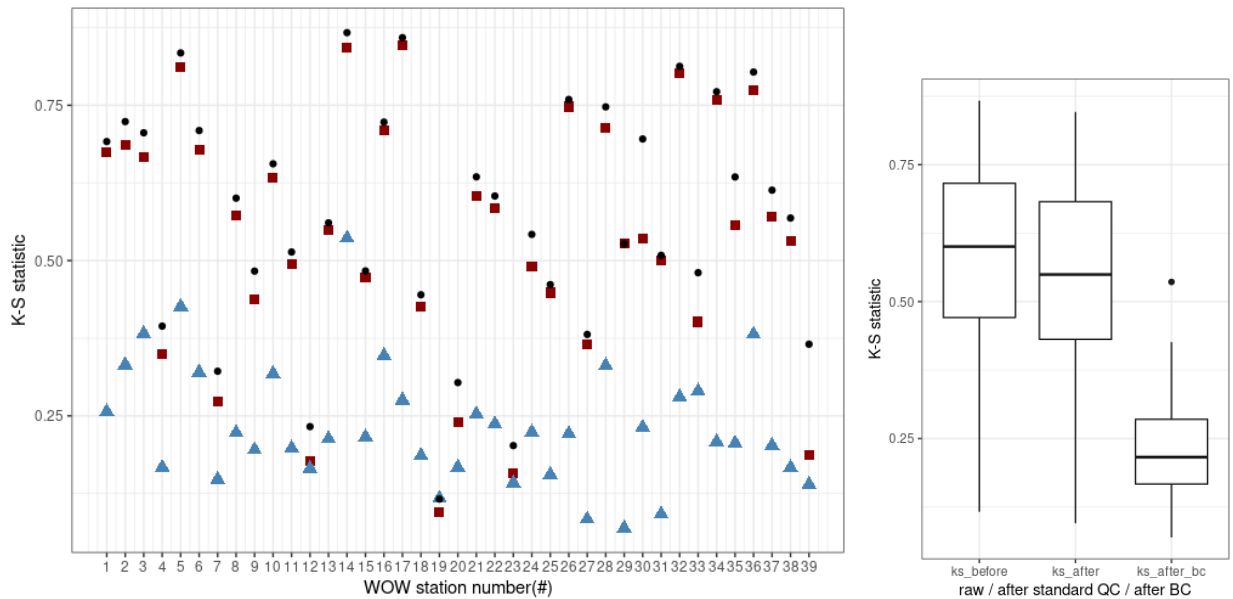


Figure 4.6: Kolmogorov–Smirnov statistic of wind speed distributions between a WOW and a *relating* KNMI, before (red squares) and after (blue triangles) bias correction, as well as before (black points) and after (red squares) standard quality control.

4.5 Discussion for future works

The empirical quantile mapping method for correcting the bias in WOW wind speeds works well from our results. However, this bias correction method still has some limitations that could be improved in future works. We list them in the following:

1. A significant problem in our WOW data set is that WOW stations record a lot of zero wind

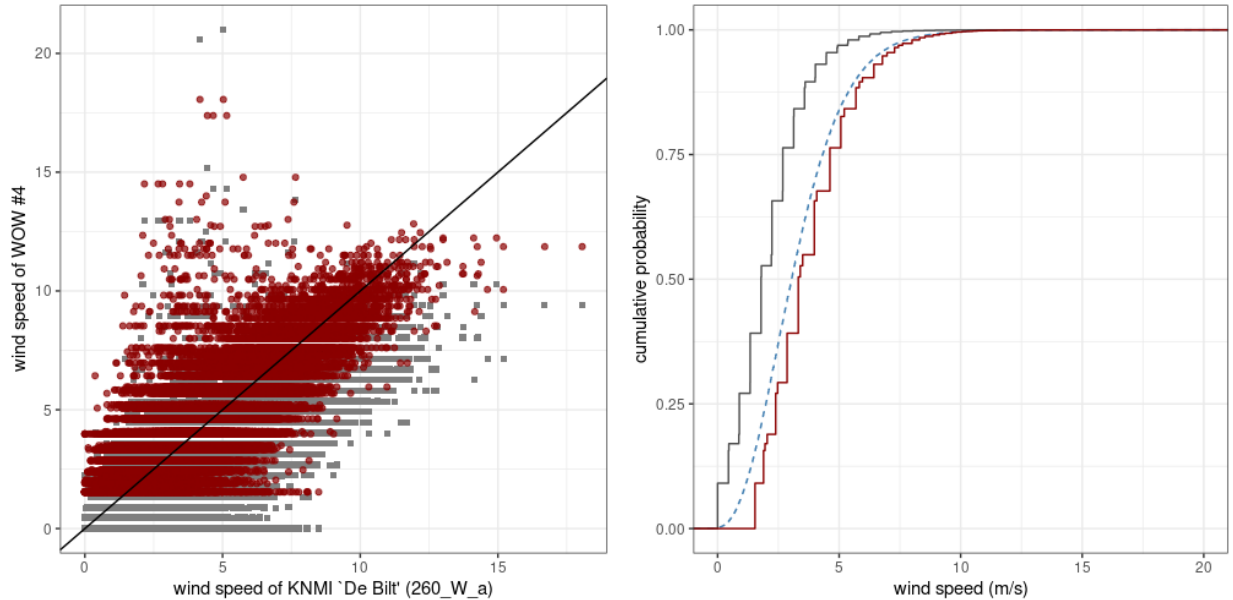


Figure 4.7: The scatterplot (left) of simultaneous wind speeds between a WOW station and the *relating* KNMI station, before (grey points) and after (red points) bias correction. The cumulative density distribution (right) of WOW wind speeds before (grey line) and after (red line) bias correction, and KNMI wind speed (blue dashed line).

speeds, while the true wind speeds from KNMI stations only have a few zero values. As we discussed before, WOW stations tend to record lower wind speeds than true wind. Therefore, when the true wind speeds are small, WOW stations tend to record values lower than the small values. Since wind speed records should be nonnegative, in such cases, WOW stations cannot record a negative value, and hence the observed wind speeds would be zero. When we are applying the bias correction, those zero wind speeds are all corrected to the same new value, which may not indicate the true wind properly.

To deal with such issues, it would be helpful to look at the behaviors of true wind speeds when WOW wind speeds are zero. If we can find some pattern in such behaviors, then we may take the zero WOW wind speeds cases for special consideration. In this way, we could construct a new bias correction method that considers two different cases; one is when WOW wind speeds are positive, and the other is when WOW wind speeds are zero.

2. As we have seen in Chapter 2.3, the wind has seasonal variations. The WOW stations may have different biases in winter and summer seasons due to the different atmospheric environments of wind sensors. Therefore, it might be a good idea to separately construct bias correction models for extended winter and summer seasons.
3. The only parametric nonlinear model that we discussed for bias correction is the power transformation. We have analyzed that it is not suitable for wind speed. However, there could exist other nonlinear models that could fit the wind speed bias correction well. Thus we may try some new models to do bias correction in the future; for instance, a piecewise polynomial or spline model may work, and it may also help correct zero wind speeds.
4. As we have analyzed in Chapter 4.2.3, the distribution mapping method works well for only a few WOW stations with high-quality. This suggests the question that do we have to use

the same bias correction method for all WOW stations. For the stations that well-fits a Weibull distribution, the distribution mapping method may work better than other methods like empirical quantile mapping.

Chapter 5

Spatial Quality Control

In Chapter 3, we implemented the standard quality control on the WOW data, which detects observations of implausible range or unrealistic variabilities with time. Standard quality control mainly checks the consistency of wind observations by a single station from the temporal point of view, while the spatial coherence for wind is rarely considered. After calibrating the bias of WOW data in Chapter 4, the corrected WOW data now indicates wind speed at 10 meters height, which means the wind speed observations from WOW and KNMI are more spatially comparable after bias correction. Therefore, we can assess the spatial consistency of WOW compared with KNMI. In this chapter, we develop methods to use spatial information in the quality control, so that we can detect inconsistent observations based on the records of their nearby stations.

The spatial quality control compares simultaneous wind speeds that are spatially distributed, and checks whether an observation (the target observation) from a WOW station (the candidate site) is abnormal compared with its surrounding wind speed records. This motivates the general steps of spatial quality control, first (i) selecting some set of nearby stations as a reference, then (ii) estimating a prediction interval for the target observation from reference records, and finally (iii) comparing the estimated interval with the target observation and determining if it is spatially consistent. These three steps lead to two main questions of spatial quality control, for a target observation, one issue is (1) to select comparable reference stations, the other is (2) to construct a valid estimate and then generate a feasible prediction interval for the target observation. We introduce our methods to solve the two questions in the following sections. After that, we provide the results of implementing spatial quality control on the WOW data and put forward discussions.

5.1 Selecting reference stations

Given a candidate site, the most direct way to choose its reference stations is to select its neighboring sites by the geographical distance. However, the closest station geographically might not be the best match for the candidate station, because of the following reasons.

- i) Local terrain and nearby obstacles can have a significant impact on wind speed measurement [Fiebrich et al., 2010]. If a CWS is placed on the roof near a tree, it is likely to record a much lower wind speed when the wind blows towards the direction where the tree can block the wind. This is also one of the reasons why a standard placement of CWS is essential. We have no guarantee that all the WOW stations are located in open areas without obstructions, and so two geographically close stations could have substantial distinctions in wind observations.

- ii) Wind behaves very differently in coastal and non-coastal areas [Gatey and Miller, 2007]. The closest station to a candidate inland station might be on the coast, and hence the wind speed distributions from those two sites would be quite different. The candidate station could have a stronger correlation with an inland station that is further away.
- iii) As we have discussed in Chapter 3.3, the low quality of WOW stations leads to large noise amplitudes of correlations between a pair of WOW stations. A nearby WOW station with low quality has a negative effect on estimating the true wind speed of the target observation.

The idea of choosing reference stations by the geographical distance does not fit our WOW data. This motivates us to seek better approaches. An alternative solution is to consider the similarity of wind observations between two stations rather than their distance.

We assume that wind speeds should be distributionally similar in the limited inland region where WOW stations located. Thus for two stations with similar wind speed distributions, it is more likely that they have comparable surroundings and measurement errors, and so we can expect the spatial consistency of observations from the two stations.

5.1.1 Pre-selecting reference stations

Previous researches have selected statistically similar reference stations for their spatial quality controls on the temperature data. Hubbard and Siva Kumar [2001] calculated the Pearson correlation between the candidate station and ten nearest neighboring stations, and set a minimum required value for the correlation. If more than five neighboring stations had correlations larger than that value, the five stations with the lowest RMSE (computed from the linear regression between the candidate station and its neighboring stations) were chosen as the reference stations. Hubbard et al. [2005] improved this reference station selection method when they proposed the new spatial weighted regression test for spatial quality control on temperature data. They first got the nearby stations within a specific radius, namely 50km in their study, and then for every target observation, they chose five stations with the lowest RMSE of linear regression [Hubbard et al., 2005]. Cheng et al. [2016] simplified the approach of Hubbard et al. [2005] so that they could obtain the results in real-time. They first got a list of nearby stations within a specific radius and then selected five stations with the highest correlation to the candidate station [Cheng et al., 2016].

We also take Pearson correlations into account when selecting reference stations for WOW. We recall from Chapter 3.2.1 that the Pearson correlation measures the linear relationship between two sequences of wind speed observations. The linear relationship matters in the estimation step of spatial quality control, as the idea is to find a prediction of the target observation based on its reference records, which is usually achieved by some interpolation methods, especially the weighted sum approaches.

We set a minimum required Pearson correlation threshold of 0.5 for the reference stations so that they could match the candidate station for estimation. Many WOW stations were active during only part of the three years 2016-2018 (as shown previously in Figure 2.4), and we need to make sure that for each target observation, there are enough valid reference records to do the estimation. Thus we set this minimum required Pearson correlation to a relatively low level such that enough reference stations (more than 12) are chosen. According to our analysis in Chapter 3.3 (also shown here in Figure 5.1), the Pearson correlation between each pair of WOW stations are lower than those between each pair of KNMI stations. Nevertheless, more than three-quarters paired WOW

stations hold a correlation higher than 0.5. Moreover, as shown in Figure 3.7 of Chapter 3.3, the Pearson correlation between a WOW station and the *relating* KNMI station are mostly larger than 0.5 with only one exception. We find that the WOW station (# 35) has low Pearson correlations (less than 0.5) with all other stations, which means it has poor linear relationships with others. We conclude that this station has a too low quality to consider, and so we remove it from our list to do spatial quality control.

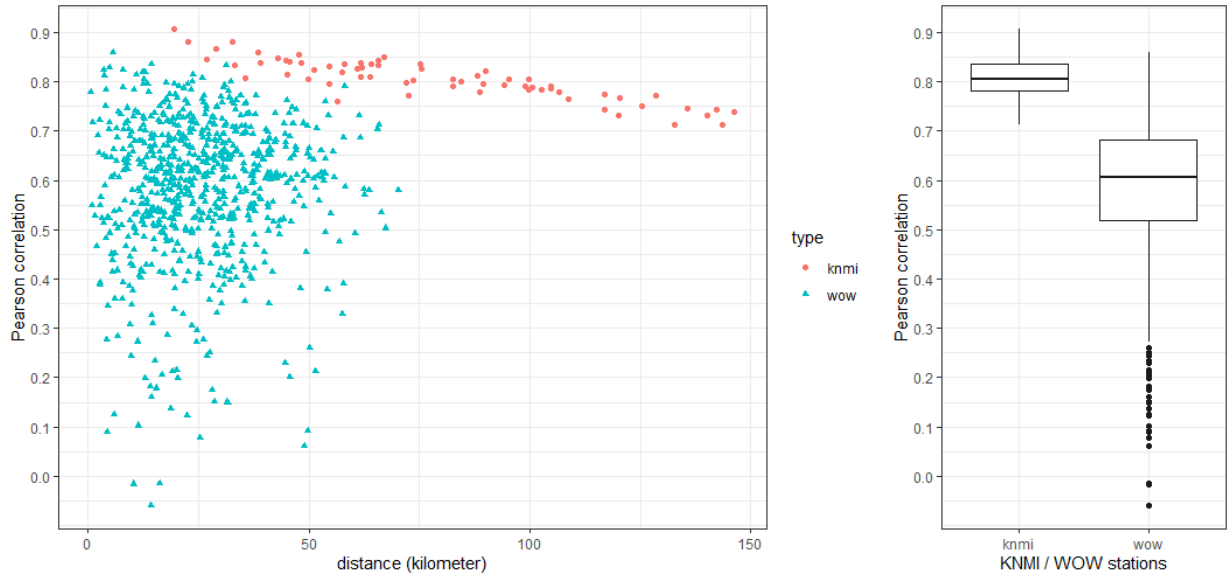


Figure 5.1: The same as Figure 3.8. Left: The relation between the distances and Pearson correlations of any two WOW stations (blue triangles) or any two KNMI stations (red points). Right: The boxplot of Pearson correlations for the left figure.

The geographical distance is not a major concern in our spatial quality control for WOW data. We do not set a limitation of radius to choose neighboring stations because the WOW stations in our data are all within 75 kilometers to each other (see Figure 5.1), and the descending trend of Pearson correlation with distance is not clear for WOW stations. The figure shows that Pearson correlations between two WOW stations vary a lot ranging from -0.1 to 0.9, and they do not drop exactly with distance. The correlations between two KNMI stations did show the trend of reducing with larger distance, but the correlations within 150 kilometers are still high compared with WOW. This also supports our former assumption that wind behaves homogeneous in the limited region. Therefore, we do not set a maximum distance threshold for the spatial quality control of our WOW data. However, we should be aware that this only suits the situation when the stations are not far away from each other. When dealing with citizen weather stations located in a more spread area, it is necessary to set at least a high limit of distance to make sure the reference stations are not too far away so that they do not have substantial climatological differences.

For a candidate WOW station, the pre-selected list of possible reference stations is chosen by stations that have a Pearson correlation larger than 0.5 with the candidate site. The list contains at least 12 sites for further selection. We do not directly select reference stations with the highest Pearson correlation because of the two reasons: (1) The correlation only measures the linear relationship between two sequences of variables, not the gaps between them; we cannot say

two sequences with high correlation must have very similar distributions. (2) It is also a vital issue to assign weights for reference observations to get the estimate; since the Pearson correlation is not a distance metric, it is not suitable to determine weights based on correlation coefficients. Thus in our spatial quality control, we attempt a new method that using the Wasserstein distance to determine reference stations, which is introduced in the following section.

5.1.2 Determining reference stations by Wasserstein distance

After pre-selecting a list of neighboring stations that have some linear relations with the candidate station, we want to choose the best reference stations from the list. Hubbard et al. [2005] suggested to compute the RMSE of linear regression in a certain time window between each neighboring station with the candidate station for every target observation. This approach selects various reference stations for a target observation at different times, and so it could provide a more accurate reference data for estimation. Cheng et al. [2016] discussed the approach from Hubbard et al. [2005], and they concluded that it is not suitable for real-time spatial quality control as it takes a long time to compute. In spatial checks, the main concern is to figure out inconsistent observations in data, and thus an estimation that is sufficient for comparison is not necessarily the optimal estimation of spatial interpolation. Cheng et al. [2016] directly selected the five stations with the highest correlation to simplify the approach and make sure the processing time is short to meet the requirements of real-time quality control.

In this section, we introduce a new mathematical distance concept, the Wasserstein distance, that measure the differences between two frequency distributions. We then use this Wasserstein distance to determine which pre-selected stations are closely distributed enough with the candidate station on wind speed, and finally generate a reference stations list for every WOW stations to do estimation afterward.

Wasserstein distance

The concept of Wasserstein distance was first introduced in the articles of Wasserstein and Dornbushin, as a characteristic of proximity of two probability measures. After that, the Wasserstein distance was further developed by Kantorovich in his study of finding an optimal solution of the transportation problems, and so it is also called the transportation distance [Kantorovich, 1960]. The transportation distance is computed as the minimum effort of transferring a density distribution into another. As shown in Figure 5.2, suppose the cost of transfer a unit area from the blue distribution to the red distribution is relevant to the length of routine; the transportation distance is the approach to transform the blue distribution to match the red distribution that has the lowest total costs. Moeckel and Murray [1997] applied this transportation distance to measure how closely the long-term behaviors of two time series resemble each other. This motivates us to consider the Wasserstein distance between two sequences of wind speed data. Since our data are recorded in the three years 2016-2018, it can be treated as a long-term behavior of wind speed in different stations. Therefore, the Wasserstein distance could help us quantify the distance between probability distributions of wind speed from two stations, which serves as a complement after considering correlations.

Vallender [1974] calculated the Wasserstein distance between two probability distributions on the line, which is also treated as the one-dimensional case. For two discrete sequences of data, we first need to generate its corresponding empirical distribution based on these discrete data points,

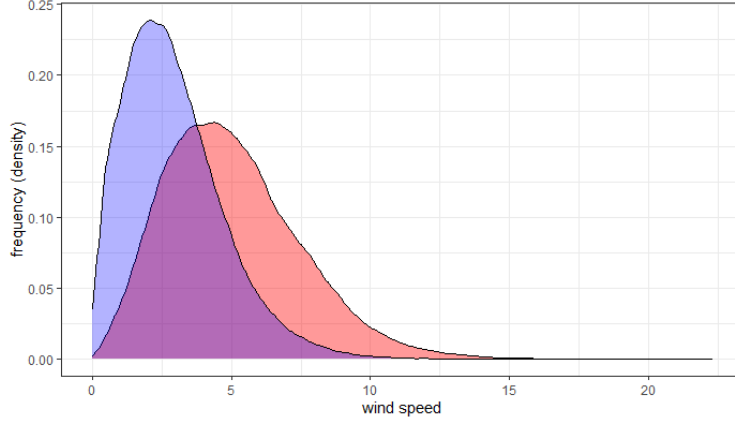


Figure 5.2: An example of transportation distance (Wasserstein distance / Earth mover’s distance).

and then calculate the one-dimensional Wasserstein distance between two empirical distributions. For this work, we only require the one-dimensional Wasserstein distance, but the details of its formulation in higher dimensions see the literature by Muskulus and Verduyn-Lunel [2011] and Panaretos and Zemel [2019].

Let X and Y be two random variables, $F_X(\cdot)$ and $F_Y(\cdot)$ are the cumulative density functions of X and Y respectively. Then the one-dimensional Wasserstein distance between X and Y are calculated by

$$W_p(X, Y) = \left(\int_0^1 |F_X^{-1}(z) - F_Y^{-1}(z)|^p dz \right)^{1/p}, \quad (5.1)$$

where $p \in \mathbb{N}$ indicates the dimension of the distance. When $p = 1$, the one-dimensional distance is also referred to as the Earth mover’s distance.

If $X = \{x_i\}_{i=1}^n$ and $Y = \{y_i\}_{i=1}^n$ are two discrete random variables with the same length, then the Wasserstein distance in discrete form can be simply estimated by

$$W_p(X, Y) = \left(\sum_{i=1}^n |x^{(i)} - y^{(i)}|^p \right)^{1/p}, \quad (5.2)$$

where $x^{(i)}$ and $y^{(i)}$ are the order statistics of X and Y respectively.

In our study, we are only concerned with the distance between two distributions of wind speeds, which is the minimum cost of turning one distribution into another. Therefore, we compute the Earth mover’s distance, the L-1 Wasserstein distance when $p = 1$, to determine whether a station is close enough in wind speed distribution with the candidate station. For two wind speed time series of a neighboring station $X = \{x_t\}_{t \in \mathcal{T}'}$ and a candidate station $Y = \{y_t\}_{t \in \mathcal{T}'}$, we consider a pairwise complete data that only takes the valid observations at the same overlapping period $|\mathcal{T}'|$. Then we can calculate the Earth mover’s distance between X and Y directly by

$$W_1(X, Y) = \sum_{t \in \mathcal{T}'} |x^{(t)} - y^{(t)}|, \quad (5.3)$$

where $x^{(t)}$ and $y^{(t)}$ are the order statistics of X and Y . We make assumptions here that the pairwise complete period is representative for the whole three years 2016-2018, and hence the Earth mover’s

distance computed from that period can be treated as the distance of wind speed distributions between two stations.

An important property of the Earth mover’s distance is that it is a mathematical metric. The mathematical definition of a *metric* (distance function) is a function that maps each pair of elements in the domain to non-negative real numbers, i.e. $d : X \times X \rightarrow [0, \infty)$, and for any $x, y, z \in X$ satisfying [NIST/SEMATECH, 2020]:

1. (identity) $d(x, y) = 0 \iff x = y$;
2. (symmetry) $d(x, y) = d(y, x)$;
3. (triangle inequality or subadditivity) $d(x, y) + d(y, z) \geq d(x, z)$.

A mathematical metric can be treated as a distance function in a space, a simple example for a metric is the physical distance on Euclidean space. If we are given the physical distances between every two points from a list of points, we can somehow reconstruct their relative locations on a plane surface. This is another reason why we choose the Wasserstein distance to select the references station. Given a matrix of distances between every two stations, we can reconstruct the relative locations in the plane so that we can have a more intuitive view of the statistical similarities between stations.

Multidimensional scaling

We reconstruct the relative locations based on the distance matrix by multidimensional scaling, which generally represent a list of techniques that model the distances as points in a Euclidean space [Muskulus and Verduyn-Lunel, 2011]. The input of the MDS should be the Earth mover’s distances between every two stations, and the output would be a list of paired coordinates that indicate the relative location of each station.

The method we applied here is the classical MDS, which is similar to the principal component analysis in the multivariate statistical analysis [Cox and Cox, 2008]. Given a matrix of distances, where the value on each row and column represents the distance between the two distributions with the row and the column numbers, the idea of classic MDS is to follow some ‘double centering’ transformations and perform ‘eigendecomposition’ after that to get the vector of coordinates [Muskulus and Verduyn-Lunel, 2011]. We have converted the Wasserstein distance into a two-dimensional distance metric space using the package ‘stats’ in R [R Core Team, 2013]. To learn more about multidimensional scaling methods, we refer to the book by Cox and Cox [2008].

After implementing the MDS on our Earth mover’s distance matrix, we get the relative locations of all WOW and KNMI stations in a new metric space, that we call it the Earth mover’s distance (EMD) space. We compare the actual locations of stations on a map with the relative locations based on distribution distances of wind speeds in Figure 5.3, and obtain a very different outcome. The origin point of the EMD space is determined by the mean values of coordinates of all these points, and so the points surround the origin in the metric space. Many stations are located near each other, which implies they are statistically similar in wind speed distributions, and this also supports our assumption that the wind speeds are distributionally similar in a limited region. There are some WOW stations that stay isolated and far from other sites in the EMD space, such as WOW #20, #35, #39, which suggests that those stations have large differences with other common WOW stations on wind speed distributions. A reason for those stations could be the severe surrounding

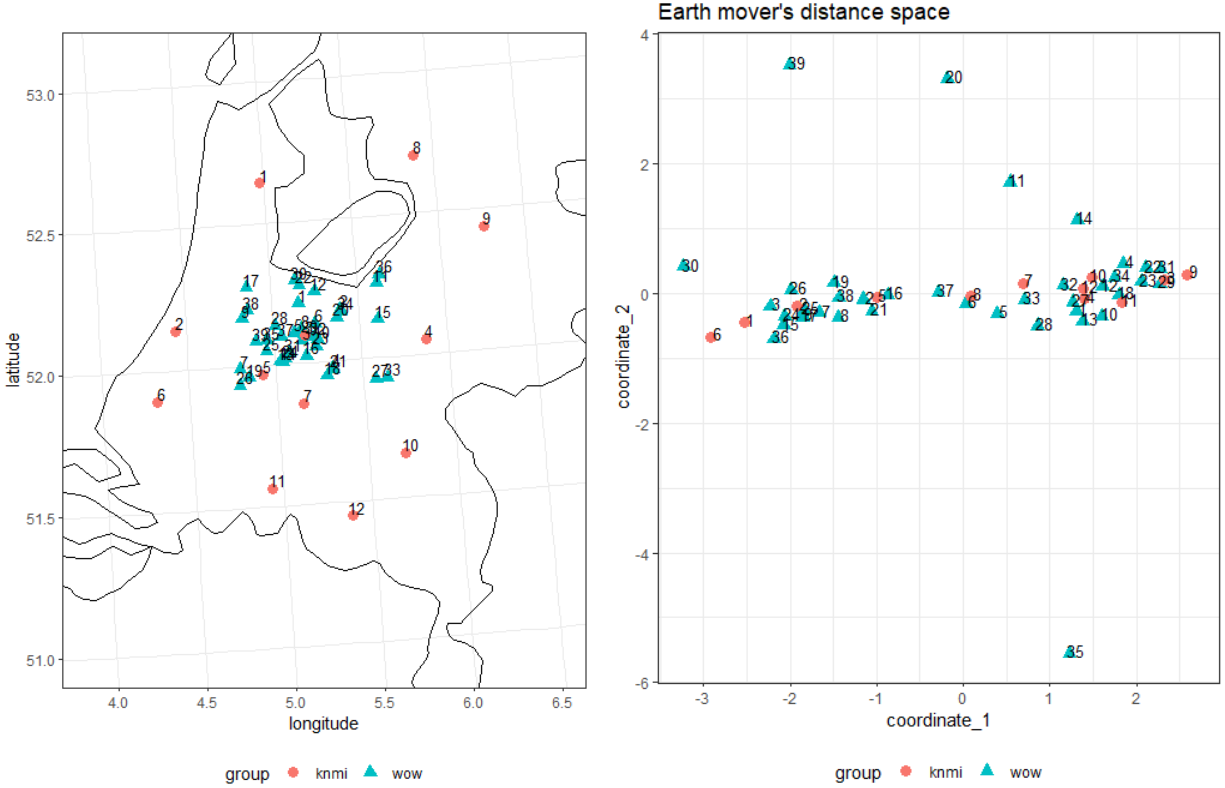


Figure 5.3: Comparing relative locations of both WOW and KNMI stations, on a map in Euclidean space (left), and on the Earth mover’s distance metric space (right), with both KNMI stations (red points) and WOW stations (blue triangles).

obstructions result in poor quality on wind speed observations. We recall that WOW #35 is the one that has too low quality that we exclude it from the data. The location of WOW #35 in the EMD space also suggests it has the most considerable difference in wind speed distribution with other stations.

5.1.3 Selection procedure of reference stations

We summarize our procedure to select reference stations as follows. For a given candidate station, we

- calculate the Pearson correlation on wind speed observations between each neighboring station and the candidate station, and obtain a pre-selected list of stations that hold a correlation higher than 0.5;
- compute the Earth mover’s distance on wind speed distributions between every station from the pre-selected list and the candidate station, then order the stations in the list by the increasing order of Earth mover’s distances;
- take the first six stations of the list as reference stations.

Hubbard et al. [2005] selected five reference stations for their spatial quality control on temperatures. After that Hubbard and You [2005] implemented the sensitivity test to discuss an optimal

number of reference stations. They have found that for their spatial regression test method, when reference stations are more than ten, the estimates do not change too much, while when there are less than ten stations, the estimates vary with the number of references. You and Hubbard [2006] suggested that the number of neighbor stations should be four to six. If there are fewer than three neighbors, the estimation calculated by neighbor observations is not sufficient. On the other hand, if there are too many neighbor observations, the computing cost would increase. This is not worthwhile when six neighbors can achieve the same goal. Cheng et al. [2016] used five reference stations, and they mentioned that at least three reference observations must be available for estimating.

We select six stations from the pre-selected list that has the lowest Earth mover’s distance with the candidate station, in order to make sure there are at least four valid reference observations for each target even if one-third of the records are missing. We summarized and followed the advice from Cheng et al. [2016], although their research is focused on temperatures, we assume that this is also reasonable for wind. However, it is necessary to do a sensitivity test to determine an optimal number of reference stations for wind speeds in future work.

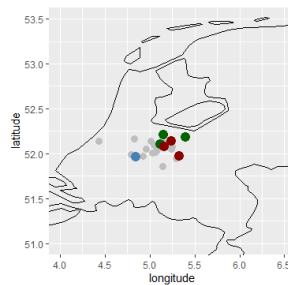


Figure 5.4: A candidate WOW station #19 (blue points), its reference stations (KNMI: green points, WOW: red points), and nearby stations (grey points) in the pre-selected list.

Figure 5.4 provides an example of the selected reference stations by a candidate station. We can see that the nearest stations to the candidate station are not chosen. A possible reason is that its nearby station has less valid simultaneous observations, which could cause a larger EMD, while some further stations have enough valid records and so the EMD estimate better.

For some target observations, there are not enough reference records with valid data. In such a situation, we flag the observation as isolated and treat it as passing the spatial quality control test. We may look back at those isolated observations if we want to go into more detail in analyzing localized wind behaviors.

The selection steps provide us a list of reference stations that not only have some linear relationship with the candidate station, but also have relatively small differences in wind speeds distributions. The observations from these reference stations form a solid basis for estimating the target record. In the next part, we develop three different ways to generate the estimation as well as the prediction interval for the target.

5.2 Spatial quality control methods

We introduce three methods to estimate the target observation here, following from the ideas of spatial quality controls on temperatures. We discuss the adaptations on wind speed data and com-

pare the approaches of different methods.

To simplify the notation, in this section, we use y_t to represent the candidate observation at time point t , and $x_{n,t}$ to represent the reference observation of reference station n at time point t , where the set of reference stations is denoted \mathcal{N} . The target estimate from a set of reference observations $\{x_{n,t} : n \in \mathcal{N}\}$ at time point t is expressed as \hat{y}_t .

5.2.1 Inverse distance weighting

The essence of inverse distance weighting (IDW) is to compute a weighted sum of the data, where the weights are generated using an inverse relation with distance. Peterson et al. [1998] tried this method in an early study of spatial interpolation methods on monthly temperature data. Estévez et al. [2011] also applied the IDW with two different ways of setting the weights, the Barnes method [Barnes, 1964] and the Cressman method [Cressman, 1959]. The supporting idea of IDW is that further stations behave less similar to the candidate station on weather observations. This idea suits for temperature observations (because they are not likely to be affected by surroundings unless you put the thermometer in a fridge or near the heating system, which citizen scientists have done in the past), and the temperature is smooth spatially.

In IDW, the estimate of target observation \hat{y}_t is given by

$$\hat{y}_t = \frac{\sum_{n \in \mathcal{N}} \omega_n x_{n,t}}{\sum_{n \in \mathcal{N}} \omega_n}, \quad (5.4)$$

where ω_n is the weight for reference station n . Let n_0 be the candidate station, then the Cressman method to compute weights is expressed as

$$\omega_n = \frac{r^2 - d^2(n_0, n)}{r^2 + d^2(n_0, n)}, \quad (5.5)$$

while the Barnes method computes weights in the way

$$\omega_n = \exp\left(-\frac{d^2(n_0, n)}{2r^2}\right), \quad (5.6)$$

where r is the radius chosen for selecting neighbors, and d is the geographical distance between the candidate station n_0 and a neighboring station n , where $n \in \mathcal{N}$. Note that r should always larger than $\max_{n \in \mathcal{N}}\{d(n_0, n)\}$ to ensure the weights are positive real numbers.

Fiebrich et al. [2010] suggested a confidence interval centered at the estimate \hat{y}_t with a half length of twice the standard deviation σ of reference observations. This means the wind speed observation x_t pass the spatial test if

$$x_t \in [\hat{y}_t - 2 \cdot \sigma_t, \hat{y}_t + 2 \cdot \sigma_t]. \quad (5.7)$$

The standard deviation σ_t of reference observations $x_{n,t}$ is estimated by

$$\sigma_t = \sqrt{\frac{1}{|\mathcal{N}|} \sum_{n \in \mathcal{N}} (x_{n,t} - \bar{x}_t)^2}, \quad \text{where } \bar{x}_t = \frac{1}{|\mathcal{N}|} \sum_{n \in \mathcal{N}} x_{n,t}.$$

If we assume that the estimate is the expected (mean) value of the observations, and the random errors of the observations are normally distributed, then this confidence interval with a twice

standard deviation radius implies that the probability that observation lies in this interval is about 95.45%, which is sufficient to determine whether to flag a target observation or not. Noting that if the left side of this prediction interval is less than zero, we take the nonnegative part of the interval for flagging.

However, the direct application of the IDW method is not suitable for estimate wind speed from neighboring observations. Unlike temperature, wind observations are sensitive to the settings and placements of sensors, and so the relation that decays with geographical distance does not fit for wind speed. As we have discussed earlier in Figure 5.1, the decrease in correlation does not correspond directly to a decrease in distance. Therefore, weights selection based on inverse distances is unlikely to be informative. This means the nearest station might not be the one that matches the candidate station well, and the classic IDW method does not fit for wind speed data.

IDW with Earth mover's distance

To improve classic IDW, we need to use a distance metric that indicates the similarity between two stations. The Earth mover's distance introduced in Chapter 5.1.2 is a nice choice, as it is a distance metric that indicates the distance between two distributions. We can therefore replace the geographical distance by Earth mover's distance in the IDW estimation. In this way, we come up with an improved IDW using EMD as the distance. The estimate formula stays the same as in Equation (5.4), as well as the weights in Equations (5.5, 5.6). The distance radius r will change to a new value, which is chosen as the highest Earth mover's distance in the pre-selected stations' list. According to our analysis, a station with the lowest Earth mover's distance with the candidate station is highly comparable, and it is reasonable that we put a larger weight on its observation when estimating.

We apply the Cressman method shown in (5.5) to compute weights in our IDW implementation. And we take $r = 3$ as the threshold of the Earth mover's distance when applying the inverse distance weighting, as it is the least integer that is larger than all Earth mover's distances by the reference stations. It would be interesting to compare the performances of Cressman and Barnes weighting methods for further studies. Furthermore, a sensitivity test that investigates the effect of choosing different r values could be taken to complement this method in the future.

The advantages of this improved IDW method are that (i) it is simple to compute, the weights for each reference stations is fixed, and hence we would get results extremely fast that perfectly meet the requirements of real-time processing, and (ii) compared with classical IDW, the estimates depend on the statistical similarities of reference stations rather than physical distances, and so the quality control flagging results are more reliable.

However, the IDW with EMD method mainly considers the distributional similarities between stations, while it lacks developing the linear regression relations on the data that might provide more information and hence generate a more credible prediction interval. This motivates us to find other approaches for the estimate that make full use of the data. An alternate way is called the spatial weighted regression test, which again was initially developed for temperature spatial quality controls. We discuss this test below that involves linear regression in a time window around our target observation.

5.2.2 Spatial weighted regression test

The spatial weighted regression test (SRT) is first introduced by Hubbard et al. [2005] on the spatial quality control of temperature data, and this method is widely used by other researchers in quality controls [You et al., 2008] [Estévez et al., 2018] [Cerlini et al., 2020]. The idea of this method is based on the linear relation between each reference station with the candidate station, which is different from IDW. In IDW, we estimate an observation from its simultaneous reference observations; While in SRT, we first get estimations from each reference site individually, and then combine those estimations to get a final estimation.

Hubbard et al. [2005] choose neighboring stations by specifying a radius, and they take a time window that has 24 past observations of daily temperature data. They perform a linear regression between the candidate station and each of the neighboring sites, based on the 24 paired past records. The five stations with best linear regression fittings are selected as reference stations, and are assigned weights based on their standard errors of linear regressions in the time window. After obtaining the linear regression by each neighboring station, they use this linear model to estimate the current target observation based on the root mean square error of the reference site. In this way, they got five estimates of the target observation from different reference stations. The weighted sum of those five estimates is the final estimate of the target observation.

The reasons why Hubbard et al. [2005] apply the SRT method on temperature data are mainly because of the following reasons: (i) The temperature observations from two nearby stations have some linear relation; (ii) They used the daily maximum/minimum temperature data in their research, and so 24 past observations mean they applied the linear regression on the temperatures during the past 24 days, this is a long time window and hence be comparable with overall linear relation; (iii) They assumed the reference station that has the least error of linear regression with the candidate site within the certain time window can be treated as the best match reference station, and so it is reasonable to assign more weights on its estimation.

For wind speed observations, there are some linear relations between two nearby stations, as we have discussed in the Pearson correlation part (Chapter 3.3). However, our wind speed data are recorded every ten minutes; if we still take 24 days as our time window, the data points would be too many to consider. Therefore, we develop a sensitivity test to determine an optimal choice for the length of the time window. After choosing the time window, we can apply similar ideas on wind speed observations and assume that the reference station with the lowest error of linear regression is the most comparable reference station in that specific time window.

Figure 5.5 helps to explain the intuition behind SRT. For each neighbor station in the pre-selected list, we consider the wind speed observations in a certain past time window (finding the length is discussed afterward) and apply a linear regression between the candidate and reference wind speed records during that time window. After getting the linear model as the blue line shown in Figure 5.5, we take the current observation from the reference station and get the corresponding estimate of target observation by the linear model. The vertical black line in Figure 5.5 is the value of current observation from the reference station, and the y -coordinate of the intersection by the black and blue lines represent the estimate of the target from this reference station.

For each neighboring station $n \in \mathcal{N}$, we obtain an estimate for the linear regression parameters α_n and β_n in (5.8), from observations within the time window of length δ . Then the estimate of y_t

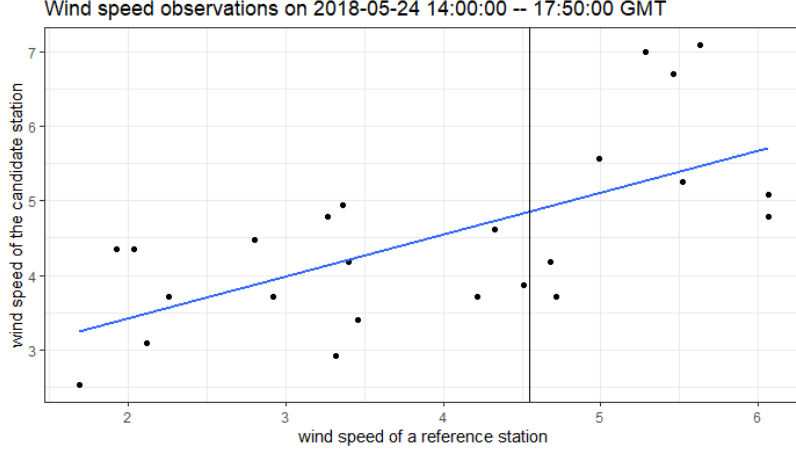


Figure 5.5: An example of linear regression (blue line) in SRT method, with a vertical (black) line showing the current reference observation.

based on individual reference observation $x_{n,t}$ is given by

$$\hat{y}_{n,t} = \alpha_n + \beta_n \cdot x_{n,t}. \quad (5.8)$$

The standard error of this linear regression is calculated from the paired list $\{(y_\tau, x_{n,\tau})\}_{\tau \in [t-\delta, t]}$, and can be written as

$$e_{n,t} = \sqrt{\frac{1}{|\delta|} \sum_{\tau \in [t-\delta, t]} (y_\tau - (\alpha_n + \beta_n x_{n,\tau}))^2}. \quad (5.9)$$

Next, we calculate a weighted sum of such estimations $\hat{y}_{n,t}$ and corresponding errors $e_{n,t}$ to get the final estimation \hat{y}_t ,

$$\hat{y}_t = \sqrt{\frac{\sum_{n \in \mathcal{N}} \hat{y}_{n,t}^2 e_{n,t}^{-2}}{\sum_{n \in \mathcal{N}} e_{n,t}^{-2}}}. \quad (5.10)$$

The weighted standard error of estimate \hat{e}_t has a similar formula with the standard deviation σ in (5.6), and we can calculate \hat{e}_t by

$$\hat{e}_t = \left(\frac{1}{|\mathcal{N}|} \sum_{n \in \mathcal{N}} e_{n,t}^{-2} \right)^{-\frac{1}{2}}. \quad (5.11)$$

Finally we construct a confidence interval for the target observation centered at the final estimate with twice weighted standard error as radius, similar to what we did in IDW. This means an observation y_t will pass this spatial regression test if

$$y_t \in [\hat{y}_t - 2 \cdot \hat{e}_t, \hat{y}_t + 2 \cdot \hat{e}_t]. \quad (5.12)$$

If the left side of this prediction interval is less than zero, we take the nonnegative part for flagging spatially inconsistent observations.

To make it clear, in IDW method, each reference record uses its own value as its estimate for the target observation, and then we calculate a weighted sum of those estimates; While in SRT method, each reference site use its past records to estimate the target observation, then we also calculate a weighted sum of those estimates.

Choosing an optimal δ for wind speed:

To find a proper value for the length of the time window in the spatial regression test, we develop this test to check the performances of linear regression for different time lengths and find the best one. We expect a time window of several hours rather than several days because (i) wind changes rapidly compared to other weather variables, and hence the reference station that matches with the best linear regression could vary in a short period; (ii) our wind speed data is recorded every ten minutes, the number of observations in several hours is enough to apply linear regression. Therefore, we test the possible choices of time length ranging from 4 hours to 24 hours, so that we have enough data points to do linear regression (no less than 24) and not a long period for the rapidly changing wind (no more than one day).

We choose a representative KNMI station ‘De Bilt’ and a WOW station (# 29) to apply linear regression at different times with the various length of time windows. For each time point, we do the linear regression with a time window of 4 to 24 hours (increasing by one hour each), and find the length such that the linear regression has the least square error. We test this for randomly picked 5000 time-points, and draw the frequency figure that shows which time length is most likely to take the best linear regression in Figure 5.6. Our test shows that the linear regression performs the best when the time window is 4 hours, then with the growth of time length, the linear regression performance goes down until when the time length is near 24 hours, the performance goes up again.

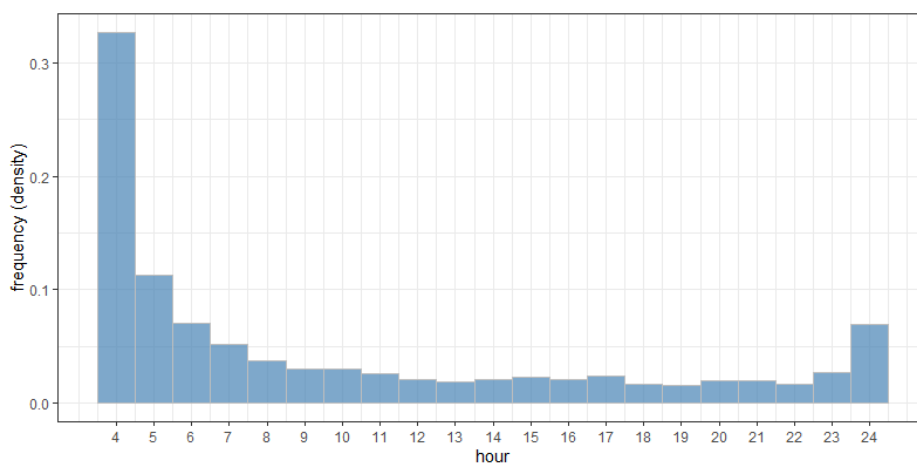


Figure 5.6: The frequency of different time lengths that achieves the least error of linear regression.

This suggest us choose the time window of 4 hours, which means we apply the linear regression in 24 pairs of simultaneous wind speed observations.

The highlight of the SRT method is the selection of reference station changes at different dates and times. If a reference station is distributionally close to the candidate station, but it does not record wind speed correctly due to some stuck issues during a certain period. Then the prediction interval could have large errors and hence influence the flagging results if we still assign large weights on this reference station. The SRT could avoid such a problem by changing reference stations, while the IDW with EMD method cannot.

Nevertheless, the SRT also has shortages. The computation procedures of SRT are too complicated that it takes much more time in processing. The method selects different reference records

for each target observation and applies linear regression for each one, which requires much more computation than fixed weights on several reference observations. For our spatial quality control study that is desired for a real-time application, it might not be practical for implementing. We try both SRT and IDW with EMD methods in Chapter 5.3, and discuss how much improvement the SRT brings.

5.3 Spatial quality control results

In this section, we implement two methods for spatial quality control on wind speeds, the inverse distance weighting with Earth mover’s distance (IDW) and the spatial regression test (SRT). We first compare the percentage of failed observations for both IDW and SRT in Table 5.1, and it shows that IDW flagged more spatially inconsistent observations than SRT. This might because some linear regressions in the four-hour time window have large errors such that the corresponding prediction intervals are too lax for the flagging conditions.

Spatial quality control method	IDW	SRT
Failed percentage	10.86%	6.34%
overall pass percent after all quality controls	75.85%	79.70%

Table 5.1: The failed percentage in spatial quality control by different methods.

To look at how the spatial quality control improves WOW data, we calculate the root mean square error (see Chapter 4.2.3), the Pearson correlation (see Chapter 3.2.1), and the K-S statistic (see Chapter 3.2.2) to quantify the performances. By comparing the RMSE after different steps of our quality control in Figure 5.7, we find that the spatial quality control reduces the error between WOW and KNMI stations, and IDE performs better than SRT. By looking at the changes in Pearson correlation (Figure 5.8), we see that the spatial quality control increases the correlation with KNMI. It is exciting that IDW filtered WOW wind speeds have correlations with KNMI larger than 0.7, which is a promising outcome on the quality of data. We also compared the K-S statistic after spatial quality control in Figure 5.9, but it shows little improvements compared to the K-S statistic after bias correction. One reason for this is that spatial quality control does not change the overall shape of wind speed distributions, and hence leads to a tiny change in K-S statistics, unlike bias correction that calibrated the distributions. As discussed at the beginning of Chapter 5, We get an estimate for every target observation to be compared, and these estimations should have a strong relation with observations. For each WOW station, we compare the time series of its estimates with the time series of its observations. The Pearson correlations for each WOW station are shown in Figure 5.10, where we try both IDW estimates and SRT estimates. The results imply that the IDW estimates match the actual observations stronger than the SRT estimates, which also suggests that the IDW method is better.

We conclude from these overall performances that IDW is more suitable for our WOW wind speeds than SRT. The possible explanation for this might be that the linear relation of neighboring simultaneous wind speeds during a short period of time is not strong enough to do linear regression. This leads to the large error of linear regressions, and thus leads to a not satisfactory enough flagging results. Another reason why IDW works well on the WOW data should be that we involve the Earth mover’s distance in selecting reference stations and assigning weights, which choose reference stations that are statistically similar to a larger weight in computing the estimate. These

findings show how mathematical concepts could be used in practical works and achieve surprisingly outcomes.

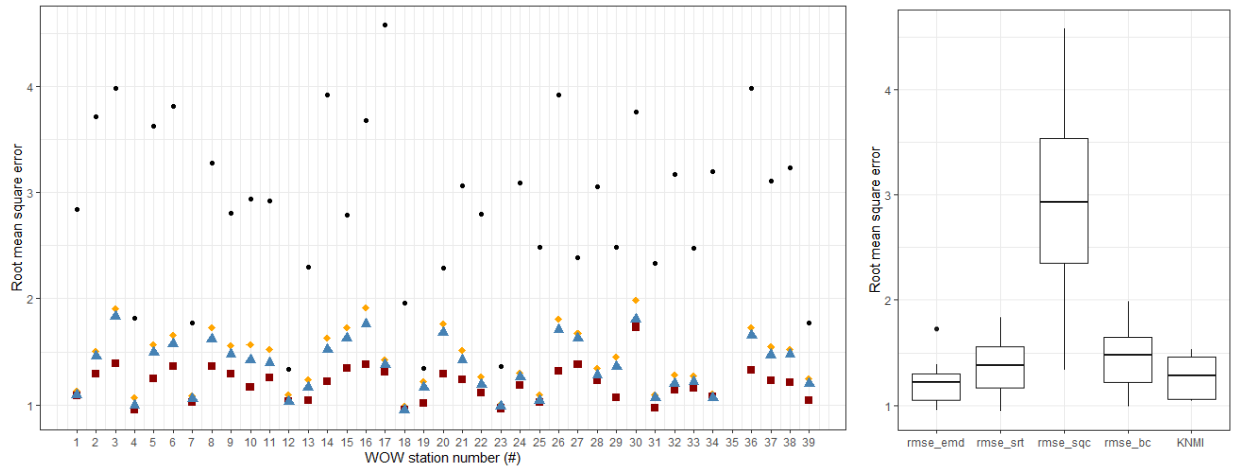


Figure 5.7: Root mean square error between a WOW station and its *relating* KNMI station, after standard quality control (black points, ‘rmse_sqc’), after bias correction (orange diamonds, ‘rmse_bc’), after IDW (red squares, ‘rmse_emd’), after SRT (blue triangles, ‘rmse_srt’), with boxplot. The RMSE between two nearby KNMI stations are also draw on ‘KNMI’ of the boxplot as a reference.

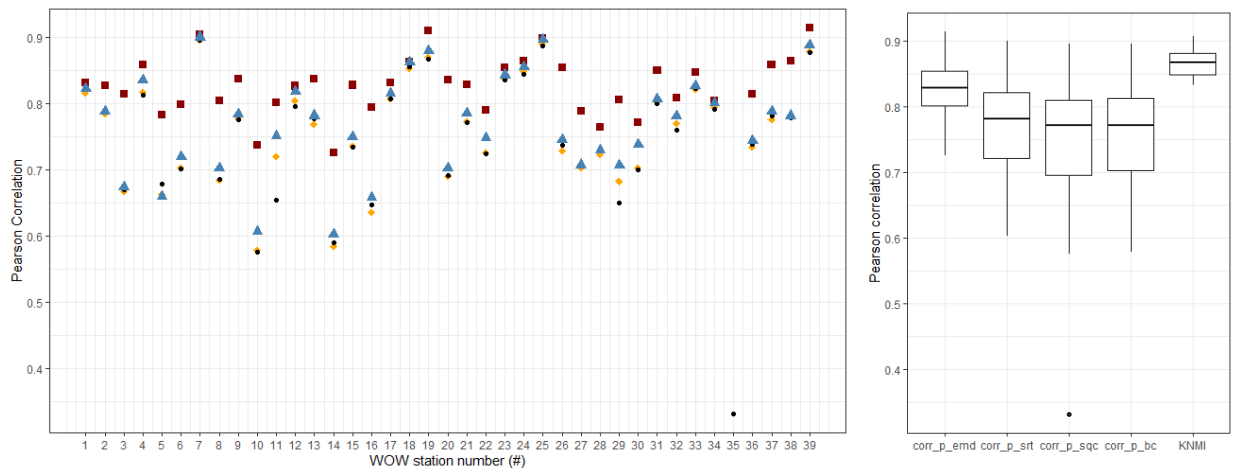


Figure 5.8: Pearson correlation between a WOW station and its *relating* KNMI station, after standard quality control (black points, ‘corr_p_sqc’), after bias correction (orange diamonds, ‘corr_p_bc’), after IDW (red squares, ‘corr_p_emd’), after SRT (blue triangles, ‘corr_p_srt’), with boxplot. The Pearson correlation between two nearby KNMI stations are also draw on ‘KNMI’ of the boxplot as a reference.

5.4 Discussion on other feasible directions

The three methods introduced in this chapter concern the relations on wind speeds between stations, but they did not investigate the distribution of estimates. Gibson et al. [2004] and Daly et al.

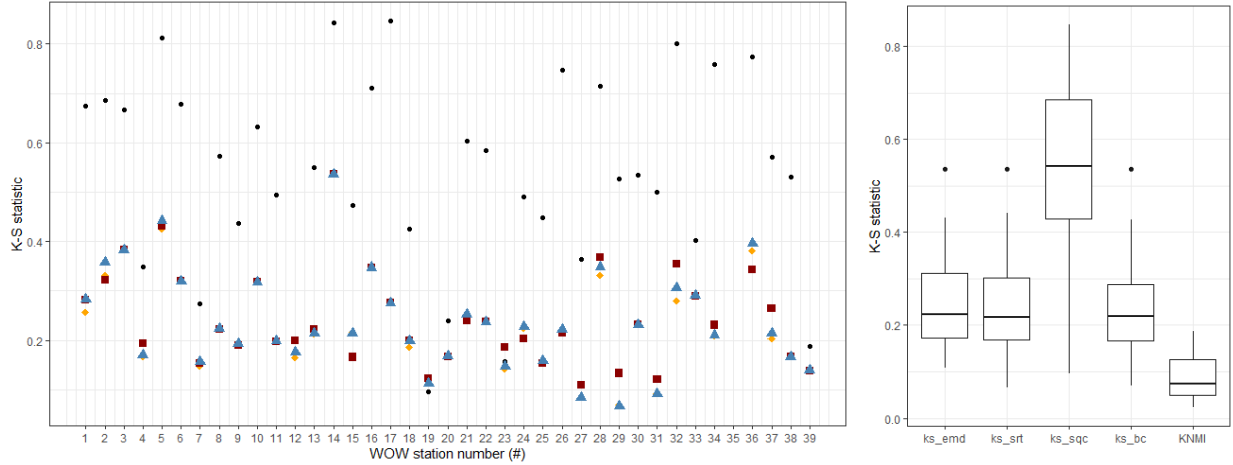


Figure 5.9: K-S statistic between a WOW station and its *relating* KNMI station, after standard quality control (black points, ‘ks_sqc’), after bias correction (orange diamonds, ‘ks_bc’), after IDW (red squares, ‘ks_emd’), after SRT (blue triangles, ‘ks_srt’), with boxplot. The K-S statistic between two nearby KNMI stations are also draw on ‘KNMI’ of the boxplot as a reference.

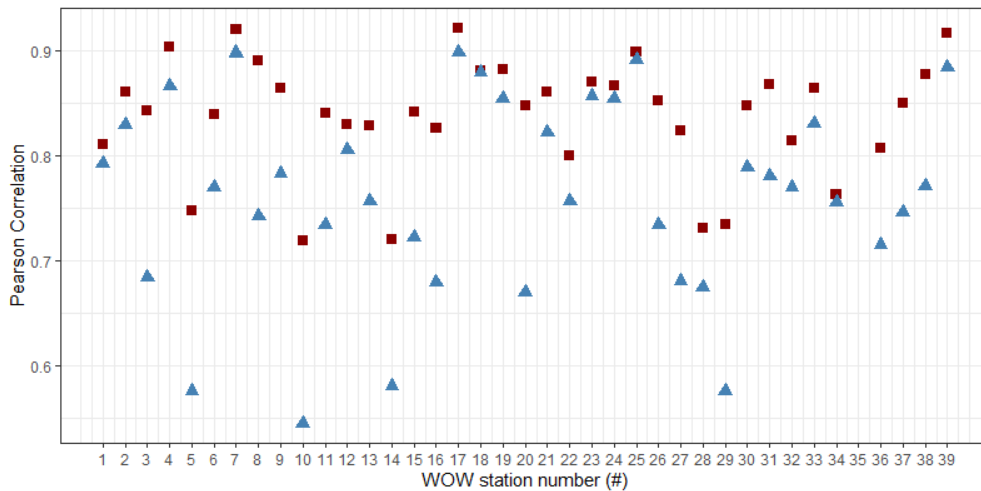


Figure 5.10: Pearson correlation between a WOW station and its estimation from reference observations by IDW (red squares) or SRT (blue triangles) methods.

[2004] constructed a probabilistic spatial quality control procedure for the temperature, based on the fact the temperatures generally follow the Normal distribution. Their approach is certainly not suited for wind speed data because wind speeds are not normally distributed. However, we may apply similar ideas by discovering another probabilistic approach to implement spatial quality control in future works, perhaps using Bayesian statistics that make a prior assumption of wind speed distributions.

Besides traditional ideas, we may also try some machine learning methods to do spatial quality control, which may provide some new aspects of understanding the data. Lee et al. [2018] have applied support vector regression methods to detect anomalies in meteorological data. They compared their results of spatial quality control on temperature data by using SVR with the tra-

ditional spatial consistency tests, and achieved a good outcome. It would be interesting to learn how these methods work on wind data, and may also take wind direction observations into concern.

Moreover, a desirable requirement of quality control is that it will not falsely flag the observations during an extreme event. Thus it is helpful to look at some localized extreme events for the case study to check whether the quality control flag suspicious observations correctly. You and Hubbard [2006] tested the observations filtered after spatial regression tests on temperature and precipitation data, and they found that their spatial quality control did not remove valuable observations during extreme events. A good direction for future works is to check the performances of spatial quality control in extreme wind events.

Chapter 6

Conclusion and Outlook

6.1 The overall quality control procedures

In this thesis, we discussed quality control of crowdsourced wind observations from the third-party platform WOW-NL. The overall procedures can be divided into four parts: (i) the pre-processing steps on the raw observation data, (ii) the standard quality control methods for general automatic weather stations, (iii) the bias correction step intended to eliminate the bias compared to official wind observations, and (iv) the spatial quality control that considers the consistency of multiple simultaneous observations. We summarize those steps in Table 6.1, and each step is referred from a certain section of the text.

Overall Quality Control Procedures		
Part	Method	Refer text
I) Pre-processing	1. Matching standard time points	Chapter 2.2
	2. Data integrity test	Chapter 2.4
II) Standard quality control	3. Internal consistency test	Chapter 3.1.1
	4. Plausible range test	Chapter 3.1.2
	5. Temporal consistency test	Chapter 3.1.3
III) Bias correction	6. Empirical quantile mapping	Chapter 4.2.4
IV) Spatial quality control	7. Inverse distance weighting using Earth mover's distance	Chapter 5.2.1

Table 6.1: The overall quality control procedures for the WOW wind observations.

Before implementing the quality control, we need to ensure that all the data are uniformly expressed. This is unnecessary for official data, as they are recorded following strict standards. However, this pre-processing procedure is essential for crowdsourced CWS observations so that we can proceed with the following steps that require comparison with official stations. We first unified the recording time sequence of different CWS data by matching the nearest observation to the standard ten-minute time sequence. This matching step excluded some observations between two time-points, and our assumptions are made to neglect their existence. Next, only stations with enough valid wind data could be further analyzed, and so we removed stations with incomplete records by performing a data integrity test. This pre-processing procedure rejected more than half the CWS, only 39 from original 93 stations are left (65 out of original 93 stations record wind). This outcome reminds us that the quality of CWS should be taken some notice and encourage us

to build more CWS for achieving a higher resolution of weather observation networks.

The standard quality control methods are mainly derived from quality control checks for official automatic weather stations. We made some adaptations such that these steps are suited to the CWS in our data set. First, we checked the internal consistency of the wind measurements by comparing synchronous wind speed and wind gust records to make sure that wind speed can not exceed wind gust at all times. If wind speed is larger than wind gust, usually both records should be removed, but in our case, we just remove wind gust records. This is an adaptation for CWS because we are less confident about their wind gust observations than wind speed. Next, we checked the plausible range of CWS wind data based on the historical highest records by official stations, where we assumed that CWS could not achieve a higher record than the all-time records. Finally, we check the variability of single CWS over time, in order to guarantee that the CWS recorded wind does not change too fast or too slow. The specific variability restrictions are also derived from official data. We take a rather strict criterion in this step so that the left CWS wind data are more sensitive to the variabilities with time and hence more reliable. These standard quality control steps filtered about 15% of the valid CWS wind speed observations after pre-processing, which implies there are still a lot of CWS observations that are trustworthy according to the standard quality control tests and thus warrant further examination.

Another big difference between CWS and official wind speed data is that the CWS records are likely to have systematic biases. This may result from the nonstandard placements of CWS, averaging over different lengths of time, or biases inherent to the measurement device. Therefore, the bias correction procedure is essential typically for CWS wind speed observations. We studied the previous bias correction methods on weather data, which are mainly applied to correct the bias for predicted observations, and we adapted them on the current wind speed data. The analysis showed that empirical quantile mapping is the most suitable bias correction method for our data, and the core idea is to find a mapping that matches the quantiles between a distribution to be corrected and a reference distribution. The bias correction procedure helps calibrate the CWS observations so that they could indicate wind speed at the same level, which fulfills the prerequisites for comparing simultaneous wind speed by different sites.

The CWS wind speed observations after standard quality control and bias correction could finally be used for comparison from site to site. The spatial quality control procedure tests the consistency of wind speed observations in the spatial perspective, unlike the temporal consistency test that checks the temporal aspect. The main idea is first to select a list of reference observations for target observation, then estimate the target by some calculation using the reference records, and finally compare the CWS wind observation with the estimated target to determine whether it is a consistent record in space. We discussed the traditional method of choosing reference stations by geographical distance, and concluded that the approach is not sufficient. Then we construct an enhanced selection method that takes additional consideration on the statistical distance between two distributions, namely the Earth mover's distance. After obtaining the reference observations, we tried different approaches to get a proper estimate that is reliable for generating a confidence interval for the target observation. Taking both the accuracy and the processing efficiency into consideration, we chose the inverse distance weighting method with Earth mover's distance at last for our spatial quality control. This advanced quality control procedure filtered about 10% of the data after bias-corrected, which indicated that many CWS observations are spatially consistent after calibration.

In total, the results of our quality control procedures showed that we have some confidence in about three-quarters of the CWS wind speed observations, which undoubtedly has a potential for future use in nowcasting and forecasting. Moreover, we take the real-time processing as a requirement when constructing the quality control steps, and so our procedures have the feasibility to work on real-time CWS observations to obtain outputs quickly. Although more improvements can be added to these quality control procedures, as discussed below, we have demonstrated for the first time at a province-level that CWS wind speed observations match well with official observations, after the application of advanced quality-control and bias-correction methods.

6.2 Outlook for further studies

The CWS in this study are located only in the Utrecht province of the Netherlands. It would be worthwhile to check how our quality control procedures performed on the crowdsourced wind records in other regions. Thus an intended direction for future research could be focused on applying the quality control on other crowdsourced wind observation data sets, and comparing the thresholds used for different regions.

Another issue arose from our pre-processing step, where we matched the CWS observations to a standard time sequence. However, this might exclude some important information, as we have discussed in Chapter 2.5. An alternate approach to get unified data is to calculate the average ten-minute wind speed from whatever observations in that ten-minute time interval. Although this may have shortages from another perspective, it is interesting to see how the quality control results would be on the new unified data and compare them with our results.

As we have discussed in Chapter 4.5, the inflated zero wind speed records by CWS make them look abnormal compared with the official observations. It is a tricky question to find the true wind speed when a CWS record zero, but it is also important to get the answer since such zero wind speeds have a significant impact on bias correction steps. Our bias correction step did not take special actions on these zero wind speeds but simply correct all of them to a positive value, which is not realistic and needed to be improved. A possible future direction for studies could be analyzing deeply in the inflated zero wind speeds of CWS observations, and work out a reasonable way to reconstruct its true records.

A considerable future direction for research should be focused on the advanced quality control and bias correction methods not only for wind speed, but also for wind gust and direction. For instance, when selecting reference stations in spatial quality control, we could also take wind direction into considerations. Due to the directional property of wind, we would expect wind behaves more similar in one direction and less alike in its perpendicular direction. Thus we would get more practical reference observations with assigned weights based on both wind speed and wind direction data. We have tried some windrose plots for different stations, as shown in Figure 6.1, and find that some stations have comparable windroses with official ones. The windroses could potentially indicate the direction of obstructions around a CWS, which provide a valuable reference for the placement information.

At last, we want to discuss the outlook of wind observations from a mathematical point of view. As we mentioned in the Introduction chapter, the wind observations form a vector field, and thus it would be useful to look at the behavior of wind vectors. We can take averages and

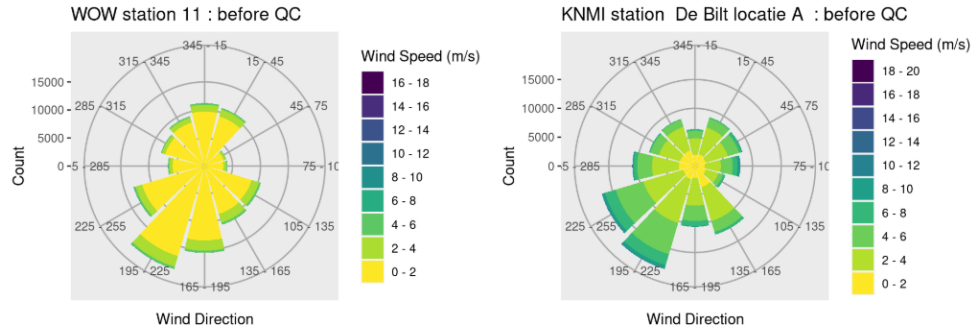


Figure 6.1: The windrose of a WOW station compared with the *relating* KNMI station, before and after quality control.

make comparisons of wind vectors in a way that considers both the speed and direction. It is also interesting to study how wind vectors change over time and give predictions of future dynamics based on specific initial values. This could potentially provide a forecasting method that fully considers the directional property of wind and perform better predictions. We think this is a meaningful direction for studying wind. We hope more researchers would be interested in the CWS wind observations field and make contributions to exploit the crowdsourced wind data effectively.

Appendix A

List of WOW stations

We list the 39 WOW stations that are used for studying quality control. We give unique numbering for each station to make it simpler to identify in the thesis. The corresponding station ID are shown in Table A.1.

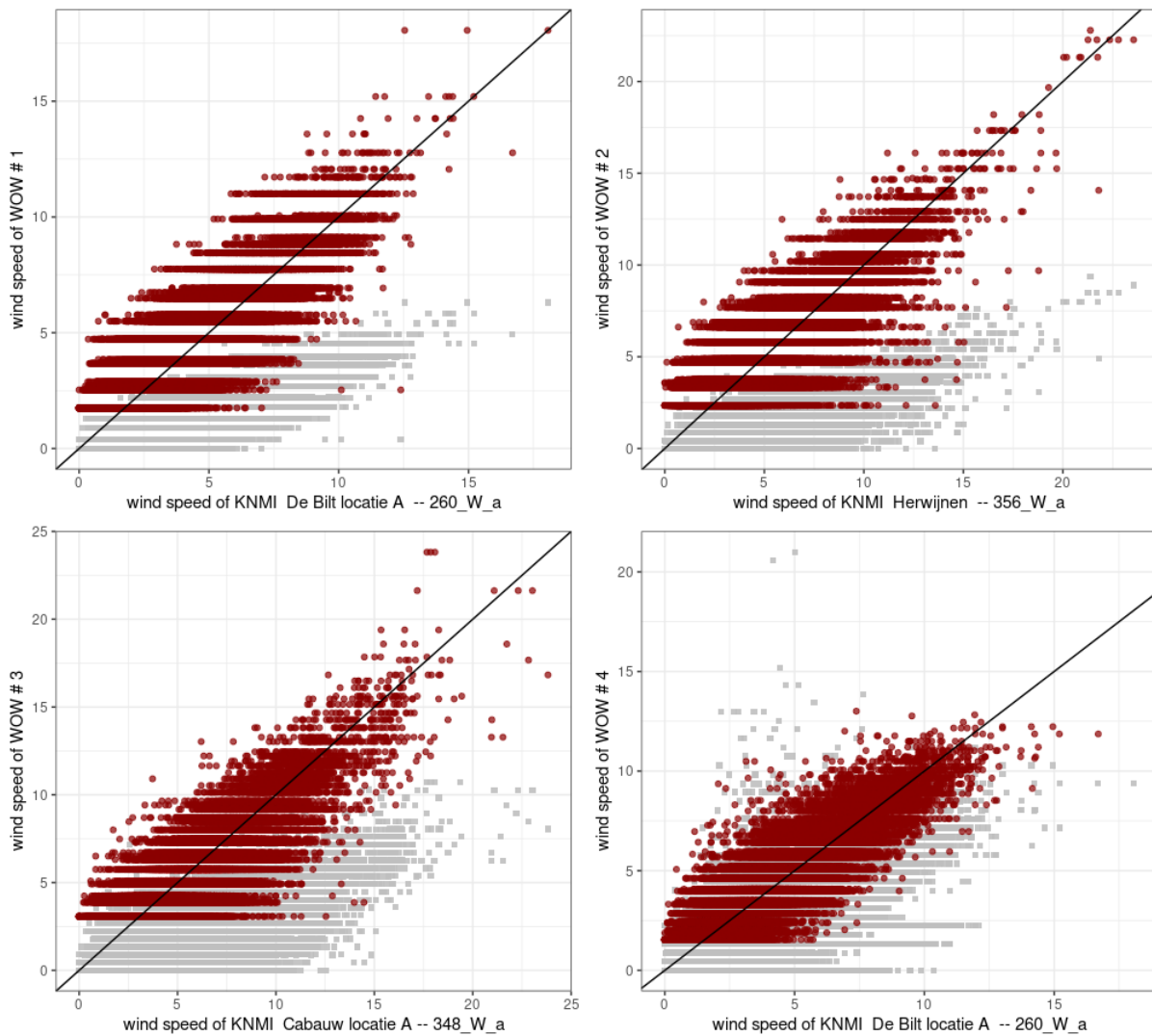
WOW number	WOW station ID
1	930246001
2	945516001
3	940386001
4	72630618
5	945566001
6	930206001
7	949436001
8	965656001
9	39813225
10	889886001
11	956276001
12	968256001
13	956296001
14	924516001
15	66385da8-7acc-e611-9400-0003ff59aed0
16	935266001
17	933736001
18	941056001
19	23571578
20	929236001
21	963146001
22	935786001
23	922476001
24	929856001
25	923976002
26	741386082
27	d1e5238b-df9d-e711-9401-0003ff59996a
28	968636001
29	288256001
30	923556001
31	933296001
32	974686001
33	935726001
34	1c063b9c-d654-e711-9400-000d3ab1cf4b
35	ce77ae9b-6161-e711-9401-0003ff598dd1
36	3a3b1f7e-b87d-e711-9402-0003ff597a5c
37	916136001
38	3845fdfc-19e6-e611-93ff-000d3ab1ce94
39	951366001

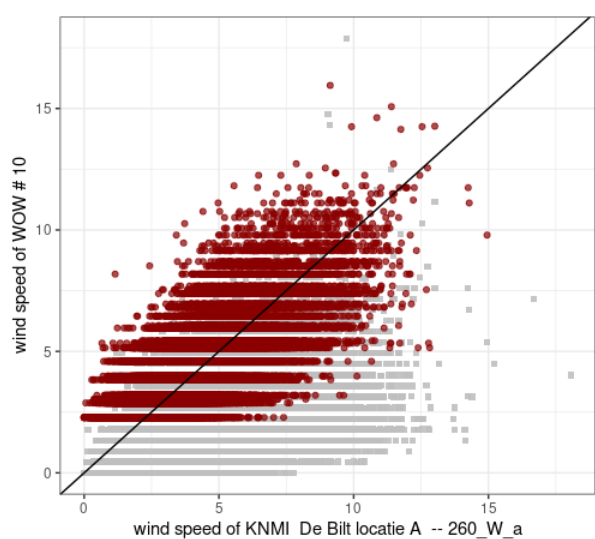
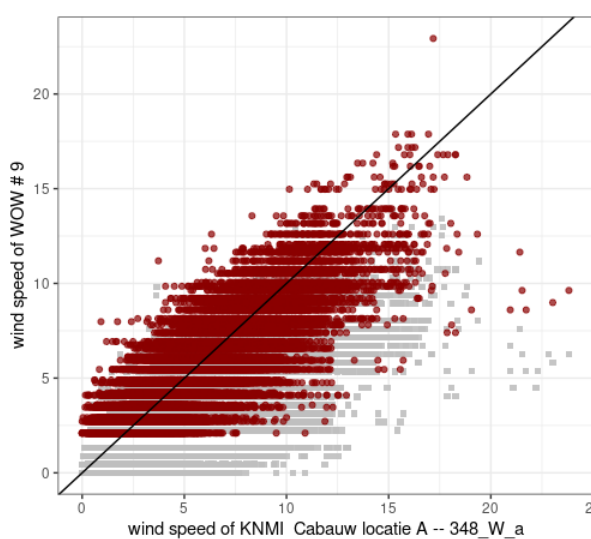
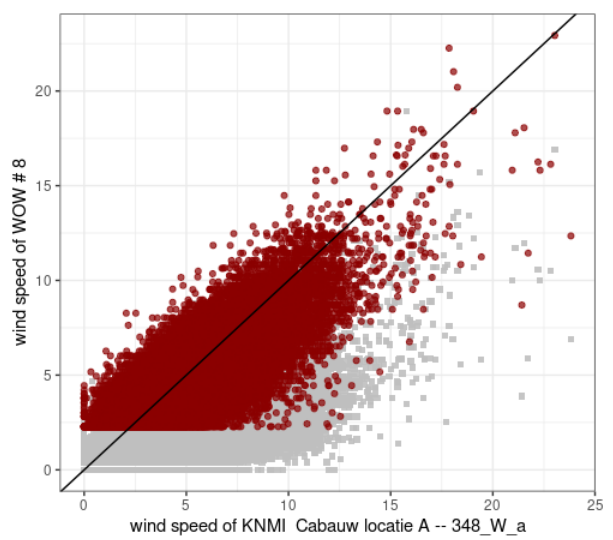
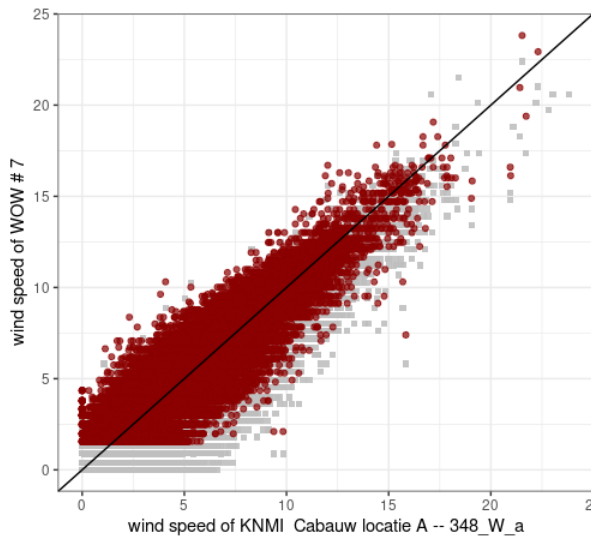
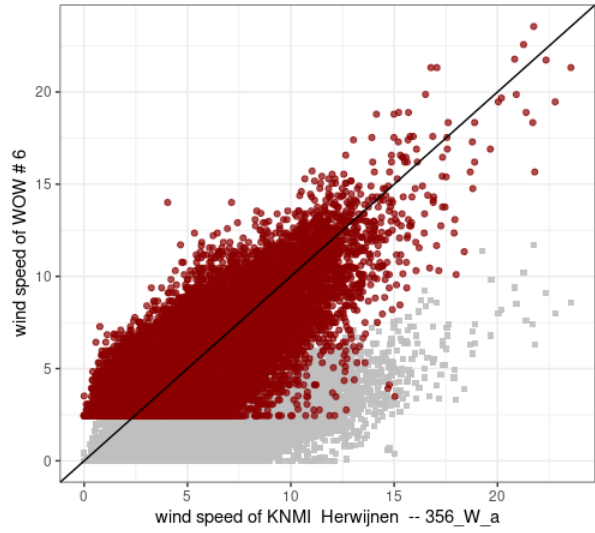
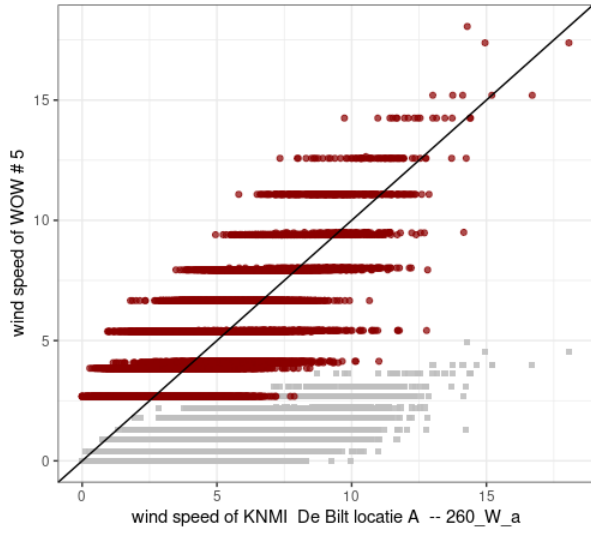
Table A.1: The numbering of each WOW station with its station ID.

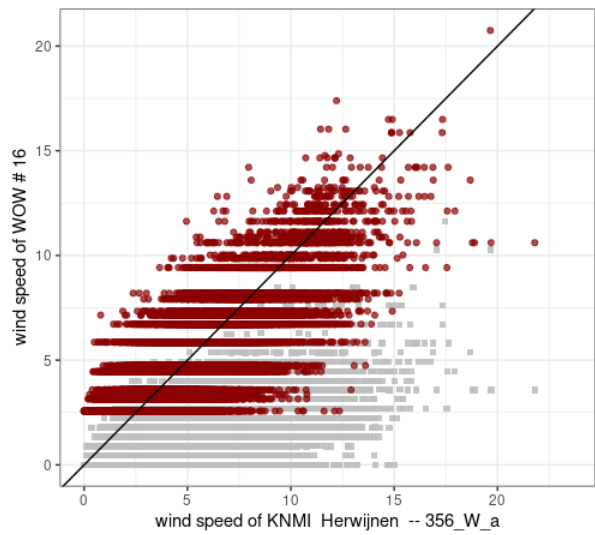
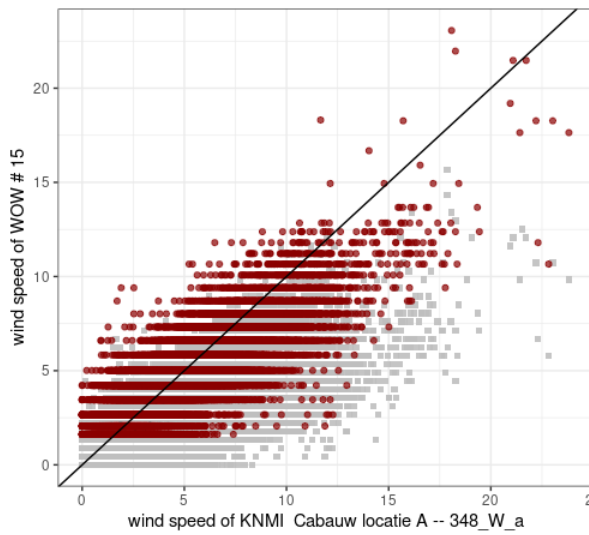
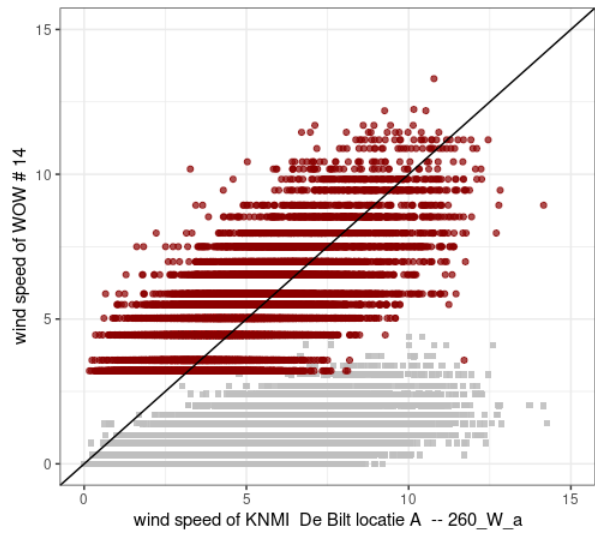
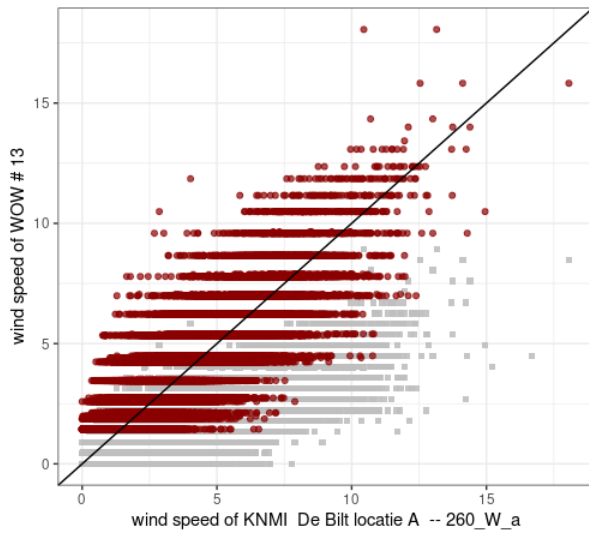
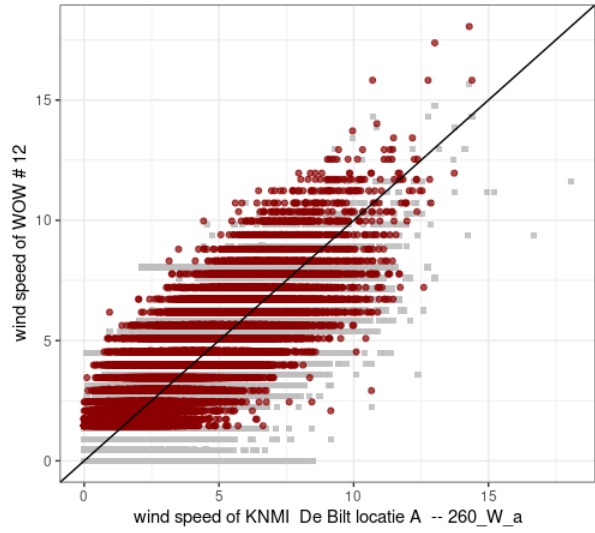
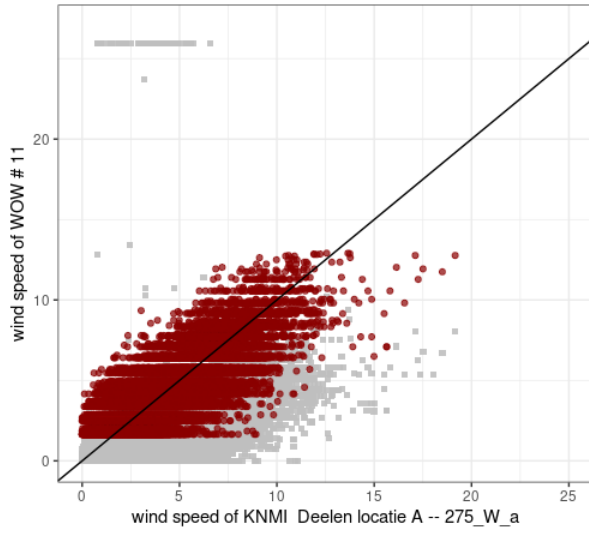
Appendix B

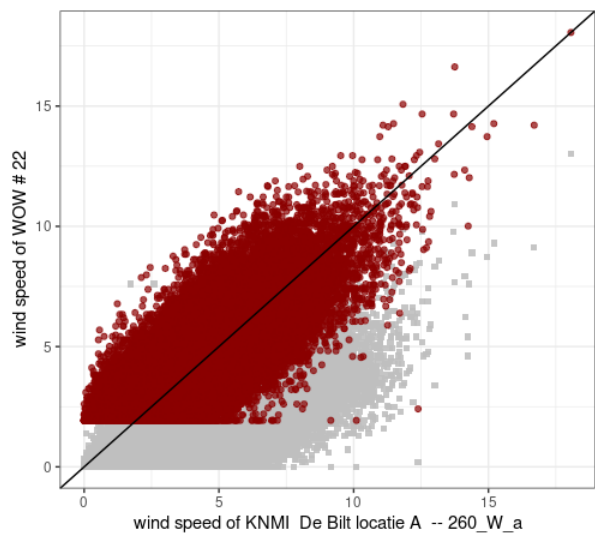
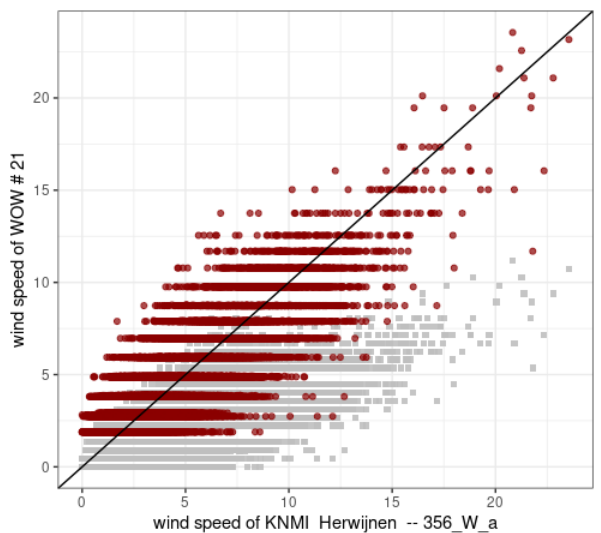
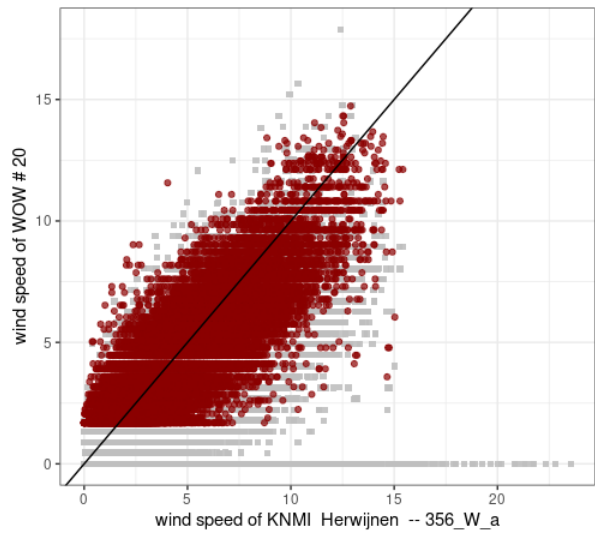
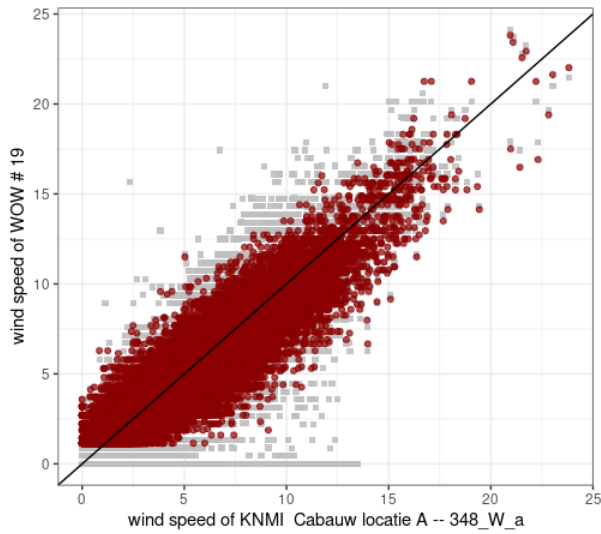
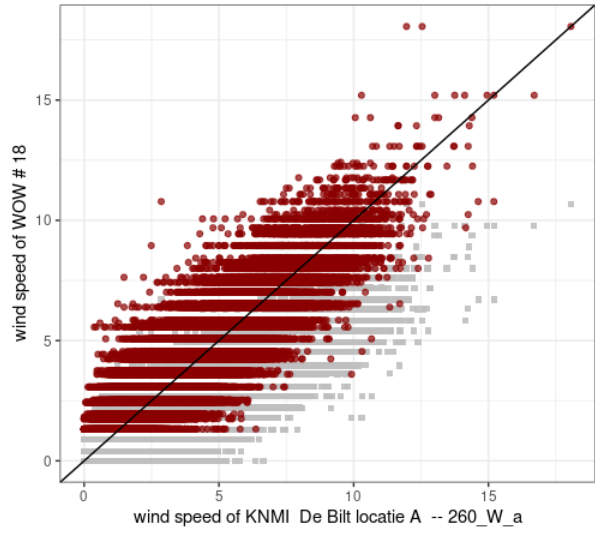
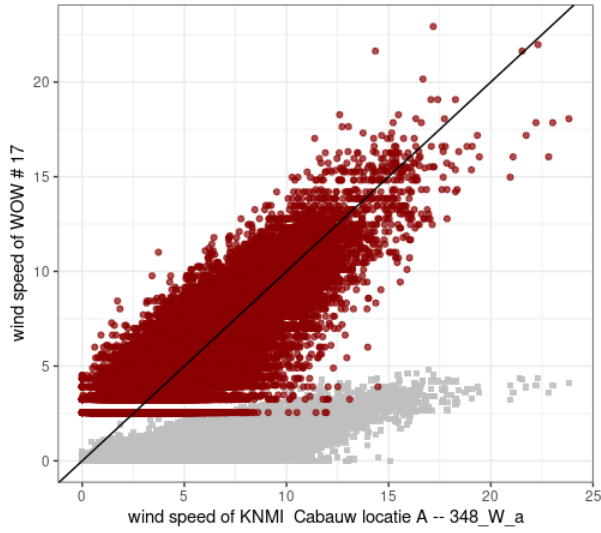
Results of each WOW station

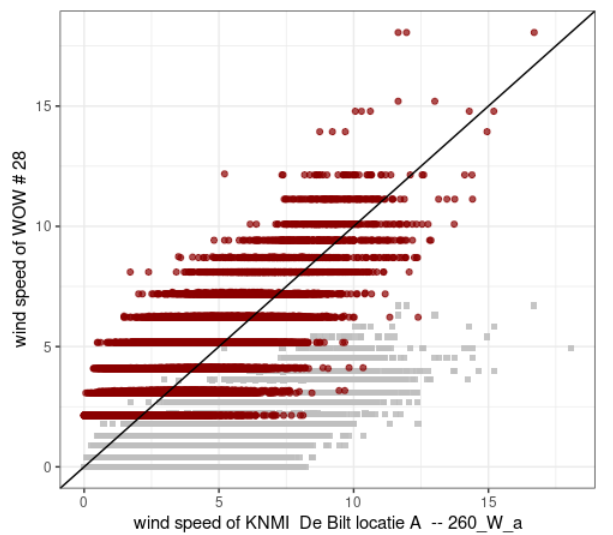
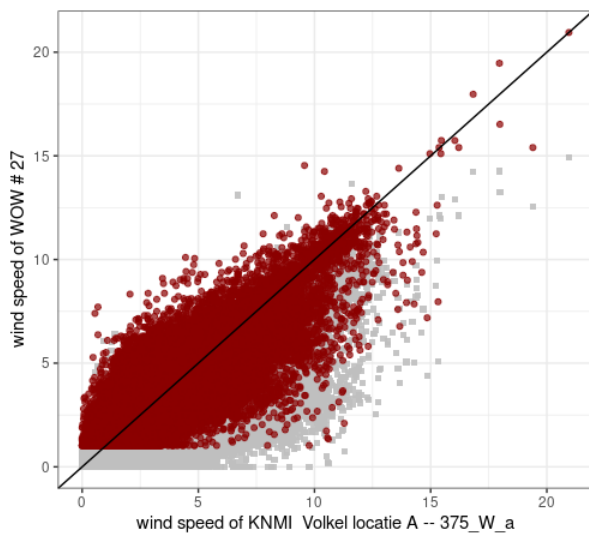
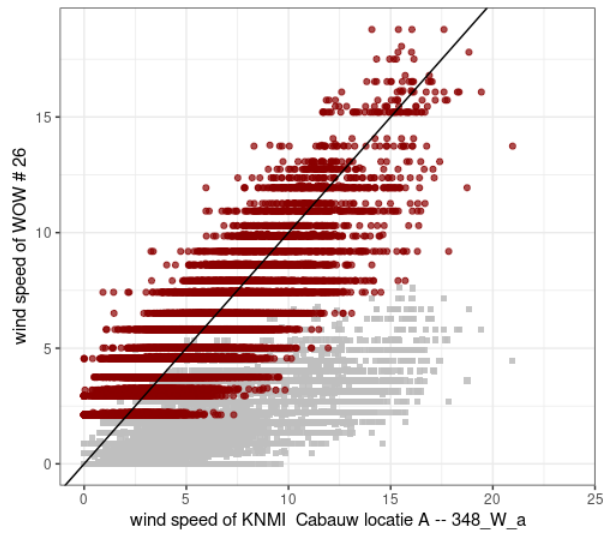
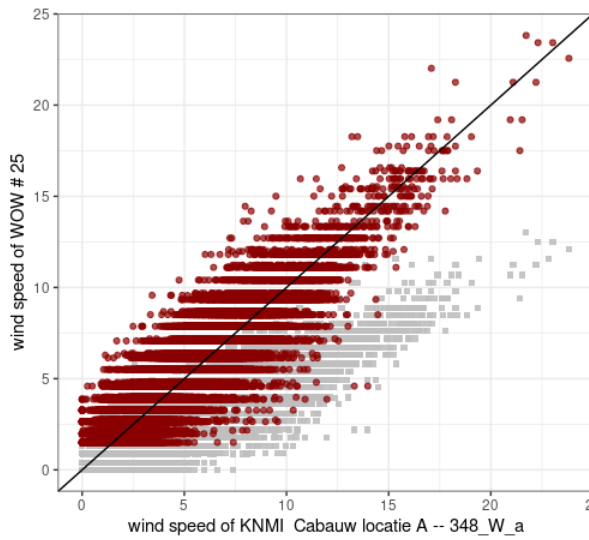
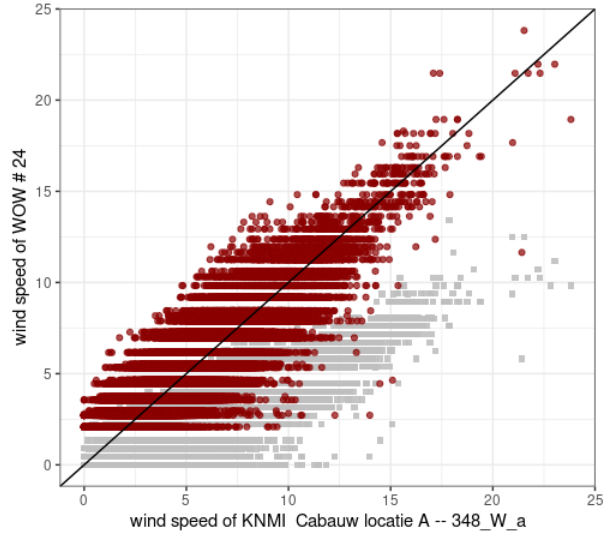
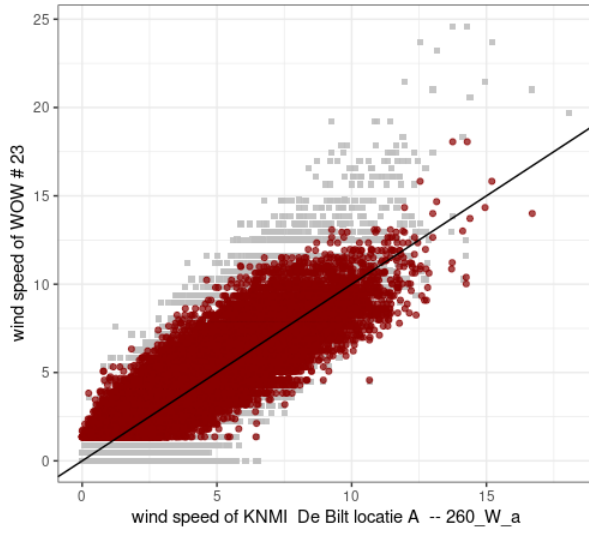
We draw scatterplots of simultaneous wind speed observations between each WOW station and its *relating* KNMI station in the following. The red points are WOW wind speeds after quality controls and bias correction, while the grey squares are WOW raw data.

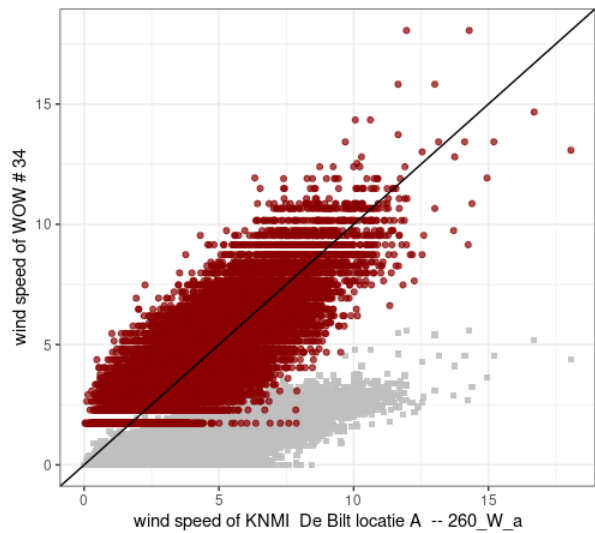
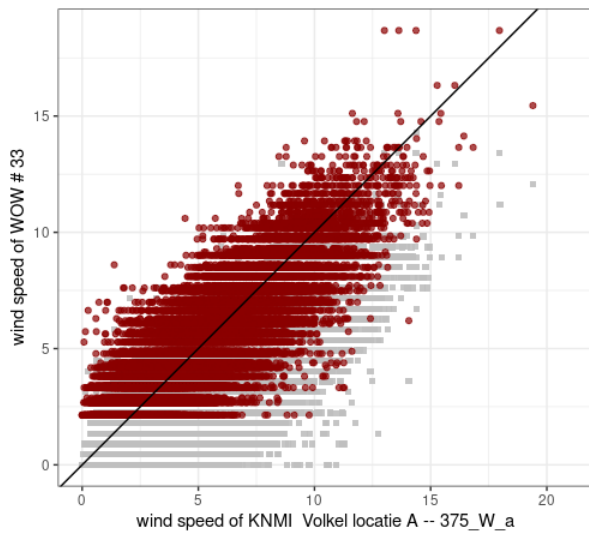
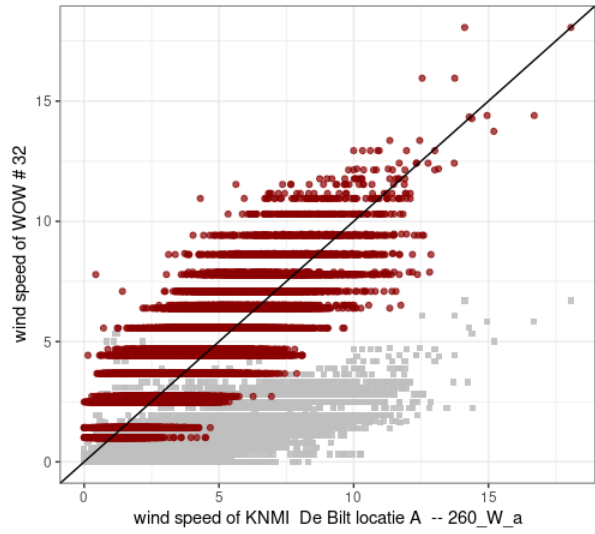
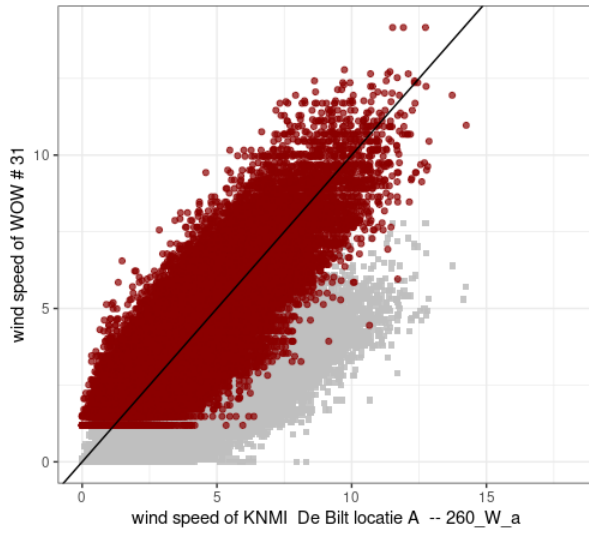
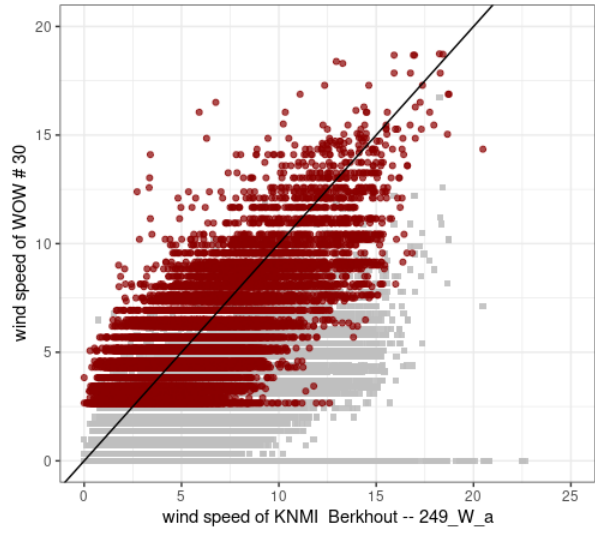
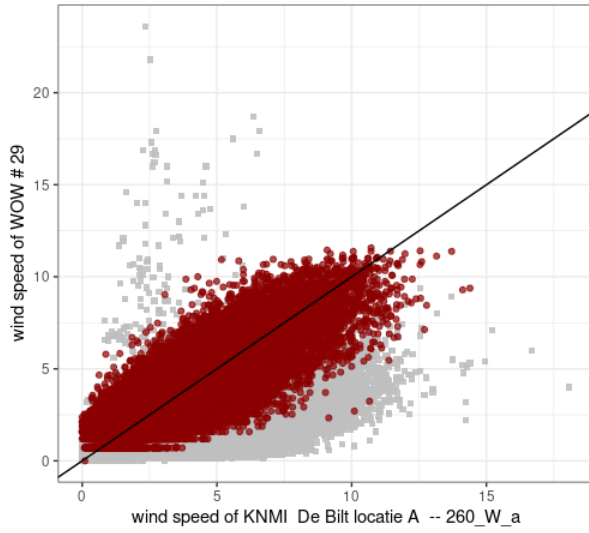


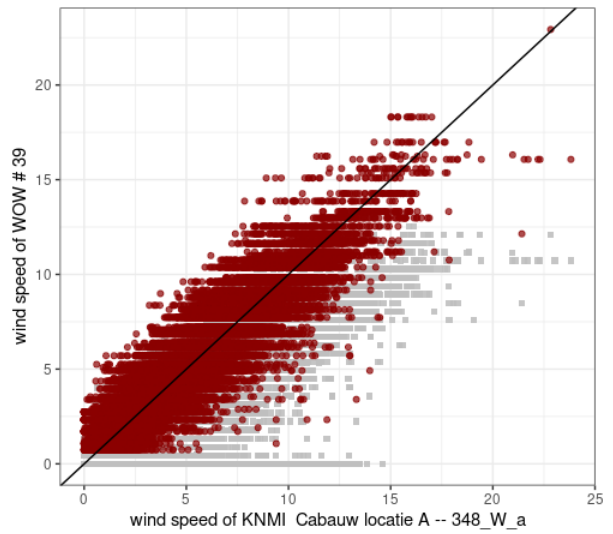
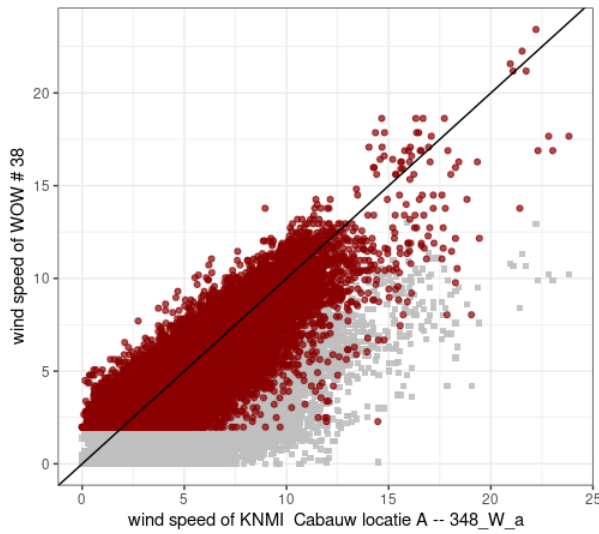
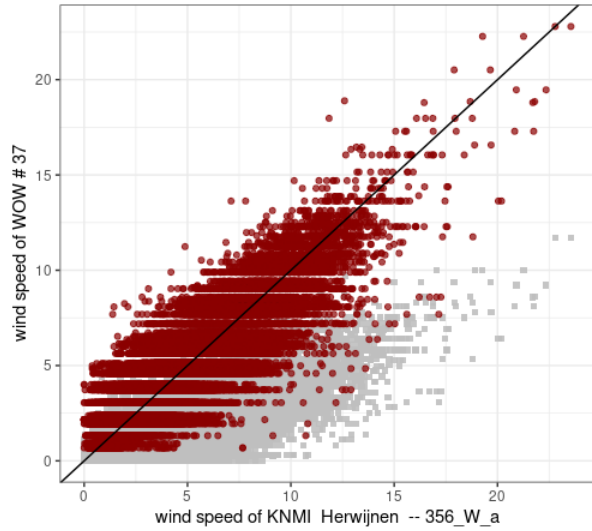
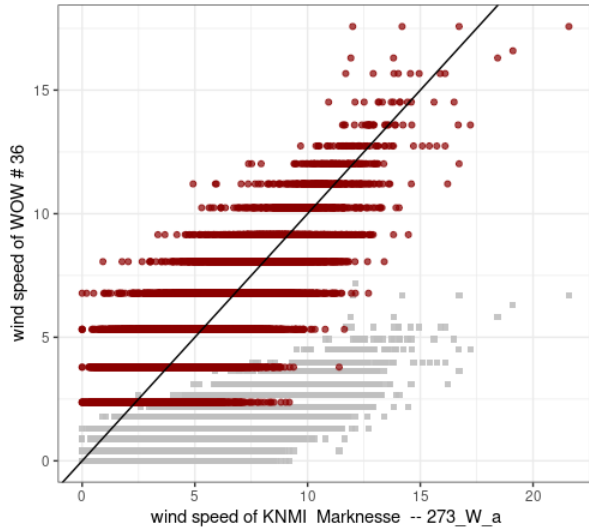




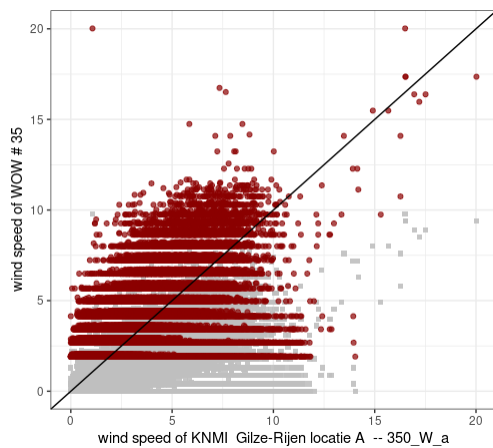








We can see from these 38 figures that most WOW stations match well with their relating KNMI after quality controls and bias correction.



WOW station #38 has too low quality of data that we do not consider in the spatial quality control, the scatterplot of its wind speed after bias correction is shown here.

Bibliography

- Akhter, J., Das, L. and Deb, A. [2017], ‘Cmip5 ensemble-based spatial rainfall projection over homogeneous zones of india’, *Climate Dynamics* **49**(5-6), 1885–1916.
- Barnes, S. L. [1964], ‘A technique for maximizing details in numerical weather map analysis’, *Journal of Applied Meteorology* **3**(4), 396–409.
- 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.
- Cerlini, P. B., Silvestri, L. and Saraceni, M. [2020], ‘Quality control and gap-filling methods applied to hourly temperature observations over central italy’, *Meteorological Applications* **27**(3), e1913.
- Cheng, A. R., Lee, T. H., Ku, H. I. and Chen, Y. W. [2016], ‘Quality control program for real-time hourly temperature observation in taiwan’, *Journal of Atmospheric and Oceanic Technology* **33**(5), 953–976.
- Cox, M. A. and Cox, T. F. [2008], Multidimensional scaling, in ‘Handbook of data visualization’, Springer, pp. 315–347.
- Cressman, G. P. [1959], ‘An operational objective analysis system’, *Mon. Wea. Rev* **87**(10), 367–374.
- Daly, C., Gibson, W., Doggett, M., Smith, J. and Taylor, G. [2004], A probabilistic-spatial approach to the quality control of climate observations, in ‘Proceedings of the 14th AMS Conference on Applied Climatology, Amer. Meteorological Soc., Seattle, WA’.
- Dann, J. and Prodata Weather Systems [2020], ‘Wind issues - Davis Weather Stations Knowledgebase’, [Windissues-DavisWeatherStationsKnowledgebasehttps://www.manula.com/manuals/pws/davis-kb/1/en/topic/wind-issues](https://www.manula.com/manuals/pws/davis-kb/1/en/topic/wind-issues). (Accessed: 9 August 2020)).
- 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.
- DeGaetano, A. T. [1997], ‘A quality-control routine for hourly wind observations’, *Journal of Atmospheric and Oceanic Technology* **14**(2), 308–317.
- Droste, A., Heusinkveld, B., Fenner, D. and Steeneveld, G.-J. [2019], Crowdsourcing the urban wind., in ‘Geophysical Research Abstracts’, Vol. 21.

- Estévez, J., Gavilán, P. and García-Marín, A. [2018], ‘Spatial regression test for ensuring temperature data quality in southern Spain’, *Theoretical and applied climatology* **131**(1-2), 309–318.
- Estévez, J., Gavilán, P. and Giráldez, J. V. [2011], ‘Guidelines on validation procedures for meteorological data from automatic weather stations’, *Journal of Hydrology* **402**(1-2), 144–154.
- Fang, G., Yang, J., Chen, Y. and Zammit, C. [2015], ‘Comparing bias correction methods in downscaling meteorological variables for a hydrologic impact study in an arid area in China’, *Hydrology and Earth System Sciences* **19**(6), 2547–2559.
- Fiebrich, C. A., Morgan, C. R., McCombs, A. G., Hall Jr, P. K. and McPherson, R. A. [2010], ‘Quality assurance procedures for mesoscale meteorological data’, *Journal of Atmospheric and Oceanic Technology* **27**(10), 1565–1582.
- Gatey, D. and Miller, C. [2007], ‘An investigation into 50-year return period wind speed differences for Europe’, *Journal of wind engineering and industrial aerodynamics* **95**(9-11), 1040–1052.
- Gibson, W., Daly, C., Doggett, M., Smith, J. and Taylor, G. [2004], Application of a probabilistic spatial quality control system to daily temperature observations in Oregon, in ‘Proc., 14th AMS Conf. on Applied Climatology’, pp. 13–16.
- Grubbs, F. E. [1969], ‘Procedures for detecting outlying observations in samples’, *Technometrics* **11**(1), 1–21.
- Gudmundsson, L., Bremnes, J., Haugen, J. and Engen-Skaugen, T. [2012], ‘Downscaling RCM precipitation to the station scale using statistical transformations: a comparison of methods’, *Hydrology and Earth System Sciences* **16**(9), 3383–3390.
- Hawkins, E., Osborne, T. M., Ho, C. K. and Challinor, A. J. [2013], ‘Calibration and bias correction of climate projections for crop modelling: an idealised case study over Europe’, *Agricultural and forest meteorology* **170**, 19–31.
- Ho, C. K., Stephenson, D. B., Collins, M., Ferro, C. A. and Brown, S. J. [2012], ‘Calibration strategies: a source of additional uncertainty in climate change projections’, *Bulletin of the American Meteorological Society* **93**(1), 21–26.
- Howe, J. [2006], ‘The rise of crowdsourcing’, *Wired magazine* **14**(6), 1–4.
- Hubbard, K. G. and Siva Kumar, M. [2001], ‘Automated weather stations for applications in agriculture and water resources management’.
- Hubbard, K. G. and You, J. [2005], ‘Sensitivity analysis of quality assurance using the spatial regression approach—a case study of the maximum/minimum air temperature’, *Journal of Atmospheric and Oceanic Technology* **22**(10), 1520–1530.
- Hubbard, K., Goddard, S., Sorensen, W., Wells, N. and Osugi, T. [2005], ‘Performance of quality assurance procedures for an applied climate information system’, *Journal of Atmospheric and Oceanic Technology* **22**(1), 105–112.
- Jiménez, P. A., González-Rouco, J. F., Navarro, J., Montávez, J. P. and García-Bustamante, E. [2010], ‘Quality assurance of surface wind observations from automated weather stations’, *Journal of atmospheric and Oceanic Technology* **27**(7), 1101–1122.

- Kantorovich, L. V. [1960], ‘Mathematical methods of organizing and planning production’, *Management science* **6**(4), 366–422.
- KNMI [2020], ‘KNMI - Wij zijn het KNMI’, <https://www.knmi.nl/over-het-knmi/over>. (Accessed: 6 Feb 2020).
- Lee, M.-K., Moon, S.-H., Yoon, Y., Kim, Y.-H. and Moon, B.-R. [2018], ‘Detecting anomalies in meteorological data using support vector regression’, *Advances in Meteorology* **2018**.
- Luo, M., Liu, T., Meng, F., Duan, Y., Frankl, A., Bao, A. and De Maeyer, P. [2018], ‘Comparing bias correction methods used in downscaling precipitation and temperature from regional climate models: a case study from the kaidu river basin in western china’, *Water* **10**(8), 1046.
- 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**, 170–191.
- Moeckel, R. and Murray, B. [1997], ‘Measuring the distance between time series’, *Physica D: Nonlinear Phenomena* **102**(3-4), 187–194.
- Muller, C., Chapman, L., Johnston, S., Kidd, C., Illingworth, S., Foody, G., Overeem, A. and Leigh, R. [2015], ‘Crowdsourcing for climate and atmospheric sciences: current status and future potential’, *International Journal of Climatology* **35**(11), 3185–3203.
- Muskulus, M. and Verduyn-Lunel, S. [2011], ‘Wasserstein distances in the analysis of time series and dynamical systems’, *Physica D: Nonlinear Phenomena* **240**(1), 45–58.
- 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**, 118.
- Navarro-Racines, C. E. and Tarapues, J. [2015], ‘Bias-correction in the ccafs-climate portal: A description of methodologies’.
- Netatmo Helpcenter [2020], ‘Smart Home Weather Station and accessories’, <https://helpcenter.netatmo.com/en-us/smart-home-weather-station-and-accessories/measures-and-calibrations/how-does-the-smart-anemometer-work>. (Accessed: 9 August 2020)).
- NIST/SEMATECH [2020], ‘e-Handbook of Statistical Methods’, <http://www.itl.nist.gov/div898/handbook/>. (Accessed: 08 July 2020).
- Panaretos, V. M. and Zemel, Y. [2019], ‘Statistical aspects of wasserstein distances’, *Annual review of statistics and its application* **6**, 405–431.
- Papoulis, A. and Pillai, S. U. [2002], *Probability, random variables, and stochastic processes*, Tata McGraw-Hill Education.
- Peterson, T. C., Vose, R., Schmoyer, R. and Razuvaëv, V. [1998], ‘Global historical climatology network (ghcn) quality control of monthly temperature data’, *International Journal of Climatology: A Journal of the Royal Meteorological Society* **18**(11), 1169–1179.
- R Core Team [2013], *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL: <http://www.R-project.org/>

- Rui Castro [2020a], ‘Lecture 1 - Introduction and the Empirical CDF’, <https://www.win.tue.nl/~rmcastro/AppStat2013/files/lecture1.pdf>. (Accessed: 31 July 2020).
- Rui Castro [2020b], ‘Lecture 2 and 3 - Goodness-of-Fit (GoF) Tests’, <https://www.win.tue.nl/~rmcastro/AppStat2013/files/lectures23.pdf>. (Accessed: 31 July 2020).
- Shafer, M. A., Fiebrich, C. A., Arndt, D. S., Fredrickson, S. E. and Hughes, T. W. [2000], ‘Quality Assurance Procedures in the Oklahoma Mesonet’, *Journal of Atmospheric and Oceanic Technology* **17**(4), 474–494.
URL: [https://doi.org/10.1175/1520-0426\(2000\)017;0474:QAPITO;2.0.CO;2](https://doi.org/10.1175/1520-0426(2000)017;0474:QAPITO;2.0.CO;2)
- Taylor, J. R. and Loescher, H. L. [2013], ‘Automated quality control methods for sensor data: a novel observatory approach’, *Biogeosciences* **10**(7), 4957.
- Terink, W., Hurkmans, R., Uijlenhoet, R., Warmerdam, P. and Torfs, P. [2008], Bias correction of temperature and precipitation data for regional climate model application to the rhine basin, Technical report, Wageningen Universiteit.
- Teutschbein, C. and Seibert, J. [2012], ‘Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods’, *Journal of hydrology* **456**, 12–29.
- Vallender, S. [1974], ‘Calculation of the wasserstein distance between probability distributions on the line’, *Theory of Probability & Its Applications* **18**(4), 784–786.
- Wasserman, L. [2013], *All of statistics: a concise course in statistical inference*, Springer Science & Business Media.
- WMO [2014], ‘Guide to meteorological instruments and methods of observation, 2014 edition’, *World Meteorological Organisation (WMO-No.8)*.
- WOW-NL [2020], ‘WOW-NL: Jouw weer op de kaart!’, <https://wow.knmi.nl/over-wow-nl>. (Accessed: 7 Feb 2020).
- You, J. and Hubbard, K. G. [2006], ‘Quality control of weather data during extreme events’, *Journal of Atmospheric and Oceanic Technology* **23**(2), 184–197.
- You, J., Hubbard, K. G. and Goddard, S. [2008], ‘Comparison of methods for spatially estimating station temperatures in a quality control system’, *International Journal of Climatology: A Journal of the Royal Meteorological Society* **28**(6), 777–787.
- Zahumenskỳ, I. [2004], ‘Guidelines on quality control procedures for data from automatic weather stations’, *World Meteorological Organization, Switzerland*.