Abstract

Climate reconstructions have contributed tremendously to our understanding of changes in the climate system and will continue to do so. However, in climate science the focus has partly shifted away from past seasonal and annual mean climate towards weather variability and extreme events. Weather events are more directly relevant for climate impacts and they capture the scale at which important processes take place. Weather reconstructions therefore help to better understand atmospheric processes, particularly during extreme events, to assess decadal-to-multidecadal climate variability through the lens of weather changes, and they allow impact modelling of past events. Consequently, attempts are currently undertaken to extend weather data sets far back into the past. In this review I discuss methods of weather reconstructions that are in use today. The methods range from expert analyses to data assimilation, from analog approaches to machine learning. Products range from weather types to four-dimensional fields. The methods complement each other as they are based on different assumptions and are based on different data sets. Weather reconstructions require more meteorological data than climate reconstructions. Additional data rescue efforts are therefore needed.

1. Introduction

Climate reconstructions are widely used to document past environments and to place current climatic changes into context [1, 2]. They yield far reaching new insights on the mechanisms underlying climatic changes and on the impacts these changes have on society. Already 150 years ago, documentary data (historical texts, but also quantitative information on grape harvest dates, river ice, or lake levels) were used to reconstruct climate hundreds of years back [3]. Over the past few decades, an increasing number of climate proxies, progress in analytical techniques to extract information from natural archives, and new numerical techniques have allowed an ever more detailed view of past climates.

However, it is often not the average conditions over a month, season, or year that matter for society, but weather events, which are not resolved in climate reconstructions. The weather scale also matters for understanding the atmospheric processes behind climatic changes. In recent years, attention has therefore increasingly shifted away from seasonal climate characterisations towards the weather scale and specifically to weather extremes. Quantitative information on past weather is therefore becoming more and more important. This calls for a shift
from climate reconstructions to weather reconstructions. The “Atmospheric Circulation Reconstructions over the Earth” Initiative (ACRE [4]), launched in 2008, exemplifies this shift.

Weather reconstructions are useful for climate scientists to study how climatic changes evolve through altering atmospheric processes. Statisticians can make use of long weather time series to study extreme events and their behaviour under different background climates. Climate risk assessment increasingly uses numerical models, and these models require quantitative weather information, including from the past. Recent extreme weather events such as the floods in 2021 also demonstrated the importance of studying individual past extreme events, including also heatwaves and drought.

Consequently, large efforts have been undertaken to produce weather reconstructions. Perhaps the most well-known effort is the “Twentieth Century Reanalysis”, whose products 20CRv2c [5] and 20CRv3 [6] are widely used. However, there are also many other efforts and the methods are constantly refined and further developed. In this review, I present a summary of techniques used for weather reconstruction. The article does not provide a comprehensive overview of all products, nor their use, e.g., for climate risk assessment [7].

The review is organised as follows. Section 2 discusses the data underlying weather reconstructions, including natural proxies, documentary data, and instrumental measurements. In Section 3 I present simple methods of climate reconstruction such as event analyses and the generation of weather types or other daily indices. Section 4 addresses the generation of weather fields. This can again be accomplished by different means: expert approaches, geostatistical techniques, analog approaches, data assimilation, and machine learning. In Section 5, I provide a short discussion. Conclusions are drawn in Section 6.

2. Data underlying weather reconstructions

2.1. Natural proxies

Single weather events are not only relevant for impacts, but they also leave traces in proxies. I therefore start the review with a brief note of palaeoweather in proxies, noting that an in-depth review of this topic is not possible in the framework of this article.

Different types of extreme weather events can be studied using climate proxies. The grain size in lake sediments can indicate storms [8], sedimentary records from coastal lakes along the Gulf of Mexico and the southern states can indicate hurricanes [9]. Zhou et al. [10] find lagoon deposits of historical storms in Hainan that align well with documented storm events. Lagoon sediments in France were demonstrated to show distinct storm events [11]. Based on similar evidence, Mann et al. [12] reconstructed land-falling Atlantic hurricanes more than a millennium back in time.

A second type of extreme events captured in proxies are floods. Flood layers have been analysed in many varved lake sediments [13–15]. Flood layers indicate heavy precipitation events, and hence sediment cores can inform about the changing frequency of heavy precipitation over the last several thousand years.

Tree rings are mostly used to reconstruct growing season mean climate, but they also record shorter extremes. Frost rings in trees indicate cold air outbreaks and corresponding synoptic weather situations [16]. Wood anatomical studies indicate that so-called “blue rings” are a proxy for early frosts at the end of the growing season [17].

Large progress in laboratory analytics allow even more detailed analyses of climate proxies. Palaeowather studies from proxies can provide information on statistics of weather events and their changes over time. However, unless there is supporting evidence, events can often not be exactly dated.
2.2. Documentary data

Historical documents have the advantage that they can mostly be dated precisely. Weather events were often noted by different observers at different locations, which allows a spatial reconstruction of an individual weather situation. Very dedicated persons observed all their life, and their diary entries allow constructing weather time series. While the oldest weather-related information in documents from China and Egypt dates back several thousand years, documentary data that could possibly be used for daily weather reconstruction reach back at most to the late medieval period. These data are mostly from Europe, China, and Japan.

Overviews of documentary weather data have been published in the 1960s and 1970s [18], and subsequent efforts have led to several data bases with historical weather data. Examples for Europe include EURO-CLIMHIST [19], Tambora.org [20] and TEMPEST [21]. The REACHES database [22] provides weather and climate observations from China back to 1644. A compilation of documentary data from Japan is provided by the Japanese Climate Data Programme, https://jcdp.jp/. A comprehensive global overview of documentary climate data is given by Burgdorf [23].

While historical (non-instrumental) weather observations are often used for studying climate impacts or extreme events, their quantitative use in climate research was mostly restricted to monthly climate indices [24–26]. Within the PAGES group CRIAS [27], historians are now collaborating more closely with climatologists, exploring numerical methods which might eventually lead to quantitative weather reconstructions.

Of particular relevance for weather reconstructions are weather diaries. An early example is the diary by Wolfgang Haller in Zurich covering 1545–1576 [28]. Important observers in the 17th century were Johannes Kepler, Tycho Brahe, or Landgraf Hermann von Hessen [18, 29]. The longest weather chronicle by Gottfried Reyger in Gdansk covers the period 1721–1786 [30]. Weather diaries also help to reconstruct weather in Japan [31] and the Caribbean [32] since the 18th century.

To be useful for quantitative approaches, the daily weather must be in categorized or categorizable form. This includes information such as the number of rain days, wind direction, or sky conditions (which however may be difficult to categorize). An example of a weather diary with categorized (pictographic) weather information from Georg Christoph Eimmart, Nuremberg, is shown in Fig 1. Weather diaries provide important and direct information on long-term change in weather [30]. Albeit very sparse, they can possibly provide additional information to early instrumental measurements for the reconstructions of individual weather events. In any case, source criticism is required to assure where, when, and by whom the observations were actually made [28]. Records from different observers may not be comparable, and even a record from one observer may very likely be inhomogeneous. An example of how weather diaries can be used in a semi-quantitative way for weather reconstruction is presented in the next Section. Their potential for quantitative weather reconstructions remains to be explored.

Rich, documentary weather information in categorized form is available from ships. Data on wind direction from ship logs reach back to the 17th century. For instance, wind indices have been generated from ship log-book entries [33–36], reaching as far back as the 17th century. Methodologies on how to abstract and process wind observations have been developed [37]. In this case, all observations were from the same source (Royal Navy), the geographic region was restricted, and only the observation closest to midday per ship was chosen. The process of combining data from different ships is described in [33]. The resulting Westerly Index is further discussed in Section 3.2. Global marine data are today compiled in the ICOADS archive, which combines the heterogeneous data from different sources in a
common collection [38], although care is still required when deriving wind products (e.g., visual versus anemometer wind, changing ship heights).

2.3. Early instrumental data

The most valuable information for weather reconstructions are instrumental measurements. Long meteorological records that can form the basis for weather reconstructions were compiled within the projects IMPROVE [39], HISTALP [40], and ADVICE [41] for Europe, EMERLAC for Latin America [42] and numerous other projects [43–49]. Several long time series exist for Europe.

However, many series, especially shorter ones, have never been digitised. A recent inventory [50] revealed hundreds of series. During the last decade, large efforts were devoted to weather and climate data rescue, coordinated within ACRE. The data digitised in these efforts are now available for weather reconstruction [51].
With respect to marine data, the ICOADS database [38] compiles global instrumental data. However, data rescue activities still have a large potential to increase coverage. Numerous data rescue projects target marine data to inform weather reconstructions [52], among them “Old Weather”, “Weather Rescue at Sea”, or projects within the Newton Fund Weather and Climate Science for Service Partnership Programme.

2.4. Documentary and instrumental data on the winter of 1683/4

To demonstrate the value of documentary and early instrumental data I use the example of the winter 1683/4. This is arguably one of the earliest winters for which not just climate, but daily weather can be reconstructed. As Manley [53] put it: “We have enough in the way of daily weather observations for 1684 to go some way towards a reconstruction of the meteorology.”

Instrumental measurements of temperature [54] and pressure [55] are available from Paris, scattered measurements of (probably indoor) temperature from London by John Downes [53]. Measurements by Reyher in Kiel are unfortunately lost. Pressure measurements were made in Oxford by Robert Plot [56], which I converted to mean sea-level pressure [57] (for the temperature reduction, in order to estimate temperature in an unheated room, I used an 11-day moving average of mean temperature from Paris, the only available record). Also shown are categorized data from weather diaries of Robert Plot in Oxford and Johann Heinrich Fries in Zurich [19] as well as information from the weather diary of John Downes in London [53].

The data (Fig 2A) show two main cold phases concerning all four sites (Oxford, London, Paris, Zurich). The first phase started in December 1683 and peaked in mid-January 1684. Temperature dropped below 10 F (-12˚C) in London and to -15˚C in Paris. The phase was accompanied by high pressure, sunny or “fair” weather. Manley [53] assumes a blocking high, which was then followed (on 20–23 January) by a low over France, leading to slightly milder temperatures. The second cold period then started at the end of January and continued to mid-February. In London, almost exclusively easterly, northeasterly, or northerly winds were observed from 10 January to 14 February. This second cold episode was abruptly terminated by a frontal passage. An extreme pressure drop occurred in Oxford and one day later in Paris, accompanied by rain, southwesterly winds, and higher temperatures. A third, but less extreme cold phase (again, at all three locations) occurred in March.

The winter of 1683/4 is known as a particularly cold winter. In fact, frost fairs were held on River Thames, which was frozen for seven weeks in that winter (Fig 2B). On a monthly scale, the winter can be analysed in climate reconstructions; anomaly fields of temperature and sea-level pressure for January are shown in Fig 2C. The cold winter affected a larger area in Central and Western Europe; sea-level pressure anomalies show a negative North Atlantic Oscillation (NAO). This winter was one of several prominent winters in this phase of the Little Ice Age (for the equally cold winter 1684/5, see Cornes [58]) which coincided with a period of low solar activity [59]. Analysing the weather scale can help to better understand the underlying mechanisms.

3. Methods of weather reconstruction

3.1. Direct analyses of individual weather events

The simplest method of weather reconstruction is the direct analysis of individual weather events in instrumental or documentary data with the help of expert knowledge. Without even drawing weather maps, individual weather systems can be captured as shown in Fig 2. As a further example, Fig 3 shows the analysis of a hurricane that hit Cuba in 1794 and that was studied in detail by Garcia-Herrera et al. [61]. The storm caused dozens of deaths in Havanna, sank
numerous ships and moved on across the Florida Keys to the Gulf of Mexico. New Orleans was hit by a storm on 31 August, which might well have been the same storm. Instrumental observations were made onboard a Spanish ship by Tomás José de Ugarte y Lyano. A copy of the data was in the possession of Alexander von Humboldt (Fig 3A), who also studied the storm. The measurements onboard the ship (Fig 3C), together with additional documentary information, allows a possible weather history to be sketched (Fig 3B).

The first hurricane studied in this way was a storm in 1680 [62]. By combining ship-based observations with data from Caribbean Islands and eventually the pressure record from Paris, the track of the hurricane could be reconstructed; perhaps the storm underwent extratropical transition and arrived in Paris as a cyclone. With only few pieces of information, these authors were able to track a weather system over a large distance. Documentary evidence on hurricanes dates even back to 1600 [61]. Numerous other studies use instrumental and documentary data directly to interpret weather patterns 300 years back in time [63].

A prominent example of a climatic event that was studied on the weather scale using direct analyses of observations concerns the winter of 1708/9. Lenke [64] compiled several temperature series, complemented with pressure series and weather observations. They allow studying this millennial extreme on the weather scale.

3.2. Weather types and daily indices

For time series analyses of daily weather, pressure indices can be defined that allow a dynamical interpretation of temperature time series. Already two well-placed time series may allow
the construction of useful weather indices. Cornes et al. [65] use pressure data from Paris and London to derive a daily, NAO-like index. This index is expected to be related to westerly or southwesterly winds over the Channel. In fact, an independent index of wind over the Channel, reaching back to 1685 was generated by Wheeler [37] and Mellado-Cano et al. [66] based on daily wind observations from ships. Both indices are shown in Fig 4A (Dec.-Mar. averages). Note that these two indices are entirely independent. Their 30-yr moving correlation (Fig 4B) is between 0.4 and 0.8 over the entire period.

A slightly more complex diagnostic are weather type classifications (or circulation type classifications), which group days according to their synoptic situation into discrete types. Many countries have their own circulation type classifications pertaining to the weather in their region, e.g., the German Grosswetterlagen [65] or the Lamb Weather types for the British Isles [67]. A harmonisation of weather type classifications for European regions was achieved within COST Action 733 [68].

Weather type classifications were developed as useful diagnostics of past weather (at that time reaching back to the 1880s) in the early 20th century [69] and were widely used in the 1980s and 1990s. In 1972, Lamb [67] defined a set of weather types for the British Isles and started to apply the concept also to the past. He assigned his weather types to days in the 16th to 19th century [70]. The earliest period studied is the summer of 1588, known as a “year without a summer”. Weather types have also been defined and reconstructed for other regions. For instance, a weather type classification exists for the Arctic, reaching back to 1891 [71].
Weather types are a simple way to reconstruct past weather if only a limited number of observations is available. They can be purpose-built, such as weather types for Geneva for 1798–1821, defined with the aim of analysing the dynamical contribution to the cold summer of 1816 [72] or the warm, dry summers around 1800 [73]. A daily series of weather types for Switzerland was provided by Schwander et al. [74] back to 1763. They allowed analysing effects of the 11-yr sunspot cycle [75] or the contribution of atmospheric circulation to flood-rich decades [76]. Delaygue et al. [77] reconstructed Lamb weather types back to 1763 and used them to construct an NAO-like index (Fig 4, note that this index uses the Paris and London series, along with five additional pressure records). For comparison, two classical NAO indices (difference of Azores and Iceland standardized Dec.-Mar. sea-level pressure) are shown, one based on the reanalysis 20CRv3 [6] and one based on the reconstruction EKF400v2 [60], both calculated on the ensemble means. These indices capture a larger spatial scale than the Paris-London Index and the Westerly Index. Yet, the agreement of all indices is very good (in the early years the Westerly Index has a large amplitude due to few observations, whereas NAO_EKF400v2 has a small amplitude as it is based on the ensemble mean). Except for EKF400v2, all indices are daily and hence these time series allow a detailed view of day-to-day weather variability over the North-Atlantic-European sector and its decadal-to-multidecadal change.

Different methods are used for obtaining weather types for the past. Lamb’s reconstructions were originally based on subjective analyses, which later was developed into an objective

Fig 4. (a) Indices of westerly flow based on Paris-London pressure (PL), the Westerly Index (WI) and NAO indices based on the Lamb weather types (NAO_LWT), 20CRv3 (NAO_20CRv3) and EKF400v2 (NAO_EKF400v2), all for Dec.-Mar. All indices are standardized with respect to the 1850–1900 period (the right panel shows a zoom-in to the period 1750–1830). (b) 30-yr moving correlations between all indices (except NAO_20CRv3, which is the shortest).

https://doi.org/10.1371/journal.pclm.0000034.g004
approach [78]. If gridded data sets are available, quantitative definitions [68] can directly be applied to these fields. Alternatively, weather types reconstructions can be calibrated in a recent period, e.g., by determining the centroid of the weather types within a given set of variables (typically station series). For any realization of these variables in the past, a distance measure to each centroid can be defined (e.g., the Euclidian distance or Mahalanobis distance) and days can be assigned according to the shortest distance. The distances can also be converted into a probability [77].

Classification is a key strength of machine learning algorithms. Algorithms such as random forests, neural networks, or similar can be used. They need to be trained in a recent period and can then be applied to the past, this is further discussed in Section 4.6.

Weather type reconstructions are simple and therefore can be tailored. For instance, Schwander et al. [74] provide a version that also includes temperature series among the predictors and one that does not, which may be preferable in some applications that require independence of temperature. None of the approaches mentioned includes the time sequence in the reconstruction explicitly other than using pressure tendency as a variable, but it would be straightforward to include time dependence in such an approach, if desired.

A disadvantage of the circulation type classifications is that they use disjunct, discrete classes without an associated measure of strength (this lies in the nature of a classification). Further, circulation types are assumed stationary, i.e., dominating types in the 20th century are assumed to meaningfully present the weather also in the past. Depending on the classification scheme used, this assumption may not hold. Measures based on 500 hPa geopotential height, for instance, may be sensitive to temperature trends. Obviously, when applying weather type classifications defined in the 20th century to the past, the approach will never extrapolate (i.e., it will not generate new types). When using calibration and if seasons are treated separately, some types may occur too rarely, and one has to revert to classifications with fewer types. In fact, Schwander et al. [74] had to reduce the nine types of the MeteoSwiss classification “CAP9” to seven types. Finally, different circulation type classifications are hardly comparable [79].

4. Daily weather fields

4.1. Expert approaches

Ever since Heinrich Wilhelm Brandes drew the first synoptic weather maps in the 1810s [80], experts have drawn weather maps retrospectively to analyse past weather and climate events [24, 70, 81]. This approach essentially combines different types of observations with expert experience both on the physics of the atmosphere and on the reliability of the observations. It can incorporate time dependence and it entails a step of generalisation, which requires a notion of weather systems and their dynamics.

At the Climatic Research Unit, Hubert Lamb and John Kington derived hand-drawn pressure fields from available information [82, 83]. Kington [81] drew daily weather maps for the years 1781–1785, following a detailed procedure by extending the technique of drawing synoptic charts to include also data at non-standard times and descriptive data [82]. These hand-drawn maps are still used in the interpretation of extreme events [84]. However, there is obviously the wish to have fields in numerical form, generated with objective methods. Such approaches are presented in the following.

4.2. Geostatistical approaches

How can we provide daily fields of pressure or temperature in a quantitative way? If sufficient data of the target variable are available, geostatistical interpolation techniques can be used.
Station data can be interpolated in space and time, perhaps using additional information such as topography or other surface characteristics. For instance, daily fields of temperature and precipitation at 1 km resolution were produced for North America back to 1980 [85]. Some methods proceed in two steps (e.g., first interpolating monthly fields or fit a smooth trend and then interpolate daily anomalies from the monthly means). E-Obs, for instance, provides daily gridded data set for Europe back to 1950 [86] using a similar approach. It has recently been extended back to 1920 for temperature (minimum, mean and maximum) and precipitation at 0.1˚ and 0.25˚ resolution.

Daily data sets such as E-Obs are widely used for climate impact studies or for hydrological or agricultural modelling. It would therefore be desirable to extend these data sets back in time. However, the further back into the past the methods are extended, the more they suffer from data sparsity. Daily sea-level pressure fields have been interpolated for the North Atlantic-European region back to 1850. [87] For temperature, producing daily fields that far back has only rarely be attempted. The EUSTACE data set provides global daily temperature fields based only on temperature measurements back to 1851 [88] at a resolution of 0.25˚. The methods assumes that the observation is composed of the true temperature and a bias (informed by breakpoints). The former consists of a moving long-term average, large-scale interannual variability and daily, weather-related changes, each of which is expressed as a linear combination of further variables. The equation system is then solved iteratively. As an example, Fig 5B shows daily mean temperature on 18 August 1858 during a heatwave in Western Europe. This summer became known as “The Great Stink” in London. It triggered the construction of the London sewer system. In the EUSTACE data, the heatwave was centred over the Low Countries, but high values (daily mean temperature above 20˚C) extended to London. The figure also shows the fields from the 20CRv3 reanalysis (see Sect. 4.4). The comparison is interesting as the two approaches not only use completely different techniques but also are based on different input data (they share the sea-surface temperatures). The figure also indicates the ensemble spread (standard deviation) of both products, which however samples only part of the uncertainty. Both are ca. 1–2 K over land areas. Results from the two approaches agree

![Figure 5](https://doi.org/10.1371/journal.pclm.0000034.g005)
with each other. Differences are consistent with the indicated ensemble spread. The larger differences over central-southern France coincide with a region in which the ensemble spread in 20CRv3 exceeds 2 K. 20CRv3 additionally provides data on atmospheric circulation and shows a pronounced ridge over Central Europe and stretching to northern Scandinavia (Fig. 5A).

Geostatistical methods that interpolate a variable in space and time are interesting for weather reconstruction as they rely only on the variable itself and not on other meteorological variables. Some approaches make use of time-independent covariates such as altitude, exposition or distance to coast, or land-surface variables such as roughness length for wind speed. However, these approaches need a dense network, which is often not available far back in time. Often the procedure is complex and involves estimates at monthly resolution, or large-scale fields which serve as a first guess for infilling anomalies. Some of these steps may require calibration. Using stochastic modelling (e.g., Gaussian Random Fields [86]), ensembles can be generated to quantify the uncertainty, or parts thereof. For situations with only few stations (but perhaps several variables), other approaches such as the analog approach may be more suitable.

4.3. Analog approaches

The analog method is a non-parametric approach that was introduced in the context of weather forecasting by Lorenz [89] and was popular from the 1970s to mid-1990s. The analog approach is relatively well studied and is also used for downscaling of climate model output [90]. The same approach can be used for local to regional reconstructions of daily weather. You et al. [91] describe how the approach can be applied for weather reconstruction, and it has been used, e.g., for generating daily data for hydrological applications from the 20CR reanalysis [92, 93]. It has also been used to generate daily temperature and pressure fields for Europe for the 1780s [94], which can be compared against Kington’s drawings.

The basis of the approach is simple—select the day that fits best with all available (multi-variable) observations from a given day in the past. However, this can be achieved in various ways depending on the application, the size of the analog pool and other factors. Moreover, the obtained analogs can be further improved using postprocessing techniques as described below.

The arguably most important factor is the size of the pool of analogs. This pool often consists of gridded products such as E-Obs. As these products grow in length, also the analog approaches get more powerful. The selection of analogs can proceed in a hierarchical way or in one step. For instance, often a seasonal window is used to exclude analogs from other seasons. Pfister et al. [95] restrict potential analogs to days with similar weather types (this is one way how categorical variables can be included; in addition, they can also be included in a distance measure). These restrictions are introduced to provide physical consistency, they exclude analogs that may be well suited in terms of a statistical distance, but stem from different weather situations. Time dependence can easily be included, if desired.

The closest analog day provides physically consistent fields, but an extrapolation is not possible. The size of the analog pool therefore remains a problem, particularly for extreme events, as few analogs are typically found for such days. This leads to a systematic underestimation of extremes. If analogs are selected based on anomalies from a mean seasonal cycle or anomalies from a trend, the pool of close analogs for extremes is larger. Adding back the climatology may then lead to an extrapolation, but also to slight physical inconsistencies.

Analog approaches essentially provide a distance matrix. Rather than selecting the closest analog, some approaches average over the best $k$ analogs (corresponding to a $k$ nearest neighbour approach), which can also be weighted. This smooths the fields and may yield better reconstruction skill, but physical consistency is no longer guaranteed. Weighting all analogs
can be framed as a Bayesian approach that converts the prior distribution (pool of analogs) into a posterior [96].

The closest analog can be further augmented. For instance, as extremes are underestimated (especially for precipitation), approaches such as quantile mapping can be used to correct for biases in the distribution [95]. Moreover, data assimilation approaches such as the offline Ensemble Kalman Filter can be used to post-process the analogs [97] (Sect. 4.4). In short, the best analog is considered as a first guess and is corrected towards the observations based on the covariances of the gridded field with the station data. Covariances can be calculated, e.g., from the best 50 analogs to capture “flow dependence” of the covariance structure. The Ensemble Kalman Filter works best for variables that are close to Gaussian such as temperature and pressure [95]. It helps to further augment the reconstruction skill and can lead to extrapolation. Further combinations of methods are possible. For instance, Devers et al. [98] combine yearly and daily time scales in their approach that combines analogs with a Kalman Filter.

As an example, Fig 6 shows analyses of daily weather reconstructions from an analog approach [99] relating to a severe flood event in Switzerland in 1817. Although large amounts of snow had piled up in the high Alps due to reduced melting during the “Year Without a Summer” of 1816, hydrological simulations [100] based on these daily analog fields indicate that, outside the Alpine valleys, the melting of this snowpack in late spring 1817 was not the decisive factor. Rather, the spring of 1817 was cold and snow fell down to the Swiss Plateau (Fig 6A), leading to very high water levels. On top of this, several rainfall events occurred in late June and early July 1817 (Fig 6B). The last of these is analysed in 20CRv2c [101] in Fig 6C. The figure indicates significant amounts of moisture (850 hPa specific humidity) in a frontal system stretching across Europe on 1 July 1817. Heavy precipitation is described in documentary data [102]. A weather diary from the monastery of Einsiedeln, Switzerland (Fig 6D) indicates heavy rainfall and thunderstorms from end of June (for unknown reasons the source has a 31 June) to 5 July, the peak of the flood event. Analyses of such events are important to assess the contribution of snow melt to flood events also in the present.

4.4. Data assimilation and reanalyses

A new era of weather reconstruction started with the „Twentieth Century Reanalysis“, demonstrating that surface-only reanalyses, i.e., systems that assimilate only surface pressure information, are feasible. This opened the door to extend reanalysis products backward by one century from the start of a global radiosonde network to the start of National Weather Services. Version 1 of 20CR covered 1908–1956, which was subsequently replaced by Versions 2 and 2c [5], reaching back to 1871 and 1851 respectively (Version 2c also included the years 1815–1817, [100]). The current Version 3 [6] reaches as far back as 1806. The European products ERA-20C [104] and CERA-20C [105] reach back to 1901. These data sets are widely used for studying past weather events [106], statistics of extremes such as storms [107] and their decadal variations [108], for heatwaves and drought, or for modelling impacts [109].

Data assimilation approaches combine observations with a numerical weather forecast model. Different techniques are in use. 20CRv2c and v3 use a variant of the Ensemble Kalman Filter [5, 6] with a 6-hour assimilation cycle. ERA-20C and CERA-20C use 4D-Var with a 24-hour assimilation window [105]. The latter data set is a coupled ocean-atmosphere reanalysis, the other products are atmospheric reanalyses. All data sets are ensembles; 20CRv3 has 80 members, CERA-20C 10 members. For further details the reader is referred to the corresponding publications.

Historical reanalyses were a game changer in weather reconstruction. They provide global, 3-d fields for any day in the past 150–200 years, along with a measure of uncertainty (ensemble
spread), which, depending on the product, is related to initial conditions, or initial and boundary conditions (e.g., by using different realisations of sea-surface temperature reconstructions). As an example, Fig 7 shows the ensemble mean sea-level pressure as well as the ensemble standard deviation for 14 October 1881. On this day, a storm known as “Great Gale” hit the north of the British Isles, 129 fishermen lost their lives [110, 111]. The storm appears in 20CRv3. The figure also illustrates that reanalyses provide global fields which are not everywhere as well constrained as over the British Isles. Hence, care and expertise are required to analyse these products. Historical reanalyses still have problems in the early years when observations were sparse [112].

Further examples from 20CRv3 are shown in Figs 4, 6 and 7. Reanalyses are physically consistent and thus allow dynamical interpretations of weather events and weather systems. The products allow calculating specific diagnostics which facilitate comparison with present-day events and with climate model simulations. At the same time, it should be kept in mind that these 3-d-products merely provide sophisticated interpretations of surface observations. It is therefore important to compare different products. According to Slivinski [113], an ensemble

Fig 6. Daily analog reconstructions and reanalyses help to assess whether the record flood in Switzerland in early July 1817 were caused by snow melt or heavy precipitation [103]. (a) Potential snow (precipitation on days with mean temperatures below 4°C) from 16 April to 15 May 1817. (b) Precipitation total in the two weeks before the flood event, (c) 850 hPa height and specific humidity in 20CRv2c (ensemble mean) on 1 July 1817. (d) Excerpt of a weather diary from the monastery of Einsiedeln (Klosterarchiv Einsiedeln: KAE.A.39/6 [100]).

https://doi.org/10.1371/journal.pclm.0000034.g006
of ensembles from multiple reanalyses can be used to estimate meta-confidence (see also [114]). However, historical data are societal products whose full uncertainty cannot easily be sampled or expressed probabilistically [115]. Another route is therefore to search for multiple lines of independent evidence, for basing conclusion on separate analyses of independent or partly independent sources of information as indicated in Figs 4, 6 and 7.

Further improvements of historical reanalyses are possible. From the observation side, the quality could be improved by assimilating additional observations, which requires more data rescue. Perhaps, the products could be extended even further back in time, particularly as global data coverage did not change much between the 1770s and 1806, which is the current start year (coverage peaked in the 1780s, see [50]). Improvement could come from assimilating additional variables. ERA-20C and CERA-20C also assimilate marine winds, but how to best assimilate this data source remains an open question. Toride et al. [116] explored using historical weather diaries to categorize cloud cover into three categories, which then can be assimilated into a global model using a Kalman Filter approach. This is interesting as historical documentary data provide categorical data rather than numeric data. Eventually, assimilating temperature or precipitation could lead to improved products. From the assimilation system side, improvements have targeted specific aspects such as localization, outlier screening, and online bias correction. Machine learning techniques could not only be used to directly reconstruct weather (Sect. 4.6) but also to support other steps in data assimilation approaches. Future assimilation systems for historical reanalyses may be coupled and include a better land-surface.

4.5. Dynamical downscaling

Reanalyses such as 20CR provide global weather reconstructions on a relatively coarse grid. For many applications such as climate risk assessment, higher-resolved weather data are
required. This can be achieved with dynamical downscaling, i.e., using the reanalyses as boundary conditions to run regional weather forecast models, typically in a sequence of nesting. Regional models provide a more realistic topography, higher resolved land surface variables, and a better representation of small-scale processes such as in the planetary boundary layer. Dynamical downscaling per se does not add meteorological information, but additional meteorological information can be incorporated in a data assimilation framework.

Dynamical downscaling of 20CR has been performed for long periods [117] as well as for a number of historical weather extremes including blizzards, flash floods or storms [118–121]. Fortunato et al. [109] downscaled a strong storm that hit the Iberian Peninsula in 1941 and simulated water levels and inundated area in the Tagus estuary. These examples show that downscaled historical extremes can contribute to better understand processes or to derive quantitative information, which then can feed into well-established impact model chains [7]. However, it may not always work and careful testing of all settings is required [118].

As an example, Fig 8 shows dynamical downscaling of the reanalysis ERA-20C using the regional model WRF [122] for a snow fall event in midst of the first world war that triggered massive avalanches on 13 December 1916. [123]. The map compares the WRF output with station observations for total precipitation on that preceding (Fig 8A) and for the change in snow depth in the preceding week (Fig 8B). An excellent agreement is found not only when compared with station data, but also with the location of avalanches. This and other studies point to the potential of dynamical downscaling of historical reanalyses for weather and climate research.

Long dynamical downscaling runs could be used for generating regional weather reconstructions back to the early 19th century. For instance, Amato et al. [124] downscaled 20CRv2c for China back to 1851. Additional data might even be assimilated during the downscaling.

4.6. Machine learning

The last method discussed briefly in this review is the use of machine learning algorithms. These techniques have found wide applications in meteorology and climatology, including in
weather forecasting [125]. Machine learning algorithms are also used, e.g., for the detection of circulation regimes [126], for filling-in missing climate data [127], or for climate reconstruction [128]. These applications share important characteristics with the problem of weather reconstructions. At the time of writing, no weather reconstruction has yet been produced using machine learning approaches, but this will undoubtedly change.

A typical application of machine learning approaches are classification problems. Methods such as random forests (which train a multitude of decision trees) or neural networks (in which several layers of nodes connect inputs and outputs, where one node is a weighted function of its inputs) can be trained using meteorological station data as well as existing circulation type classifications. Once trained, the methods can then be applied to past meteorological data. Weather type reconstruction would therefore be a classical application of machine learning approaches.

Machine learning algorithms can also be used to directly reconstruct fields, or they can be a part in the workflow as a component (e.g., quality control) of other techniques, leading to hybrid approaches. The question whether deep learning can eventually replace numerical weather forecasting is currently discussed [125]. Given the fact that basically all techniques in weather reconstruction were developed in the context of weather forecasting, some applications of machine learning in weather forecasting will likely be taken up in weather reconstruction. Deep neural networks can substitute data assimilation, postprocessing, and downscaling.

In particular, machine learning can be an interesting option when dealing with categorical data such as wind direction, rain/no rain/snow, or cloud cover, which can be found in weather diaries back to the 16th century. Information such as that shown in Figs 1 and 2 are often difficult to incorporate in other approaches. Furthermore, many challenges that are addressed in the context of weather forecasting (e.g., how to enforce physical consistency) will also benefit weather reconstruction.

However, the challenges listed in Schultz et al. [125] with respect to using deep learning for weather forecasts—scarcity of extremes in training data, autocorrelation, non-Gaussian variables, periodicities, dynamic correlations, coupling across a range of spatiotemporal scales—also hold for weather reconstructions. Weather reconstructions face even more challenges. First, the problem of missing values and outliers is even more pronounced for weather reconstructions. Machine learning can already be used in the phase of data preparation to fill in missing values or detect outliers, but this makes the interpretation more complex.

An even larger problem is the fact that observations that have the same qualities as their historical counterparts are typically not available for a more recent period that can be used as training data set. Weather diaries are not available from the late 20th century, and instrumental data have different properties. Therefore, to train machine learning approaches, “synthetic weather diaries” and “pseudo early instrumental data” need to be generated from other data sources [129]. However, weather reconstructions based on machine learning techniques will undoubtedly become frequent in years to come and will provide an additional data source.

5. Discussion

Weather reconstructions have been performed ever since Brandes’ first retrospective synoptic weather maps. Today, the effect of climate change on weather leads to a new focus on past weather and therefore new techniques for weather reconstruction. This review briefly discusses a number of them.

All approaches need to deal with sparse observations, often in different form (categorial, numeric—in this review I do not even touch upon descriptive data at all), with many missing values, biases and errors. Fortunately, the approaches differ widely in their underlying
concepts and assumptions, such that a wide variety of methods and products is available. The approaches differ, e.g., in whether or not other variables than the target variable are included. They may or may not incorporate the time sequence, they may be able or unable to extrapolate. Simple approaches that require few input data may be applied much further back in time than more complex approaches. Some of the approaches provide physically consistent fields, others provide statistically consistent fields. Some allow dynamical interpretation, others are targeted at providing input data for climate impact modelling. Hence, there is no best approach, but the approach must be suitable for the purpose.

The plurality of approaches is an advantage. Different methods provide different views of the same weather. It is therefore advisable to combine information from different approaches.

What can we learn from weather reconstructions? First, the past provides a very long sample of weather events. This is relevant for extreme events, which occur rarely by definition, and those that did occur should be studied. Analysing past weather, such as during extreme events, may help to better understand the dynamical causes even if the underlying observations are sparse. Weather reconstructions allow linking weather changes to climate changes. As such, they may serve as a testbed for climate model simulations. For instance, decadal-to-multidecadal variability in climate may or may not be expressed in changing weather depending on the contribution of thermodynamic and dynamic changes. Weather reconstructions may allow to address the underlying contributions.

Finally, weather reconstructions are not only closer to the atmospheric processes than climate reconstructions, but also closer to the impacts on society, economy, and nature. An important application of weather reconstruction therefore is climate impact modelling, e.g., for flood events, windstorms, droughts, or similar. Weather reconstructions may also help to better understand historical events.

Given that global climate enters a new era that differs from anything observed in the past millennium, the question is sometimes posed whether we can still learn from studying past climate. Weather reconstructions demonstrate that this is still the case. However, weather reconstructions require massively more data than climate reconstructions. This calls for data rescue efforts such as those coordinated by ACRE. Data rescue is not only important to extend reconstructions back in time, but also to improve reconstructions in the 19th century. Note that requirements are not the same as for climate reconstructions. Short records, of which there are many, may be fruitfully assimilated in a reanalysis while their added value for long-term climate reconstructions is small.

6. Conclusions

Climate reconstructions have a long tradition in climate science. They have shaped our understanding of climatic changes and will continue to do so. However, to better understand climatic changes, the weather scale needs to be addressed. Weather events are more relevant for climate impacts, and they are more directly related to atmospheric processes. Particularly, in the context of extreme events such as storms or droughts, past weather records have been reanalysed. Therefore, weather reconstruction has gained prominence over the last decade.

Weather reconstructions are nothing new but reach back to the very beginning of atmospheric sciences. However, with new numerical techniques and renewed interest in the weather scale, new approaches have been developed that are reviewed in this paper.

Today we have a plurality of methods and products, ranging from expert analyses, analyses of proxies, construction of simple index series or weather type classifications, to the reconstruction of daily fields using geostatistical approaches, analog approaches, data assimilation, or machine learning. Each method provides a different view of past weather. Therefore, the
methods complement each other as they are based on different assumptions, data sets, and techniques.

Weather reconstructions are used in manifold ways, demonstrating that society can still learn from the past climatic record even in a situation in which climate change leads us to an unprecedented regime.

Acknowledgments

Simulations underlying EKF400v2 were conducted at the Swiss Supercomputer Centre CSCS under a CPU grant to SB. Support for the Twentieth Century Reanalysis Project version 3 data-set is provided by the U.S. Department of Energy, Office of Science Biological and Environmental Research (BER), by the National Oceanic and Atmospheric Administration Climate Program Office, and by the NOAA Physical Sciences Laboratory. We acknowledge the collaboration within PAGES group CRIAS as well as the tremendous effort of the ACRE (Atmospheric Circulation Reconstructions over the Earth) community led by Rob Allan.

References


46. Plot R. A letter from Dr. Robert Plot of Oxford, to Dr. Martin Lister F. of the R. S. concerning the use which may be made of the following history of the weather, made by him at Oxford throughout the year 1684. Phil Trans Roy Soc. 1684; 15(169). https://doi.org/10.1098/rstl.1685.0022


82. Lamb HH. Climate, history and the modern World. 2nd ed. Routledge; 1996.


