RMetS

Royal Meteorological Society

Identifying, attributing, and overcoming common data quality issues of manned station observations

Stefan Hunziker,[a,b]* Stefanie Gubler,[c] Juan Calle,[d] Isabel Moreno,[d] Marcos Andrade,[d]
Fernando Velarde,[d] Laura Ticona,[d] Gualberto Carrasco,[e] Yaruska Castellón,[e] Clara Oria,[f]
Mischa Croci-Maspoli,[c] Thomas Konzelmann,[c] Mario Rohrer[g] and Stefan Brönnimann[a,b]

[a] *Institute of Geography, University of Bern, Switzerland*
[b] *Oeschger Centre for Climate Change Research, University of Bern, Switzerland*
[c] *Federal Office of Meteorology and Climatology MeteoSwiss, Zurich, Switzerland*
[d] *Laboratorio de Física de la Atmósfera, Instituto de Investigaciones Físicas, Universidad Mayor de San Andrés, La Paz, Bolivia*
[e] *Servicio Nacional de Meteorología e Hidrología de Bolivia (SENAMHI), La Paz, Bolivia*
[f] *Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI), Lima, Peru*
[g] *Meteodat GmbH, Zurich, Switzerland*

ABSTRACT: *In situ* climatological observations are essential for studies related to climate trends and extreme events. However, in many regions of the globe, observational records are affected by a large number of data quality issues. Assessing and controlling the quality of such datasets is an important, often overlooked aspect of climate research. Besides analysing the measurement data, metadata are important for a comprehensive data quality assessment. However, metadata are often missing, but may partly be reconstructed by suitable actions such as station inspections. This study identifies and attributes the most important common data quality issues in Bolivian and Peruvian temperature and precipitation datasets. The same or similar errors are found in many other predominantly manned station networks worldwide. A large fraction of these issues can be traced back to measurement errors by the observers. Therefore, the most effective way to prevent errors is to strengthen the training of observers and to establish a near real-time quality control (QC) procedure. Many common data quality issues are hardly detected by usual QC approaches. Data visualization, however, is an effective tool to identify and attribute those issues, and therefore enables data users to potentially correct errors and to decide which purposes are not affected by specific problems. The resulting increase in usable station records is particularly important in areas where station networks are sparse. In such networks, adequate selection and treatment of time series based on a comprehensive QC procedure may contribute to improving data homogeneity more than statistical data homogenization methods.

KEY WORDS   quality control; error attribution; station observations; data rescue; metadata; data homogenization; Bolivia; Peru

Received 28 July 2016; Revised 25 January 2017; Accepted 25 January 2017

1. Introduction

Assessing climatological trends and extremes requires data that meet high quality standards. If data are affected by quality issues, a consequence may be misleading and erroneous results (Nese, 1994; Petrovic, 2001; Viney and Bates, 2004; Mahmood *et al.*, 2006; Westerberg *et al.*, 2009; Zhang *et al.*, 2009; Rhines *et al.*, 2015). In order to eliminate errors, quality control (QC) should be applied before analysing observational records (Westerberg *et al.*, 2009; WMO, 2011). The World Meteorological Organization (WMO) provides guidelines on QC in various documents (WMO, 1993; Plummer *et al.*, 2003; WMO, 2008; Klein Tank *et al.*, 2009; WMO, 2010; WMO, 2011). A wide range of different QC approaches are used by National Weather Services and research projects. In some cases, only a few basic methods are applied that may

not ensure the desired quality of the data. According to Štěpánek *et al.* (2013), there is a lack of a generally accepted methodology for QC. Furthermore, if data are retrieved from a standard dataset, the suitability of the data for the intended purpose is often not questioned (Brönnimann, 2015).

Although satellite and reanalysis data find widespread use in climate science, they are not a substitute for ground-based observations due to limitations such as a lower precision, shorter record lengths, calibration problems, and the need for *in situ* data for validation (Plummer *et al.*, 2003; Bližňák *et al.*, 2015; Mantas *et al.*, 2015; Schmocker *et al.*, 2016). Therefore, quality issues in observational records must be addressed.

Data quality issues may occur at any point of the data generation chain, for example, due to unsuitable station configuration or siting, poor station maintenance, erroneous instrument reading, or inaccurate data digitization and post processing (Brönnimann, 2015). The characteristics and frequencies of errors may vary between

* Correspondence to: S. Hunziker, Institute of Geography, University of Bern, Hallerstrasse 12, Bern 3012, Switzerland. E-mail: Stefan.hunziker@giub.unibe.ch

different networks depending on factors such as the grade of automatization, instruments used, observer training, station maintenance, and climate type. Hence, a QC approach may be suitable for one network but not for another.

Not all data quality issues are detectable by controlling measurement data. Metadata are another important source in the quality assessment. According to the Global Climate Observing System monitoring principles (WMO, 2002), metadata (i.e. the details and history of local conditions, instruments, operation procedures, data processing algorithms, etc.) should be documented and treated with the same care as the measurements. Metadata are also needed to ensure that end-users are able to extract accurate conclusions from their analyses, and they play an important role in the detection and correction of inhomogeneities and are essential to managing networks and support systems (Aguilar *et al.*, 2003; Plummer *et al.*, 2003; WMO, 2008). Nevertheless, in many networks worldwide, metadata are fractional, of poor quality, or simply not available (e.g. Westerberg *et al.*, 2009; Trewin, 2010; Fall *et al.*, 2011), and therefore measures should be taken to address this shortcoming.

Assuring the quality of a dataset is more difficult for sparse than for dense station networks. Important QC tests depend on the availability of suitable neighbouring stations (Plummer *et al.*, 2003; Kunkel *et al.*, 2005; Durre *et al.*, 2010). Furthermore, records in dense networks may be excluded from the dataset without losing much information (e.g. Vertačnik *et al.*, 2015), which allows removing time series that are incomplete or whose quality is uncertain. This, however, may not apply to sparse network datasets. In order to obtain the greatest benefit from weather observations in regions of low station density, identification of specific errors and, if possible, error corrections are particularly important.

In this study, we investigate data quality issues in daily maximum/minimum temperature and precipitation observations in Bolivia and Peru as exemplary cases for predominantly manned station networks. According to Vuille *et al.* (2008), data quality problems and the lack of long-term weather observations hinder reliable trend analyses in the Central Andean area. The same applies to other regions, such as Central America (Westerberg *et al.*, 2009) or eastern Africa (Kizza *et al.*, 2012). Analysing weather observations from the Central Andean area is challenging, as the region is characterized by complex terrain and strong climate gradients. Consequently, the impacts of climate change vary (Vuille *et al.*, 2000; Diaz *et al.*, 2014; Wang *et al.*, 2014). This would require investigations on relatively small spatial scales, but station networks are sparse in the region. Furthermore, metadata records are fragmentary or completely missing. Hence, Bolivian and Peruvian station records are good case studies to address various challenges in climate monitoring.

This study describes the current situation of station networks in Bolivia and south-eastern Peru (Section 2). Measurement conditions, practices, and instrumentation are outlined, and efforts on QC and metadata gathering are described. Because metadata are largely missing,

we suggest strategies to reconstruct and generate metadata (Section 3). The close collaboration of various organizations, extensive investigations on site, and detailed data analyses enabled us to characterize common errors found in the data and to identify error sources (Section 4). In order to describe the impact on results, we mention for each major data quality issue particularly affected and unaffected climate change indices. Finally, this study shows methods for error detection, and discusses how station records may potentially be used for certain purposes despite existing quality issues (Section 5).

2. Data

In order to investigate common data quality issues of manned station observations, data from the Central Andean area are analysed. The study area includes Bolivia and the Peruvian districts Puno, Cuzco, and Junín (Figure 1(a)). In the following, the 'Peruvian data' refers to data from these districts. Note, however, that the data of the Peruvian districts used in this study is not complete and does not cover the full area of the districts. The climate variables investigated are maximum temperature (TX), minimum temperature (TN), and precipitation (PRCP), which are key variables in climatology (Plummer *et al.*, 2003) and are the most frequently measured parameters in the Central Andean area.

Data were accessed between 2014 and 2016 from the National Meteorological and Hydrological Service (SENAMHI) Bolivia, the project Pilot Program for Climate Resilience (PPCR) for Bolivia, and SENAMHI Peru. The dataset of PPCR includes transcriptions that originate from SENAMHI Bolivia and the civil airport administration (AASANA). There are several different digitized datasets that are based on weather observations of SENAMHI Bolivia. They are the result of multiple transcription programs implemented at different times. Even though they have the same primary data source, these datasets differ regarding their station coverage, time series completeness, and transcription quality. This study was conducted within the framework of the projects 'Data on climate and Extreme weather for the Central AnDEs' (DECADE) and 'Servicios CLIMáticos con énfasis en los ANdes en apoyo a las DEcisioneS' (CLIMANDES). As both projects aim to increase the quality of observational data in the Central Andes, the available datasets were evaluated, merged, corrected if possible, and updated. An output of the DECADE project will be a new and improved dataset for Bolivia and south-eastern Peru.

2.1. Station networks in Bolivia and Peru

The most important organizations conducting weather observations are the National Meteorological and Hydrological Services, i.e. SENAMHI Bolivia and Peru. Both institutions work with limited resources, which may increase the risk of the occurrence of quality issues. According to WMO (2008), successful observing practices require skill, training, equipment, and support, which is
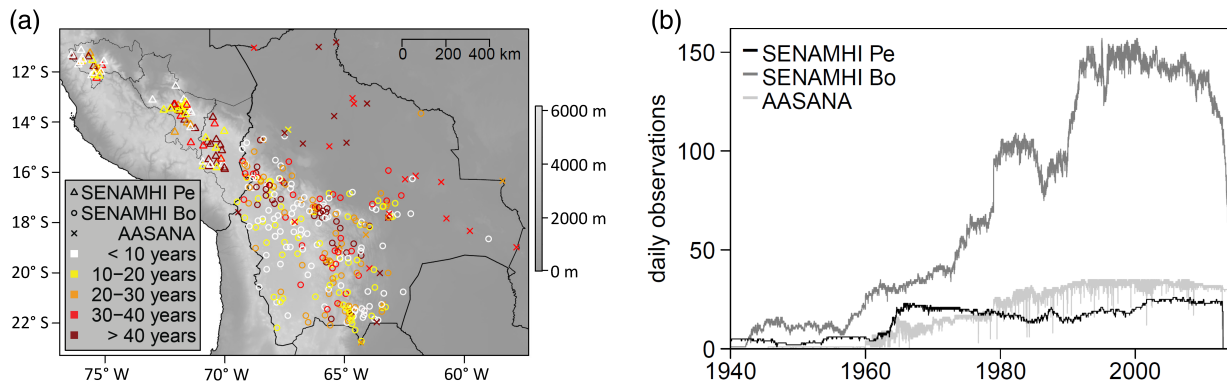
Figure 1. (a) Maximum temperature (TX) observations of SENAMHI Peru in the Peruvian districts Junín, Cusco, and Puno (from north to south), and of SENAMHI Bolivia and the Bolivian civil airport administration (AASANA). Note that the Peruvian stations shown in this figure do not reflect the complete station coverage of the three districts. The duration of the records are classified by the total number of daily observations at each station in years. (b) Temporal development of the sum of daily TX observations. Note that the strong decrease of observations in the last few years results mainly from the delay of data transcription of recent measurements. [Colour figure can be viewed at wileyonlinelibrary.com].

not always available to a sufficient degree. Furthermore, organizational weaknesses may contribute to observational data quality issues in many networks (Westerberg *et al.*, 2009). A large fraction of the Central Andean station records cover only a short measurement period or contain large data gaps, which is also an issue in many other networks worldwide (e.g. Kizza *et al.*, 2012; Trewin, 2013; Brunet *et al.*, 2014a) and drastically lowers the amount of time series suitable for climatological studies.

Currently, the station network of SENAMHI Bolivia contains around 255 manned and 33 automatic stations (F. Villalpando, 2016, pers. comm.). Note that the latter are not included in this work. For the digitized time series that were used in this study, the median length of the sum of observations at each station of SENAMHI Bolivia is slightly below 20 years for TX (Figure 1(a)), TN, and PRCP. The spatial and temporal distributions of TX and TN observations are nearly identical. The spatial distributions of temperature and PRCP measurements are similar, but the total number of PRCP time series is about twice as high. Since the year 2000, there has been a decline in the number of daily temperature measurements (Figure 1(b)). For PRCP, the decline starts already in the early 1990s, resulting in about half as many observations in the last decade compared with the maxima in the 1980s and 1990s. This pattern is observed in many weather observation networks in the developing world (Sene and Farquharson, 1998). The reasons for this development are manifold, and include inadequate institutional frameworks and the lack of appreciation of the worth of long-term data (Sene and Farquharson, 1998). The oldest digitized measurements of SENAMHI Bolivia date back to 1917. However, in the context of this study, a non-digitized record dating back to 1892 was detected. Historic weather data are often not completely transcribed (e.g. Brunet and Jones, 2011, Trewin, 2013, Brönnimann, 2015). Currently, a program of the International Environmental Data Rescue Organization (IEDRO) is addressing this issue in Bolivia (http://iedro.org/activities/bolivia-data-rescue-program).

Besides SENAMHI Bolivia data, nearly all station records of AASANA were included this study. Observations go back to 1942. In contrast to the SENAMHI Bolivia network, the amount of daily measurements remains nearly constant since the 1990s (Figure 1(b)). The median observation length is 40 years for temperature and 52 years for precipitation. Collecting and analysing data from different networks of the same region has been shown to be beneficial to the accuracy of meteorological analyses (e.g. Westerberg *et al.*, 2009).

The first measurements in the Peruvian time series available for this study were taken in 1931. In contrast to SENAMHI Bolivia, the number of daily measurements in SENAMHI Peru is more constant (Figure 1(b)), and no clear decline in daily observations occurs in recent years, either for temperature or for PRCP. The median observation duration is 20 years for temperature and 28 years for PRCP.

2.2. Measurement practices and instrumentation

The vast majority of the stations in the Central Andes are manned. According to Fiebrich and Crawford (2009), the automatization of a network results in a clear increase in data quality. Automatic weather stations (AWS) have proven to be more accurate than manned observations for many variables (e.g. temperature), but for others (e.g. PRCP), there are measurement issues that have to be addressed (Ciach, 2003, Plummer *et al.*, 2003, Tokay *et al.*, 2010). However, AWS may be exposed to disturbances, such as technical defects, thievery, or vandalism. Such data losses may make time series unusable for climatological analyses (e.g. Westerberg *et al.*, 2009). Furthermore, the capabilities, personnel, and equipment must be adequate to run the network (WMO, 2008). For instance, AWS require station visits at least twice a year (Plummer *et al.*, 2003). This would be a great challenge in the Central Andean region, where access to remote areas is seasonally impeded and station inspections are currently done very infrequently. Without strict adherence to prescribed inspection and maintenance programs and schedules, the

quality of AWS observations can deteriorate dramatically (Plummer *et al.*, 2003). Hence, a sudden automatization of the networks may not only cause considerable inhomogeneities in the measurement records (e.g. Hubbard and Lin, 2006, WMO, 2008, Trewin, 2010), but also a likely quality decrease of Central Andean weather observations.

Observers in Bolivia and Peru are largely laypersons, who conduct the measurements besides their main occupation. At some stations, the official observers can also be temporarily replaced by aides (usually family members), which results in changes in the way data is collected.

Measurements of SENAMHI Bolivia stations are taken at 0800 LST (1200 UTC). Partially, a second daily measurement is taken at 1700 or 1800 LST. Bolivian airports measure hourly during operation times, but only daily summaries are transcribed. At stations of SENAMHI Peru, measurements are normally taken three times a day at 0700 (1200 UTC), 1300, and 1900 LST. Data are largely transmitted to SENAMHI in the form of the original documents (in Bolivia every second month), as it is the actual or former practice in other networks (e.g. Kunkel *et al.*, 2005). Currently, the observers at some Peruvian stations transmit the data immediately to SENAMHI by mobile phone.

In Bolivia, TX and TN are mostly measured with Six's thermometers. These thermometers are considered as robust and simple to use, but were superseded for professional field work in many networks a long time ago (Austin and McConnell, 1980). In general, only high quality instruments should be used for meteorological purposes (WMO, 2008). However, as instruments should also be suitable regarding reliability, durability, simplicity of design, the resources for observer training, and the costs (WMO, 2008), Six's thermometers may still be the best choice for certain stations. At some Bolivian and most Peruvian stations, liquid in glass TX and TN thermometers are used. PRCP is mostly measured with Hellmann rain gauges without wind shields. The use of different types of instruments to measure a single parameter may cause data continuity problems. (WMO, 2008).

2.3. Metadata records

Metadata analysis may reveal quality issues that are not detectable in the measurement data. For instance, station siting has a great impact on temperature and PRCP measurements (Aguilar *et al.*, 2003, Mahmood *et al.*, 2006, WMO, 2008, Yilmaz *et al.*, 2008). According to WMO guidelines (Plummer *et al.*, 2003, WMO, 2008), an observation site should be representative of the climatic regime for which it is intended. If this requirement is not met, the measured data may not be meaningful for comparisons (Plummer *et al.*, 2003, Brönnimann, 2015). However, good station exposure as defined by WMO (2008) is difficult to achieve and compromises are necessary. Questionable station siting occurs in the Central Andes, as well as in many networks worldwide (Trewin, 2010). For instance, the vast majority of the United States National Weather Service's cooperative observer network (COOP) was reported to be poorly sited (Fall *et al.*, 2011). The specific effects of poor station exposure on the measurements

are complex (Mahmood *et al.*, 2006, Kumamoto *et al.*, 2013), but definitely increase the uncertainty in the data (Vose *et al.*, 2005, Mahmood *et al.*, 2006, Pielke *et al.*, 2007). Hence, high quality metadata allow users to assess the circumstances at which the data were recorded, and time series that seem inadequate for a certain purpose can be removed.

In Bolivia and Peru, metadata are not systematically recorded. In Bolivia, efforts have been made to collect metadata for SENAMHI and AASANA stations. The resulting information is available through a webpage (http://www.senamhi.gob.bo/sige/). Even though this is a helpful tool, the metadata are very basic and in some cases incorrect or outdated. SENAMHI Peru operates a website that provides similar information (http://www.senamhi.gob.pe/include_mapas/_map_data_tesis.php).

2.4. QC practices

No systematic and consistent QC is applied on Bolivian and Peruvian station data, as it is the case for other networks (e.g. Fiebrich and Crawford, 2009; Westerberg *et al.*, 2009). However, there are some procedures that should guarantee the quality of the data. In Bolivia, a supervisor of SENAMHI verifies the monthly data sheets provided by the observers. Suspicious values are highlighted by handwritten notes. However, these notes are not always taken into account in the subsequent digitization of the original documents. At SENAMHI Peru, data sheets are partly digitized twice in order to avoid typing errors. Furthermore, some QC has been done in the frame of different projects. According to WMO (2008), the provision of meteorological data of good quality is not a simple matter and is impossible without a quality management system. The lack of such a system may increase the danger of inadequate actions that may not only result in the loss of original data, but also may introduce errors (Reek *et al.*, 1992; Aguilar *et al.*, 2003). Because no quality management system is established in the Central Andean networks, original time series should be treated as non-quality controlled data. The same applies for other networks where QC processes are not clearly documented, or if the adequacy of the applied QC is questioned.

3. Metadata generation

There are different sources for reconstructing station information if metadata are partly or completely missing. To get the best possible set of metadata, the information of all sources should be combined. This is a very time and resource consuming process. However, the methods described in the points 1–3 in the following paragraphs can be applied rather easily, because there is no need to be on site. Accessing the other data sources (points 4 and 5) requires on site investigation, except if photos of the required documents are available (point 4).

Using the 5 steps below, we successfully generated and collected metadata from Bolivia and Peru:

1. Collection of all transcribed information: Digitized metadata may not be stored in only one document.

Hence, the different information sources should be compiled. Discrepancies (e.g. different coordinates for the same station) may give evidence to changes in the station's history (such as station relocations). In Bolivia and Peru, these sources are available at the aforementioned websites (see Section 2.3), in the headers of the data files, or in separate station lists.

2. Investigation of the station surroundings: Station surroundings can be investigated and reported based on 'Google Earth' images. This or similar methods were used in previous studies (Mahmood *et al.*, 2006; Fall *et al.*, 2011). The approach furthermore allows the identification of erroneous or suspicious station coordinates such as station locations in the middle of a lake.

3. Extracting metadata from the data: A detailed QC reveals many indicators for events in the history of the stations. For example, a transition in the measurement precision (see Section 4.3.1) is a clear indication for an undocumented event (Rhines *et al.*, 2015) such as an observer or instrument change. Reporting such observations creates additional metadata that can be used, *inter alia*, for homogeneity estimations and breakpoint corroboration (Rhines *et al.*, 2015). This source is particularly valuable if documentation about the station's history is missing.

4. Screening of original documents: Search for non-digitized notes on archived data sheets, station inspection reports, working memorandums, and similar documents. Furthermore, some information is hidden by the digitization process. For instance, a change in handwriting is a strong indication for an observer change. In addition, comparing original records with digitized data enables to assess the quality of the data transcription.

5. Station visits: Station condition and environmental surroundings can be determined from in person station visits that should include measuring the station's coordinates with GPS and photographing the station from all directions. The station's history may be, at least partially, reconstructed by interviewing observers or inhabitants living in the vicinity. Furthermore, station visits allow investigating previously detected discrepancies and errors in metadata and measurement records. It is reasonable to create a station visit form in order to gain objective and comparable metadata. Station inspections aiming to generate metadata were performed in previous projects (e.g. Davey and Pielke, 2005; Mahmood *et al.*, 2006; Westerberg *et al.*, 2009; Fall *et al.*, 2011).

4. Description of common quality issues

This section describes common data quality issues that do not just affect single measurements but may appear systematically during several years. The list is not complete, but it covers the most common errors found in the Central Andean time series. The cases demonstrated in the following sections are exemplary. Note that these data quality issues often manifest in a less pronounced and/or somewhat different form. In such cases, the impact on results of data analyses might be less drastic, but error detection and attribution are made more difficult. Furthermore, errors frequently occur intermittently, have individual specificities, and overlap. Various approaches may be suitable to detect data quality issues, but the models suggested here have been proven to be effective. In some cases, errors could be clearly attributed to a specific cause, while in other cases, error sources could only be assumed. The effect of data quality issues on specific analyses will be mentioned by means of selected climate change indices defined by the Expert Team on Climate Change Detection and Indices (ETCCDI, http://etccdi.pacificclimate.org/). Thereby, we demonstrate that time series affected by data quality issues may still be usable for some analyses, but not for others. The ETCCDI indices are widely used in studies on changing climate extremes (e.g. Alexander *et al.*, 2006; New *et al.*, 2006; You *et al.*, 2011; Zhang *et al.*, 2011; Donat *et al.*, 2013; Sillmann *et al.*, 2013) and are included in the software package RClimDex (Zhang and Yang, 2004).

4.1. Temperature

4.1.1. Missing temperature intervals

In some time series, measurements in a certain temperature range are missing or the frequency of records is reduced. Such missing temperature intervals may extend up to several degrees. Values around 0 °C are predominantly affected. Because nighttime temperatures at high altitude stations in the Central Andes frequently fall below 0 °C, TN records are more often affected than TX measurements. The occurrence of the error is usually not associated with an increase of missing values in the time series.

In Bolivia and Peru, approximately 10–20% of the station records are affected by missing temperature intervals. A very clear example of the error is found in the TN observations of the station Progreso in the Peruvian district of Puno (Figure 2(a)). The instrument used at this station is a liquid-in-glass thermometer with a dumbbell-shaped rod (Figure 2(b)), from which TN is read at the right side of the rod. When discussing the issue with the observer, it was found that he read the temperature on the left side of the rod whenever the rod's centre was below 0 °C. The length of the rod corresponds to about 4.2 °C, and hence determines the size of the missing temperature interval.

Even though different causes may induce missing temperature intervals such as the misinterpretation of zeros as invalid numbers, erroneous instrument reading is often the most likely source of this type of quality issue. The error may bias means as well as extremes. In the case of TN time series of Progreso, the error affects all ETCCDI indices [except the 'number of tropical nights' (TR) that is irrelevant for this station]. However, as the cause of the problem was clearly identified and the observer did the same error consistently for many years, the biased negative temperature values may be corrected by adding 4.2 °C corresponding to the rod's length. The carefully corrected time series may then be used for any analysis.
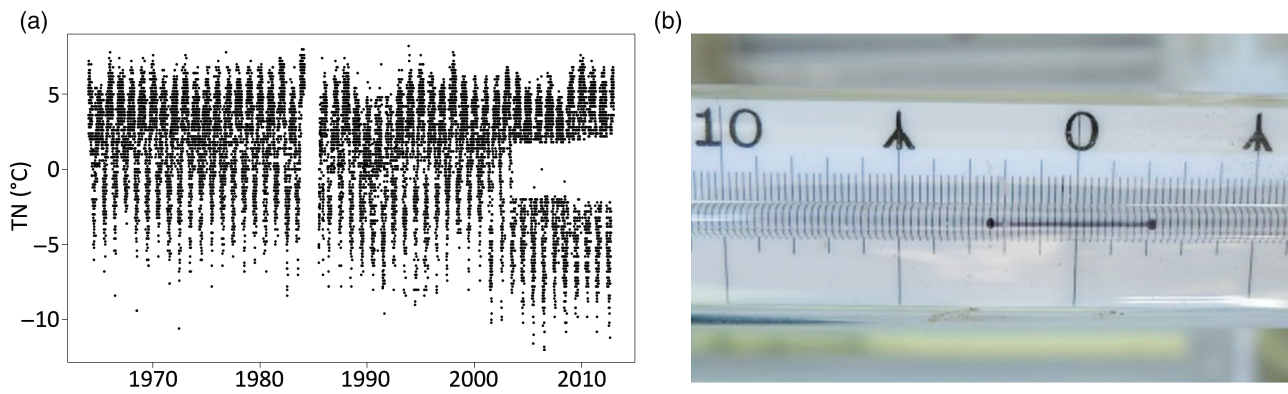
(a)

(b)

Figure 2. (a) Minimum temperature (TN) time series of Progreso (Peru). Since 2003, a missing temperature interval between approximately −2 and +2 °C is observed. (b) Liquid-in-glass minimum thermometer at the station Progreso. Values to the right of the zero point are positive, to the left negative. The temperature must be read on the right side of the rod (slightly above +2 °C for the case shown). However, if the centre of the rod was below 0 °C, the observer erroneously read the temperature on the rod's left side. Hence, the length of the rod (corresponding to about 4.2 °C) determines the missing temperature range. [Colour figure can be viewed at wileyonlinelibrary.com].

4.2. Precipitation

4.2.1. Heavy precipitation truncation

PRCP records may have a deficiency of heavy PRCP events above a certain threshold value. This may manifest as a complete truncation or a frequency reduction. Often, there is an accumulation of data points around the threshold value. The occurrence of the error does normally not come along with an increase of missing values.

A relatively large fraction of about 35% of the Bolivian time series are affected by a temporary lack of PRCP values above 20 mm (sometimes 40 mm). In Peru, about 15% of the stations are affected by the same issue, but the truncation occurs usually at 10 mm. A typical example of heavy PRCP truncation is found in the records of the station Aguirre in Bolivia (Figure 3(a)). The source of the error was detected when inspecting the Hellmann rain gauge that is generally used in Bolivia (Figure 3(b)). Observers measure the PRCP in the instrument's inner container that captures 20 mm of PRCP before overflowing. Rainwater in the outer container is measured by refilling the inner container. Finally, the measured PRCP amounts are totalled. Apparently, various observers failed to measure the rainwater collected in the outer container.

Even though other error sources are theoretically possible such as leaking outer containers, the error can generally be attributed to incorrect measurement practices. Very similar errors are known from other networks, e.g. from Brazil (P. V. da Costa Pereira, 2016, pers. comm.). Much higher threshold values could be found at stations located in regions where PRCP amounts of extreme events are very large. For instance, at some stations in the Bolivian Amazon basin, the instrument's storage capacity of 200 mm was too small for several 1-day PRCP events. The same may happen if observers do not take measurements with the required frequency that may be increased during intense PRCP events. At a station of MeteoSwiss, for example, this issue lead to a partial heavy precipitation truncation in the 1970s and 1980s (experienced personally by co-author Mario Rohrer).

Missing heavy PRCP events are a severe data quality problem, and affected time series are not suitable for most analyses. However, some of the ETCCDI indices, such as the 'annual count of days when PRCP ≥ 10 mm' (R10mm) or the 'maximum length of dry and wet spells' (CDD and CWD, respectively), are not affected by a 20 mm PRCP truncation. In the case of station records of Aguirre, however, the accumulation of values around 10 mm and missing values between 1 and 2 mm (see Section 4.2.2) impede the application of such indices.

4.2.2. Small precipitation gap

PRCP records may have missing measurements of small PRCP events. This may affect a range of very low values (e.g. 0–1 mm), but may also reach up to two-digit numbers. Small PRCP gaps do not usually come along with an increase of missing values, but with an increase of precipitation free days. In some cases, an often uniform accumulation of very low values (≤1 mm) is found instead of an increase of zeros.

A large fraction of about 60–65% of the Bolivian and Peruvian time series is affected by small PRCP gaps. A typical example of the error is found in the records of Chorocona, Bolivia (Figure 4(a)). As in many cases, the size of the gap varies in time. Presumably, the most frequent cause of this error is attributable to observer behaviour because the detected pattern occurs if measurements are taken only after substantial rain events. As a result, small PRCP events remain unrecorded and get accumulated. Most observers will fill in a zero for the missed measurements, but some observers seemingly recorded a uniform low number (≤1 mm) as a rough estimation instead. Another cause that may lead to small PRCP gaps is the degradation of the scales used for the measurements. Some of the scales are weathered and the marks are faded, especially towards the scale's ends (Figure 4(b)). Unrecognizable scale and water marks may get interpreted as zero PRCP.

Small PRCP gaps are also found in other station networks. For example, the error occurs quite frequently in
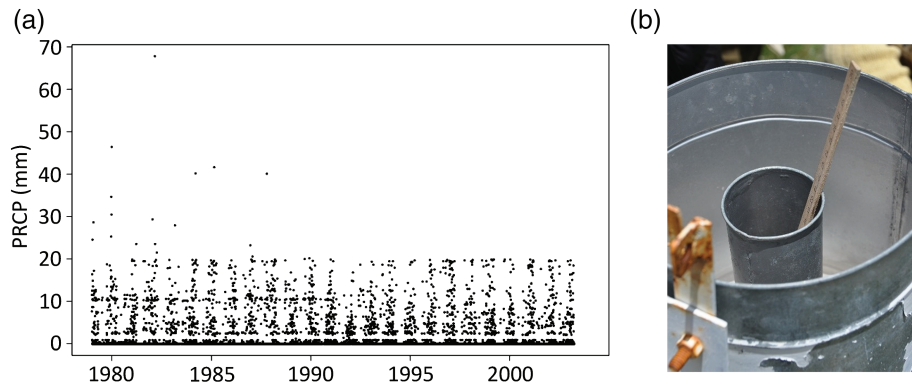
(a) (b)

Figure 3. (a) Precipitation (PRCP) observations of the station in Aguirre (Bolivia). Only a few PRCP values above 20 mm are reported until 1988, followed by a complete truncation. The time series is temporarily affected by further errors, namely an excess of values around 10 mm and a missing interval of values between 1 and 2 mm. (b) Hellmann rain gauge as usually used by SENAMHI Bolivia. It is topped by a cone that directs the precipitation into the inner container of the instrument. Overflowing rainwater of the inner container is collected by the outer container. The observer measures precipitation by sticking a scale into the inner container and reading the water mark on the scale. The volume of the 20 cm high inner container corresponds to 20 mm of precipitation. [Colour figure can be viewed at wileyonlinelibrary.com].
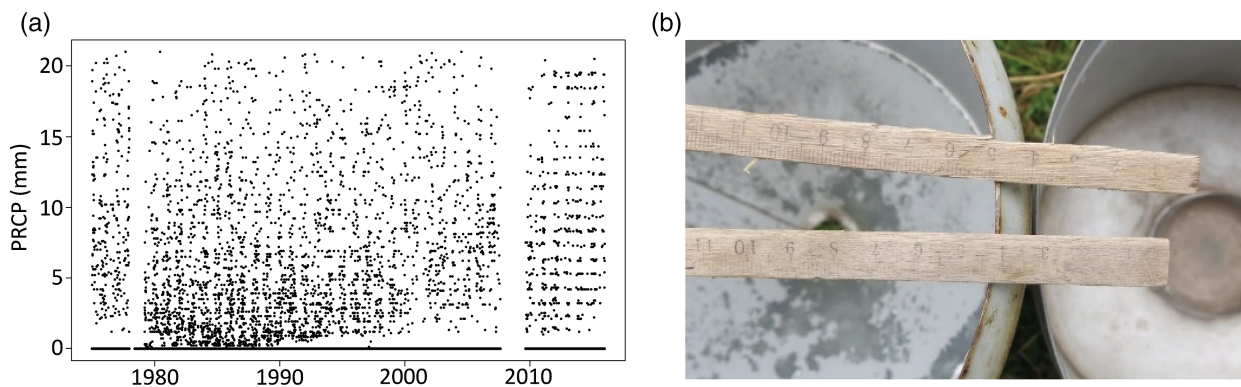
(a) (b)

Figure 4. (a) In the data of the Bolivian station Chorocona, precipitation (PRCP) values between 0 and up to about 3 mm are largely missing until the early 1980s and from the 1990s onward. Values >21 mm are not shown in the plot. (b) Precipitation in Bolivia is measured by sticking a scale in the inner container of the Hellmann rain gauge and reading the water mark on the scale (1 cm on the scale corresponds to 1 mm of PRCP). Some scales are aged and have faded scale marks, especially towards the ends. [Colour figure can be viewed at wileyonlinelibrary.com].

Brazilian station records (P. V. da Costa Pereira, 2016, pers. comm.) and the United States COOP network (Daly *et al.*, 2007). If we have good reasons to assume that PRCP amounts of small rain events are accumulated and accredited to the next measurement, the impact of the error on yearly or monthly PRCP sums may be insignificant. Hence, the time series may be used to calculate the ETC-CDI index 'Annual total precipitation in wet days' (PRCP-TOT). However, note that wet days (WDs) are defined as daily PRCP ≥ 1 mm (Klein Tank *et al.*, 2009). If rain events get accumulated, also PRCP events <1 mm get integrated in the index calculation. On the other hand, evaporation losses are increased if collected rainwater remains longer than 1 day in the rain gauge. In arid climates, these two effects may not be negligible. Other indices that depend on the count of WDs, such as the CDD and CWD or the 'simple precipitation intensity index' (SDII), are strongly affected by small PRCP gaps. Indices on extreme PRCP events, such as the 'monthly maximum 1-day precipitation' (Rx1day) or the R10mm, are altered by the accumulation of rainfall. To what extend these indices are biased depends on the specific peculiarity of the error.

4.2.3. Weekly precipitation cycles

PRCP frequency in time series may occur in weekly cycles. This may happen because no observations are taken on a certain day of the week. Mostly, weekends are affected, as daily routines on those days usually differ from workdays. The typical resulting pattern is a PRCP lack followed by a PRCP excess because of the multi-day rainfall accumulation. Missed measurements should be marked by an unambiguous code (e.g. a horizontal line '-'). Furthermore, the observation practice should be mentioned in the metadata. Hence, data users are informed about the issue and can take appropriate measures. Some relevant methods are discussed by Viney and Bates (2004).

However, weekly PRCP cycles are also found in station records without indications of non-daily measurements. Any manned station may be affected if the observer uses an inappropriate missing value code such as a zero (Durre *et al.*, 2010). Furthermore, missed measurements could be associated with the negligence of the observer's task, which may tempt observers to insert a zero rather than a missing value code.
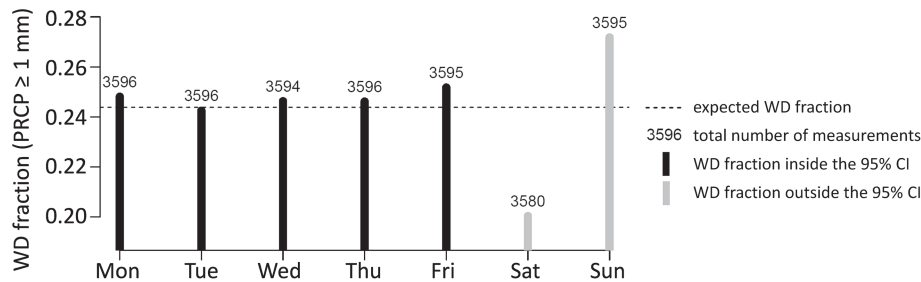
Figure 5. Fraction of wet days (WD) in San Calixto (Bolivia). WD fractions outside the 95% confidence interval (CI) are marked in grey. The fraction of WDs on Saturdays (Sundays) is significantly lower (higher) than the expectation. The number of reported measurements on Saturdays (3580) is only negligibly lower (0.4%) than on the other days of the week.

Weekly PRCP cycles were detected in various manned station networks worldwide. Viney and Bates (2004) found that the majority of Australian high quality PRCP time series are affected, and Schmocker *et al.* (2016) detected the error in about 20% of station records in the Mount Kenya area.

However, not only observer routines, but also air pollution caused by human activity may cause weekly PRCP cycles (Simmonds and Keay, 1997; Cerveny and Balling, 1998; Wilby and Tomlinson, 2000). These findings are yet often indistinct and not confirmed by other studies (DeLisi *et al.*, 2001; Schultz *et al.*, 2007; Stjern, 2011). Roughly half of the relevant publications have found pollution-induced weekly PRCP cycles, while the other half did not (Tuttle and Carbone, 2011). Generally, the signal of observer behaviour on weekly PRCP cycles is much stronger than the potential influence of air pollution. Furthermore, the two causes lead typically to fundamentally different patterns of weekly cycles.

Identification of weekly PRCP cycles requires specific checks. Results may well be indistinct, because the missed measurements may occur irregularly and stretch over more than 1 day of the week. Furthermore, checks should be made not only for the whole time series but also for shorter time periods, as the problem usually occurs temporarily. The weekly distribution of WDs is a much better indicator than the highly variable weekly distribution of PRCP amounts. Viney and Bates (2004) suggest to compare the number of wet Sundays with the assumingly unaffected mid-week days (Tuesday to Friday). This method seems accurate for the Australian dataset, where stations were predominantly run by public service employees with regular working hours on weekdays only. However, it may not be transferable to other networks and countries, where the observers' occupations are manifold and Sundays are not generally days off. Furthermore, date shifts (see Section 4.3.3) may misalign the weekly cycle pattern. We therefore suggest a more general approach that does not assume affected Sundays and unaffected weekdays. Instead, each day of the week is expected to have the same probability for PRCP, which is calculated by dividing the total number of WDs by the total count of measurements. The number of WDs of each day of the week is then tested by a two-sided binomial test on the 95% confidence level. The same test should be applied for shorter time periods such as

single years. A row of consecutive years with days outside the confidence interval indicates a time period affected by weekly PRCP cycles.

Weekly PRCP cycles are found in roughly 15 and 5% of the station records in Bolivia and Peru, respectively. A typical case of the error was detected in the time series of San Calixto, Bolivia. A significant WD lack on Saturdays is followed by a significant WD excess on Sundays (Figure 5). The signal on Saturdays is more pronounced than on Sundays. This pattern is explained due to rain events that extend over several days, contributing to the WD lack on Saturdays but not to the WD excess on Sundays. The test on annual time scale revealed that the problem started after a measurement interruption in 1992. Since 1993, the fraction of WDs differs strongly between Saturdays (0.1) and Sundays (0.3), while the period before 1992 is unaffected. The rain gauge of San Calixto is located at a Jesuit geophysical observatory (www.osc.org.bo). Until 1991, the observatory's director Rev. Ramón Cabré Roige (Udias, 2003) was responsible for the measurements. Since the resumption of the observations in 1993, measurements are taken by the observatory's caretaker who has his days off on Sundays. Because measurements are taken in the morning, they are ascribed to the previous day, which shifts the WD lack from Sundays to Saturdays.

In some station records, significant weekly PRCP cycles are found that do not follow explicable patterns. For instance, there are several cases where the WD fraction is evenly distributed among the days of the week except for one day with a much higher WD fraction. Such a pattern may occur if measurements are taken very infrequently (often just once a week) or if external factors disturb the observations. For instance, water from irrigation of the station's compound or vicinity may fill the rain gauge, as it happened at a station in Italy (Y. Brugnara, 2016, pers. comm.). Hence, if the pattern of a weekly PRCP cycle cannot be attributed to a certain cause, the records should be considered suspicious. However, if the error is explicable such as in the case of San Calixto, it is not unreasonable to expect that total rainfall on larger time scales will be more or less preserved (Viney and Bates, 2004). Hence, the disturbance of rain sums is most likely negligible on time scales ≥7 days. However, the data should not be used on shorter time scales, such as for the ETCCDI indices Rx1day or Rx5day. Furthermore, indices using thresholds,
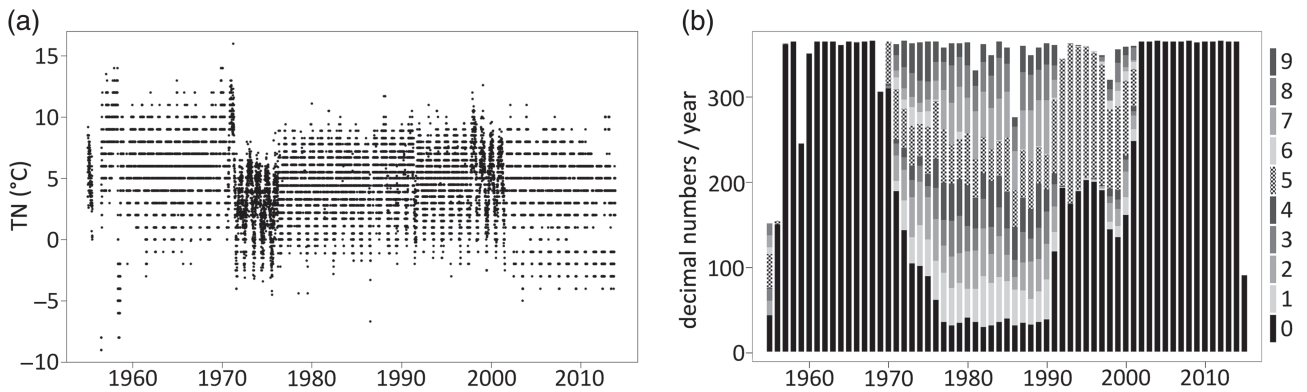
Figure 6. (a) The minimum temperature (TN) time series of Tiraque (Bolivia) reveals six major measurement precision transitions in 1956, 1970, 1976, 1991, 1998, and 2002. Precisions switch between 0.1 °C, 0.5 °C, 1 °F, and 1 °C. (b) Yearly frequencies of decimal numbers of the TN time series of Tiraque. The state of the measurement precision cannot be clearly defined in some segments (e.g. 1971–1975 or 1998–2002). Between 1976 and 1991, measurements were originally recorded in full degrees Fahrenheit.

such as the R10mm or the CWD, are also affected by weekly PRCP cycles.

4.3. Climate variable independent errors

4.3.1. Measurement precision inconsistencies

In this section, the term 'precision' is not used in a mathematical way, but rather as it is used in previous studies (Zhang *et al.*, 2009; Rhines *et al.*, 2015). Hence, the measurement precision is defined here as the precision of the reported observations, which may be influenced by the instrument type, the observing practice, unit conversions, or transcription errors.

Coarse measurement precision and switches of the precision state are known from a large number of datasets (Schaal and Dale, 1977; Nese, 1994; Trewin, 2010; Rhines *et al.*, 2015). This is an important source of uncertainty that has not been adequately addressed (Rhines *et al.*, 2015). Therefore, Rhines *et al.* (2015) introduced a precision decoding algorithm based on a hidden Markov model that was successfully applied on a large number of time series. Currently, the algorithm requires that the observers follow deterministic rules for rounding and conversion of the data. If the data structure diverges from this idealized behaviour, the algorithm may not perform satisfactorily. This was the case for the vast majority of the Bolivian station records.

The most frequent measurement precisions in the Bolivian and Peruvian datasets are 0.1, 0.2, 0.5, and 1.0 °C for temperature, and 0.1, 0.5, and 1.0 mm for PRCP. Nearly all records contain precision transitions and segments of low precisions (such as 1 °C or 1 mm). Often, the measurement precisions are indistinct and cannot be assigned to one precision state. Furthermore, odd frequencies of decimal numbers are detected in many time series. For instance, in several cases only the decimal numbers 0–4 are reported, while the decimal numbers 5–9 are completely missing, and sometimes only the decimal numbers 0 and 2 are reported.

An illustrative example of precision inconsistencies is found in the time series of the Bolivian station Tiraque

(Figure 6(a)). Note the large inhomogeneities that may come along with precision switches (e.g. in 1976). Between 1976 and 1991, the precision structure can be attributed to a unit conversion from full degrees Fahrenheit to 0.1 °C. A typical fingerprint of this conversion is the absence of 'fives' in the decimals of the Celsius scale. However, in the time series of Tiraque, 'sixes' are missing instead (Figure 6(b)). Apparently, degrees Fahrenheit values that convert to Celsius decimals of 5/9 (e.g. 42 °F) were erroneously rounded down.

As precision transitions occur usually simultaneously in TX and TN (sometimes also in PRCP), the most likely cause is observer changes. However, instrument replacements (identified as the main cause in Tiraque), new observer instructions, or new measurement forms may also cause precision inconsistencies. Odd frequencies of decimal numbers could be explained by erroneous instrument reading. For instance, if thermometer tick marks of 0.2 °C are just counted or interpreted as 0.1 °C, only the decimal numbers 0–4 will be reported. Furthermore, transcription errors (see also Section 4.3.3) may influence the measurement precision. For example, a consistent misinterpretation of observer's handwritten 'fives' by 'twos' leads to data records containing only decimal numbers of 'zeros' and 'twos' if the original precision state is 0.5 °C. As previously demonstrated, unit conversions may also lead to precision inconsistencies. Even though the Celsius scale has always been in use in the Central Andean area, several time series were apparently converted from Fahrenheit to Celsius scale. Similar cases were found for PRCP, where original measurements in 0.05 or 0.10 inches were converted to the metric scale. In Tiraque, original documents confirmed that measurements were reported in full degrees Fahrenheit. Most likely, the issue emerged due to the use of thermometers having both Celsius and Fahrenheit scales, which misled some observers to report the values of the wrong scale.

Odd frequencies of decimal numbers and, to a lower extent, unit conversions introduce a bias in the time series and are therefore the most problematic form of precision inconsistencies. Low and varying precisions, on the other
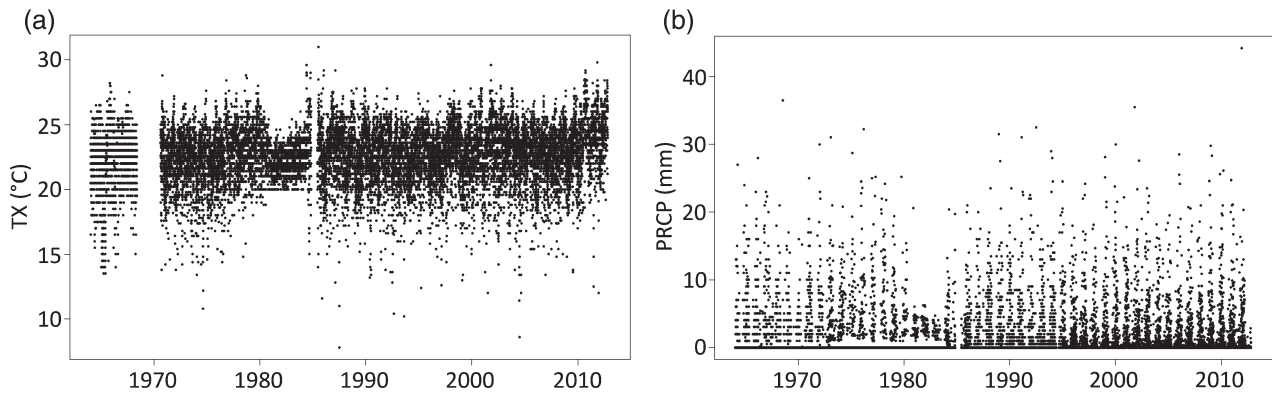
(a)

(b)

Figure 7. (a) Maximum temperature (TX) measurements from Urubamba (Peru). In the early 1980s, the variability is strongly reduced. Furthermore, values below 20 °C seem to be truncated. (b) Simultaneously to the TX measurements, the variability of precipitation (PRCP) records in Urubamba is diminished during the same years. The error coincides with the years of an internal conflict in Peru.

hand, do most likely not adversely affect means and standard deviations of large and long term datasets (Nese, 1994). However, many statistical methods assume that data are continuously distributed (Rhines *et al.*, 2015). For instance, the effect of precision inconsistencies is not negligible for threshold statistics (Schaal and Dale, 1977; Nese, 1994; Zhang *et al.*, 2009; Rhines *et al.*, 2015). Hence, the error affects a large number of the ETCCDI indices, such as the 'number of frost days' (FD), 'number of summer days' (SU), or the CDD and CWD.

4.3.2. Reduced variability in time series segments

A segment in a time series may have clearly reduced variability compared to the rest of the records. Different climate variables may be affected simultaneously. The error is detected rather easily by data visualization or by algorithms searching for inconsistencies of the standard deviation.

Roughly 20–30% of the Bolivian and Peruvian stations records are affected by this error, whereas TX and TN are more frequently affected than PRCP observations. An exemplary case of reduced variability in time series segments is found at the station Urubamba in Peru (Figure 7). The figure only shows TX (Figure 7(a)) and PRCP (Figure 7(b)) records, but TN as well as temperature measurements at 0700, 1300, and 1900 LST are affected too. As all measured variables are affected, the most likely cause is an observer error. In Urubamba, the affected time period coincides with the outbreak of the internal conflict between the Peruvian government and Maoist and socialistic guerrilla groups in the early 1980s.

During social unrests, the frequency of quality issues and data gaps in weather observations is increased. For example, the period of World War II is data-poor in many datasets (e.g. Brunet *et al.*, 2014a, 2014b). Data quality may be affected by station relocations to unsuitable, but easier accessible sites, changes to new and barely instructed observers, or pure estimation of measurements (Westerberg *et al.*, 2009). It has also been found that broken thermometers may cause a variability reduction. Furthermore, if the issue occurs abruptly and uniformly, it may

be caused by a decimal point error or an untransformed unit change. However, a clear error attribution is usually not possible such as in the case of Urubamba, and the affected time series segments cannot be used for any data analyses.

4.3.3. Transcription errors

Data in a digital dataset may differ from original handwritten values due to errors in the transcription process that are hard to avoid completely. If the digitization of the data is done in a double-keying process, comparison of the independent transcriptions allows for the detection of inconsistencies (Kunkel *et al.*, 2005; WMO, 2007). This method is partly applied at SENAMHI Peru. If inconsistencies between two transcriptions are found or if there are other doubts about the transcription, comparing the digital data with the original documents is mostly the only way to clearly detect and correct transcription errors. The benefit of such comparisons was mentioned in earlier studies (e.g. Durre *et al.*, 2008; Fiebrich and Crawford, 2009; Brunet *et al.*, 2014a). The most common transcription errors in the datasets of SENAMHI Bolivia and AASANA are described below.

Incomplete availability of original data sheets at the time of digitization caused some considerable data gaps in the digital Bolivian datasets. Furthermore, insufficient understanding of the original document structure by data transcribers resulted in various errors. For instance, in the original data forms used at the airports, the sign of temperature measurements is reported in a separate column. In some cases, the sign was not transcribed, leading to missing negative temperature records. Similarly, decimals numbers of PRCP in the same forms are stored in a separate column and were therefore not always transcribed. In some transcriptions, negative temperature values were interpreted as either negative or positive decimals places.

According to WMO guidelines (Plummer *et al.*, 2003), TX and PRCP measurements should be credited to the previous day if only a morning measurement is taken at a station. It is reported from many networks that such shifts were done directly when reporting the observation, either manually by the observer (Reek *et al.*, 1992;

Fiebrich and Crawford, 2009) or by means of specially keyed data forms (Menne *et al.*, 2012). In other cases, data shifts were introduced in post processing (Turco *et al.*, 2013). However, shifting practices were often not applied uniformly and they changed in time, resulting in undocumented inhomogeneous shifts among many datasets (Reek *et al.*, 1992; Kunkel *et al.*, 2005; Menne *et al.*, 2012; Turco *et al.*, 2013). The comparison between original documents and digitized data in Bolivia revealed many discrepancies of dates, varying between back shifts of time series of up to 3 days and forward shifts of 1 day. Note that shifting practices are not recorded in Bolivian metadata, as it is the case for other networks (e.g. Kunkel *et al.*, 2005). Besides accessing the original documents, detection of shifts may also be possible by comparing measurements of neighbouring stations (Reek *et al.*, 1992; Menne *et al.*, 2012) or using other datasets as references (Turco *et al.*, 2013). However, these methods rely on high correlations between time series, and the shifting state of the reference data must be known. This hardly applies for the Bolivian station observations.

Other, less frequent transcription errors in the digitized Bolivian datasets are the misinterpretation of the observer's handwriting (see also Section 4.3.1) and swapping of time series segments (e.g. the months June and July). Similar problems are reported from many other manned networks (Feng *et al.*, 2004; Fiebrich and Crawford, 2009; Westerberg *et al.*, 2009).

5. Discussion

5.1. Detection of systematic errors

Many QC methods have problems in detecting the data quality issues described in this study. The error detection of most approaches focuses rather on suspicious single values than on patterns in the data. The frequently used spatial consistency test relies on the availability of suitable neighbouring stations, and the test's efficiency is hence impeded by complex terrain and sparse station networks (Plummer *et al.*, 2003; Kunkel *et al.*, 2005; WMO, 2008). Furthermore, the application of internal consistency tests that draw on the physical relationships between different climatological variables (Plummer *et al.*, 2003; Trewin, 2013) may be limited due to a low number of observed variables.

As QC tests should be designed to detect known data problems (Durre *et al.*, 2008), existing methods should be enhanced for the detection of systematically occurring data quality issues. However, there are three main difficulties: in datasets such as those from the Central Andes, data quality issues occur in a highly variable manner in time, the errors have a large amount of individual specificities, and several errors may occur simultaneously. Hence, even methods that have been successfully applied on other datasets [e.g. the method for precision decoding by Rhines *et al.* (2015), see Section 4.3.1] may be ineffective in such cases. Possibly, novel error detection approaches that do not depend on previously defined assumptions such as machine learning

(Bishop, 2006; Alpaydin, 2014; Lakshmanan *et al.*, 2015) could successfully address the issue. However, the use of such approaches in error detection of station observations is a new field, and to the authors' knowledge, no such work has yet been published in a peer reviewed journal.

According to WMO (1993), machine and man together can achieve much more in quantity and quality of work than either can alone. For the Bolivian and Peruvian station networks, this statement certainly still holds true. Data visualization is an excellent tool that can be of great help, as the human mind is especially skilful at identifying spatial patterns (Plummer *et al.*, 2003; Klein Tank *et al.*, 2009). Hence, visual QC allows the detection of all sorts of quality issues, independent of specific error characteristics. Furthermore, visual QC allows for the assessment of systematic errors, which is crucial for the further use of the data (see Section 5.2). Various QC software tools, such as RClimDex (Zhang and Yang, 2004) or RClimTool (Llanos, 2014), offer graphic data outputs. However, these plots are not necessarily suitable to detect systematic patterns in the data. For instance, time series are often visualized by lines that often hide certain errors such as missing temperature intervals. A more illustrative general overview is achieved by point plots.

The methods used in this work to detect and identify common data quality issues may be applied to any observational time series, as they do not depend on neighbouring stations or measurements of other, physically related climate variables. Of course, the suggested methods are not a self-contained QC approach, but they should be combined with other test methods. As systematically occurring data quality issues may reduce the effectiveness of detection algorithms, we recommend identifying and treating (see Section 5.2) these errors before applying other QC tests. Furthermore, data quality issues may require adapting the parameters of statistical QC tests. For instance, the threshold for consecutive equal values must be higher if measurement precisions are low in order to avoid overdetection.

5.2. Overcoming systematic errors in time series

The best way to overcome systematically occurring errors is to prevent their occurrence. As in other manned station networks (e.g. Fiebrich and Crawford, 2009), a large number of data quality issues in the Bolivian and Peruvian records can be traced back to observer errors. Hence, the importance of well-trained and informed personnel cannot be stressed enough (Plummer *et al.*, 2003). Strengthening observer training and providing clear and simple measurement and maintenance guidelines may have a large positive impact on data quality with relatively little effort. Therefore, in the context of this study, readily comprehensible guidelines for station observers were produced. However, in order to maintain the quality of the measurements, further actions are necessary. WMO (2008) suggests near real-time QC for manned stations, because error sources can thereby be identified and eliminated. Data transmission methods such as the data transfer by mobile

phone at some Peruvian stations offer new options for such approaches. Also, controlling the quality of station records that are transmitted conventionally directly after delivery, and intervening immediately if needed, could reduce the error rate drastically.

However, data users can generally not influence measurement practices, and such efforts are an investment for the future rather than of use for today's climatological research. Hence, data users have to deal with data quality issues in many datasets. The rejection of time series affected by errors may result in substantial loss of available records (e.g. Westerberg *et al.*, 2009; Kizza *et al.*, 2012), and the removal of suspicious values causes data gaps. However, the completeness of time series is very important, as any dataset containing a large amount of missing data is of little use to climatologists (Plummer *et al.*, 2003). According to WMO (1989), not more than five missing temperature measurements in a month (only three if occurring in succession), and no missing daily PRCP totals (if rain is suspected) are permitted for calculating monthly means and sums, respectively. Hence, strategies should be defined on how time series affected by data quality issues may still be usable for climatological analyses. This is particularly important for regions of sparse station networks.

Besides rejecting station records affected by systematically occurring errors from further use, there are four options that may restore the data's usability:

1. Error assessment: Identifying the specific error characteristics and potentially detecting the error source enables users to decide for what analyses the data are still usable. This is demonstrated by means of ETCCDI indices in Section 4. As the specificities of the data quality issues vary significantly, each case needs to be evaluated separately. The most effective tools for the error assessment are suitable data visualizations.
2. Data correction: If the source of an error is clearly identified and quantified, the data may be correctable, as in the case of the TN time series of Progreso (see Section 4.1.1). Whenever possible, corrections should be made in order to maintain the largest possible applicability of the data. However, corrections must be done carefully in order to avoid inappropriate alteration of single measurements. Ideally, other data sources should be used to confirm the correction. In the case of Progreso, for instance, parallel thermograph measurements may allow a cross control.
3. Time series trimming: If an error is not correctable and compromises the intended purposes, the affected time series segment may be removed. The remaining unaffected records, if still of useful length, can be used for the intended study.
4. Data adjustment for certain analyses: Some data quality issues can be addressed by approximating the original data structure in order to make the data adequate for specific analyses. For instance, Zhang *et al.* (2009) introduced a method to artificially restore a data precision of 0.1 °C, which is the measurement precision recommended by WMO (1996, 2008). The approach

makes data of low precision suitable for threshold statistics, and hence for various ETCCDI indices (see Section 4.3.1). As good results are achieved even if measurement records are over-adjusted (Zhang *et al.*, 2009), the method is also applicable to time series of inconsistent measurement precisions and indefinable precision states.

6. Conclusions

Data quality issues are often not adequately addressed and contribute to the uncertainty in results of climatological studies. As many factors affecting the quality of observational records (e.g. poor station siting) are hardly detectable in the measurement data, metadata are an important source for assessing data quality. However, in many station networks, metadata are largely missing and/or of poor quality. For such cases, this study suggests different methods on how to generate or improve metadata.

The most common systematically occurring errors in the predominantly manned Bolivian and Peruvian station networks were identified and attributed. The same or similar errors are found in many other station networks worldwide, especially if observation conditions are comparable. Most of the data quality issues are related to observer errors. Therefore, the most effective way to prevent the occurrence of systematic errors is to improve observer training and to establish a near real-rime QC that allows quick intervention if an error occurs.

Many QC approaches have trouble in detecting systematically occurring data quality issues. These methods should be enhanced in order to address this shortcoming. However, the detection of such quality issues is made difficult by the high temporal variability of error occurrence, the vast number of error specificities, and error interactions. As data quality issues described in this article affect the efficiency of many QC algorithms, these systematic errors should be treated first, and parameters of statistical QC tests should be adapted if required.

Data visualization has been shown to be an effective tool to identify and attribute data quality issues. Providing this information enables data users to decide for what purposes a station record is still usable, although affected by a systematic error. In ideal cases, the defective data are even correctable. Furthermore, affected time series segments may be removed from the station records, and for some specific applications, methods to adjust data can be applied. The gain of usable time series achieved by this approach is particularly important in sparse station networks.

All factors that affect data quality may also cause break points in time series (WMO, 2008) and may hence hamper the detection and correction of other inhomogeneities (Domonkos, 2013). This is particularly critical in areas with low station density such as the Central Andes, where low station correlations strongly reduce the performance of relative data homogenization methods, and where data adjustments potentially increase the inhomogeneity of

the data (Gubler *et al.*, 2017; personal communication). Because homogenization of monthly or annual means may not be sufficient for daily time scales (e.g. Costa and Soares, 2009; Trewin, 2013; Brönnimann, 2015), daily homogenization would be needed to study extremes that are often most strongly affected by data quality issues. This, however, would require even higher station correlations due to the lower signal-to-noise ratio on shorter time scales. Furthermore, statistical data adjustments of the common data quality issues presented in this article would often not be adequate because of the discrete character of the errors. In summary, the selection and treatment of time series based on a comprehensive QC reduces the number of inhomogeneities. This may thus contribute to improving data homogeneity in the Central Andean and comparable networks more than statistical data homogenization methods. As a next step, nevertheless, carefully applying statistical data homogenization may further increase the data homogeneity. The findings on the systematically occurring data quality issues of this study may be useful for the creation of realistic benchmark datasets for QC and homogenization algorithms.

Acknowledgements

References

Aguilar E, Auer I, Brunet M, Peterson TC, Wieringa J. 2003. *Guidelines on Climate Metadata and Homogenization*. WCDMP No 53, WMO/TD No 1186. World Meteorological Organization, Geneva, Switzerland.

Alexander LV, Zhang X, Peterson TC, Caesar J, Gleason B, Klein Tank AMG, Haylock M, Collins D, Trewin B, Rahimzadeh F, Tagipour A, Rupa Kumar K, Revadekar J, Griffiths G, Vincent L, Stephenson DB, Burn J, Aguilar E, Brunet M, Taylor M, New M, Zhai P, Rusticucci M, Vazquez-Aguirre JL. 2006. Global observed changes in daily climate extremes of temperature and precipitation. *J. Geophys. Res. Atmos.* **111**: D05109, doi: 10.1029/2005JD006290.

Alpaydin E. 2014. *Introduction to Machine Learning*, 3rd edn. MIT Press: Cambridge, MA, London.

Austin JF, McConnell A. 1980. James Six F.R.S. – Two hundred years of the six's self-registering thermometer. *Notes Rec. R. Soc. Lond.* **35**: 49–65, doi: 10.1098/rsnr.1980.0004.

Bishop C. 2006. *Pattern Recognition and Machine Learning*. Springer-Verlag: New York, NY.

Bližňák V, Valente MA, Bethke J. 2015. Homogenization of time series from Portugal and its former colonies for the period from the late 19th to the early 21st century. *Int. J. Climatol.* **35**: 2400–2418, doi: 10.1002/joc.4151.

Brönnimann S. 2015. *Climatic Changes Since 1700*. Springer International Publishing: Cham, Switzerland, doi: 10.1007/978-3-319-19042-6_4.

Brunet M, Jones P. 2011. Data rescue initiatives: bringing historical climate data into the 21st century. *Clim. Res.* **47**: 29–40, doi: 10.3354/cr00960.

Brunet M, Gilabert A, Jones P, Efthymiadis D. 2014a. A historical surface climate dataset from station observations in Mediterranean North Africa and Middle East areas. *Geosci. Data J.* **1**: 121–128, doi: 10.1002/gdj3.12.

Brunet M, Jones PD, Jourdain S, Efthymiadis D, Kerrouche M, Boroneant C. 2014b. Data sources for rescuing the rich heritage of Mediterranean historical surface climate data. *Geosci. Data J.* **1**: 61–73, doi: 10.1002/gdj3.4.

Cerveny RS, Balling RC. 1998. Weekly cycles of air pollutants, precipitation and tropical cyclones in the coastal NW Atlantic region. *Nature* **394**: 561–563, doi: 10.1038/29043.

Ciach GJ. 2003. Local random errors in tipping-bucket rain gauge measurements. *J. Atmos. Oceanic Tech.* **20**: 752–759, doi: 10.1175/1520-0426(2003)20<752:lreitb>2.0.co;2.

Costa A, Soares A. 2009. Homogenization of Climate Data: Review and New Perspectives Using Geostatistics. *Math. Geosci.* **41**: 291–305, doi: 10.1007/s11004-008-9203-3.

Daly C, Gibson WP, Taylor GH, Doggett MK, Smith JI. 2007. Observer bias in daily precipitation measurements at United States cooperative network stations. *Bull. Am. Meteorol. Soc.* **88**: 899–912, doi: 10.1175/bams-88-6-899.

Davey CA, Pielke RA Sr. 2005. Microclimate exposures of surface-based weather stations: implications for the assessment of long-term temperature trends. *Bull. Am. Meteorol. Soc.* **86**: 497–504, doi: 10.1175/BAMS-86-4-497.

DeLisi MP, Cope AM, Franklin JK. 2001. Weekly precipitation cycles along the northeast corridor? *Weather Forecast.* **16**: 343–353, doi: 10.1175/1520-0434(2001)016<0343:WPCATN>2.0.CO;2.

Diaz HF, Bradley RS, Ning L. 2014. Climatic changes in mountain regions of the American Cordillera and the Tropics: historical changes and future outlook. *Arct. Antarct. Alp. Res.* **46**: 735–743, doi: 10.1657/1938-4246-46.4.735.

Domonkos P. 2013. Measuring performances of homogenization methods. *Q. J. Hung. Meteorol. Serv.* **117**: 91–112.

Donat MG, Alexander LV, Yang H, Durre I, Vose R, Dunn RJH, Willett KM, Aguilar E, Brunet M, Caesar J, Hewitson B, Jack C, Klein Tank AMG, Kruger AC, Marengo J, Peterson TC, Renom M, Oria Rojas C, Rusticucci M, Salinger J, Elrayah AS, Sekele SS, Srivastava AK, Trewin B, Villarroel C, Vincent LA, Zhai P, Zhang X, Kitching S. 2013. Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: the HadEX2 dataset. *J. Geophys. Res. Atmos.* **118**: 2098–2118, doi: 10.1002/jgrd.50150.

Durre I, Menne MJ, Vose RS. 2008. Strategies for evaluating quality assurance procedures. *J. Appl. Meteorol. Climatol.* **47**: 1785–1791, doi: 10.1175/2007JAMC1706.1.

Durre I, Menne MJ, Gleason BE, Houston TG, Vose RS. 2010. Comprehensive automated quality assurance of daily surface observations. *J. Appl. Meteorol. Climatol.* **49**: 1615–1633, doi: 10.1175/2010JAMC2375.1.

Fall S, Watts A, Nielsen-Gammon J, Jones E, Niyogi D, Christy JR, Pielke RA. 2011. Analysis of the impacts of station exposure on the U.S. Historical Climatology Network temperatures and temperature trends. *J. Geophys. Res. Atmos.* **116**: 15, doi: 10.1029/2010JD015146.

Feng S, Hu Q, Qian W. 2004. Quality control of daily meteorological data in China, 1951–2000: a new dataset. *Int. J. Climatol.* **24**: 853–870, doi: 10.1002/joc.1047.

Fiebrich CA, Crawford KC. 2009. Automation: a step toward improving the quality of daily temperature data produced by climate observing networks. *J. Atmos. Oceanic Tech.* **26**: 1246–1260, doi: 10.1175/2009JTECHA1241.1.

Hubbard KG, Lin X. 2006. Reexamination of instrument change effects in the U.S. Historical Climatology Network. *Geophys. Res. Lett.* **33**: L15710, doi: 10.1029/2006GL027069.

Kizza M, Westerberg I, Rodhe A, Ntale HK. 2012. Estimating areal rainfall over Lake Victoria and its basin using ground-based and satellite data. *J. Hydrol.* **464–465**: 401–411, doi: 10.1016/j.jhydrol.2012.07.024.

Klein Tank AMG, Zwiers FW, Zhang X. 2009. *Guidelines on Analysis of extremes in a changing climate in support of informed decisions for adaptation.* WCDMP No 72, WMO/TD No 1500. World Meteorological Organization, Geneva, Switzerland.

Kumamoto M, Otsuka M, Sakai T, Hamagami T, Kawamura H, Aoshima T, Fujibe F. 2013. Field experiment on the effects of a nearby asphalt road on temperature measurement. *Sci. Online Lett. Atmos.* **9**: 56–59, doi: 10.2151/sola.2013-013.

Kunkel KE, Easterling DR, Hubbard K, Redmond K, Andsager K, Kruk MC, Spinar ML. 2005. Quality control of pre-1948 Cooperative Observer Network Data. *J. Atmos. Oceanic Tech.* **22**: 1691–1705, doi: 10.1175/JTECH1816.1.

Lakshmanan V, Gilleland E, McGovern A, Tingley M. 2015. Machine Learning and Data Mining Approaches to Climate Science. In *Proceedings of the 4th International Workshop on Climate Informatics.* Springer International Publishing, Cham, Switzerland.

Llanos L. 2014. *RClimTool Manual del Usuario.* Clima y Sector agropecuario Colombiano: Colombia.

Mahmood R, Foster SA, Logan D. 2006. The GeoProfile metadata, exposure of instruments, and measurement bias in climatic record revisited. *Int. J. Climatol.* **26**: 1091–1124, doi: 10.1002/joc.1298.

Mantas VM, Liu Z, Caro C, Pereira AJSC. 2015. Validation of TRMM multi-satellite precipitation analysis (TMPA) products in the Peruvian Andes. *Atmos. Res.* **163**: 132–145, doi: 10.1016/j.atmosres.2014.11.012.

Menne MJ, Durre I, Vose RS, Gleason BE, Houston TG. 2012. An overview of the global historical climatology network-daily database. *J. Atmos. Oceanic Tech.* **29**: 897–910, doi: 10.1175/JTECH-D-11-00103.1.

Nese JM. 1994. Systematic biases in manual observations of daily maximum and minimum temperature. *J. Clim.* **7**: 834–842, doi: 10.1175/1520-0442(1994)007<0834:SBIMOO>2.0.CO;2.

New M, Hewitson B, Stephenson DB, Tsiga A, Kruger A, Manhique A, Gomez B, Coelho CAS, Masisi DN, Kululanga E, Mbambalala E, Adesina F, Saleh H, Kanyanga J, Adosi J, Bulane L, Fortunata L, Mdoka ML, Lajoie R. 2006. Evidence of trends in daily climate extremes over southern and west Africa. *J. Geophys. Res. Atmos.* **111**: D14102, doi: 10.1029/2005JD006289.

Petrovic P. 2001. In *Selection of Data Sets by Quality and its Role in Climate Research*, India M, Bonillo D (eds). Springer Berlin: Heidelberg, Berlin, Germany.

Pielke R Sr, Nielsen-Gammon J, Davey C, Angel J, Bliss O, Doesken N, Cai M, Fall S, Niyogi D, Gallo K, Hale R, Hubbard KG, Lin X, Li H, Raman S. 2007. Documentation of uncertainties and biases associated with surface temperature measurement sites for climate change assessment. *Bull. Am. Meteorol. Soc.* **88**: 913–928, doi: 10.1175/BAMS-88-6-913.

Plummer N, Allsopp T, Lopez JA. 2003. *Guidelines on climate observations networks and systems.* WCDMP No 53, WMO/TD No 1186. World Meteorological Organization, Geneva, Switzerland.

Reek T, Doty SR, Owen TW. 1992. A deterministic approach to the validation of historical daily temperature and precipitation data from the cooperative network. *Bull. Am. Meteorol. Soc.* **73**: 753–762, doi: 10.1175/1520-0477(1992)073<0753:ADATTV>2.0.CO;2.

Rhines A, Tingley MP, McKinnon KA, Huybers P. 2015. Decoding the precision of historical temperature observations. *Q. J. R. Meteorol. Soc.* **141**: 2923–2933, doi: 10.1002/qj.2612.

Schaal LA, Dale RF. 1977. Time of observation temperature bias and "climatic change". *J. Appl. Meteorol.* **16**: 215–222, doi: 10.1175/1520-0450(1977)016<0215:TOOTBA>2.0.CO;2.

Schmocker J, Liniger HP, Ngeru JN, Brugnara Y, Auchmann R, Brönnimann S. 2016. Trends in mean and extreme precipitation in the Mount Kenya region from observations and reanalyses. *Int. J. Climatol.* **36**: 1500–1514, doi: 10.1002/joc.4438.

Schultz DM, Mikkonen S, Laaksonen A, Richman MB. 2007. Weekly precipitation cycles? Lack of evidence from United States surface stations. *Geophys. Res. Lett.* **34**: L22815, doi: 10.1029/2007GL031889.

Sene KJ, Farquharson FAK. 1998. Sampling errors for water resources design: the need for improved hydrometry in developing countries. *Water Resour. Manag.* **12**: 121–138, doi: 10.1023/a:1007952319495.

Sillmann J, Kharin VV, Zwiers FW, Zhang X, Bronaugh D. 2013. Climate extremes indices in the CMIP5 multimodel ensemble: Part 2. Future climate projections. *J. Geophys. Res. Atmos.* **118**: 2473–2493, doi: 10.1002/jgrd.50188.

Simmonds I, Keay K. 1997. Weekly cycle of meteorological variations in Melbourne and the role of pollution and anthropogenic heat release. *Atmos. Environ.* **31**: 1589–1603, doi: 10.1016/S1352-2310(96)00344-5.

Štěpánek P, Zahradníček P, Farda A. 2013. Experiences with data quality control and homogenization of daily records of various meteorological elements in the Czech Republic in the period 1961–2010. *Q. J. Hung. Meteorol. Serv.* **117**: 123–141.

Stjern CW. 2011. Weekly cycles in precipitation and other meteorological variables in a polluted region of Europe. *Atmos. Chem. Phy.* **11**: 4095–4104, doi: 10.5194/acp-11-4095-2011.

Tokay A, Bashor PG, McDowell VL. 2010. Comparison of rain gauge measurements in the mid-Atlantic region. *J. Hydrometeorol.* **11**: 553–565, doi: 10.1175/2009jhm1137.1.

Trewin B. 2010. Exposure, instrumentation, and observing practice effects on land temperature measurements. *Clim. Change* **1**: 490–506, doi: 10.1002/wcc.46.

Trewin B. 2013. A daily homogenized temperature data set for Australia. *Int. J. Climatol.* **33**: 1510–1529, doi: 10.1002/joc.3530.

Turco M, Zollo AL, Ronchi C, De Luigi C, Mercogliano P. 2013. Assessing gridded observations for daily precipitation extremes in the Alps with a focus on northwest Italy. *Nat. Hazards Earth Syst. Sci.* **13**: 1457–1468, doi: 10.5194/nhess-13-1457-2013.

Tuttle JD, Carbone RE. 2011. Inferences of weekly cycles in summertime rainfall. *J. Geophys. Res. Atmos.* **116**: D20213, doi: 10.1029/2011JD015819.

Udias A. 2003. *Searching the Heavens and the Earth: The History of Jesuit Observatories.* Springer Netherlands: Dordrecht, The Netherlands, 286, doi: 10.1007/978-94-017-0349-9.

Vertačnik G, Dolinar M, Bertalanič R, Klančar M, Dvoršek D, Nadbath M. 2015. Ensemble homogenization of Slovenian monthly air temperature series. *Int. J. Climatol.* **35**: 4015–4026, doi: 10.1002/joc.4265.

Viney NR, Bates BC. 2004. It never rains on Sunday: the prevalence and implications of untagged multi-day rainfall accumulations in the Australian high quality data set. *Int. J. Climatol.* **24**: 1171–1192, doi: 10.1002/joc.1053.

Vose RS, Easterling DR, Karl TR, Helfert M. 2005. Comments on "microclimate exposures of surface-based weather stations". *Bull. Am. Meteorol. Soc.* **86**: 504–506, doi: 10.1175/BAMS-86-4-504.

Vuille M, Bradley RS, Keimig F. 2000. Mean annual temperature trends and their vertical structure in the tropical Andes. *Geophys. Res. Lett.* **27**: 3885–3888, doi: 10.1029/2000GL011871.

Vuille M, Francou B, Wagnon P, Juen I, Kaser G, Mark BG, Bradley RS. 2008. Climate change and tropical Andean glaciers: past, present and future. *Earth Sci. Rev.* **89**: 79–96, doi: 10.1016/j.earscirev.2008.04.002.

Wang Q, Fan X, Wang M. 2014. Recent warming amplification over high elevation regions across the globe. *Clim. Dyn.* **43**: 87–101, doi: 10.1007/s00382-013-1889-3.

Westerberg I, Walther A, Guerrero J-L, Coello Z, Halldin S, Xu C-Y, Chen D, Lundin L-C. 2009. Precipitation data in a mountainous catchment in Honduras: quality assessment and spatiotemporal characteristics. *Theor. Appl. Climatol.* **101**: 381–396, doi: 10.1007/s00704-009-0222-x.

Wilby RL, Tomlinson OJ. 2000. The 'Sunday Effect' and weekly cycles of winter weather in the UK. *Weather* **55**: 214–222, doi: 10.1002/j.1477-8696.2000.tb04064.x.

WMO. 1989. *Calculation of monthly and annual 30-year standard normals.* WCDP No. 10, WMO/TD No. 341. World Meteorological Organization, Geneva, Switzerland.

WMO. 1993. *Guide on the Global Data-Processing System.* WMO No. 305. World Meteorological Organization, Geneva, Switzerland.

WMO. 1996. *Guide to Meteorological Instruments and Methods of Observation* (6th edn), WMO No. 8. World Meteorological Organization, Geneva, Switzerland.

WMO. 2002. Report of the Eighth Session of the GCOS/WCRP Atmospheric Observation Panel for Climate (AOPC), WMO/TD No. 1125, GCOS-76. World Meteorological Organization, Geneva, Switzerland.

WMO. 2007. *Climate Data Management Systems*, WMO/TD No. 1176. World Meteorological Organization, Geneva, Switzerland.

WMO. 2008. *Guide to Meteorological Instruments and Methods of Observation* (7th edn), WMO No. 8. World Meteorological Organization, Geneva, Switzerland.

WMO. 2010. *Manual on the Global Data-processing and Forecasting System.* WMO No. 485. World Meteorological Organization, Geneva, Switzerland.

WMO. 2011. *Guide to Climatological* Practices (3rd edn). WMO No. 100. World Meteorological Organization, Geneva, Switzerland.

Yilmaz H, Toy S, Irmak MA, Yilmaz S, Bulit Y. 2008. Determination of temperature differences between asphalt concrete, soil, and grass surfaces of the city of Ezurum, Turkey. *Atmosfera* **21**: 135–146, doi: 10.1029/2005JD006290.

You Q, Kang S, Aguilar E, Pepin N, Flügel W-A, Yan Y, Xu Y, Zhang Y, Huang J. 2011. Changes in daily climate extremes in China and their connection to the large scale atmospheric circulation during 1961–2003. *Clim. Dyn.* **36**: 2399–2417, doi: 10.1007/s00382-009-0735-0.

Zhang X, Yang F. 2004. *RClimDex (1.0) User Guide*. Climate Research Branch Environment Canada, Downsview, Ontario, Canada.

Zhang X, Zwiers FW, Hegerl G. 2009. The influences of data precision on the calculation of temperature percentile indices. *Int. J. Climatol.* **29**: 321–327, doi: 10.1002/joc.1738.

Zhang X, Alexander L, Hegerl GC, Jones P, Tank AK, Peterson TC, Trewin B, Zwiers FW. 2011. Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Clim. Change* **2**: 851–870, doi: 10.1002/wcc.147.