Computers & Graphics 104 (2022) 1–10

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Special Section on EnvirVis

A visualization technique to assist in the comparison of large meteorological datasets

Dana K. Urribarri, Martín L. Larrea *

Department of Computer Science and Engineering, Universidad Nacional del Sur (UNS), Bahía Blanca, CP8000, Argentina
Computer Graphics and Visualization R&D Laboratory, Universidad Nacional del Sur (UNS) - CIC Prov. Buenos Aires, Bahía Blanca, CP8000, Argentina
Institute for Computer Science and Engineering, Universidad Nacional del Sur (UNS) - CONICET, Bahía Blanca, CP8000, Argentina

ARTICLE INFO

Article history:
Received 28 July 2021
Received in revised form 22 February 2022
Accepted 28 February 2022
Available online 4 March 2022

Keywords:
Visualization
Comparison task
Comparative visualization
Meteorological data
Time series
Time misalignment

ABSTRACT

The comparison of meteorological data is a fundamental task within the techniques of meteorological forecasting due to the cyclical nature of the climate. However, interpretation of meteorological datasets can be difficult due to their large size. The goal of information visualization is to support a better understanding of the data what includes assisting the user in the data comparison process. However, few visualization techniques have been specifically developed to support the comparison process of big meteorological data. In this paper, we improved a visualization technique for large time-series comparison. The new technique is more suitable for the comparison of meteorological data.

1. Introduction

The comparison of meteorological data is a fundamental task within meteorology analysis. Climate models comparison [1,2], city weather comparison to compare travel destinations [3], or comparison between different meteorological data sources [4,5] are just a few examples of meteorological data comparison. Since meteorological measurements over time generate a large volume of data, processing and managing these datasets is a challenging task compared to traditional methodologies and platforms.

The goal of information visualization is to facilitate a better understanding of the data. Therefore, it can assist the user in this comparison process. However, few visualization techniques have been developed specifically to support the comparison process of meteorological data. In this context, new tools and techniques are required.

Our goal is to assist in the analysis of meteorological data by facilitating the comparison process. Therefore, we propose improvements to an existing visualization for large meteorological datasets [6]. That visualization is based on a visualization designed to compare karate movements [7,8]. The proposed technique is an overview+detail visual analysis tool, where the overview visualization shows a quantitative summary of the misalignment between pairs of sequences. The detailed visualization is based on Dynamic Time Warping (DTW) [9,10] and mainly shows how a sequence should be transformed to match another one. The DTW technique is used in time series analysis to measure the dissimilarity between two time sequences. To calculate this measure, DTW searches for an optimal mapping between the series. This mapping induces a warping function that allows the transformation of one of the sequences to obtain two time-aligned sequences. In this work, we updated the alignment calculations in the technique. We also addressed some detected issues in the visualization. This work was conceived for a specific task, the comparison, of a specific dataset, the meteorological data. However, we believe that the results obtained not only show a contribution to the meteorological data visualization area but also to the visualization area in general.

This paper continues with a discussion on related works. The following section, Section 3, presents the meteorological dataset used in this work. Section 4 presents the proposed improvements to the original work, and Section 5 a description of the updated overview+detail visualization. Section 6 presents a case study and discusses the achieved visualization. Finally, Section 8 draws some conclusions and presents intended future work.

2. Related work

Today in meteorology, remote sensing and location data observations are visualized to improve weather forecasting and

* Corresponding author at: Department of Computer Science and Engineering, Universidad Nacional del Sur (UNS), Bahía Blanca, CP8000, Argentina.
E-mail addresses: dku@cs.uns.edu.ar (D.K. Urribarri), mll@cs.uns.edu.ar (M.L. Larrea).

https://doi.org/10.1016/j.cag.2022.02.011
0097-8493/

atmospheric research. Rautenhaus et al. [11] present a comprehensive review of the state-of-the-art of weather data visualizations, especially for data comparison. Comparison is one of the essential tasks that motivate the exploration of datasets. This task becomes arduous when it involves large datasets. Few studies have been conducted with comparison-oriented visualizations of large datasets using meteorological data as a test case.

Within the bibliography analysis, we can observe that certain general-purpose visualization techniques are used for the specific task of comparing climate data. In addition, there are other techniques designed and developed specifically for this end. Since this work enhances a connected heatmap visualization of climate data, we focus this related work on three types of visualization: general-purpose and ad-hoc visualizations used for climate data comparison and visualization for time comparison on time series.

2.1. General-purpose visualization

One of the most widely used general-purpose techniques is side-by-side visualization [1,3,12–15]. This technique is suitable for comparison but does not scale well on large datasets. Another problem with this data-comparison strategy is that the comparison process falls on the user as a cognitive load. This means that the user must observe two visual representations and calculate for himself the differences or similarities between them. A variant of this type of technique is referred to as small multiples [16]. Poco et al. [17] used this later visualization, and they recognize it is not suitable for handling large datasets. They proposed a visual analytic tool based on linked views which allows the user to dynamically create, modify and observe the interaction among the grouping of data. Their solution targets model comparison in climate science. Mason et al. [5] used the same technique for comparing meteorological data in New Zealand.

Nocke et al. [18] address the challenges of comparing climate-related datasets using 3D visualization, which they claimed is a suitable approach. Unfortunately, there are known problems with the visual representation of information in 3D, for example, information occlusion and distortion and the difficulties in perceiving depth in two-dimensional devices, like a typical computer monitor or a sheet of paper. The interaction mechanisms can also be challenging [19]. Squillacote et al. [12] also rely heavily on 3D visual representations.

The overlay of graphic elements is another approach for visually comparing the information with similar problems to 3D visualization, like occlusion and distortion. Tang et al. [20], Sauber et al. [21] and Quinan and Meyer [22] all used an overlay of visual information. Tang et al. [20] encode weather data by using a multilayer texture system. The proposal is suitable for large datasets but does not involve time information. Also, their results are limited to only the three textures used in the article. Sauber et al. [21] presented a visualization of the correlations between different scalar fields in an atmospheric multifield dataset. The work does not consider the time variable within the datasets, and there is no mention of the support of large datasets. Quinan and Meyer [22] proposed WeaVER, an interactive tool to visualize how different weather features relate across an ensemble of possible forecast outcomes. The system used current meteorological conventions and visualization principles. The authors do not mention the volume of data that the system handles, so it cannot be determined if the proposal supports large datasets.

Spaghetti plots are another commonly used general-purpose technique [2,4,23–25]. This technique does not scale well when the data has a large number of dimensions. The accumulation of lines on the viewing area generates occlusion problems and interference in the information.

2.2. Ad-hoc visualization

Among the specific techniques for comparing meteorological data, the Hovmöller diagram [26–28] is a common technique for plotting and comparing meteorological data. The diagram depicts the temporal–spatial variation of a chosen meteorological parameter, such as temperature, density, and other values. One axis of the diagram represents time while the other represents a spatial variable; different color scale encodes the meteorological parameter. Although this technique is commonly used in the area and is known for its ability to allow comparison through side-by-side visualization [29,30] of different values over time, its use and interpretation require some expertise.

2.3. Time comparison visualization

Larrea and Urribarri [6] proposed a visualization based on DTW and connected heatmaps to analyze the misalignment between two years of weather measurements. The support of large datasets rests mainly on offline preprocessing. Even though the original proposal [7,8] was intended for karate movements comparison, the authors demonstrated that it is also feasible for comparing meteorological data. However, DTW aims to compare time series that differ mainly in speed and, although weather data follow an annual cycle, the patterns of the series differ in more than just speed which generates a noisy alignment. Moreover, the records are composed of different kinds of measurements but are compared using Euclidean distance, the misalignment is classified as early, late, or on time, regardless of its magnitude (then, this classification maps to a discrete color palette), and heatmaps are connected using a function-like strategy that generates gaps at one end of the connectors.

That work [6] improves upon existing techniques by allowing the comparison of large datasets, facilitating the comparison of time misalignment with a DTW visualization approach, and improving the comparison of multidimensional records. The technique enables the comparison of time information, which is not possible with the visualization techniques developed by [20,21]. Unlike [5,17,22], it was conceived to support large datasets. Not to rely only on a side-by-side comparison, as in [1,3,12–15], the technique includes connections between related records and color-coded information about the relationship. The visualization is completely represented in 2D without overlaying information, which facilitates the user's interactions, unlike [18,20–22]. Although there are studies that support this last statement [19], usability tests are still pending.

Several works use side-by-side heatmaps to compare time-series data [31–37]; however, the intention of those visualizations is not to analyze time misalignment. Hao et al. [31,32] proposed a framework for intelligent time- and data-dependent visual aggregation of data along with multiple resolution levels. The result is a multi-resolution visualization that resembles connected heatmaps. The connections between those heatmaps provide the time-alignment information lost by displaying time-series at different resolutions.

The visualization used in this article uses connected timelines. This is a visual tool that facilitates the perception relationships between visual elements. Other works have used this same visual tool like [38–41], but their application domain, objective, and visualization technique are different. Knipp et al. [42] use connected timelines to visualize the effects on the technology of solar and geospace events and of solar and geomagnetic activities. Their visualizations are infographics, with high textual content and without interactions.

Table 1
Meteorological measures gathered throughout each year.

Name	Description	Units
Tmax	Maximum temperature	°C
Tmed	Average temperature	°C
Tmin	Minimum temperature	°C
Td	Dewpoint temperature	°C
STP	Pressure at station	hPa
Vis	Visibility	km
Vmed	Average wind velocity	km/h
Vmax	Maximum wind velocity	km/h
Prec	Precipitation	mm
Rad	Incident solar radiation	W/m^2

2.4. Contribution

Our proposal is an improvement on the work of [6]. We present improvements for the data preprocessing and the visualization. For the former, we improved the usage of DTW to compare roughly periodic time-series and the distance used to compare the meteorological records. For the latter, we measured the magnitude of the misalignment, which leads to a continuous color palette, and we defined a mapping between the heatmaps to avoid gaps between connectors' ends.

3. Meteorological data

We obtained a series of meteorological data for the outskirts of Bahía Blanca from Meteobahía [43] for the range of years 2010 to 2019. The dataset consists of 3856 daily records. Each record contains the date (day, month, and year) and a weather summary of the ten meteorological measurements described in Table 1.

Bahía Blanca City is located in the southwest of the province of Buenos Aires (38°44'S; 62°16'W), at 20 m above sea level, near to the Atlantic coast [44]. Climate is temperate/mesothermal with constant precipitation throughout the year and hot summers. The annual mean temperatures are between 14° and 20 °C, being markedly different between seasons. Annual rainfall is between 500 and 600 mm, with a high degree of variability. The Atlantic Ocean influences this climate which has a particularly modulating effect on air temperature. However, the area is on the edge of an estuary with continental features in the region. There are well-differentiated summers and winters and moderate springs and autumns. During the hot season, temperatures range around 40 °C. In winter, marked cold waves are usual with minimum temperatures near −12°C. The movement of air masses of different origin and characteristics generates high variability of atmospheric conditions and determines a transition climate for the area, i.e., between the cold and dry influence from Patagonia and the hot and humid effect from the east of Buenos Aires province [45].

4. Proposal

This new work continues the one done by Larrea and Urribarri [6]. It keeps the main idea of the overview+detail visualization but proposes improvements for the alignment calculation and the visualization:

- The usage of a more appropriate distance function to compare records with measurements with incompatible units;
- The correction of the noise generated by aligning time spans from a roughly periodic phenomenon;
- To measure the degree of misalignment to obtain a more consistent color code between the overview and all the detailed views;
- To represent the visualization of the parallel time relationship more as a mapping than as a function.

This section starts presenting the DTW technique, particularly the warping and the misalignment functions. Then, it continues with three alignment-calculation changes to the original work: the distance function, the data smoothing, and the calculation of the degree of misalignment.

4.1. Dynamic time warping

To calculate the dissimilarity measurement between two time series, DTW looks for an optimal mapping between them. This mapping induces a warping function that allows the transformation of one of the sequences to be time-aligned with the other.

As described in Algorithm 1, given two sequences R and S of length L and N respectively, DTW computes an accumulated-distance matrix $M \in \mathbb{R}^L \times \mathbb{R}^N$ between them. $M_{1,1}$ is the distance between the two first records of the sequences, while $M_{L,N}$ is the accumulated distance for the two last records and the dissimilarity measure between the sequences. The accumulated distance for every pair of records is the distance between those two records plus the minimum between the accumulated distances for the three immediately preceding pairs of records.

Algorithm 1 Accumulated distance matrix M between sequences R and S

Input Two sequences R and S of length L and N respectively.
Output The accumulated distance matrix M between the two sequences.

// The distance between the initial record
$M_{1,1} \leftarrow \text{dist}(R_1, S_1)$
// Accumulated distance between S_1 and R's records.
for all $i \in [2, L]$ **do**
  $M_{i,1} \leftarrow \text{dist}(R_i, S_1) + M_{i-1,1}$
// Accumulated distance between R_1 and S' records.
for all $j \in [2, N]$ **do**
  $M_{1,j} \leftarrow \text{dist}(R_1, S_j) + M_{1,j-1}$
// Minimum accumulated distance between every other pair of records
for all pair $i, j \in [2, L] \times [2, N]$ **do**
  $M_{i,j} \leftarrow \text{dist}(R_i, S_j) + \min\{M_{i-1,j-1}, M_{i-1,j}, M_{i,j-1}\}$

4.1.1. The warping function

Calculating measure $M_{L,N}$ implies finding the minimal accumulated-distance path from (1, 1) to (L, N) (see Algorithm 2). This path P is a list of pairs (i, j) that defines a warping between sequences the records of R and S. Functions that warp one of the sequences into the other can be inferred from the warping path P. The warping function $F : [1, L] \rightarrow [1, N]$ that warps sequence R into S is defined as:

$$F(n) = \lfloor \text{mean}_{(n,j)\in P}\{j\} \rfloor \quad (1)$$

4.1.2. The misalignment function

Given a record number n in the sequence R, the function $F(n)$ returns the number of the best matching record from S. The misalignment function $G(n) = F(n) - n$ is the distance from $F(n)$ to the perfect alignment. The slope of $G(n)$ gives a hint about whether S is delayed, on-time, or early with respect to R. Depending on whether S is delayed, early, or on time, $G(n)$ has a negative, positive, or flat slope, respectively.

Algorithm 2 Warping path P between sequences R and S

Input Accumulated-distance matrix M of $L \times N$ elements.
Output Warping path P between the two sequences.
1: P starts empty.
2: $i \leftarrow L, j \leftarrow N$
3: **while** $i \geq 1$ and $j \geq 1$ **do**
4: Add (i, j) to the path P
5: Let m be $\min\{M_{i-1,j-1}, M_{i-1,j}, M_{i,j-1}\}$
6:
$$(i,j) \leftarrow \begin{cases} (i, j-1) & \text{if } i = 1, \\ (i-1, j) & \text{if } j = 1, \\ (i, j-1) & \text{if } M_{i,j-1} = m, \\ (i-1, j) & \text{if } M_{i-1,j} = m, \\ (i-1, j-1) & \text{otherwise.} \end{cases}$$

4.2. The distance function

Algorithm 1 can use any distance function, and previous work [6] used the Euclidean distance to calculate the difference between two records. However, since each record is composed of measurements of several meteorological variables (for example, precipitation, temperature, and pressure) and not all have the same units, only the magnitudes can be considered when operating. This forces a unitless distance and implies that the variable with the highest magnitude tends to dominate the Euclidean distance. A more appropriate distance is the Karl-Pearson distance [46], where each value is weighted by the reciprocal of the corresponding standard deviation. The distance between records a and b with V variables (Eq. (2)) is the Euclidean distance between the standardized records. Note that the result of the Karl-Pearson distance is a unitless value.

$$\text{dist}(a, b) = \sqrt{\sum_{v=1}^{V} \left(\frac{a_v - b_v}{\sigma_v} \right)^2} \tag{2}$$

Fig. 1 compares the difference between the aligned sequences using the Euclidean distance (Fig. 1(a)) and the Karl-Pearson distance (Fig. 1(b)). The Euclidean distance results in larger values and more dispersion because the magnitudes of pressure and solar radiation are higher than the ones of temperature, wind, and precipitation.

4.3. Data smoothing

Even though the weather follows similar patterns over the years, it is not a periodic phenomenon. Therefore, it is more relevant to focus on trends than on the exact match of the daily changes between years. For this reason, the misalignment function G is smoothed along a week with a 0-mean Gaussian kernel of radius 3.

$$\tilde{G}(n) = K_3 * G(n) = K_3 * (F(n) - n) = \tilde{F}(n) - n \tag{3}$$

Given that the convolution between a 0-mean Gaussian function and the identity function results in the identity function, calculating the function $G(n)$ from the smoothed $F(n)$ is equivalent to smoothing $G(n)$.

Fig. 2 compares the G and $\tilde{G}$ functions. Since smoothing reduces the details of the alignment function, this is no longer optimal, and the difference between the aligned sequences may be larger (as can be seen in Fig. 4); however, it gives a better understanding of the late, on-time, or delayed alignment trend between the two sequences.

4.4. Degree of misalignment

The strategy described in Section 4.1.2, looking at the sign of the slope, is useful to classify the misalignment of a record into one of the three categories: delayed, on-time, or early. However, it does not contemplate degrees of misalignment. Moreover, the smoothed function $\tilde{G}$ presents very few records with exactly 0-slope. In this scenario, we defined a function $H(n)$ to quantify the misalignment (see Eq. (4)). If $\tilde{G}$ is the misalignment function and $d\tilde{G}$ is the discrete derivative of that function, $H(n)$ is a function between -1 and 1 resulting from normalizing $d\tilde{G}$ by its maximum or minimum, depending on its sign. Positive (negative) values of $H(n)$ indicate how early (delayed) the record is.

$$H(n) = \begin{cases} \frac{d\tilde{G}(n)}{\max(d\tilde{G})} & \text{if } d\tilde{G}(n) > 0, \\ 0 & \text{if } d\tilde{G}(n) = 0, \\ \frac{d\tilde{G}(n)}{|\min(d\tilde{G})|} & \text{if } d\tilde{G}(n) < 0. \end{cases} \tag{4}$$

The value $|H(n)|$ indicates how early or delayed a record is, depending on $H(n)$ being positive or negative, respectively. Finally, $1 - |H(n)|$ indicates how on-time the record is.

Fig. 4 shows the difference between the sequences for the alignments of Fig. 2. The color under the curve encodes the misalignment between the records. The upper graph visualizes the slope of function G described in Section 4.1.2 with the discrete color palette of Fig. 3(a); however, the bottom graph visualizes the function H described by Eq. (4) with the color scale of Fig. 3(b).

5. The visualization

The visualization itself is a visual comparative analysis of the misalignment between a set of time series. There is an overview of the misalignment between the data corresponding to n different datasets and an on-demand detailed view focusing on the comparison of two of them. This section shows the result of the improvements detailed in the previous section on the visualization of the whole dataset.

5.1. The input data

Since the DTW is time and memory-consuming, it was performed offline in a previous preprocessing step, and the result of the computation is part of the data to visualize. The input data includes a file per year with a daily meteorological record with the Table 1 measurements and a 3D matrix M with the preprocessed warping functions. If A and B are two data series, and i is the number of a record from A then, $M(A, B, i)$ corresponds to the number of the record from B equivalent to the record i from A. In this case, the DTW was performed computing the Karl-Pearson distance between records, and, for each record, it considered all available meteorological variables.

5.2. Overview: a summary-box matrix

Given a dataset of n-years meteorological measurements, the overview contains all the pairwise summary boxes of the two-series comparison.

The summary box shows an overview of how delayed, on time, or early is a sequence with respect to another by showing three values. Unlike previous work [6–8], the summary box does not show percentages indicating *how many* records there are for each type of misalignment but instead shows values representing *how much* misaligned the records are. Given a sequence Q of degrees of misalignment, Eq. (5) calculates those three values, where v_1, v_2, and v_3 represent how much delayed, on-time, and early the

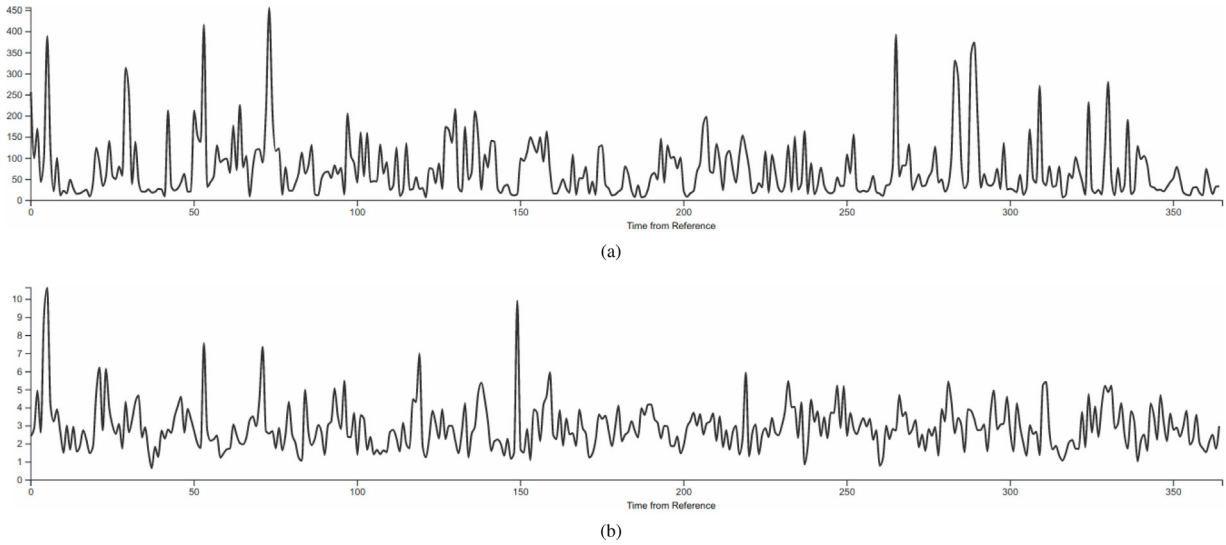

Fig. 1. The distance between sequences using two different distance functions. Figure (a) shows the difference between the sequences using Euclidean distance. The distance values are scattered over a wide range. Figure (b) shows the difference between the same sequences using Karl-Pearson distance. In this case, the distance shows much less dispersion. In both images, the *x*-axis represents the timeline as "record number", and the *y*-axis represents a unitless distance between records.

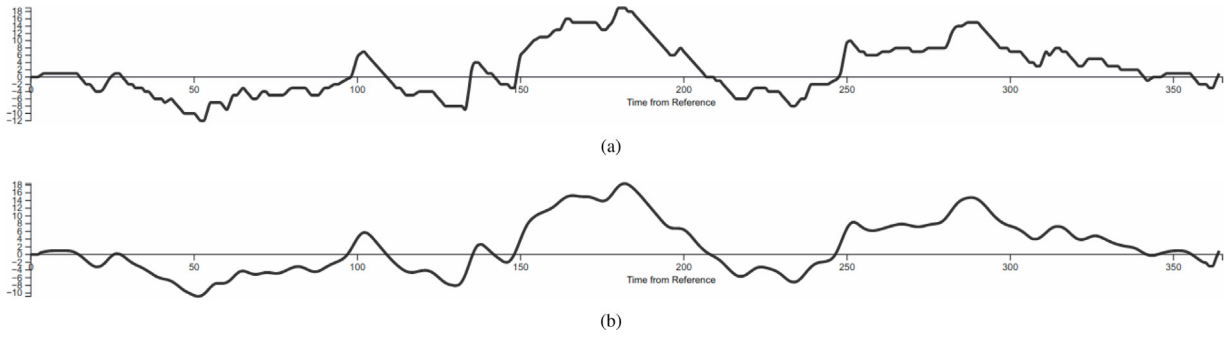

Fig. 2. Smoothed misalignment function. In both images, the *x*-axis represents the timeline as "record number", and the *y*-axis represents the time shift as "number of records". Figure (a) shows the raw misalignment function corresponding to Fig. 1(b). Figure (b) shows the same misalignment function but smoothed over a week. Note that the function allows focusing on trends rather than daily changes.

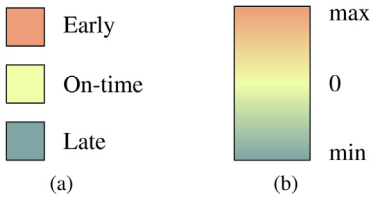

Fig. 3. Color scales. Figure (a) shows a discrete palette to encode whether a record is delayed, on time, or early. Figure (b) shows a continuous palette encoding the amount of misalignment. The *min* value corresponds to how far is the most delayed record of being on time, while the *max* value corresponds to how far is the earliest record of being on time.

records are, respectively. This equation takes into account how far is a misaligned record for being on time.

$$\{v_1, v_2, v_3\} = \frac{\sum_{q\in Q}\{|\min(q,0)|,\ 1-|q|,\ \max(q,0)\}}{\|\sum_{q\in Q}\{|\min(q,0)|,\ 1-|q|,\ \max(q,0)\}\|_{L_1}} \quad (5)$$

Fig. 5 shows a comparison between two summary boxes. One is the summary of a raw misalignment and summaries just the number of records with a negative, flat, or positive slope in function *G*. The other summaries the smoothed misalignment and uses the values from Eq. (5).

5.3. Detailed view

The detailed view consists of three graphs. The first one (Fig. 4(b)) shows the distance between the records of the reference sequence and the correspondents in the target (represented by the height of the curve), plus the time alignment between the two sequences (represented by the area's color under the curve). The second graph is a representation of the misalignment function $\tilde{G}$ (Fig. 2(b)). The third graph is a *parallel heat-maps visualization* [6]. This visualization (see Fig. 6) combines two *heat maps* and a *parallel time-relationship visualization* to visualize the whole sequences and the time warping.

A *parallel time-relationship visualization* [7,8] is an explicit representation of the time-warping that the records undergo when aligning the two sequences. Two parallel axes represent the timelines of the sequences, and the lines connecting those axes represent the deformation. The temporal relationship between the sequences is emphasized with the color of the line, where red means early, yellow on-time, and blue, early. Besides, the red–yellow and blue–yellow gradients encode the amount of misalignment.

If $W(t) : T_1 \rightarrow T_2$ is a warping function between two sequences of timelines T_1 and T_2, then the Parallel time-relationship visualization consist of a set of lines connecting points t from T_1 with $W(t)$ from T_2. When $W(t+1) > W(t)+1$ then, to avoid

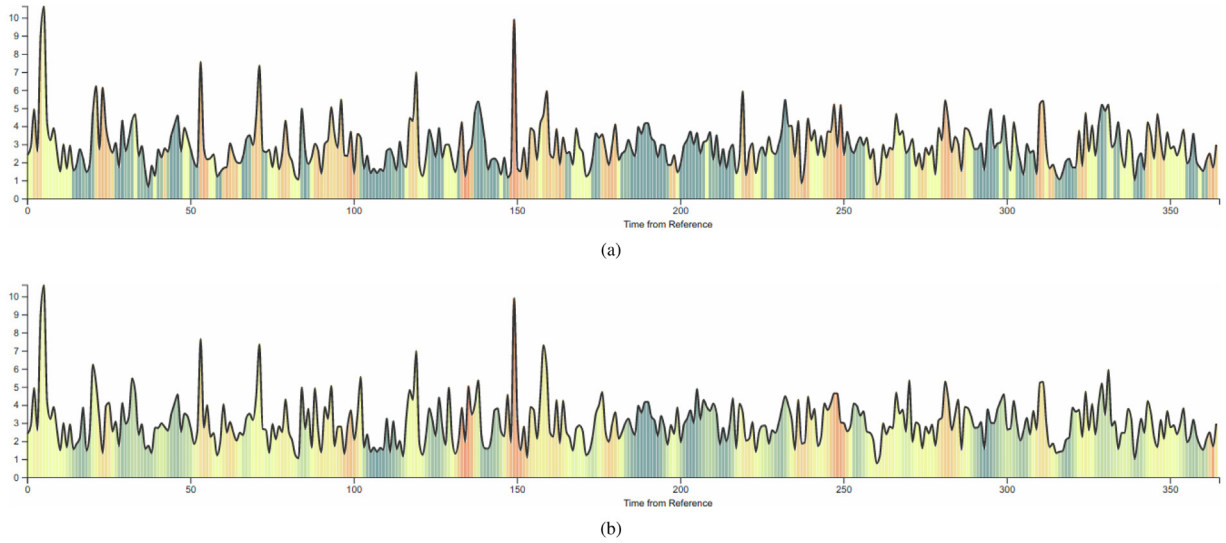

Fig. 4. Distance and degree of misalignment between sequences using Karl-Pearson distance with and without smoothing. Figure (a) corresponds to the alignment using the Karl-Pearson distance and no smoothing. The color encodes whether the slope of the function G is negative, flat, or positive. Figure (b) corresponds to the alignment using the Karl-Pearson distance and smoothing. The color encodes how early, on-time, or delayed the record is.

Fig. 5. Figure (a) corresponds to the summary box of the same comparison as Fig. 4(a). This box encodes just the number of records for each type of misalignment. Figure (b) corresponds to the summary box of the same comparison as Fig. 4(b) encodes not a number of records but an indicator of how much misalignment there is. Both graphs use the discrete color palette from Fig. 3(a).

holes in the visualization, the set includes the lines connecting t with the points of T_2 between $W(t)$ and $W(t+1)$.

In a *parallel heat-maps visualization*, the heat maps serve as temporal axes, which are connected to represent the warping function, i.e. the misalignment, between the sequences. *Heat maps* are a pixel-oriented visualization where the position of the pixel and its color encode the desired information. In this visualization, the x-axis encodes time, the y-axis encodes the meteorological measurement, and the color of the pixel encodes the value of the measurement at a particular time. As shown in Fig. 7, the original visualization uses different color scales depending on the units of the measurement. Even though there are no standard colors, these color scales intend to be meaningful and maintain meteorologists' familiarity with them. This choice expects to reduce the visualization learning curve.

6. Use case

Agricultural activity is one of the main drivers of the economy of the Bahía Blanca region, and the analysis of meteorological data is essential for making decisions in this area. One of the biggest problems for farmers is frost since it endangers all the effort invested in crop production. The frosts that cause the greatest damage to crops are not those that occur during the normal period of occurrence of this phenomenon, but those that appear outside that term. Autumn and spring frosts are the most damaging to crops since they occurred in the birth, flowering, or fruiting stages.

Depending on the point of view, there are two definitions of frost [47]. Meteorological frost, or just frost, is any temperature less than or equal to 0 °C measured in a meteorological shelter. Agrometeorological frost is any temperature less than or equal to 3 °C measured in the meteorological shelter, equivalent to 0 °C or less outdoors. These different definitions of frost are necessary since not all crops are damaged at 0 °C; although some are more resistant to low temperatures, others are less resistant and are damaged at higher temperatures.

A tool to compare frost patterns over the years is a first step in providing farmers with a way to predict future frost behavior. It would be useful for them to know that one year will be similar to another, especially in terms of early or late frosts and intensity. This comparison tool can help forecast future frost by comparing minimum temperatures from previous years. The frosts of one year could be estimated assuming that they will behave similarly to the year following the one most similar to the previous one.

In this context, the DTW is calculated using only the minimum temperature. For example, frost from 2018 can be predicted looking for the year most similar to 2017. Analyzing the minimum temperature and calculating the DTW with only that meteorological variable, from the overview (see Fig. 8), the year more similar to 2017 (target) is the year 2011 (reference): the visualization shows a high percentage, 70%, of on-time minimum temperatures, that is, high synchronicity between both years. Then, it is expected that in 2018 the minimum temperature will behave similarly to 2012. To analyze if there is any trend that would have justified this prediction, we compare 2017 with 2012 and 2011 and 2012 with 2011. Finally, we compare the prediction with the actual temperature of the year 2018.

- *2017 vs. 2011:* By comparing the years 2017 and 2011 (see Fig. 9), the visualization tool highlights that meteorological frosts start in mid-April in both years; however, in 2017, the last meteorological frost occurs a month later than in 2011.
- *2012 vs. 2011:* From the comparison of 2012 and 2011 (see Fig. 10), in 2012, the first meteorological frost occurred a week earlier than the previous year, and the last one occurred 10-days later. However, those temperatures are not well aligned with respect to the whole year minimum temperature since the years did not behave similarly.
- *2017 vs. 2012:* The comparison between 2017 and 2012 follows a similar pattern (see Fig. 11). In 2017, the first meteorological frost occurred just 4-days later, while the last one occurred more than 2-weeks later.

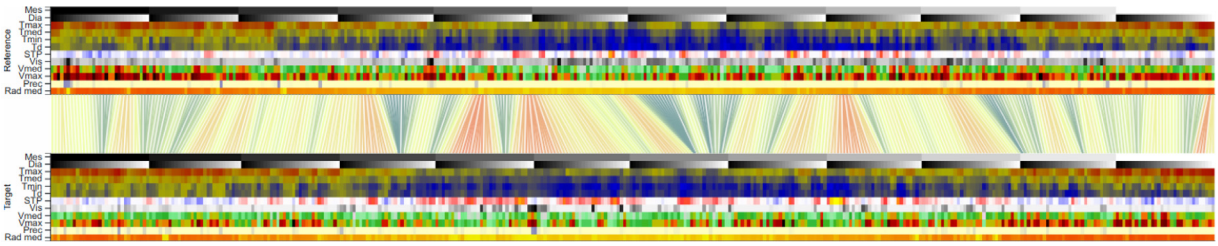

Fig. 6. A *parallel heat-maps visualization* combines two parallel *heat maps* that serve as the timelines of the sequences and a *parallel time-relationship visualization* to visualize the whole sequences and the time warping.

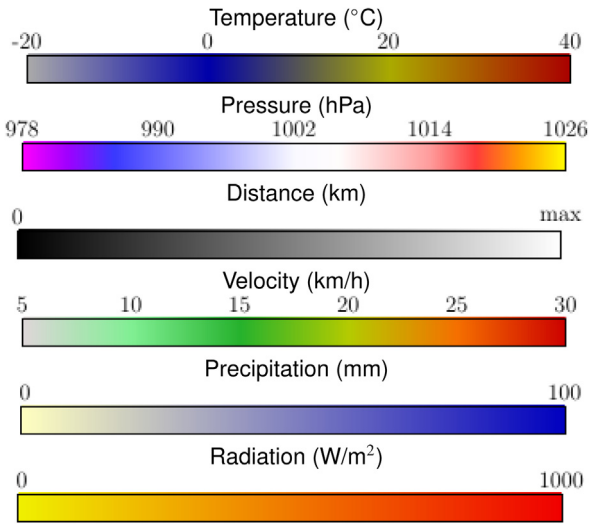

Fig. 7. Color scale for each type of measure.

- *Prediction:* With this information, one can expect that, in 2018, the first meteorological frost occurs nearby the date of the first one in 2017; however, the last one can be expected to occur much later.
- *Result:* Knowing the minimum temperatures of 2018 and comparing them with the ones from 2017, we can conclude that the prediction would have been correct (see Fig. 12). The first frost occurred just 2-days later, but the last one occurred 20-days later.

The color-coding of the minimum temperature (see Fig. 12) differentiate values smaller or equal to 0° with dark blue, temperatures between 0° and 5° with light blue, temperatures between 5° and 15° with yellow, and over 15° degrees with orange. The scale distinguishes temperatures between 0° and 5° since not all crops are vulnerable to the same values of low temperatures (see Fig. 13).

7. Implementation notes

The prototype tools presented in this work were implemented using Javascript and the D3.js library. Section 5 prototype is available at https://cs.uns.edu.ar/~dku/vis/vis-temp-datos-meteorologicos/v02/index_todo.html, and Section 6 prototype is available at https://cs.uns.edu.ar/~dku/vis/vis-temp-datos-meteorologicos/v02/index_tmin.html.

8. Conclusions and future work

Within the analysis of meteorological data, and in particular, in the forecast of the climate, the comparison of information is an essential task. One that is an especially arduous task when it comes to large datasets. The objective of information visualization is to help gain insight into the data, and in this context, it should assist the user in the task of comparison. However, few visualization techniques have been specifically developed to support the comparison process, even less for large datasets.

In this article, we present enhancements to a previously proposed visualization technique for comparing time series of meteorological data. That technique was designed to compare karate movements, and its generality was shown by applying it to meteorological data. However, the nature of the meteorological data makes it not entirely correct to apply the technique with the original methodology. We addressed these flaws by proposing a change in the distance function used to compare records and smoothing the misalignment series to consider that although the weather is a cyclic phenomenon, it does not define a perfect periodic function. In addition, we proposed improvements to the visual data representation. The potential of this technique was tested through the case study in which a comparison of the historical meteorological data of the city of Bahía Blanca in Argentina was visualized.

This proposal intends to be one more input to the work of the professional in charge of predicting meteorological behavior. We consider this an input of great value because it benefits both from the inherent advantages of information visualization and from the advantages of being a technique specifically designed for the task of meteorological comparison.

The experience of working in the meteorology area allowed us to survey the needs that exist in terms of visualization tools. For future works, we identify at least the following tasks:

- To transform the current prototype into a complete application available to experts in the field. That means providing a platform to upload their datasets and explore them freely using our technique.
- To improve the color scales used for different meteorological magnitudes. We will intend to keep them meaningful and familiar to meteorologists but reduce multi-hued scales, surprisingly popular in meteorology visualizations [48], to a minimum.
- To improve the temporal $x-$axes with more meaningful labels.
- To add more interactions to the technique. One interesting interaction is to compare periods besides calendar years. This interaction needs the improvement of the calculation of the DTW to be computed online.

These improvements will help with the usability of the technique and its ability to compare data. Although preliminary tests carried out with professionals in the area did not show perception problems, controlled experiments to investigate the usability of the technique are still pending. As Fuch et al. showed [49] and Yan Liu stated [50], these types of controlled experiments are a challenge in terms of their design and users involved.

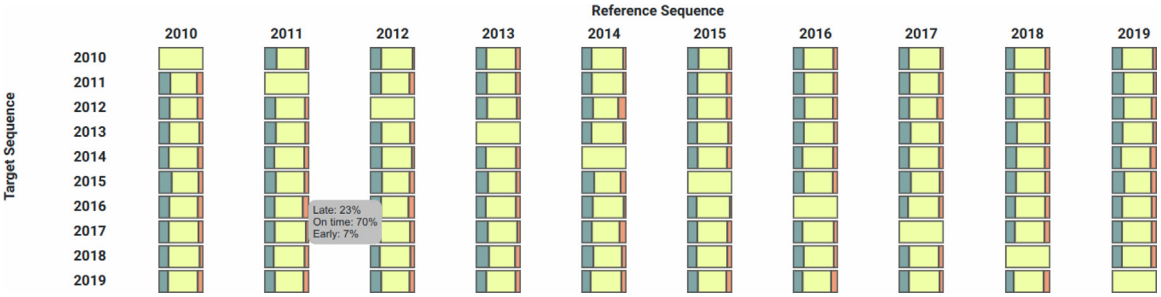

Fig. 8. The overview shows the summary boxes for the pairwise comparison of minimum temperatures. In general, as expected, all comparisons show high synchronicity between them. In particular, the comparison between the years 2017 and 2011 shows high synchronicity between them.

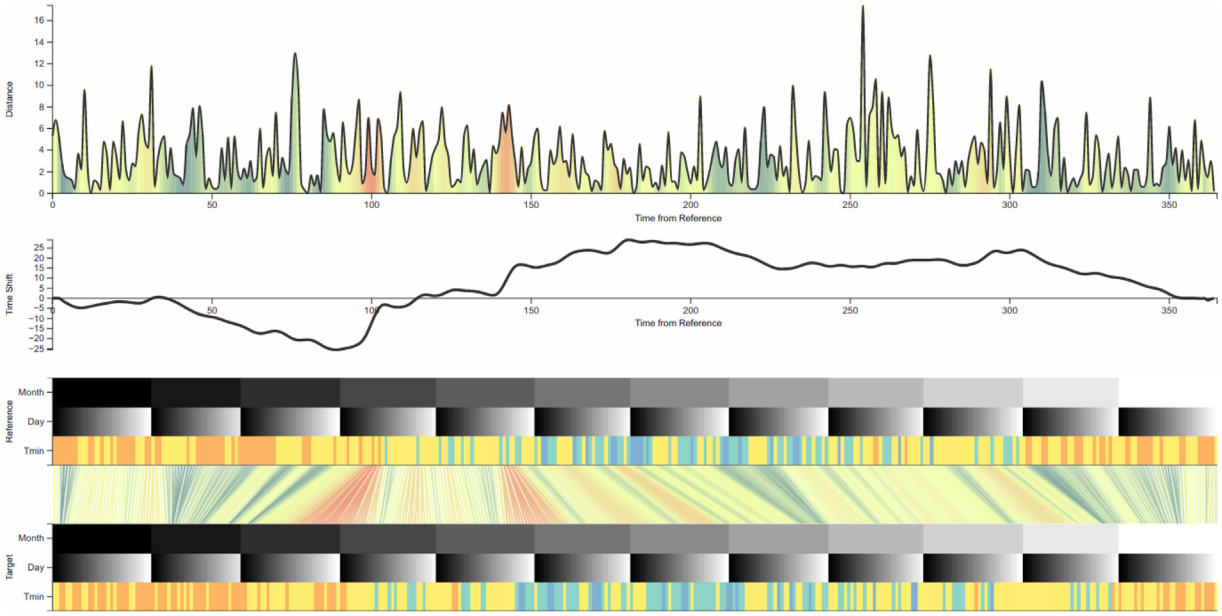

Fig. 9. Detailed view of the 2017 vs. 2011 comparison. The upper graph shows the distance that measures the error of the aligned sequences. The middle graph reveals the trend of the misalignment, for example, how before the first meteorological frost, 2011 tends to delay with respect to 2017. The bottom graph compares both heatmaps and explicitly depicts the warping function. In this last graph, the heatmaps exhibit that meteorological frost started around the same time but in 2017 ended later than in 2011.

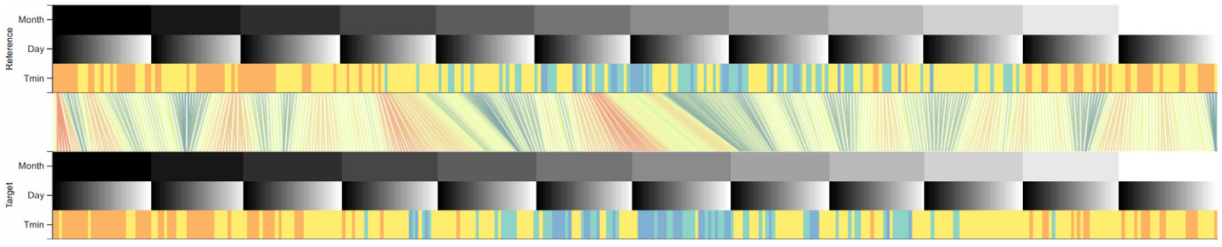

Fig. 10. Heatmaps from the 2012 vs. 2011 comparison. In 2012, the first meteorological frost occurred a week earlier than the previous year, and the last one occurred 10-days later. However, those temperatures are not well aligned with respect to the whole year minimum temperature since the years did not behave similarly.

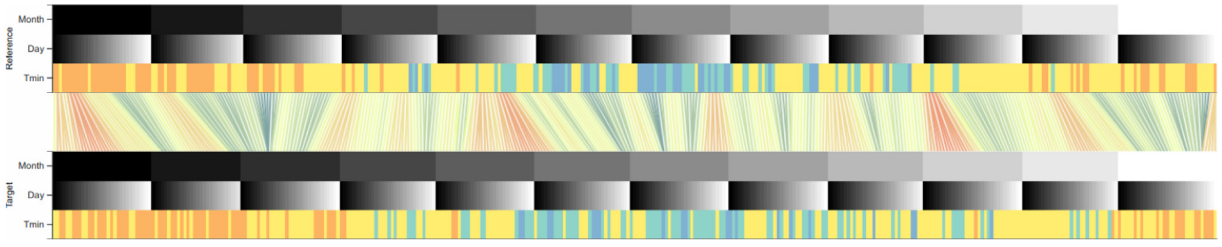

Fig. 11. Heatmaps from the 2017 vs. 2012 comparison. In 2017, the first meteorological frost occurred just 4-days later, while the last one occurred more than 2-weeks later.

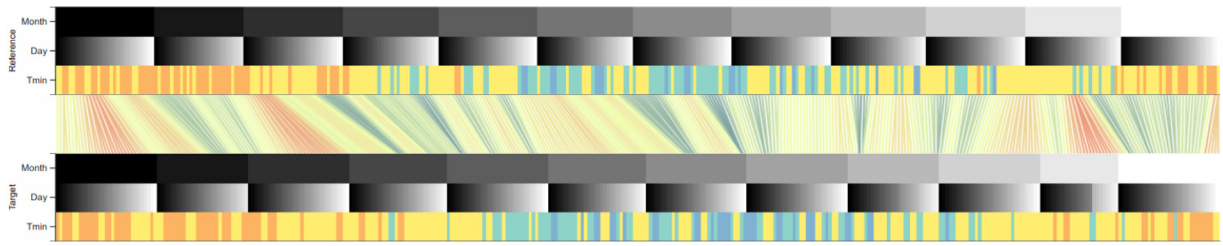

Fig. 12. Corroboration of 2018 forecast. Comparing the 2018 vs. 2017 comparison heatmaps, it can be noted that the first meteorological frost occurred just 2-days later, but the last one occurred several days later.

Fig. 13. Color-coding of the minimum temperature. Dark and light blues differentiate frost and temperatures below 5°.

CRediT authorship contribution statement

Dana K. Urribarri: Conceptualization, Implementation of the technique, Test case data curation, Test case implementation, Manuscript writing. **Martín L. Larrea:** State-of-the-Art investigation, Conceptualization, Manuscript writing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Thanks to Carlos Zotelo del *Centro de Recursos Naturales Renovables de la Zona Semiárida* from Universidad Nacional del Sur, CONICET for his valuable contribution in the interpretation and analysis of the dataset. This work was partially funded by the following research projects: PGI 25/N050, PGI 24/ZN29, and PGI 24/ZN35 from the *Secretaría General de Ciencia y Tecnología*, Universidad Nacional del Sur, Argentina.

References

[1] Cui Y, Yan D, Hong T, Xiao C, Luo X, Zhang Q. Comparison of typical year and multiyear building simulations using a 55-year actual weather data set from China. Appl Energy 2017;195:890–904. http://dx.doi.org/10.1016/j.apenergy.2017.03.113.

[2] DelSole T, Tippett MK. Comparing climate time series – part 1: Univariate test. Adv Stat Climatol Meteorol Oceanogr 2020;6(2):159–75. http://dx.doi.org/10.5194/ascmo-6-159-2020.

[3] Weather 2 travel. 2021, https://www.weather2travel.com/weather-comparison/. (Accessed 11 November 2021).

[4] Schüller L, Burfeind O, Heuwieser W. Short communication: Comparison of ambient temperature, relative humidity, and temperature-humidity index between on-farm measurements and official meteorological data. J Dairy Sci 2013;96(12):7731–8. http://dx.doi.org/10.3168/jds.2013-6736.

[5] Mason EG, Salekin S, Morgenroth JA. Comparison between meteorological data from the New Zealand national institute of water and atmospheric research (NIWA) and data from independent meteorological stations. N Z J For Sci 2017;47(7). http://dx.doi.org/10.1186/s40490-017-0088-0.

[6] Larrea ML, Urribarri DK. Visualization technique for comparison of time-based large data sets. In: Naiouf M, Rucci E, Chichizola F, De Giusti L, editors. IX conference on cloud computing conference, big data & emerging topics. Communications in computer and information science series, vol. 1444, Cham: Springer International Publishing; 2021, p. 179–87.

[7] Urribarri DK, Larrea ML, Castro SM, Puppo E. Overview+detail visual comparison of karate motion captures. In: Argentine congress of computer science. Communications in computer and information science series, Springer International Publishing; 2019, p. 139–54.

[8] Urribarri DK, Larrea ML, Castro SM, Puppo E. Visualization to compare karate motion captures. In: XXV congreso argentino de ciencias de la computación. 2019, p. 446–55.

[9] Berndt DJ, Clifford J. Using dynamic time warping to find patterns in time series. In: Proceedings of the 3rd international conference on knowledge discovery and data mining. AAAIWS'94, AAAI Press; 1994, p. 359–70.

[10] Rabiner L, Juang B-H. Fundamentals of speech recognition. Prentice Hall, Inc. 1993.

[11] Rautenhaus M, Böttinger M, Siemen S, Hoffman R, Kirby RM, Mirzargar M, et al. Visualization in meteorology—a survey of techniques and tools for data analysis tasks. IEEE Trans Vis Comput Graphics 2017;24(12):3268–96.

[12] Squillacote AH, Ahrens J, Law C, Geveci B, Moreland K, King B. The paraview guide. Vol. vol. 366. NY: Kitware Clifton Park; 2007.

[13] Ayachit U, Geveci B, Moreland K, Patchett J, Ahrens J. The ParaView visualization application. 2012.

[14] Wang Y, Fan C, Zhang J, Niu T, Zhang S, Jiang J. Forecast verification and visualization based on Gaussian mixture model co-estimation. In: Computer graphics forum. Vol. 34. No. 6. Wiley Online Library; 2015, p. 99–110.

[15] Legarretaetxebarria A, Alonso K, Serrano M, Aginako N, Olaizola I. Web-based platform for data management, analysis and visualization of weather and environmental monitoring stations. In: 5TH International conference on cartography and gis. 2014, p. 651.

[16] Tufte ER. The visual display of quantitative information. J Healthc Qual (JHQ) 1985;7(3):15.

[17] Poco J, Dasgupta A, Wei Y, Hargrove W, Schwalm CR, Huntzinger DN, et al. Visual reconciliation of alternative similarity spaces in climate modeling. IEEE Trans Vis Comput Graphics 2014;20(12):1923–32.

[18] Nocke T, Flechsig M, Bohm U. Visual exploration and evaluation of climate-related simulation data. In: 2007 Winter simulation conference. IEEE; 2007, p. 703–11.

[19] Imielinska C, Laino-Pepper L, Thumann R, Villamil R. Technical challenges of 3D visualization of large color data sets. In: The second visible human project conference proceedings. 1998.

[20] Tang Y, Qu H, Wu Y, Zhou H. Natural textures for weather data visualization. In: Tenth international conference on information visualisation. IEEE; 2006, p. 741–50.

[21] Sauber N, Theisel H, Seidel H-P. Multifield-graphs: An approach to visualizing correlations in multifield scalar data. IEEE Trans Vis Comput Graphics 2006;12(5):917–24.

[22] Quinan PS, Meyer M. Visually comparing weather features in forecasts. IEEE Trans Vis Comput Graphics 2015;22(1):389–98.

[23] Sanyal J, Zhang S, Dyer J, Mercer A, Amburn P, et al. Noodles: A tool for visualization of numerical weather model ensemble uncertainty. IEEE Trans Vis Comput Graphics 2010;16(6):1421–30.

[24] Ozdenefe M, Dewsbury J. Simulation and real weather data: A comparison for cyprus case. Build Serv Eng Res Technol 2016;37(3):288–97.

[25] Astsatryan H, Grogoryan H, Gyulgyulyan E, Hakobyan A, Kocharyan A, Narsisian W, et al. Weather data visualization and analytical platform. Scalable Comput Pract Exp 2018;19(2):79–86.

[26] Hovmöller E. The trough-and-ridge diagram. Tellus 1949;1(2):62–6.

[27] Persson A. The story of the hovmöller diagram: An (almost) eyewitness account. Bull Am Meteorol Soc 2017;98(5):949–57.

[28] Martius O, Schwierz C, Davies HC. A refined hovmöller diagram. Tellus A 2006;58(2):221–6.

[29] Galí M, Levasseur M, Devred E, Simó R, Babin M. Sea-surface dimethylsulfide (DMS) concentration from satellite data at global and regional scales. Biogeosciences 2018;15(11):3497–519. http://dx.doi.org/10.5194/bg-15-3497-2018.

[30] Zhou H, Liu X, Xu P. Sensitivity of sverdrup transport to surface wind products over the tropical north Pacific ocean. Ocean Dyn 2019;69(5):529–42. http://dx.doi.org/10.1007/s10236-019-01260-8.

[31] Hao MC, Dayal U, Keim DA, Schreck T. Multi-resolution techniques for visual exploration of large time-series data. In: EUROVIS 2007. 2007, p. 27–34.
[32] Hao MC, Dayal U, Keim D, Schreck T. A visual analysis of multi-attribute data using pixel matrix displays. In: Visualization and data analysis 2007. Vol. 6495. SPIE; 2007, p. 44–52.
[33] Wittenburg K, Pekhteryev G. Multi-dimensional comparative visualization for patent landscaping. In: Proceedings of businessvis workshop. 2015.
[34] Lammarsch T, Aigner W, Bertone A, Gärtner J, Mayr E, Miksch S, et al. Hierarchical temporal patterns and interactive aggregated views for pixel-based visualizations. In: 2009 13th International conference information visualisation. IEEE; 2009, p. 44–50.
[35] Pleil JD, Stiegel MA, Madden MC, Sobus JR. Heat map visualization of complex environmental and biomarker measurements. Chemosphere 2011;84(5):716–23.
[36] Bagwan WA. An assessment of rainfall-induced land degradation condition using erosivity density (ED) and heatmap method for urmodi river watershed of maharashtra, India. J Sediment Environ 2020;5(3):279–92.
[37] Kumatani S, Itoh T, Motohashi Y, Umezu K, Takatsuka M. Time-varying data visualization using clustered heatmap and dual scatterplots. In: 2016 20th International conference information visualisation. IEEE; 2016, p. 63–8.
[38] Lévesque F, Hurtut T. Muzlink: Connected beeswarm timelines for visual analysis of musical adaptations and artist relationships. Inf Vis 2021;20(2–3):170–91.
[39] Jensen M. Visualizing complex semantic timelines. 2003, Citeseer. Derived from the World Wide Web: http://newsblip.com/tr.
[40] Huang F, Zhang D, Chen X. Vegetation response to groundwater variation in arid environments: Visualization of research evolution, synthesis of response types, and estimation of groundwater threshold. Int J Environ Res Public Health 2019;16(10):1849.
[41] Chen C. Citespace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. J Am Soc Inf Sci Technol 2006;57(3):359–77.
[42] Knipp DJ, Bernstein V, Wahl K, Hayakawa H. Timelines as a tool for learning about space weather storms. J Space Weather Space Clim 2021;11:29.
[43] Zotelo C, Martín E. Meteobahia. Datos meteorológicos. 2014, https://meteobahia.com.ar/. (Accessed 10 November 2021).
[44] Murray MG, Galan C. Effect of the meteorological parameters on the olea europaea l. pollen season in bahía blanca (Argentina). Aerobiologia 2016;32(3):541–53.
[45] Gentili JO, Fernández ME, Gil V. Influence of topography on local atmospheric features in a peri-urban area of bahía blanca (Argentina). Environ Process 2020;7(1):23–40. http://dx.doi.org/10.1007/s40710-019-00408-4.
[46] Wilks D. Statistical methods in the atmospheric sciences. International geophysics, 2. vol. 91, Academic Press; 2005.
[47] Fernandez-Long M, Barnatan I, Dominici C, Murphy G. Información agroclimática de las heladas en la Argentina: generación y uso. Meteorologica 2016;41:7–31.
[48] Stauffer R, Mayr GJ, Dabernig M, Zeileis A. Somewhere over the rainbow: How to make effective use of colors in meteorological visualizations. Bull Am Meteorol Soc 2015;96(2):203–16. http://dx.doi.org/10.1175/bams-d-13-00155.1.
[49] Fuchs J, Fischer F, Mansmann F, Bertini E, Isenberg P. Evaluation of alternative glyph designs for time series data in a small multiple setting. In: Proceedings of the SIGCHI conference on human factors in computing systems. 2013, p. 3237–46.
[50] Liu Y. Multivariate data visualization: A review from the perception aspect. In: Symposium on human interface. Springer; 2011, p. 221–30.