Received: 3 February 2021 | Revised: 30 June 2021 | Accepted: 28 July 2021

DOI: 10.1002/qj.4146

Quarterly Journal of the Royal Meteorological Society

RESEARCH ARTICLE

Quality control and bias adjustment of crowdsourced wind speed observations

Jieyu Chen[1,2,3] | **Kate Saunders**[1,4,5] | **Kirien Whan**[1]

[1]Weather and Climate Modelling (RDWK), Royal Netherlands Meteorological Institute (KNMI), De Bilt, the Netherlands

[2]Mathematical Institute, Utrecht University, Utrecht, the Netherlands

[3]Chair of Statistics and Econometrics, Karlsruhe Institute of Technology, Karlsruhe, Germany

[4]Institute of Applied Mathematics, Delft University of Technology, Delft, the Netherlands

[5]School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia

Correspondence
J. Chen, Chair of Statistics and Econometrics, Karlsruhe Institute of Technology, Karlsruhe 76185, Germany.
Email: jieyu.chen@kit.edu

Abstract

Wind observations collected at citizen weather stations (CWSs) could be an invaluable resource in climate and meteorology studies, yet these observations are underutilised because scientists do not have confidence in their quality. These wind speed observations have systematic biases, likely caused by improper instrumentation and station sitings. Such systematic biases introduce spatial inconsistencies that prevent comparison of these stations spatially and limit the possible usage of the data. In this paper, we address these issues by improving and developing new methods for identifying suspect observations and adjusting systematic biases. Our complete quality control and bias adjustment procedure consists of four steps: (a) performing within-station quality control tests to check the plausible range and the temporal consistency of observations, (b) adjusting the systematic bias using empirical quantile mapping, (c) implementing between-station quality control to compare observations from neighbouring stations to identify spatially inconsistent observations, and (d) providing estimates of the true wind when CWSs falsely report zero wind speeds, as a complement to the bias adjustment. We apply these methods to CWSs from the Weather Observation Website (WOW) in the Netherlands, comparing the crowdsourced data with official data, and statistically assessing the improvements in data quality after each step. The results demonstrate that the crowdsourced wind speed data are more comparable with official data after quality control checks and bias adjustment steps. Our quality assessment methods therefore give confidence in CWSs, converting their observations into a usable data product and an invaluable resource for applications in need of additional wind observations.

KEYWORDS

bias adjustment, crowdsourcing, citizen weather stations, data quality, quality control, Weather Observation Website, wind observations

1 INTRODUCTION

Wind is an important meteorological component for regular weather forecasts, as well as being the source of many weather warnings (KNMI, 2019). Observations of small-scale wind features can help meteorologists to identify extreme wind events and enable them to issue skilful warnings. However, the official wind observing network in the Netherlands is sparsely distributed, partially due to the high costs of construction and maintenance. The growing popularity of citizen weather stations (CWSs) provides wind observations at many more locations than the official observing network. For example, the network of CWSs used in this research had more than 900 stations in the Netherlands between 2015 and 2019 compared to the 47 official sites operated by the Royal Netherlands Meteorological Institute (KNMI 2020a). The weather observations collected by CWSs, also known as the crowdsourced observation data (Muller *et al.*, 2015), have the potential to greatly increase the spatial density of the official network. However, their data quality is not known or guaranteed because of their non-standard equipment settings and sitings. For instance, many urban CWSs have nearby obstructions, such as buildings and trees, which might block the wind from certain directions. Also, while lower manufacturing costs make CWSs accessible to the general public, these lower costs are often associated with lower precision instruments. The lower quality of the devices and the common problem of substandard station siting have meant that the quality of crowdsourced observations is viewed with scepticism. To date this has inhibited the use of crowdsourced observations in meteorological studies, as also addressed by Chapman *et al.* (2017).

The attention on crowdsourced observations has become more prominent in recent years, triggering researchers to assess the data quality and analyse potential uses (Bell *et al.*, 2013; 2015; Muller *et al.*, 2015; De Vos *et al.*, 2020). Bell *et al.* (2015) developed field studies to investigate the quality of CWS observations of temperature, rainfall and humidity. They concluded that the crowdsourced data could contain significant instrument biases, and therefore any application of those observations would require a quality control (QC) system that can both remove errors and correct biases. Other recent studies into the QC of crowdsourced observations showed that there is much potential for the application of temperature data (Meier *et al.*, 2017; Napoly *et al.*, 2018; Nipen *et al.*, 2019), precipitation data (de Vos *et al.*, 2017; 2019), and surface wind data (Droste *et al.*, 2020). Droste *et al.* (2020) were the first to assess the quality of wind observations from CWSs located in urban areas, based on devices from the same manufacturer Netatmo. This first attempt for assuring wind data quality benefited from the homogeneity of device settings and a relatively small and smooth study region, where the systematic biases were adjusted in a linear way. To date, there has been no comprehensive study on the quality assurance of crowdsourced wind observations from a large-scale, inhomogeneous network. This research aims to increase the spatial density of the official observing network by making use of crowdsourced wind speed observations by application of a detailed QC procedure.

For official automatic weather stations, the QC of observations is common practice and there is a great deal of research into the QC methods for official meteorological observations (Shafer *et al.*, 2000; Durre *et al.*, 2010; Fiebrich *et al.*, 2010; Estévez *et al.*, 2011; Taylor and Loescher, 2013; Otop *et al.*, 2018). Some researchers studied the QC for official wind observations in detail (DeGaetano, 1997; Jiménez *et al.*, 2010; Lucio-Eceiza *et al.*, 2018a; 2018b). In an early study, DeGaetano (1997) built a QC routine based on hourly wind speed and direction data which focused on examining the variation of wind with time, identifying excessively varying or inordinately constant observations. Jiménez *et al.* (2010) proposed more detailed quality assurance procedures for surface wind speed and direction observations, and their output dataset had been proved helpful in wind climatology studies. They mainly considered three aspects – range limits, temporal consistency, and systematic biases – together with some necessary manual checks (Jiménez *et al.*, 2010). Zahumenskỳ (2004) summarised general QC steps for wind observations measured at automatic weather stations in the WMO guidelines, including range and temporal checks. In our QC system for CWS wind speed observations, we build the first part of the procedure based on these classical steps, and we refer to them as 'within-station QC' as the checks are performed on individual stations.

The use of only within-station QC methods, which were developed for official stations, is not sufficient to guarantee the quality of crowdsourced data. According to WMO standards for automatic weather stations and wind observations, the wind sensor should be placed at 10 m above ground, and the measurement site should be flat and unobstructed by trees or buildings (WMO, 2014). CWS users generally do not follow these standards rigorously in practice, and this introduces widespread systematic biases in their recorded wind speed observations. An additional bias adjustment (BA) step is therefore needed as part of the quality assurance of crowdsourced data. While this is not common practice for official wind data, there are numerous other studies within the climate literature (Terink *et al.*, 2009; Gudmundsson *et al.*, 2012; Ho *et al.*, 2012; Teutschbein and Seibert, 2012; Hawkins *et al.*, 2013; Fang *et al.*, 2015; Navarro-Racines and Tarapues, 2015; Akhter *et al.*, 2017; Luo *et al.*, 2018). These are mainly developed for correcting the bias of climatological model output, particularly for temperature and precipitation data. We extend the traditional empirical quantile mapping method found in the literature to adapt the method to be suitable for bias adjustment of crowdsourced wind data.

While the within-station QC cannot detect all inaccurate observations, some could be identified by comparing with nearby concurrent observations. A ‘between-station QC’ is therefore needed to assess the spatial consistency of wind observations at an individual station based on surrounding stations. The between-station QC is particularly advisable for quality assurance of crowdsourced data which have a denser observation network but are of lower quality. However there have been no such studies for crowdsourced wind data to date. Many researchers developed methods to check spatial consistency in their QC studies, mostly for official temperature observations (Eischeid *et al.*, 1995; Daly *et al.*, 2004; Hubbard and You, 2005; Durre *et al.*, 2010; Fiebrich *et al.*, 2010; Lussana *et al.*, 2010; Estévez *et al.*, 2011). In addition to seasonal and diurnal variability, wind observations can change rapidly in space and time due to both small-scale and synoptic-scale effects. Droste *et al.* (2020) noted that urban wind speed and direction are particularly hard to quantify because of the strong turbulent nature of wind. All these features of wind make it difficult to assess the quality, compared to more slowly varying variables such as temperature. To address these unique issues for crowdsourced wind data and assess the spatial consistency, we extend the classical inverse distance weighting method for official stations (Eischeid *et al.*, 1995).

Finally, we develop a novel approach to interpolate the inflated zero wind speeds that typically appear in crowdsourced data during periods of low wind speeds. This approach is an extension of the bias adjustment, and it attempts to fill the missing gaps in the underestimated zero-valued observations. The inflated zero wind speeds were excluded from analysis in previous research on crowdsourced wind quality assurance (Droste *et al.*, 2020), given that light winds are of less interest for climatology studies. Our novel approach provides estimates for the censored observations instead of excluding them, and that is advantageous in a robust QC procedure and its evaluation. The combination of within-station QC and between-station QC, together with BA and the estimates of zero-valued censored data, makes our four-step QC system applicable to wind speed observations from any CWSs.

The remainder of the manuscript is structured as follows. In the next section, we introduce the data required. In Section 3, we develop the four-step QC and BA procedure, which results in a new quality controlled and bias adjusted crowdsourced dataset. Then in Section 4, we compare the output crowdsourced data with official wind speed observations, and we use statistical evaluation metrics to quantify the improvements in the data quality. Finally in Section 5, we outline our conclusions, discuss the limitations of the study and propose avenues of further research.

2 | DATA

We focus on wind speed observations during the three years 2016–2018 (times reported in UTC) from CWSs mainly located in the province of Utrecht, a land area of about 1,500 km^2 with relatively smooth terrain. The crowdsourced wind data are provided by the third-party platform, Weather Observations Website (WOW-NL 2020), and we refer to these CWSs as WOW stations throughout the paper. There are 93 WOW stations in our raw dataset. We exclude stations that do not have enough wind data, and so our study is based on 39 WOW sites that report wind observations for more than one year. For the WOW data obtained for this study, the wind measurements were provided in knots rounded to one decimal place regardless of whether the original unit of measurement was knots or m·s^{-1}. An additional column exists in the dataset to show wind speed in m·s^{-1} adjusted from knots. We also use wind observations from official weather stations as a reference indicating true wind. The official data are accessed from the data centre at KNMI (KNMI, 2020b), and we refer to the 47 official sites spread over the Netherlands as KNMI stations. These stations follow the standard guidelines in WMO (2014), and they report wind speed every 10 min in m·s^{-1}, rounded to two decimal places. We show the locations of 47 KNMI stations in the Netherlands and 39 WOW stations considered in this study in Figure 1, and it clearly demonstrates the higher spatial density of the WOW stations compared to the KNMI sites.

The WOW dataset also contains some metadata information, including longitude, latitude, and sensor height for each station. It is necessary that we trust the location information on longitudes and latitudes as the basis of our QC research. The sensor height information is not always correct as we have found cases of one station having multiple height values, which might indicate an inhomogeneity in the station associated with a change in location and/or sensor height, or it could simply indicate that the metadata have been changed. However, a summarised analysis on the sensor heights suggested that most WOW wind sensors (around 76.7%) are placed at 0–6 m above ground, and those low sensor heights could cause underestimation of wind speed relative to the official 10 m. On the WOW website of each station, there are artificial grades for locations and weather variables, and some stations include an additional description of instrument brand and placement.

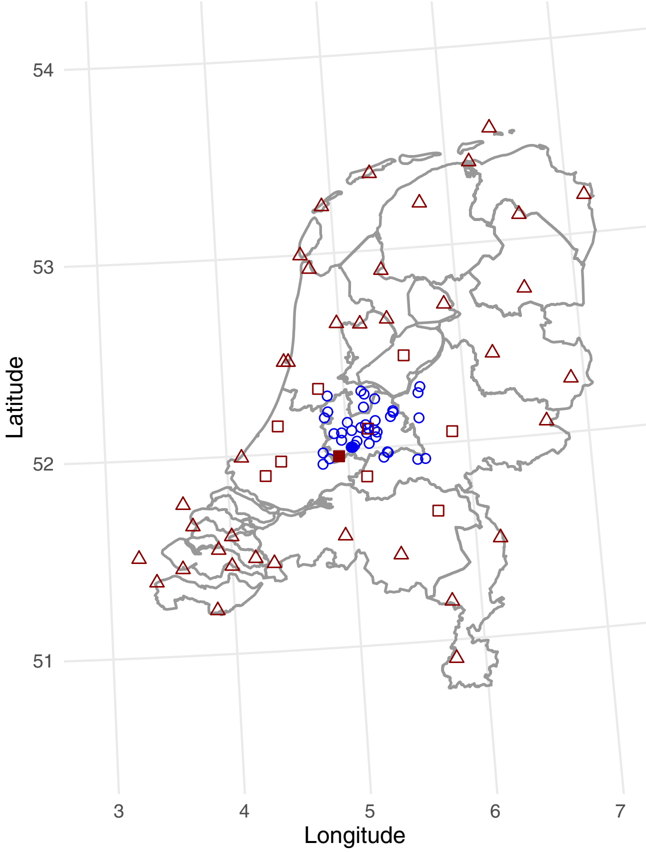

FIGURE 1 Locations of the 39 WOW stations considered in this study (circles) and the 47 KNMI stations (squares and triangles). Squares represent KNMI stations in or around the study region, and triangles are the rest of the KNMI stations in the Netherlands. The two KNMI and WOW stations with filled colour are for the case-study in this work. Grey lines show the provinces of the Netherlands [Colour figure can be viewed at wileyonlinelibrary.com]

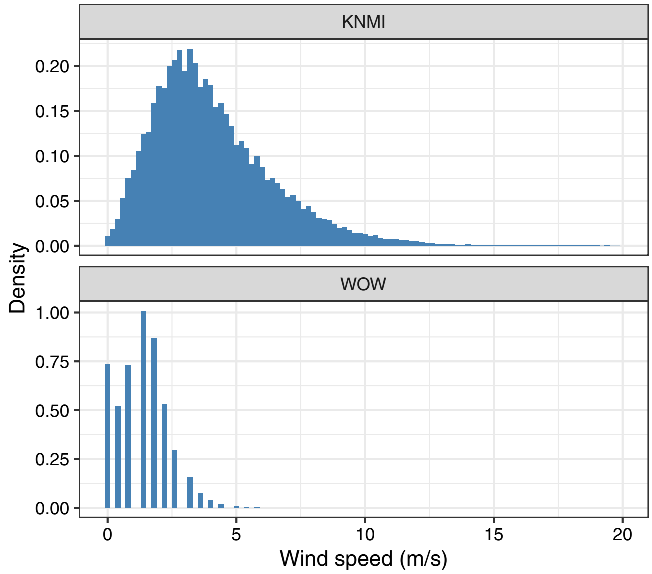

FIGURE 2 Density frequency histograms of wind speed observations from a KNMI station (Cabauw) and an example WOW station (serial number: 956296001). The two stations are highlighted with filled colour in Figure 1 [Colour figure can be viewed at wileyonlinelibrary.com]

All the information is contributed by CWS owners. A previous analysis report on WOW data suggested that the grades for station location and measurement are not always reliable and have little association with the average error (Koole, 2016). Only a few CWS owners provide the manufacturer information, and that makes it impossible to tailor a QC approach to different types of stations. Given this lack of dependable metadata, we develop a data-driven QC and BA procedure that is flexible enough to adjust individual station biases, and is not hindered by a lack of metadata.

According to WMO (2014), the standard reporting frequency is every 10 min with integer time indices (MM:SS; 00:00, 10:00, 20:00, 30:00, 40:00, 50:00), but most WOW stations do not follow such a frequency or timestamp. Popular brands like Netatmo and Davis report observations every 5 min, while some others report every 2 or 10 or 15 min, etc. To pre-process WOW data to be matched with the standard, we aggregate observations every 10 min and take the average as the wind speed at the nearest standard time index. The number of aggregated observations in each 10 min interval depends on the time resolution at different stations, and for stations that report every 15 min or longer we have to match the observation with its nearest standard timestamp. This pre-processing step is necessary as our QC and BA procedure requires 10 min wind speed observations. The within-station QC checks are developed for 10 min average wind speeds, and comparisons with simultaneous official observations are involved in the bias adjustments and between-station QC. The benefits of pre-processing WOW data outweigh any drawbacks, such as the introduction of rounding errors during the averaging.

We notice three kinds of erroneous or suspicious observations in the WOW data based on a first comparison of wind speed density histograms at a KNMI station (Cabauw) and an example WOW station (serial number 956296001), shown in Figure 2. First is the underestimation of the WOW data compared to the KNMI data; we develop a bias adjustment step to address this in Section 3.2. Second is the inflated zero observations in light-wind cases which are reported as zero as a result of underestimation. We introduce a novel step to provide estimates for these censored observations in Section 3.4. Last is the suspect observations that are usually invisible in histograms because they could take on any values. However in our example histogram (Figure 2) the overlarge density peak of WOW wind speed at around 2 m·s^{-1} might indicate

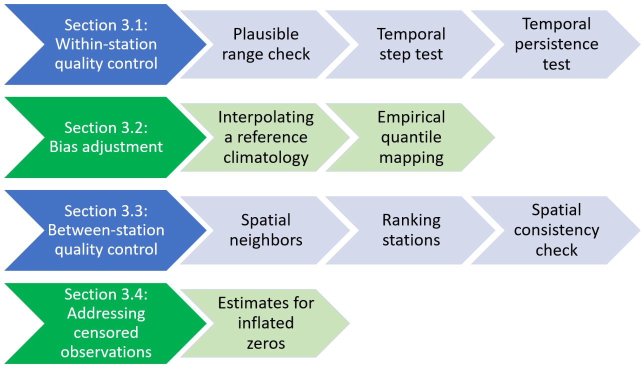

FIGURE 3 Diagram of the overall quality control and bias adjustment system for WOW wind speed observations. The four parts correspond to the four sections in the Methodology; The first and third parts are quality control checks that can be performed on any wind speed data, including official observations; The second and fourth parts are additional bias adjustment steps required for crowdsourced data [Colour figure can be viewed at wileyonlinelibrary.com]

some suspect observations. We identify these suspect values using multiple QC checks in Section 3.1 and 3.3. In addition, WOW wind speed data typically have stratified values (Figure 2). This is likely caused by the low-quality devices, as values from cheap wind sensors are more likely to be rounded. In our study, we focus on detecting inaccurate wind speeds and adjusting instrumentation biases to address these three erroneous types. Correcting for the stratification in the data is left to future work.

3 | METHODOLOGY

In this section, we introduce the standard QC checks for official wind speed observations and extend these steps so they can be applied to crowdsourced observations. Our QC and BA procedure takes the following steps (Figure 3):

(a) within-station (or intra-station) QC (Section 3.1),
(b) bias adjustment (Section 3.2),
(c) between-station (or inter-station) QC (Section 3.3), and
(d) addressing the censored data (Section 3.4).

Following existing QC methods for official observations, we seek to perform both within-station and between-station QC on the WOW data. In the within-station QC, we follow the guidelines by Zahumenský (2004) to check the plausible range and temporal consistency of WOW wind speeds. However, direct application of existing methods to check spatial consistency is not possible due to differences in how WOW data are recorded. WOW stations are typically positioned at different heights, and are usually lower than 10 m. This means WOW observations systematically underestimate the actual wind at 10 m, are not comparable in space with KNMI data which report wind at 10 m. Adjusting the underestimated WOW data is therefore necessary before checking the spatial consistency, which is addressed in the BA step. Once the WOW data have been adjusted, the between-station QC can be performed among neighbouring stations. The systematic underestimation in WOW data feeds into a secondary censoring problem – most WOW stations fail to record low wind speeds and instead report a zero observation. In the last step of our QC and BA procedure, we provide estimates for the missing low wind speeds by spatial interpolation which is specifically developed for WOW wind data.

3.1 | Within-station quality control

The standard within-station QC steps for automatic weather stations check the plausible range and the temporal consistency of observations, where the temporal checks consist of a step test and a persistence test (Zahumenský, 2004). These steps can be applied to WOW data without alteration from the standard guidelines, and we summarise these in Table 1. If an observation fails one of these within-station QC checks, it is flagged and will be filtered, i.e., it is not considered in any further analysis.

3.1.1 | Plausible range check

The range check examines whether an observation is within acceptable limits and detects implausible values. For instance, a WOW wind speed observation higher than the world record, or a negative value should be flagged and rejected. We set the upper bounds based on historical KNMI highest wind speed records in the Netherlands (red triangles and squares in Figure 1). This serves as an intelligent upper bound as WOW stations systematically underestimate the actual wind speeds.

3.1.2 | Temporal consistency check – Step test

The step test is part of the temporal consistency check and aims to identify excessive variability with time. Wind speed observations that have an overly large rate of change could indicate erroneous recording values of the wind sensor. For example, a WOW station might report an observation with an incorrect decimal place, making the value ten times

Within-station QC	Conditions to pass
Plausible range check	Between 0 and 35.0 $m{\cdot}s^{-1}$
Temporal consistency check – step test	Changes of no more than 13.88 $m{\cdot}s^{-1}$ in 10 min
Temporal consistency check – persistence test	Changes of at least 0.05 $m{\cdot}s^{-1}$ in 160 min

TABLE 1 Summarised within-station quality control steps for WOW wind speed observations

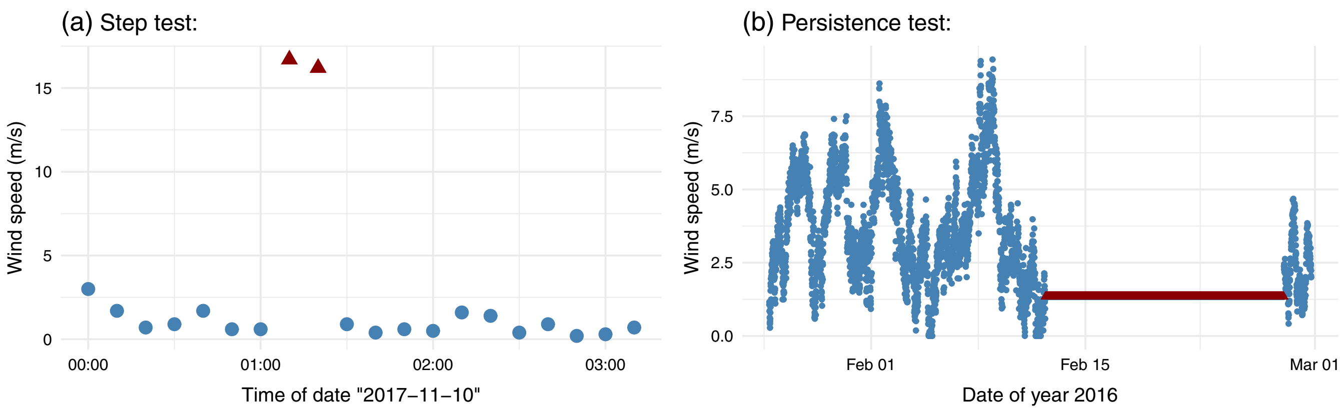

FIGURE 4 Examples of accepted (dot points) and failed (triangle points or bar) wind speed observations in the two parts of the temporal consistency checks: (a) the step test, and (b) the persistence test [Colour figure can be viewed at wileyonlinelibrary.com]

larger than it should be, e.g., from 1.698 to 16.98 $m{\cdot}s^{-1}$, as shown in Figure 4a. There are two parameters needed for the step test – the length of the time window and the maximum variation. We determine these parameters from the wind data of ten KNMI stations in and around the study region (squares in Figure 1). The biggest step change in the standard 10 min KNMI wind data during the three years 2016–2018 is 13.88 $m{\cdot}s^{-1}$. WOW wind speeds are expected to change less rapidly than KNMI data due to the low precision of cheap devices. We therefore take 13.88 $m{\cdot}s^{-1}$ and 10 min as the two parameters.

3.1.3 | Temporal consistency check – Persistence test

The final step of within-station QC is the persistence test which aims to detect abnormally low variability in wind observations with time, particularly consecutive constant values over a long period. In Figure 4b we show an example of WOW observations keeping constant for about 15 days, which is likely caused by an equipment error. To perform the persistence test, we require two parameters, a time window and the minimum change acceptable during this time window. The WOW wind speeds are given in knots rounded to one decimal place, meaning that the minimum gap between two different observation values is 0.1 knot ($\approx$ 0.05 $m{\cdot}s^{-1}$). We therefore use 0.05 $m{\cdot}s^{-1}$ as the minimum change. The time window is determined based on wind data at the ten KNMI stations which are also used in

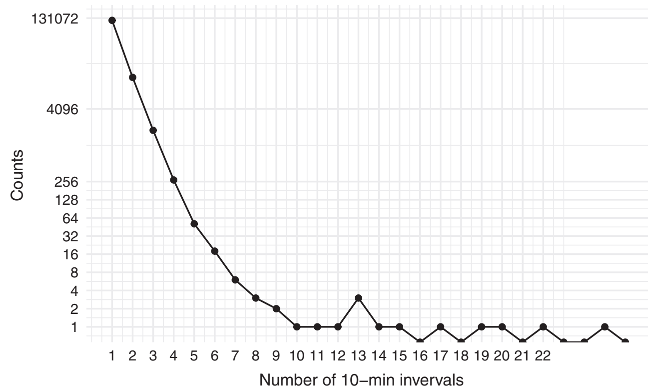

FIGURE 5 The lengths of persistent periods (wind speed change of less than 0.05 $m{\cdot}s^{-1}$) in KNMI data

the step test. We analyse the time periods in which KNMI wind speeds fluctuate less than 0.05 $m{\cdot}s^{-1}$, and show the frequency regarding the persistent time length in Figure 5. We find that the persistent periods are mostly less than 30 min, and the frequency drops to zero at 160 min (i.e., sixteen 10 min intervals). A corresponding window length of 160 min is therefore conservatively chosen based on our analysis of KNMI data.

For the persistence test, there are two situations that need to be treated differently. One is spurious data, likely caused by equipment errors, and the other is censored data, as a result of censoring of low wind speeds. Persistent spurious wind speed observations are usually non-zero or extensive long-lasting zeros, while persistent censored wind speed observations are consecutive zeros over short time periods. We reject only the first case in our persistence test, flagging non-zero observations that fail the condition, and flagging consecutive zero observations if the period lasts for more than 2 days. The 2-day time period comes from our investigation of the same ten KNMI stations during light-wind periods. The time periods when KNMI wind speed observations are below 2.5 m·s^{-1} are mostly shorter than 2 days, with only ten cases exceeding 2 days. Therefore, WOW zero observations repeated for more than 2 days are much more likely to be erroneous rather than underestimated. Differentiation of these two cases is important, as the different mechanisms behind the missing data need to be treated differently in the later BA step. The spurious data should be excluded while the censored data should be kept as they hold an indispensable probability mass in the WOW wind speed distribution (Section 3.2.1).

3.1.4 | Filtering low-quality stations

After the plausible range and temporal consistency checks, the quality of some WOW stations is too low to be considered further. We identify those stations by considering three aspects:

(a) the Pearson correlation with nearby KNMI stations (less than 0.5),

(b) the percentage of failed observations in the persistence test (more than 15%), and

(c) the percentage of zeros after the persistence test (more than 35%).

Also we manually check the selected low-quality stations to determine whether they should be excluded from further analysis. The 0.5 correlation threshold and 15% percentage threshold are determined based on a general inspection of the data, and we find that most WOW stations and all KNMI stations satisfy these conditions. A station has either poor association with the true climatology or low-quality data in terms of temporal persistence if these conditions are not met. The 35% zero percentage threshold is determined based on the percentage of low wind speeds (less than 2.5 m·s^{-1}) at two KNMI stations in the study region. We consider it is normal that WOW stations report 0 m·s^{-1} during low wind speeds, but the >35% situation indicates excessive censored data. For a different dataset or another observing region, one could change the thresholds accordingly. After the manual checks, we removed seven low-quality WOW stations out of the original 39 stations, leaving 32 WOW stations with acceptable quality for further analysis.

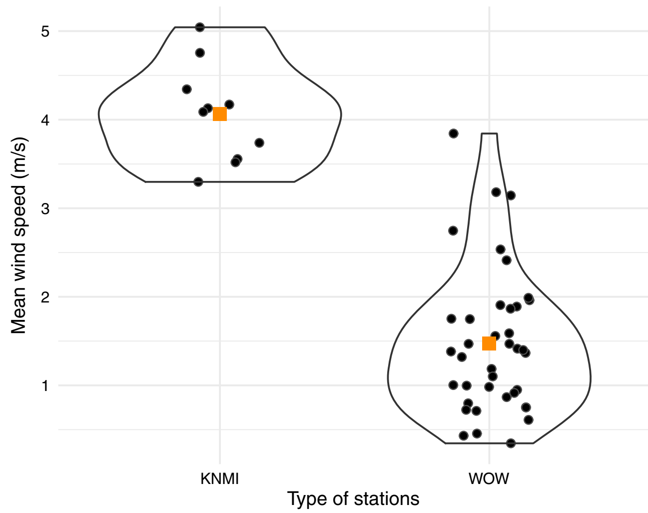

FIGURE 6 The violin plot shows the mean wind speed observations at WOW stations (right) and ten nearby KNMI (squares in Figure 1) stations (left) during the 2016–2018 period. The solid lines show the distribution of each group, while the square points indicate the average of each group [Colour figure can be viewed at wileyonlinelibrary.com]

3.2 | Bias adjustment

Many WOW stations measure wind speeds that are systematically lower than those observed by KNMI stations, as can be seen from the mean wind speed at each station in Figure 6. This underestimation is likely caused by the typically low siting of the wind sensors (usually about 0–6 m above ground compared to the 10 m standard recommended by WMO (2014)) and surrounding obstructions, especially in urban locations. In order to complement the KNMI official network with WOW wind speed observations, WOW data should be adjusted so that their climatology is comparable to that of KNMI stations. This is important for the spatial consistency check and for any future uses of WOW data. The step is commonly known as bias-correction in meteorological studies, and has been largely developed for temperature and precipitation data output from climate models (Terink *et al.*, 2009; Gudmundsson *et al.*, 2012; Ho *et al.*, 2012; Teutschbein and Seibert, 2012; Hawkins *et al.*, 2013; Fang *et al.*, 2015; Navarro-Racines and Tarapues, 2015; Akhter *et al.*, 2017; Luo *et al.*, 2018). We call it bias adjustment in our study because there are no truth data as a correct reference at the WOW locations.

Among existing methods for BA, there are both parametric and non-parametric models developed. We investigated multiple parametric methods but found that they were not well-suited for WOW wind data, partially due to poor distributional assumptions (Chen, 2020). We therefore use the non-parametric method of empirical quantile mapping (EQM). EQM has been applied for temperature and precipitation data in previous research (Gudmundsson *et al.*, 2012; Fang *et al.*, 2015; Navarro-Racines and Tarapues, 2015; Akhter *et al.*, 2017; Luo *et al.*, 2018), and we adapt the method here for crowdsourced wind speed observations.

3.2.1 | Empirical quantile mapping

EQM is used to map the quantiles from the WOW cumulative probability distribution to the corresponding quantiles in a target distribution that represents the true 10 m wind speed. In this way, we calibrate the climatology but retain day to day variability. Let the empirical cumulative distribution function (CDF) of the WOW observations be $\hat{F}_X(x)$, and let the empirical CDF of the true 10 m observations be $\hat{F}_Y(y)$. The WOW data are bias adjusted using the mapping

$$\tilde{x} = \hat{F}_Y^{-1}\left(\hat{F}_X(x)\right). \tag{1}$$

Note that the 10 m wind speed is not observed at the WOW locations, and therefore the target distribution is not given directly. Instead, we determine the target distribution by interpolating the climatological quantiles, details of which are given in Section 3.2.2. We use empirical CDF for WOW data to avoid distributional assumptions, and to easily address the presence of the inflated zeros. Manual inspection reveals that the zero inflation holds a distributional mass representing the cumulative probability of actual low wind speeds. This probability mass cannot be ignored when estimating the empirical CDF. However, EQM maps all the zeros to the same value that corresponds to the appropriate quantile in the true distribution, despite the fact that they actually represent various low wind speed values. We develop a supplementary step after the between-station QC to provide estimates for these missing low wind speeds (Section 3.4).

The wind climatology of the Netherlands suggests that wind varies seasonally and diurnally, and so it is necessary to perform the EQM bias adjustment for different seasons and times of day as the statistical distributions would be different. Investigation of the wind speed distributions for both WOW and KNMI in different months and hours of the day confirmed that wind speed is typically lower during the night than the day and it is stronger in winter than in summer. Based on inspection, we concluded that splitting the data into three seasons and two diurnal periods is a suitable choice. We set December to March as the extended winter season, and June to September as the extended summer season, and the remaining months as the transition season; we also define 0700–1800 UTC as the daytime period, and the remainder as the night-time period. Following such division, we perform EQM BA on six different periods separately and combine the output data afterwards. This extra splitting step introduces minor discontinuities, but manual inspection reveals that general trends in observations are retained. We intend to avoid these discontinuities in future work, using approaches such as quantile regression, but it was beyond the scope of this study. A quantile regression approach could also be adapted to account for variation in sunrise time throughout the year. For WOW observations in each period, we apply five-fold cross-validation to avoid overfitting, dividing the observations into five consecutive groups, each containing the same length of valid data. In each round, we take four groups as the training set to train the EQM model, and then use this to perform BA on the remaining one group. We aggregate the results of those five groups, and in this way, the biases in WOW data are adjusted not entirely according to the true climatology to be compared afterwards. The six implementing periods together with five-fold cross-validation ensure the EQM BA is able to be generalised to any observation dataset.

3.2.2 | Spatial interpolation of the wind speed climatology

In EQM BA, the true wind speed at a WOW location is required to estimate the target distribution. However, the sparsely distributed KNMI stations are not capable of providing a true wind reference at the locations of many WOW stations. We therefore interpolate spatially the wind speed climatology of KNMI stations to estimate the true wind that accounts for spatial variability at WOW stations. We interpolate the quantiles of KNMI wind speed distributions rather than every single observation in the wind speed time series as it is more stable with much lower computational cost. This approach also fits with the EQM method.

We take 100 quantiles (the 1st to 100th percentile) from wind speed distributions at 46 KNMI stations (stations drawn in triangles and squares with no filled colour in Figure 1, i.e., KNMI stations excluding Cabauw) in the three years 2016–2018, and spatially interpolate the 46 data points to estimate a smooth surface of each percentile in wind climatology over the Netherlands. We then extract the corresponding percentile at a WOW location from the estimated quantile surface. Occasionally, there are crossovers among these 100 quantiles. We reorder them to avoid crossing quantiles, and the resulting empirical CDF is used as the true target distribution at the WOW location in EQM.

To interpolate the spatial quantile surface, there are multiple geostatistical approaches, including kriging, inverse distance weighting, and thin-plate spline regression, which have been tested and compared on temperature or precipitation data (Stahl *et al.*, 2006; Hofstra *et al.*, 2008; Moral, 2010; Contractor *et al.*, 2015). We choose the ordinary kriging method for this application, as our testing demonstrated it reliably captured the wind behaviour along the coasts of the Netherlands, generating smooth interpolated surfaces. Ordinary kriging is an approach to provide unbiased linear estimates for unsampled values (Wackernagel, 2003). It predicts the target observation at an unknown location by taking a weighted average of surrounding reference data, where the weights are based on the spatial variance of the input data with location information. To fit the relationship between the spatial variances and physical distances, we need a variogram model. We use the exponential model for the variogram due to its stability in optimisation. To perform the ordinary kriging we use the function *autoKrige* in the R package *automap* (Hiemstra *et al.*, 2008). Within the kriging method, a roughness map could be incorporated to account for terrain or urban effects, but it is beyond the scope of this work.

In Figure 7a we show an example interpolated surface of the 50th percentile from wind speed distribution during the transition season in the daytime period. We see that wind speed is higher in coastal areas as well as around the inland bay IJsselmeer, which agrees with the known climatology of the Netherlands. The 50th percentile of observed wind speeds at the 46 KNMI stations are also shown, and it can be seen that most KNMI percentiles are very close to the interpolated estimates. In this way, we can obtain a sequence of 100 estimated quantiles that indicate the true climatological wind speed distribution at any location given the longitude and latitude. We leave one KNMI station (Cabauw) out of the analysis to illustrate the agreement between the kriging interpolated and real quantiles, shown in Figure 7b.

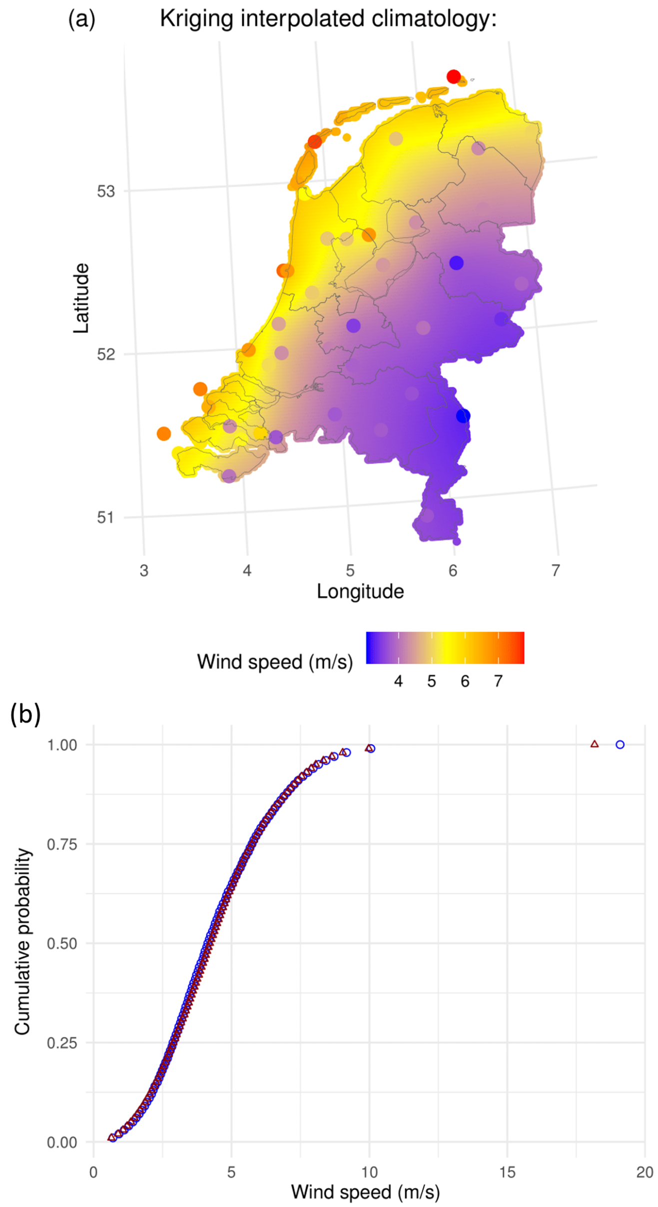

FIGURE 7 (a) The interpolated 50th percentile surface of the wind speed distributions during the transition season in the daytime period, which are estimated with ordinary kriging from 46 KNMI stations (excluding Cabauw). (b) Estimated quantiles at the location of Cabauw returned by interpolation (triangles) and quantiles of the observed wind speed distribution (circles) at Cabauw, both during the transition season in the daytime period [Colour figure can be viewed at wileyonlinelibrary.com]

3.3 | Between-station quality control

The between-station QC compares simultaneous observations between multiple neighbouring stations to detect observations that are spatially inconsistent. It should be performed when the BA has mapped WOW observations at various heights to match the wind climatology of KNMI stations, meaning the observations are now comparable in space. The between-station QC for crowdsourced wind data has not been studied before, to our knowledge, likely due to a lack of robust BA. We adapt and extend existing methods developed for temperature and official wind observations to check the spatial consistency of WOW data.

In the between-station QC, we refer to the WOW station to be examined as the candidate station and its observation as the target observation. To check spatial consistency, we estimate a confidence interval for the target observation based on a set of its neighbouring observations, and identify the target as spatially inconsistent if it lies outside the confidence interval. Two challenges therefore arise; one is to select comparable neighbouring stations, the other is to estimate reliable confidence intervals.

3.3.1 | Selecting neighbouring stations

The most common way to choose the reference neighbouring stations for a candidate station is to select nearby sites by geographical distance. However, local terrain and nearby obstructions have an impact on wind speed measurements (Fiebrich *et al.*, 2010), which is more frequently seen for WOW stations with non-standard sitings. Another uncertainty comes from the coastal areas, where wind behaves differently from inland locations (Gatey and Miller, 2007). These factors suggest that simply choosing nearby stations using geographical distance for a candidate's reference will not always be suitable. Instead it is prudent to consider the statistical similarity between the wind speed distributions at two stations.

The neighbouring stations for a candidate site should be both geographically close and distributionally similar. We limit the potential neighbouring stations to those within 75 km and consider two statistical indicators of similarity. The two indicators are Pearson correlation and Earth mover's distance (EMD; the shaded area in Figure 8). We introduce the EMD in selecting neighbouring stations for the first time, while Pearson correlation has been considered to determine reference stations for spatial consistency check on temperature data in previous work (Hubbard and Sivakumar, 2001). The EMD considers the scaling and shifting differences between two distributions. Additionally, the EMD is a distance metric while the Pearson correlation is not. The distance metric property is advantageous in replacing the Euclidean distance in the later inverse distance weighting (Section 3.3.2).

The EMD, also known as the transportation distance or Wasserstein distance, is a mathematical metric that measures the total costs in transporting a distribution of mass into another distribution with an optimal strategy (Kantorovich, 1960). The EMD indicates the total distance between two probability distributions, as shown in Figure 8. Given empirical CDF of two wind speed distributions, $\widehat{F}_X(\cdot)$ and $\widehat{F}_Y(\cdot)$, the EMD is given by

$$\mathrm{EMD}(X, Y) = \int_{-\infty}^{\infty} \left| \widehat{F}_X(x) - \widehat{F}_Y(x) \right| \, \mathrm{d}x. \tag{2}$$

We use the function *Wasserstein1d* in the R package *transport* to compute the EMD (Schuhmacher *et al.*, 2020).

Use of EMD and Pearson correlation ensures both statistical closeness in distributions and a linear association for two stations. Our criteria for selecting neighbours could therefore be stated as: taking nearby stations (within 75 km) that hold a correlation >0.7 and an EMD <1 between the candidate site and its reference neighbours. The constraints for correlation and EMD are determined based on the ten KNMI stations nearby or in our study region (squares in Figure 1). The correlation of wind speed observations between any two KNMI stations is higher than 0.7. The EMD of wind speed distributions between any two KNMI stations is lower than 2, with 80% lower than 1. Occasionally some WOW stations do not have enough neighbours that satisfy our criteria. If the number of selected neighbouring stations is less than 10, we relax the correlation constraint from 0.7 to 0.6. Ideally, the nearest reference stations are chosen as neighbours, but it is not always the case for WOW stations or stations near the coast which may have different wind regimes; this is why we do not choose neighbours based only on geographical distances.

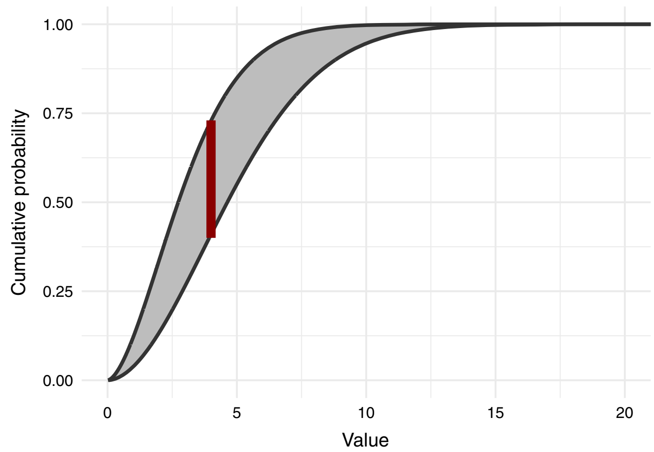

FIGURE 8 Two cumulative probability distributions, together with their Earth mover's distance (the shaded area between two lines) and their KS distance (the vertical bar, i.e., the maximum vertical distance between the two lines) [Colour figure can be viewed at wileyonlinelibrary.com]

3.3.2 | Estimating confidence intervals

By considering three factors, the geographical distance, the Pearson correlation, and the EMD, we get a list of comparable neighbouring stations for each candidate WOW station. The confidence interval for the target observation can then be estimated based on simultaneous wind speed observations at neighbouring stations. We extend previous work on the spatial consistency checks for temperature data, by first investigating the classical inverse distance weighting (IDW; Eischeid *et al.*, 1995; Peterson *et al.*, 1998) and the spatial weighted regression test (SRT; Hubbard *et al.*, 2005, 2007; You and Hubbard, 2006; Estévez *et al.*, 2018), but find that their adaptations to wind speed data are not satisfactory. The classical IDW depends on only geographical distance making it not suitable for wind variables with directional dependence, and the SRT is not suitable for wind as it is a more stochastic variable than temperature. The ordinary kriging method introduced in Section 3.2.2 is not suitable here either, because it is computationally intensive and the high variability of wind does lead to unstable interpolated surfaces.

We develop the inverse EMD weighting (IEMDW) method, as an adjusted version of the classical IDW. IDW estimates the mean of the target by a weighted average of its neighbouring observations, where the weights are inversely proportional to their geographical distance from the target. In IEMDW we incorporate statistical similarity into the equation to avoid shortages of IDW that assign more weights to the nearest stations regardless of the actual climatology difference. We replace the geographical distance with EMD, and so the weights are determined by statistical similarities, giving larger weights to neighbours that are distributionally similar with the target. Compared to SRT and kriging, the IEMDW is computationally cheap and generates results efficiently in real-time processing, since the weight for each neighbouring station is fixed and recycled.

A neighbouring station with a small EMD, meaning statistically similar with the target station, is assigned with a large weight, and vice versa. Given a target observation y and its neighbouring observations $\{x_n : n = 1, 2, \dots\}$, the expression of the estimated target mean $\hat{y}$ is

$$\hat{y} = \frac{\sum_n \omega_n x_n}{\sum_n \omega_n}, \tag{3}$$

where ω_n is the weight for each neighbouring station. We use the Cressman method to assign the inverse EMD weights (Cressman, 1959),

$$\omega_n = \frac{r^2 - \mathrm{EMD}^2(n_0, n)}{r^2 + \mathrm{EMD}^2(n_0, n)}, \tag{4}$$

where $\mathrm{EMD}(n_0, n)$ is the EMD between the candidate station n_0 and a neighbouring station n. We set $r = 1$ to ensure the weights are positive real numbers, as the selected neighbouring stations are restricted to having an EMD value less than 1.

We also need the standard deviation to determine a confidence interval, and that is estimated from the variance of neighbouring observations. To avoid non-physical negative wind speed values, we use a truncated normal distribution which truncates the left-side lower tail of the normal distribution at zero and is restricted to positive values, as seen in other studies of wind speed data (Thorarinsdottir and Gneiting, 2010). The confidence interval is between the 0.005-quantile and 0.995-quantile of the truncated normal distribution. If the target observation lies outside this confidence interval, it is considered spatially inconsistent and is rejected by the between-station QC.

We want to start with WOW stations that have more high-quality neighbouring stations, and after checking spatial consistency, these WOW stations are more reliable as a reference. In practice, we define a rating for each candidate station that indicates the average quality of its neighbours, and the implementation order for WOW stations is then determined by descending order of their ratings. KNMI stations are assigned the highest quality grade of 2, along with WOW stations that have been assessed for spatial consistency. WOW stations that are not yet assessed are assigned quality grades based on their highest correlation with KNMI stations; those with correlation larger than 0.8 get a quality grade of 2, those with correlation larger than 0.7 get 1, while the rest get 0. The rating for a candidate station is given as the mean of its neighbours' quality grades. We keep updating the WOW data with observations that have passed the between-station QC for subsequent rounds, ensuring the reliability of reference neighbouring observations.

3.4 | Interpolating the missing low wind speeds

A common issue with the WOW wind speed data is the inflated zero observations. There are two types of zero observations that we deal with differently. The observations that keep constant zeros for more than 2 days are rejected by the temporal persistence test in Section 3.1.3. The remaining inflated zeros are censored data that results from an underestimation during periods of light winds. This could be caused by poor exposure of devices (e.g., near a building or tree), overly low siting (e.g., at 2 m above ground), or measurement devices that are not sensitive enough. We provide estimates for these censored low wind speeds by interpolation, rather than mapping them to a fixed value in the EQM BA. We apply a similar approach as in the IEMDW method to estimate the wind speed at the times when WOW stations recorded suspect zero values. Addressing the censored low wind speeds is the final step in our QC and BA procedure, and can be treated as an extension of the EQM BA. For each zero wind speed observation, we extract the existing selected neighbouring stations together with corresponding inverse EMD weights, and estimate the mean (weighted average) to fill in the censored zero values. A few IEMDW estimates are higher than the EQM bias adjusted values, suggesting the observations are not caused by systematic underestimation. These estimates are therefore not accepted, and labelled as missing data in the final output.

4 | RESULTS

We assess the quality of crowdsourced wind data and the overall performance of our complete QC and BA procedure (Figure 3) in two ways. First, we assess the percentage of observations that fail each step of the QC checks. Second, we apply three statistical evaluation methods to quantify the improvements in data quality after implementing the procedure. To demonstrate the method, we leave one KNMI station (Cabauw) out of all stages in the QC and BA procedure so that we have a more fair assessment of the improvement of WOW stations. Later in this section, we compare Cabauw with one nearby WOW station (about 6 km away) as an example to illustrate that WOW observations are more comparable after QC and BA.

Few observations are rejected by the range check and the temporal step test compared to the temporal persistence test, as the left two light grey bars of each station are barely visible in Figure 9. This demonstrates that one of the main error sources on wind speed observations by WOW stations is the constant recordings. The results show that many WOW stations have less than 5% of their observations flagged by the within-station QC, which means more than 95% of the WOW wind data have plausible ranges and acceptable variability over time. There are some WOW stations with too many flagged observations (Figure 9). We manually check those stations and find that two of them (serial numbers 923556001 and 951366001) kept recording zero wind speed for more than one month. However, the remaining observations that pass the within-station tests are useful to consider. These two stations are therefore retained in the dataset for further steps by truncating their constant observations. We do remove seven other stations from further analysis because of their low quality (the starred stations in the grey box in Figure 9), as outlined in Section 3.1.4.

Based on the percentage of wind speed observations that fail the between-station QC (black bars in Figure 9), we find that for most WOW stations, about 2% to 8% of observations are rejected for spatial inconsistency. We examine the locations of WOW stations, and find that sites in the open countryside tend to reject fewer observations, which suggests that surrounding obstructions especially in urban areas might be a cause of spatially inconsistent observations. For most WOW stations that are considered in our study, more than 85% of wind speed observations pass all QC checks. This is a promising outcome showing that WOW wind data after QC can provide an informative dataset.

We consider three statistical evaluation indicators, root mean square error (RMSE), Kolmogorov–Smirnov statistic (KS distance; Figure 8), and Pearson correlation, to analyse the similarity of wind speeds between stations and quantify improvements in data quality. We match each WOW station with the KNMI station that has the highest Pearson correlation with the WOW station's final data. In this way, every WOW station is compared to an official site with credible wind speed observations as a reference. We plot the change of the three statistical indicators at each station after performing the QC and BA procedure in Figure 10. In addition, we calculate the same indicators for each KNMI station (ten sites located nearby or in our study region, squares in Figure 1) with its highest correlated KNMI station, so we have a benchmark of natural variability. All three indicators are greatly improved after our QC and BA steps, and they are much closer to the KNMI benchmarks. For most WOW stations, the final Pearson correlation is higher than 0.7 (Figure 10c), indicating a solid linear association between the final WOW data and the official KNMI data. Both RMSE and KS distance are much reduced in the final WOW data, and they are close to the KNMI benchmarks (Figure 10a,b). A detailed analyses (not shown) finds that the two indicators achieve their greatest improvement after EQM BA and estimating the inflated zeros, as the systematic underestimation is one of the primary sources of biases. The results on the three evaluation indicators show that the final WOW data, after the QC and BA procedure, are statistically similar to the official data and that they can be used in climatological and meteorological studies.

We choose an example WOW station to demonstrate each step of the QC and BA procedure. We compare this WOW station[1] (serial number: 956296001) to its nearby KNMI station (Cabauw), which is also the most correlated KNMI station. In Figure 11a we compare the cumulative density distributions of WOW wind speeds at different stages of the QC and BA procedure, together with the KNMI Cabauw distribution as a reference. There is a substantial horizontal gap in cumulative distributions between the raw WOW data and KNMI data (Figure 11a), indicating a sizeable systematic bias. Our EQM BA eliminates the gap, as the WOW distribution after BA coincides with the distribution of KNMI data, although the left censoring with zeros is still evident. After estimating those zeros in the final step of the QC and BA procedure, the lower tail in the cumulative distribution is interpolated

[1]The station (serial number 956296001) can be visited at https://wow.knmi.nl/#956296001 and https://wow.metoffice.gov.uk/observations/details/20210520o7tw4cf3jre6znxqyyb96scefw

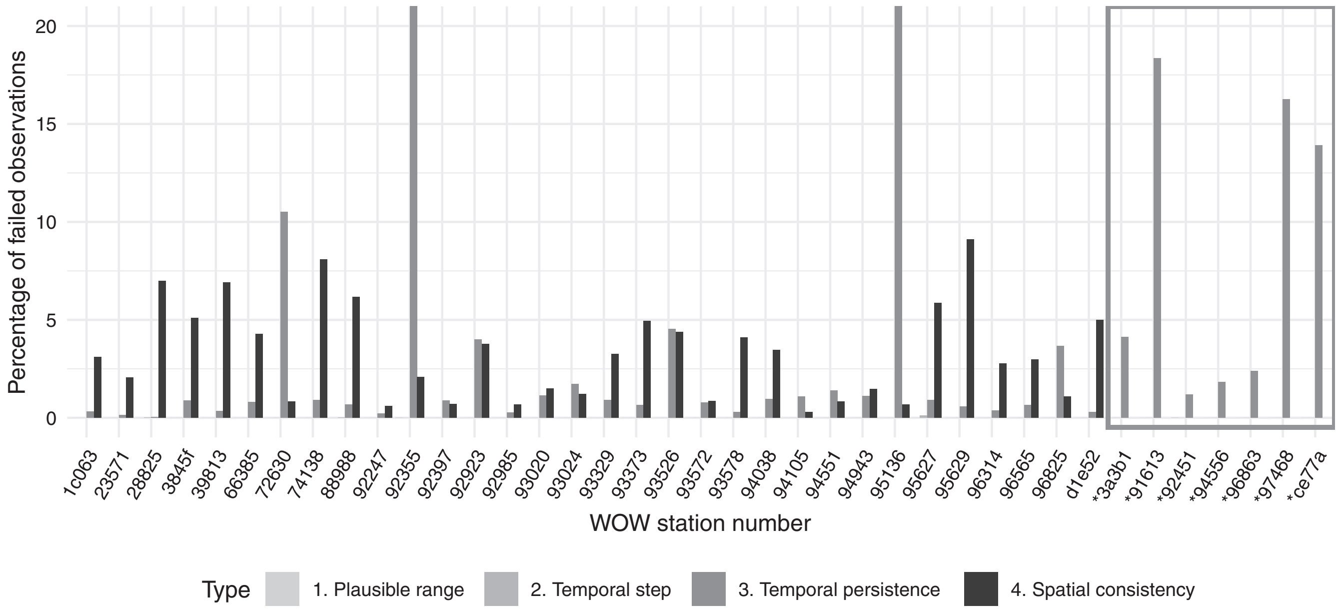

FIGURE 9 The percentage of flagged observations at each WOW station by (1) the plausible range check, (2) the temporal step test, (3) the temporal persistence test, and (4) the spatial consistency check. At each station on x-axis, the four bars are placed in the same order. The light grey bars for plausible range and temporal step are very small. The seven starred stations inside the grey box are low-quality stations which are excluded before checking spatial consistency. We truncate the serial number of WOW stations to the first five letters for a better view

and is found to be similar with the KNMI lower tail, suggesting the censored low wind speeds are estimated nicely in terms of climatology. The stratification in the raw WOW data has been reduced after EQM BA (Figure 11a). This is mainly because we perform BA in six periods separately, and so the same wind speed in different periods could then be calibrated to distinct values due to their separate quantile mappings. We also compare the simultaneous relationship between WOW and KNMI wind speeds before and after performing the QC and BA procedure in Figure 11b. There is a bias towards lower wind speeds in the raw WOW data than in the KNMI data, as most data points are below the diagonal. After the QC and BA steps, the final WOW data points are mostly spread around the diagonal, suggesting the final WOW wind data are comparable with KNMI data.

Finally, we plot the cumulative probability distribution of all 32 WOW stations in the final dataset in Figure 12, for a comparison of the climatologies. The distributions of the raw WOW wind speed are quite dissimilar, with empirical CDFs of various shapes, stemming from various levels of systematic biases and stratified values. Our generalised QC and BA procedures that are customised for individual stations improves the distribution of the raw WOW wind speed data considerably. We choose ten KNMI stations that are located nearby or in our study region (squares in Figure 1) to represent the true climatology of wind in that area. The cumulative distributions of the final WOW data are mostly inside the range of the KNMI wind speed distributions, showing that WOW wind speeds after the QC and BA procedures have a similar climatology to the KNMI stations. This is a promising outcome suggesting that the final WOW data coincide well with KNMI wind climatologies, and it demonstrates that WOW data have the potential to be of added value in a climatology analysis.

5 | CONCLUSION AND DISCUSSION

In this study, we developed a comprehensive procedure to improve the data quality of crowdsourced wind observations. This is the first study to examine the quality of wind speeds observed by many citizen weather stations from a heterogeneous network comprised of various manufacturers' devices, distributed over a large area (about 1,500 km^2). The procedure includes both QC and BA, as summarised in Figure 3. The QC follows from existing methods for official data, and we made essential adaptations and enhancements to make the methods applicable to crowdsourced data. We performed both within-station and between-station QC tests, identifying observations with implausible ranges, too much or too little temporal variability, and spatial inconsistencies. The BA step is an extension specifically for crowdsourced wind data, due to the systematic underestimation of wind at citizen weather stations. We applied different approaches to adjusting zero and non-zero wind speed observations separately.

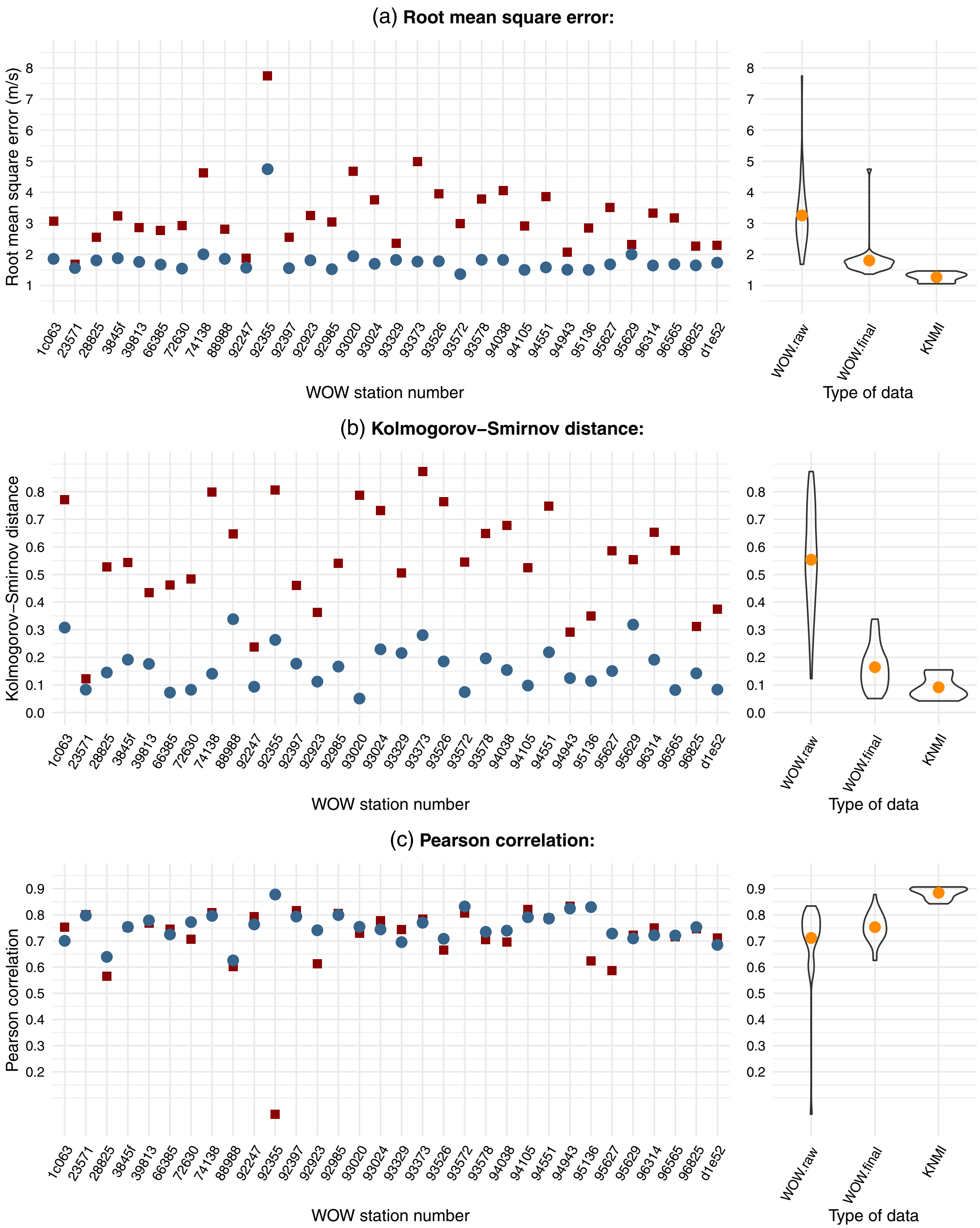

FIGURE 10 Improvements in (a) the RMSE, (b) the KS distance, and (c) the Pearson correlation between WOW stations and their reference KNMI stations after the complete quality control and bias adjustment. The dot plots (left) show the evaluation indicators for each WOW station. The indicators are evaluated for the raw WOW data (square points) and final WOW data (dot points). The violin plots (right) show the distribution of indicators of all WOW stations, including their average (dot points), together with KNMI benchmarks [Colour figure can be viewed at wileyonlinelibrary.com]

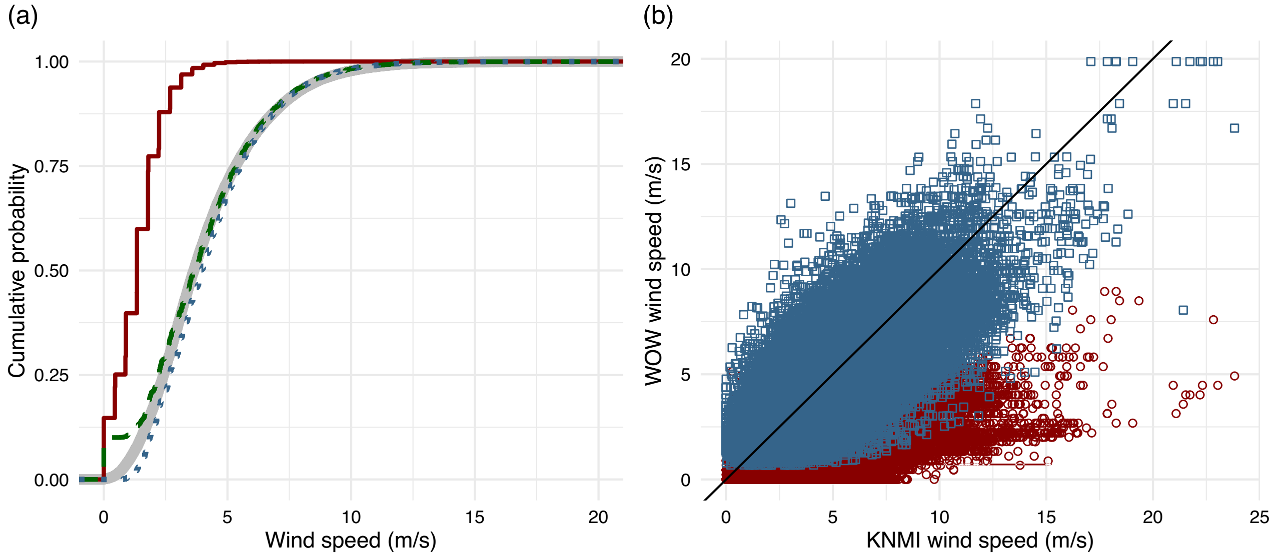

FIGURE 11 QC and BA results of an example WOW station (serial number: 956296001) compared with a KNMI station (Cabauw). (a) Cumulative density distribution of WOW wind speed in raw data (solid line), after the within-station QC and the EQM BA (dashed line), and after the between-station QC and estimating censored low wind speeds (dotted line), with KNMI wind speed (grey thick line) as a reference. (b) Scatter plot of simultaneous observations between the WOW and the KNMI stations, showing both the raw WOW data (circles) and final WOW data (squares) [Colour figure can be viewed at wileyonlinelibrary.com]

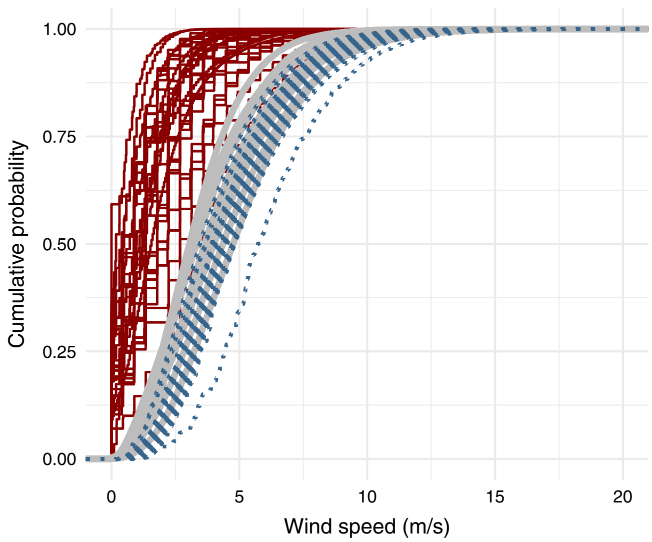

FIGURE 12 Cumulative distribution of all WOW stations with raw data (solid lines) and after the overall QC and BA procedure (dotted lines), with the distribution of ten nearby KNMI stations (grey thick lines) as reference [Colour figure can be viewed at wileyonlinelibrary.com]

For non-zero wind speeds, we used the empirical quantile mapping (EQM), a common BA method for climate models' outputs (Gudmundsson *et al.*, 2012). For the censored low wind speeds that are falsely reported as zero, we provide estimates of the true values by interpolating from surrounding observations. This is a necessary extension of the EQM BA as otherwise the zeros would be mapped to a fixed value. Our methods are more data-driven than in classical approaches, making them notably appropriate when lacking metadata information.

Our research is the most comprehensive study of crowdsourced wind data to date, following on from the pioneering work of Droste *et al.* (2020) who examined wind speed observations in a single city with an urban setting. Their study had the advantage that CWSs are from the same brand (Netatmo) and a network of high-quality stations in the city is provided as a reference, and so there are lower root mean square errors in their results. The quality assurance protocol by Droste *et al.* (2020) filters implausible locations and unfavourable meteorological circumstances, and it truncates all the low wind speeds. In comparison, our QC system is more data-driven and can reject inconsistent observations temporally or spatially, with the capability to be applied to various brands of devices. We perform the spatial interpolation to correct falsely inflated zeros, which improves upon their method which truncates the censored low wind speeds. Overall, our results show a more generalised and comprehensive assessment of crowdsourced wind data.

Our results show that the quality controlled and bias adjusted wind speed observations are more comparable with official data, whereas the raw WOW data are not always reflective of wind speed at 10 m. We find that the processed crowdsourced wind data are much closer to official observations in terms of Pearson correlations, RMSE, and KS distances. There are 32 high-quality WOW stations in the remaining wind speed data from the initial 93 sites in the dataset, with the systematic biases adjusted and around one-sixth of observations rejected as suspect. These results show that our WOW dataset has the potential to complement the official observing network with higher spatial resolution. Compared to traditional downscaling approaches, the crowdsourced observations allow more stochastic variability in space since the observations are actually measured by devices, which could be informative in wind field analysis. For example, during a regional thunderstorm there can be strong winds in regions that are unobserved by the KNMI stations. A spatial interpolation is not able to detect these winds, while detection is possible with the CWS data. An enhanced, dense observing network with high-quality data is advantageous for capturing small-scale extreme wind events. The crowdsourced data are also valuable to be an extra input for numerical weather prediction models, as seen in other recent research (Hintz *et al.*, 2019; Nipen *et al.*, 2019).

This work is a very promising first attempt at ensuring the quality of crowdsourced wind data for operational usage, but there is still scope to improve the methods in future work. Our methods are developed based on a lack of metadata, and so they could be generalised to any CWS without knowing the type of device. However we could expect a more accurate quality assurance if the methods could be tailored to different types of device in terms of specific measurement settings. Future work could try to extract more information from the metadata, such as inferring device type from the measurement frequency or stratification, and then to adapt our QC and BA methods accordingly. The inflated zero-valued observations is a common issue in crowdsourced wind speeds. Our approach rejected erroneous zeros that are persistent for longer than two days, and estimated the actual wind speed for the censored zeros. It is worthwhile to see in future studies if there are better solutions to deal with the problem, for instance comparing with the nearest official observation to decide whether to reject zeros. Another issue is about splitting different periods to perform EQM BA. Our approach splits different seasons and day or night times in a fixed way, but it is highly recommended that a more flexible splitting criterion be set up such as quantile regression since sunrise times differ throughout the year.

We will extend our study region to the whole Netherlands in a subsequent study. When applying this methodology to other regions, users should be aware that all parameters in the within-station QC and the neighbours selection step need modification according to local climatology and the specific crowdsourced data. The spatial interpolations in our QC and BA procedure would need to be adapted to areas with more complicated terrain, although we have considered the statistical similarities between stations. For instance, when performing kriging in the bias adjustment, it is advisable to incorporate roughness maps as an extra input. When selecting neighbouring stations to check spatial consistency in the between-station QC, it is suggested that more components be considered such as wind direction to pick more reliable neighbours and generate more precise estimates. Users should be aware that the interpolation of the censored zero-valued observations is limited under complicated situations, and necessary adjustments such as extra inputs are required in some cases. Further, future researchers can explore the value of other wind speed climatologies as the target for the bias adjustment, such as a climatology from a reanalysis product or a more advanced spatial interpolation that incorporates surface roughness. Our QC and BA procedure has been developed for wind speed only, but it would be worthwhile in future work to extend our methods to wind gusts and direction, and bias adjust the wind speed by wind direction.

Performing our four-step QC and BA procedure for crowdsourced wind speed data gives confidence in observations collected at citizen weather stations. The final product is a high-quality, spatially dense observing network that has many possible applications in meteorological studies.

AUTHOR CONTRIBUTIONS

Jieyu Chen: Kate Saunders: conceptualization; methodology; project administration; supervision; writing – original draft; writing – review and editing. **Kirien Whan:** conceptualization; data curation; funding acquisition; methodology; project administration; supervision; writing – original draft; writing – review and editing.

ACKNOWLEDGEMENTS

This research was partly funded by the KNMI Early Warning Centre programme. We thank our colleagues at KNMI for their help in this research, Irene Garcia-Marti for providing the dataset and useful comments, Ine Wijnant and Maurice Schmeits for helpful suggestions, and Marijn de Haij and Andrew Stepek for providing the detailed metadata of each KNMI station. We would also like to thank Jason Frank at Utrecht University for his contributions and role as thesis supervisor. We also thank our three anonymous reviewers for their valuable suggestions.

ORCID

Jieyu Chen https://orcid.org/0000-0002-8151-5916

REFERENCES

Akhter, J., Das, L. and Deb, A. (2017) CMIP5 ensemble-based spatial rainfall projection over homogeneous zones of India. *Climate Dynamics*, 49, 1885–1916.

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

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

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

Chen, J. (2020). Quality control and verification of citizen science wind observations. Master's thesis, Utrecht University, Netherlands. Available at: https://dspace.library.uu.nl/handle/1874/398742; accessed 18 November 2020.

Contractor, S., Alexander, L.V., Donat, M.G. and Herold, N. (2015) How well do gridded datasets of observed daily precipitation compare over Australia?. *Advances in Meteorology*, 2015. https://doi.org/10.1155/2015/325718.

Cressman, G.P. (1959) An operational objective analysis system. *Monthly Weather Review*, 87, 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, Seattle, WA. Boston, MA: American Meteorological Society.

De Vos, L., Droste, A., Zander, M., Overeem, A., Leijnse, H., Heusinkveld, B., Steeneveld, G. and Uijlenhoet, R. (2020) Hydrometeorological monitoring using opportunistic sensing networks in the Amsterdam metropolitan area. *Bulletin of the American Meteorological Society*, 101, E167–E185.

DeGaetano, A.T. (1997) A quality-control routine for hourly wind observations. *Journal of Atmospheric and Oceanic Technology*, 14, 308–317.

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, 765–777.

de Vos, L.W., Leijnse, H., Overeem, A. and Uijlenhoet, R. (2019) Quality control for crowdsourced personal weather stations to enable operational rainfall monitoring. *Geophysical Research Letters*, 46, 8820–8829.

Droste, A.M., Heusinkveld, B.G., Fenner, D. and Steeneveld, G.-J. (2020) Assessing the potential and application of crowdsourced urban wind data. *Quarterly Journal of the Royal Meteorological Society*, 146, 2671–2688.

Durre, I., Menne, M.J., Gleason, B.E., Houston, T.G. and Vose, R.S. (2010) Comprehensive automated quality assurance of daily surface observations. *Journal of Applied Meteorology and Climatology*, 49, 1615–1633.

Eischeid, J.K., Bruce Baker, C., Karl, T.R. and Diaz, H.F. (1995) The quality control of long-term climatological data using objective data analysis. *Journal of Applied Meteorology*, 34, 2787–2795.

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, 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, 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, 2547–2559.

Fiebrich, C.A., Morgan, C.R., McCombs, A.G., Hall, P.K. and McPherson, R.A. (2010) Quality assurance procedures for mesoscale meteorological data. *Journal of Atmospheric and Oceanic Technology*, 27, 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, 1040–1052.

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, 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.

Hiemstra, P.H., Pebesma, E.J., Twenhöfel, C.J.W. and Heuvelink, G.B.M. (2008) Real-time automatic interpolation of ambient gamma dose rates from the Dutch radioactivity monitoring network. *Computers & Geosciences*, 35, 1711–1721. https://doi.org/10.1016/j.cageo.2008.10.011.

Hintz, K.S., O'Boyle, K., Dance, S.L., Al-Ali, S., Ansper, I., Blaauboer, D., Clark, M., Cress, A., Dahoui, M., Darcy, R., Hyrkkanen, J., Isaksen, L., Kaas, E., Korsholm, U.S., Lavanant, M., Le Bloa, G., Mallet, E., McNicholas, C., Onvlee-Hooimeijer, J., Sass, B., Siirand, V., Vedel, H., Waller, J.A. and Yang, X. (2019) Collecting and utilising crowdsourced data for numerical weather prediction: propositions from the meeting held in Copenhagen, 4–5 December 2018. *Atmospheric Science Letters*, 20. https://doi.org/10.1002/asl.921.

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, 21–26.

Hofstra, N., Haylock, M., New, M., Jones, P. and Frei, C. (2008) Comparison of six methods for the interpolation of daily, European climate data. *Journal of Geophysical Research: Atmospheres*, 113(D21). https://doi.org/10.1029/2008JD010100.

Hubbard, K.G. and Sivakumar, M.V.K. (2001). Automated Weather Stations for Applications in Agriculture and Water Resources Management; in Proceedings of an International Workshop, Lincoln, NE, USA, 6–10 March 2000. Geneva, Switzerland: WMO. WMO/TD No. 1074.

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, 1520–1530.

Hubbard, K.G., 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, 105–112.

Hubbard, K.G., Guttman, N.B., You, J. and Chen, Z. (2007) An improved QC process for temperature in the daily cooperative weather observations. *Journal of Atmospheric and Oceanic Technology*, 24, 206–213.

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, 1101–1122.

Kantorovich, L.V. (1960) Mathematical methods of organizing and planning production. *Management Science*, 6, 366–422.

KNMI (2019). Archief code oranje & rood in 2018. Available at: https://www.knmi.nl/kennis-en-datacentrum/uitleg/archief-code-oranje-rood-in-2018; accessed 5 August 2021.

KNMI (2020a). KNMI – Wij zijn het KNMI. Available at: https://www.knmi.nl/over-het-knmi/over; accessed 6 February 2020.

KNMI (2020b). KNMI Data Centre – wind. Available at: https://dataplatform.knmi.nl/group/wind; accessed 5 August 2021.

Koole, M. (2016). Rapportage WOW-NL evaluatie, Intern Rapport IR-2016-08. De Bilt, Netherlands: KNMI. Available at: https://cdn.knmi.nl/knmi/pdf/bibliotheek/knmipubIR/IR2016-08.pdf; accessed 5 August 2021.

Lucio-Eceiza, E.E., González-Rouco, J.F., Navarro, J. and Beltrami, H. (2018a) Quality control of surface wind observations in northeastern North America. Part I: data management issues. *Journal of Atmospheric and Oceanic Technology*, 35, 163–182.

Lucio-Eceiza, E.E., González-Rouco, J.F., Navarro, J., Beltrami, H. and Conte, J. (2018b) Quality control of surface wind observations in northeastern North America. Part II: measurement errors. *Journal of Atmospheric and Oceanic Technology*, 35, 183–205.

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. https://doi.org/10.3390/w10081046.

Lussana, C., Uboldi, F. and Salvati, M.R. (2010) A spatial consistency test for surface observations from mesoscale meteorological networks. *Quarterly Journal of the Royal Meteorological Society*, 136, 1075–1088.

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.

Moral, F.J. (2010) Comparison of different geostatistical approaches to map climate variables: application to precipitation. *International Journal of Climatology*, 30, 620–631.

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, 3185–3203.

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, 1–16. https://doi.org/10.3389/feart.2018.00118.

Navarro-Racines, C.E. and Tarapues, J.E. (2015) *Bias-Correction in the CCAFS-Climate Portal: A Description of Methodologies. Decision and Policy Analysis (DAPA) Research Area*. Cali: International Center for Tropical Agriculture.

Nipen, T.N., Seierstad, I.A., Lussana, C., Kristiansen, J. and Hov, Ø. (2019) Adopting citizen observations in operational weather prediction. *Bulletin of the American Meteorological Society*, 101(1), E43–E57.

Otop, I., Szturc, J., Ośródka, K. and Djaków, P. (2018). Automatic quality control of telemetric rain gauge data for operational applications at IMGW-PIB. In: ITM Web of Conferences.

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*, 18, 1169–1179.

Schuhmacher, D., Bähre, B., Gottschlich, C., Hartmann, V., Heinemann, F. and Schmitzer, B. (2020). transport: computation of Optimal Transport Plans and Wasserstein Distances. R package version 0.12-2. Available at: https://cran.r-project.org/package=transport; accessed 5 August 2021.

Shafer, M.A., Fiebrich, C.A., Arndt, D.S., Fredrickson, S.E. and Hughes, T.W. (2000) Quality assurance procedures in the Oklahoma Mesonetwork. *Journal of Atmospheric and Oceanic Technology*, 17, 474–494. https://doi.org/10.1175/1520-0426(2000)017<0474:QAPITO>2.0.CO;2.

Stahl, K., Moore, R., Floyer, J., Asplin, M. and McKendry, I. (2006) Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density. *Agricultural and Forest Meteorology*, 139, 224–236.

Taylor, J.R. and Loescher, H.L. (2013) Automated quality control methods for sensor data: a novel observatory approach. *Biogeosciences*, 10, 4957–4971.

Terink, W., Hurkmans, R., Uijlenhoet, R., Warmerdam, P. and Torfs, P. (2009) Bias correction of temperature and precipitation data for regional climate model application to the Rhine basin. *Hydrology and Earth System Sciences Discussions*, 6(4). https://doi.org/10.5194/hessd-6-5377-2009.

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.

Thorarinsdottir, T.L. and Gneiting, T. (2010) Probabilistic forecasts of wind speed: ensemble model output statistics by using heteroscedastic censored regression. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 173, 371–388.

Wackernagel, H. (2003). Ordinary Kriging, pp. 79–88 in Multivariate Geostatistics, Berlin.

WMO (2014) *Guide to Meteorological Instruments and Methods of Observation (WMO-No.8)*. Geneva: World Meteorological Organisation.

WOW-NL (2020). WOW-NL: Jouw weer op de kaart!. Available at: https://wow.knmi.nl/over-wow-nl; accessed 7 Febraury 2020.

You, J. and Hubbard, K.G. (2006) Quality control of weather data during extreme events. *Journal of Atmospheric and Oceanic Technology*, 23, 184–197.

Zahumenskỳ, I. (2004). Guidelines on Quality Control Procedures for Data from Automatic Weather Stations. Report CBS/OPAG-IOS/ET AWS-3/Doc. 4(1), Geneva, Switzerland: World Meteorological Organization.

How to cite this article: Chen, J., Saunders, K. & Whan, K. (2021) Quality control and bias adjustment of crowdsourced wind speed observations. *Quarterly Journal of the Royal Meteorological Society*, 1–18. Available from: https://doi.org/10.1002/qj.4146