

Similarity assessment and adjustment of integrated temperature databases. Cuyo region, Argentina

Evaluación de la similitud y ajuste de bases de datos integrados de temperatura. Región de Cuyo, Argentina

Luis Bernardo Bastidas*, Alberto Ismael Juan Vich** y María Cintia Piccolo***

Recibido: 12/11/2020. Aceptado: 13/1/2021. Publicado: 10/5/2021.

Abstract. Given the frequent spatial-temporal limitations and deficiencies of instrumental meteorological records, the use of alternative information sources, such as integrated databases, are important for analyses and studies of diverse nature. The research aim was to evaluate the accuracy of integrated databases of monthly temperature, belonging to Climate Research Unit, University of Delaware and Global Historical Climatology Network, gridded with a pixel size of 3,098.01 km² (0.5° x 0.5°), surface area of 151,802.5 km² and temporary length of 22 years (1993-2014), through the modified structural similarity index (mSSIM). The study area is located in central-western Argentina (between 30° and 35° S, and 71° and 66° W). The University of Delaware grid showed the best fit of the data series from 10 weather stations located in the study area. Therefore, a proposal was presented to increase similarity indices, especially for those cells without instrumental reference information. The study determined that by applying this modification, the gridded datasets increases the similarity of the measured data, es-

pecially in mountainous areas, where originally there were differences of more than 7.5 °C between the gridded data and observed one. The proposal decreases these differences to average values below 1 °C. The use and subsequent adjustment of these integrated databases, allows access to information in areas without meteorological records.

Key words: gridded datasets, structural similarity, monthly temperature, central-western Argentina, concordance analysis.

Resumen. Dadas las limitaciones y carencias espacio-temporales de los registros meteorológicos instrumentales, el uso de fuentes de información alterna, como las bases de datos climáticas integradas, resultan de gran importancia para análisis y estudios de diversas índoles. El objetivo de la investigación fue evaluar la precisión de las bases de datos integradas de temperatura mensual de la Climate Research

* Instituto Argentino de Nivología, Glaciología y Ciencias Ambientales IANIGLA – Consejo Nacional de Investigaciones Científicas y Técnicas CONICET. Ruiz Leal s/n, Parque San Martín, CP 5500, Mendoza, República Argentina. Departamento de Geografía y Turismo DGyT – Universidad Nacional del Sur UNS. 12 de Octubre y San Juan, Bahía Blanca, CP 8000, República Argentina. ORCID: <https://orcid.org/0000-0002-8570-9734>. Email: lbastidas@mendoza-conicet.gov.ar

** Instituto Argentino de Nivología, Glaciología y Ciencias Ambientales IANIGLA – Consejo Nacional de Investigaciones Científicas y Técnicas CONICET. Ruiz Leal s/n, Parque San Martín, CP 5500, Mendoza, República Argentina. Facultad de Filosofía y Letras FFyL – Universidad Nacional de Cuyo UNCuyo. Centro Universitario, Mendoza, República Argentina. CP M5502JMA. ORCID: <https://orcid.org/0000-0002-0477-2476>. Email: aijvich@mendoza-conicet.gov.ar

*** Instituto Argentino de Oceanografía IADO - Consejo Nacional de Investigaciones Científicas y Técnicas CONICET. Florida 450, CP 8000, Bahía Blanca, República Argentina. Departamento de Geografía y Turismo DGyT – Universidad Nacional del Sur UNS. 12 de Octubre y San Juan, CP 8000, Bahía Blanca, República Argentina. ORCID: <https://orcid.org/0000-0002-5184-9149>. Email: ofpiccol@criba.edu.ar

Unit, University of Delaware y del Global Historical Climatology Network, organizados en grillas con pixel de 3 098.01 km² (0.5° x 0.5°), extensión superficial de 151 802.5 km² y longitud temporal de 22 años (1993-2014), a través del índice modificado de similitud estructural (mSSIM). La zona de estudio se localiza en el centro-occidente argentino (entre los 30° y 35° S, y los 71° y 66° O). La grilla de la University of Delaware mostró el mejor ajuste de las series de datos de 10 estaciones meteorológicas localizadas en el área de estudio. Por tanto, se presenta una propuesta para elevar los índices de similitud, en especial para aquellas celdas sin información instrumental de referencia. El estudio permitió determinar que, aplicando

dicha modificación, la base de datos grillada aumenta la similitud de los datos medidos, sobre todo en áreas montañosas, donde originalmente se encontraban diferencias de más de 7.5 °C entre el dato grillado y el observado. La propuesta disminuye esas diferencias a valores promedios inferiores a 1 °C. El uso y posterior ajuste de estas bases de datos integradas, permite acceder a información en zonas sin registros meteorológicos.

Palabras clave: bases de datos grilladas, similitud estructural, temperatura mensual, centro-occidente de Argentina, análisis de concordancia.

INTRODUCTION

The growing need for the use of meteorological data, makes its demand focus on increasing their availability in time and space (WMO, 2011). The above goes in contrast to the fact that there are fewer gauge stations with shorter temporal lengths or lower quality, both in the world (Shelton, 2008) and in Argentina and the Cuyo region (Boninsegna & Villalba, 2014; Santos, 2014; Lui, 2019). While the optimal scenario is to base studies on instrumental time series, problems of extension in space and acceptable time lengths (≥ 30 years, according to the World Meteorological Organization, WMO, 2017), make it increasingly frequent to use alternative data sources.

Integrated Climate Databases, or ICB (NOAA, 2019), are meteorological datasets, established in a grid of certain X and Y dimensions, and with generally long temporal lengths. According to the predetermined pixel dimensions, its information (Z-axis) is stored as a time series. Some, like UDEL (University of Delaware), turn out to be a compendium of data from a large number of stations, based on the Global Historical Climatology Network (GHCN), resulting in a monthly climatology of more than 100 years of temporal length and with eminently terrestrial coverage. ICB interpolations are executed using of digital elevation models, assisted interpolation, traditional interpolations, and climatologically assisted interpolations (Matsuura & Willmott, 2018). One of its main advantages is that the coverage can cover areas with no measurement stations, so it represents, in many cases, the only information source

in these areas. Since interpolations have a margin of error concerning the original nature of a given event, choosing the one with the highest adjustment is essential to increase the levels of certainty about it.

Several studies (Takido et al., 2016; Wong et al., 2017; Nashwan et al., 2019; Qiu et al., 2019) have aimed to identify the accuracy of these databases and then apply them as inputs for climate analysis, including regions of central-eastern Argentina (Rusticucci et al., 2014; Casado & Picone, 2018). The measures to assess the adjustment of the ICB are mainly based on the magnitudes of error (between these and the instrumental data) and structure or correlation indicators, such as Pearson's r coefficient (Rusticucci et al., 2014; Ferrelli et al., 2016). It should be noted that the study area presents large regions without meteorological measurements (Fig. 1), which would lead to having to use interpolations from neighboring stations or the use of ICB, with the usual uncertainty that precedes them. In this sense, the aim was to evaluate the similarity of three gridded databases of monthly temperature, with respect to ten meteorological stations located in central-western Argentina. An unconventional similarity index, frequently used in medical settings, was used to assess the statistic adjustment (Zhao et al., 2019).

METHODOLOGY

Data used

As data input, the ICB of the Climate Research Unit - CRU v. 4.04 (UIACRU et al., 2020). Uni-

versity of Delaware - UDEL v. 5.01 (Matsuura & Willmott, 2018) and the Global Historical Climatology Network - GHCN v.2 (NCARS, 2019), were freely downloaded from the data repositories of the Climatic Research Unit - CRU (University of East Anglia, 2019), Earth System Research Laboratory - ESRL (NOAA, 2019) and the National Center for Atmospheric Research Staff (NCARS, 2019), respectively. These databases have a spatial resolution grid of $0.5^\circ \times 0.5^\circ$ and a temporal resolution (selected from the total) of 22 years (1993 - 2014); the time period was chosen due the available weather data in the region. The domain designed covers from 30° S to 35° S and 71° W and 66° W. The data source for the cons-

truction of these ICB is based on meteorological stations data (Fig. 1, Tab. 1).

Study area

The study area is located in the central-western part of Argentina, specifically in the Cuyo region and near the border with Chile (Fig. 1). The grid site covers 3.5° of latitude and longitude respectively. With each pixel or cell having dimensions of $0.5^\circ \times 0.5^\circ$, or $3,098.01 \text{ km}^2$, the total area of the gridded zone is $151,802.5 \text{ km}^2$, distributed among the provinces of Mendoza, San Juan, San Luis, and La Rioja. The 13 meteorological stations (Fig. 1, Tab. 1) distributed within the area, depending on national agencies, such as the National Meteorolo-

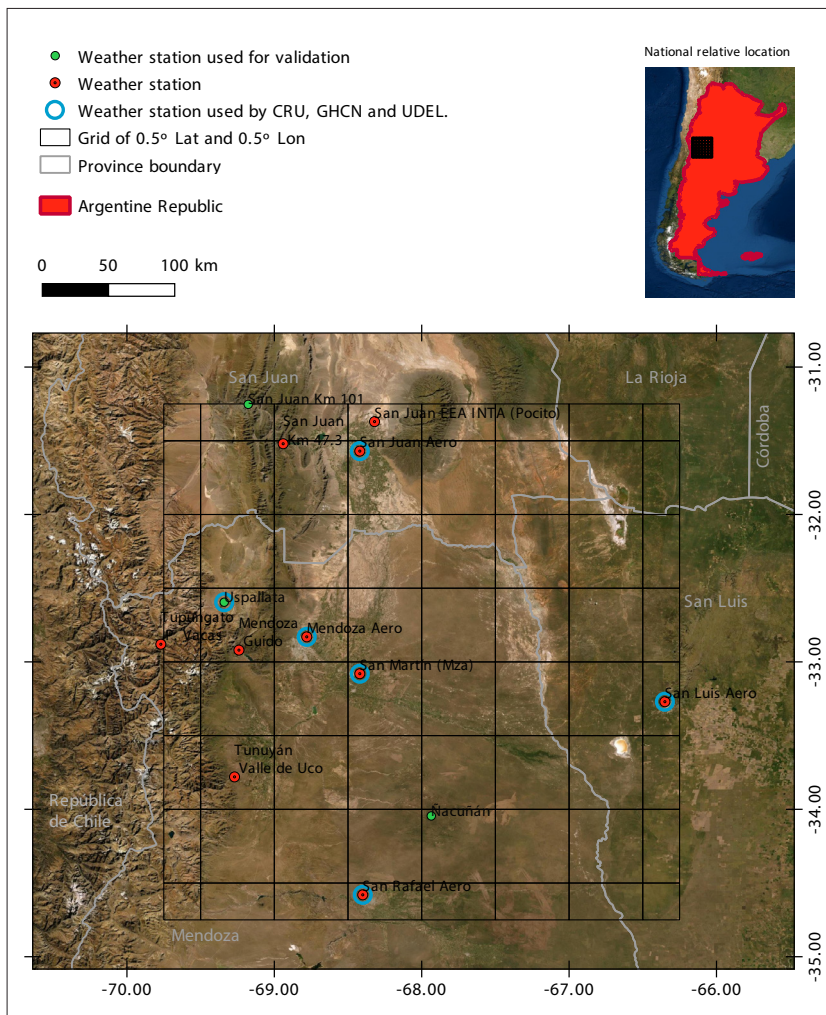


Figure 1. Geographic location of the CRU, UDEL and GHCN grid extension and the considered weather stations. Source: self-made elaboration.

Table 1. Selected meteorological stations.

Station	Agencies	Code	Stations used by / for:			Applied method	Validation
			ICB	UDEL	GHCN		
San Juan Km. 47,3	SMN	1208				X	
San Juan Km. 101	SMN	1211					X
Uspallata	SMN	1491	X				X
Tupungato Pta. de Vacas	SMN	1420				X	
Tunuyán Valle de Uco	SMN	1419				X	
Mendoza Guido	SMN	1413				X	
San Juan Aero	SIPHN	87311	X	X	X	X	
Mendoza Aero	SIPHN	87418	X	X	X	X	
San Martín (Mza)	SIPHN	87416		X	X	X	
San Luis Aero	SIPHN	87436	X	X	X	X	
San Rafael Aero	SIPHN	87509	X	X	X	X	
Ñacuán	IADIZA	-					X
San Juan EEA INTA	INTA	NH0445				X	

Source: self-made elaboration.

gical Service (SMN, 2019), the National Institute of Agricultural Technology (INTA, 2019), the National Secretariat of Infrastructure and Water Policy (SIPHN, 2019) and the Argentine Institute for Arid Zone Research (IADIZA, 2019).

Each database is stored in their respective web portals as NetCDF (Network Common Data Form) file format, where each pixel, besides its X and Y dimensions, has temporary information (Z dimension) of monthly scale (264 months). Regarding Table 1, it should be noted that only the stations managed by the SIPHN were used as a base for the construction of the gridded products, except for the Uspallata station (only used by CRU).

Extraction of data from ICB

Overlaying the weather stations over the NetCDF databases in a map (all with a common period 1993 - 2014), it is found that each one has associated a pixel where it is immersed. By extracting the time series of each associated pixel, pairs of time series (Y) were obtained, whose months are in the same chronological position as the instrumental ones (X).

Application of similarity indices

The contrast between pairs of time series (CRU, UDEL, and GHCN) versus instrumental data, was made by applying the Modified Structural Similarity Index [mSSIM] (Wang & Bovik, 2002; Mo et al., 2014). mSSIM utilizes the mean, variance, and structure (Eq. 1), to evaluate how similar or dissimilar are the contrasted time series (concordance). The range of similarity varies between 0 and 1, being 0 for completely dissimilar series, and 1 if it turns out to be the same (Fig. 2).

The interpretation of mSSIM is based on comparing the similarity between two data vectors, X and Y. Since the closeness or distance of these depends on the product of the measures mentioned above, the index will be more robust when their values are closer to 1 (with subsequent verification of statistical significance). Since the similarity is a function of the magnitude of the vectors tested, the use of the index is not recommended for variables at different scales of measurement (i.e. temperature and precipitation) (Wang & Bovik, 2002).

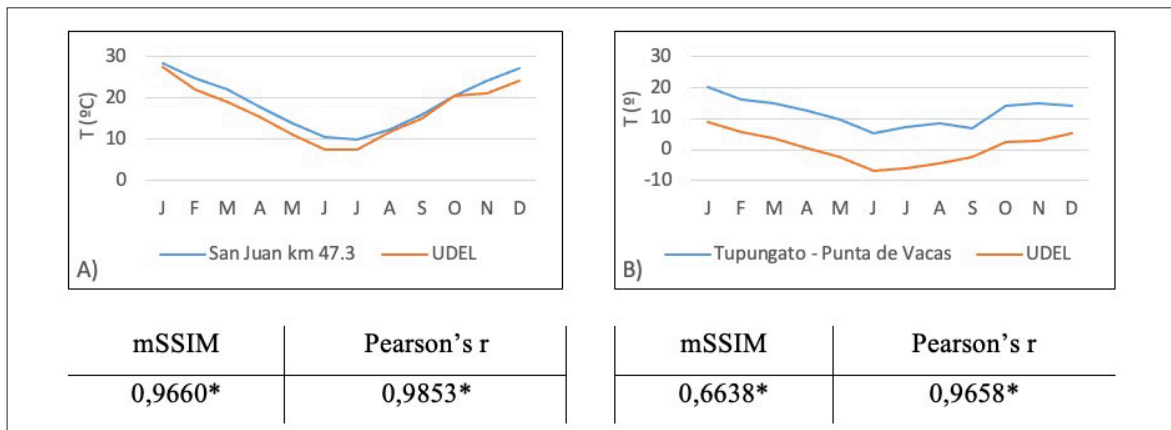


Figure 2. Comparison between the mSSIM and Pearson's r indexes, regarding data results from San Juan km 47.3 (A) and Tupungato - Punta de Vacas (B) weather stations, versus UDEL data. Year 2014. Significant values according to Student's t-test, with $\alpha=0.05$. Source: self-made elaboration.

Since several authors (i.e. Rusticucci et al., 2014; Bustos et al., 2016; Rivera et al., 2018) used Pearson's correlation coefficient (r) as an indicator of adjustment between grids and instrumental series, the mSSIM index is presented as a robust measure to evaluate similarity by structure and closeness between two data vectors X and Y (Mo et al., 2013; Mo et al., 2014). This index, which is frequently used in medical imaging comparison studies (Zhao et al., 2019), also has advantages over other similarity measures, including the Pearson's r coefficient (Mo et al., 2013). Therefore, both methods (mSSIM and r) were applied for comparison purposes, whose equations (1 and 2) are detailed below, respectively.

$$mSSIM = \left(\frac{2\tilde{x}\tilde{y}}{\tilde{x}^2 + \tilde{y}^2} \right) \left(\frac{2\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2} \right) \left(\frac{\sigma_{xy}}{\sigma_x\sigma_y} \right)$$

Where: $\tilde{x} = \bar{x} - \Psi_{xy}$; $\tilde{y} = \bar{y} - \Psi_{xy}$; $\Psi_{xy} = \min(x_i, y_i) | i = 1, 2 \dots N$. 1)

$\left(\frac{2\tilde{x}\tilde{y}}{\tilde{x}^2 + \tilde{y}^2} \right)$: mean; $\left(\frac{2\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2} \right)$: variance; $\left(\frac{\sigma_{xy}}{\sigma_x\sigma_y} \right)$: structure.

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}} \quad 2)$$

Where: x and y represent the contrasted matrices vectors, respectively.

The justification for using mSSIM (Eq. 1),

as opposed to the Pearson correlation coefficient r (Eq. 2), lies in the fact that the latter exposes the structure or trend relationships between the contrasted data series, proving to be insensitive to proportional differences in measurements and therefore, considered as an inaccurate measure of similarity (Mo et al., 2013). It should be noted that Pearson's r is contained in the mSSIM equation (as structure parameter).

Determination of ICB with higher adjustment

Subsequently, and having ten mSSIM values between the time series of each weather station (Tab. 1), versus each gridded series (CRU, UDEL, GHCN) associated with the first one, 30 results were obtained. Ranked in terms of accuracy (i.e., 1st, 2nd, and 3rd place), the objective was to find the database with the best position (highest mSSIM) concerning the others.

Although the previous step determines which ICB has the greatest similarity with respect to the measured data, since it presents gridded values, it is very unlikely that the similarity is equal to 1, so the existing differences can be adjusted to reduce the dissimilar gap between the two series.

Adjustment proposal

Once the grid that had presented the greatest similarity with the instrumental data was chosen,

the aforementioned adjustment was carried out, which was executed by:

$$C_i = M_i - \bar{X}_K \quad \therefore \quad I_i < M_i \quad 3)$$

$$C_i = M_i + \bar{X}_K \quad \therefore \quad I_i > M_i \quad 4)$$

Where, : C_i : ICB calculated in month i ; M_i : gridded value for month i ; I_i : instrumental record of the month i ; \bar{X}_K : difference average between I_i and M_i for every month K .

Prediction model for cells without information

To find a prediction model based on the adjusted or modified data (Eq. 5), the equation of the best-fit line between the gridded and modified data series for each station was calculated. The contrasting time series of each grid point were subjected to a cluster analysis, in order to define the different groupings in space. These clusters were considered as the influence area of the considered weather stations.

$$y = a + bx \quad 5)$$

Where, y : modified grid value; a : b determined intercept and slope parameters, respectively; x : original grid value.

In this sense, the equation of the line found for each weather station (gridded versus modified data) was applied in the pixels that were located within their influence area (grouped by cluster analysis). In the linear regression equation of type $y=a+bx$, x represents each month of the time series of each one of the pixels, obtaining in this way values of y (corrected by regression equation). Thus, pixels that did not have measurements from weather stations to be contrasted, could be adjusted.

Accuracy and validation

Finally, to the instrumental time series, gridded and calculated by regression (only from those pixels that comprise the weather stations), their respective arithmetic averages were calculated. In this way, as each pixel had only one value, three layers were ge-

nerated in raster format. Thus, using map algebra, the raster of instrumental values was subtracted from the chosen grid, and then the calculated or modified one. In this way, the level of precision reached by the second one was determined according to the followed methodology.

To evaluate the pixel adjustment level without data from meteorological stations, the following ones were used: Ñacuñán (IADIZA, 2019), Uspallata (SIPHN, 2019), and San Juan km 101 - code 1211 (SMN, 2019), located at 34°02'42" S, 67°56'06" W; 32°35'42" S, 69°20'24" W; and 31°15'11" S, 69°10'37" W, respectively (Fig. 1, Tab. 1). From these stations, only Uspallata was used by CRU as a base for the conformation of its grid.

The chosen periods to evaluate the level of pixel adjustment correspond to the lengths of continuous records available from these stations (Tab. 2); that is, 21.4 years - for Uspallata (08/1993 - 12/2014), five years for San Juan km 101 (01/2010 - 12/2014) and six years for Ñacuñán (01/2009 - 12/2014).

RESULTS

The mSSIM was applied for the ten gauge stations (Tab. 1) against the three ICB, resulting in 30 values, whose averages were ranked according to the results for each station. Ordered data indicated the greatest similarity as follows (Tab. 2): UDEL with 0.9491, CRU with 0.9314, and GHCN with 0.9211 (all with significant values by Student's t-test, with $\alpha=0.05$). Although the differences between these values were less than 3 %, UDEL was chosen to continue with the adjustment methodology.

Since ICB consider almost the same weather stations as a base, on the one hand, the similarity values were close to each other; however, although the differences between mSSIMs were less than 3 %, UDEL presented the highest similarity. On the other hand, the differences with Pearson's r were less than 0.2 %, giving UDEL and GHCN the highest value, but without defining in terms of similarity the ICB with the highest adjustment (Fig. 2).

Table 2. Values of mSSIM and Pearson's r between ICB and weather stations.

ICB	mSSIM	Ranked mSSIM	Pearson's r	Ranked Pearson's r
UDEL	0,9491	I	0,9885	I
CRU	0,9314	II	0,9867	II
GHCN	0,9211	III	0,9885	1

Source: self-made elaboration.

As shown in the figure above, Pearson's r does not vary, even though in Figure 2B, there are differences in magnitude up to 10 °C. Meanwhile, mSSIM is presented as a more sensitive and robust index, incorporating structural, magnitude, and variation measures. Although both data series (instrumental and gridded) are similar in structure, they are not the same, so the Pearson r coefficient does not expose the differences due to similarity. Therefore, if both time series follow the same pattern, but are different in magnitude, Pearson's r will not detect it, unlike mSSIM that makes it explicit (Fig. 2).

Applied the fitting procedure (Eq. 3 and 4), the differences between UDEL and the contrasted instrumental series decreased, thus obtaining new modified UDEL values (Fig. 3). Subsequently, the linear regression model found between X (UDEL) and Y (modified UDEL), allow to estimate the temperature for those UDEL cells that did not have associated measurements of meteorological stations.

Although it is observed that the regression model overlaps with the modified UDEL curve (Fig. 3), the aim is to apply it to cells where no weather stations exist. Therefore, a cluster analysis (Fig. 4) was applied to the UDEL cells to find the respective clusters and to apply the regression models to the regions found previously.

Once the time series were averaged for each cell, they were subjected to cluster analysis (Ward method with Euclidean distance), resulting in five clusters (optimum defined by Elbow and Silhouette methods). However, since I cluster is divided into two sections (Fig 4A), due to spatial discontinuity, its northern part was formed as a sixth cluster (Fig. 4B).

A way to compare the cluster formation above (based on UDEL data) with the relief and the relationship between them, is observing the natural conditions in the region. In the next figure (Fig. 5), are presented the mean UDEL (without adjustment) temperature contours (°C), the altitudes (SRTM) and the location of temperature stations.

Comparing figures 4 and 5, the spatial distribution of formed clusters is coherent with the temperature and relief spatial configuration. In this way, the lowest temperatures are presented at the east, according with the highest altitudes of the Andes Mountains.

After the clusters were defined, the equations generated by the linear regression for each weather station were applied. Thus, for example, the equation of the Mendoza Guido station was applied to the six cells of the corresponding region IV. Therefore, in the equation of the form , represents the corrected monthly value, and the original UDEL data.

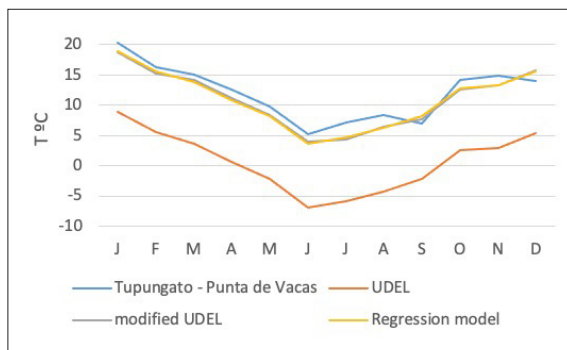


Figure 3. Differences between the instrumental value data (Tupungato - Punta de Vacas) and UDEL, its modification, and the applied linear regression model. Year 2014. Source: self-made elaboration.

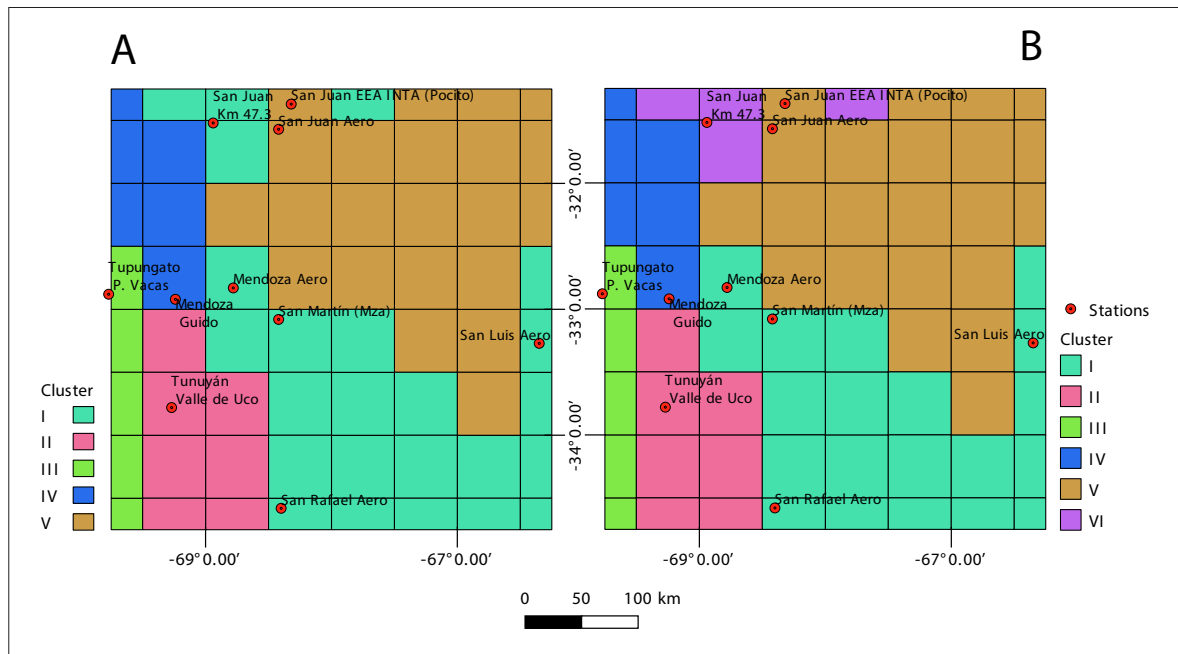


Figure 4. Cluster formation (A) and its modification (B), including the location of the considered weather stations. Source: self-made elaboration.

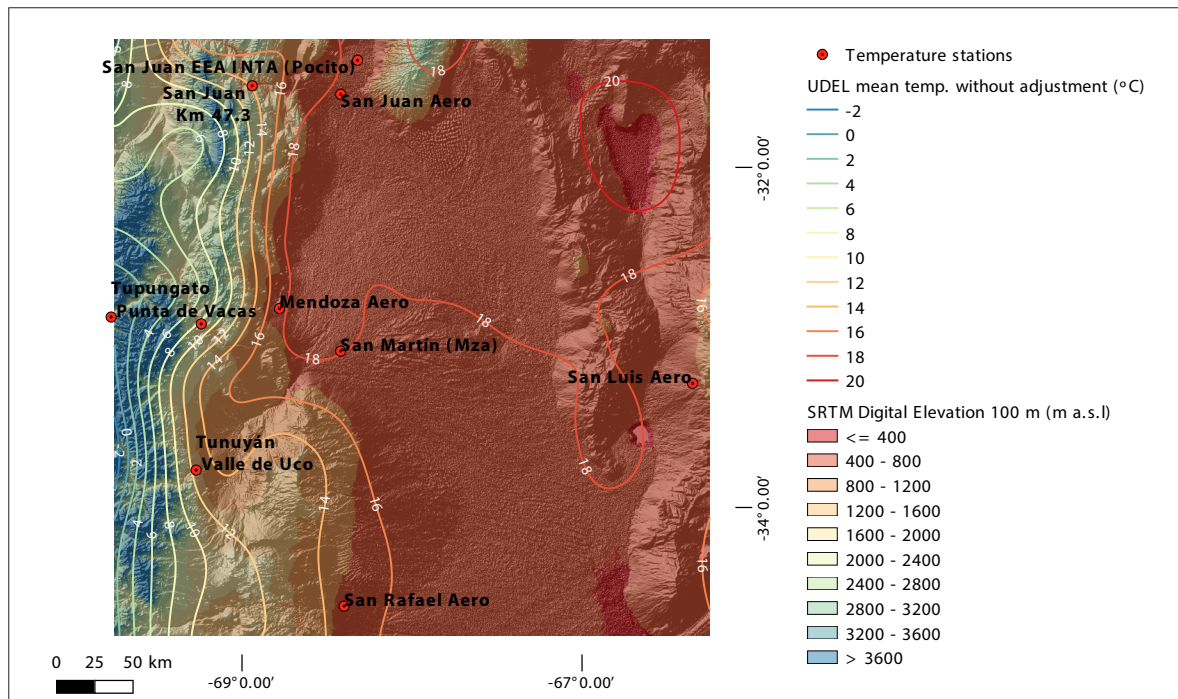


Figure 5. Temperature stations and its position in front of mean UDEL (without adjustment) temperature contours (°C) (1993–2014) and SRTM digital elevation (m a.s.l.). Source of data used: Farr et al. (2007); Matsuura & Willmott, (2018). Source of map: self-made elaboration.

Since there are cases such as clusters I and V (Fig. 4), whose data was originated from more than one weather station, the linear regression model was applied to the surrounded cluster. Then, the adjusted results were averaged as a function of the number of instrumental stations. Since each gridded cell (original UDEL) and adjusted has three-dimensional information (Z as temporal axis), the time series of each cell of these two grids were averaged to form two raster images (Fig. 6).

Although it is observed that both images (Fig. 6) have the same measurement scale, it is evident that the adjusted UDEL grid (Fig. 6B) overestimates the unmodified one (Fig. 6A). However, since only measured data can verify the reliability of the estimates, the information from the weather stations was rasterized (an average per station) with the same cell size ($0.5^\circ \times 0.5^\circ$). These cells (using map algebra) were subtracted from their averaged pairs of unmodified UDEL and modified UDEL, to evaluate where the most significant differences are found (Fig. 7).

In the figure above, it is observed how the largest differences (over 2°C and up to 10°C)

correspond to the original UDEL grid (Fig. 7A), while after modification or adjustment of UDEL (Fig. 7B), the differences for the considered pixels are less than 1°C .

Finally, as part of the validation for cells without instrumental information, it was proceeded to contrast them with the Uspallata, San Juan km 101, and Nacuñán weather stations, which were not included in the previous analyses nor were they used as a basis for the construction of UDEL (Tab. 1).

The time series of the selected stations were contrasted with the original and modified UDEL time series cells (by regression), where they were immersed. That is, having X as the observed value and Y as the gridded series with and without modification, coinciding on the same time period. The contrast and similarity values by mSSIM between both series can be seen in Figure 8 and Table 3, respectively.

The visual fit level between the curves of the three meteorological stations considered, and the UDEL curves (modified and unmodified) is shown in Figure 8. Although only the 2014 records were

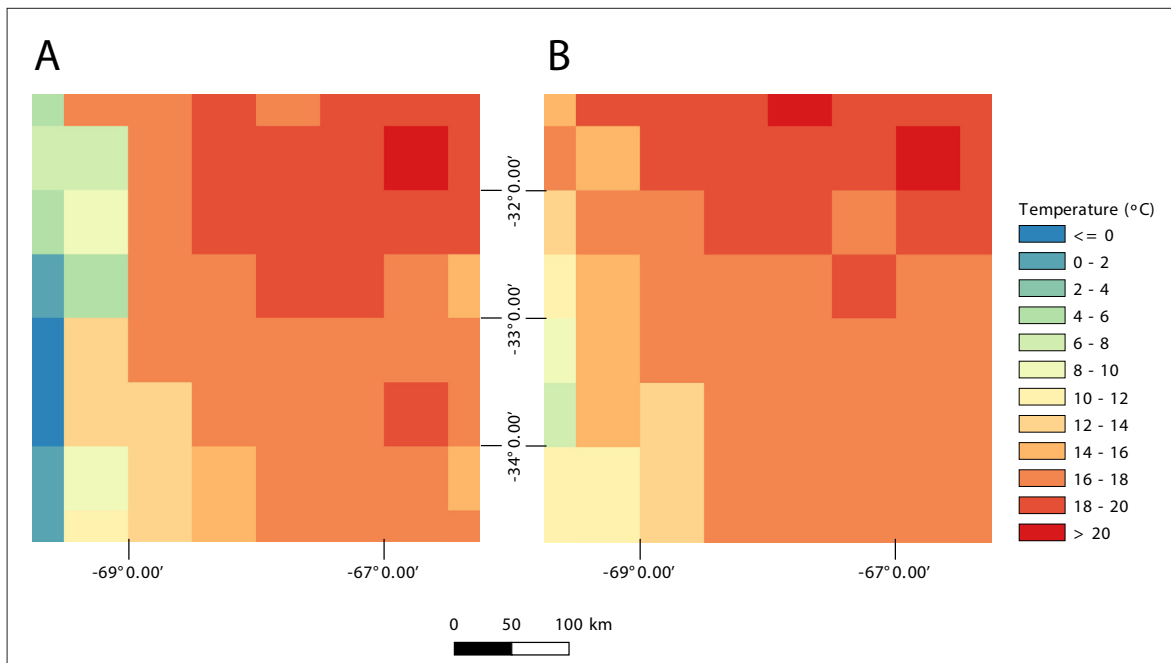


Figure 6. UDEL averages without adjustment (A) and adjusted (B). Source: self-made elaboration.

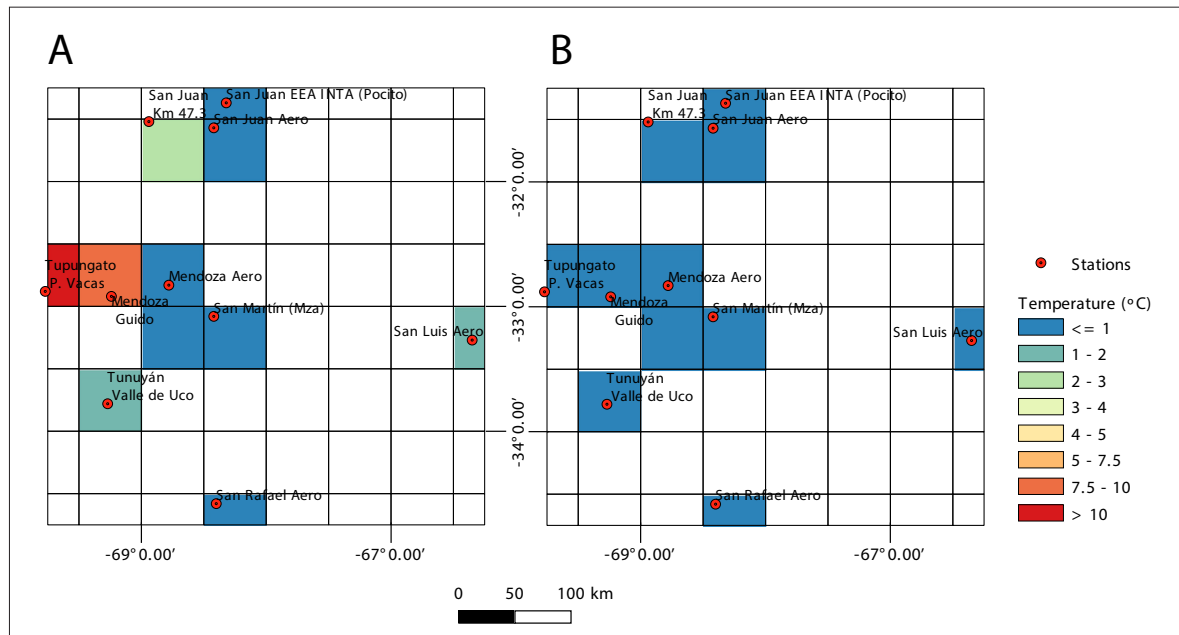


Figure 7. Differences in cell averages (where weather stations are located) between original (A) and modified (B) UDEL. Source: self-made elaboration.



Figure 8. Contrast between instrumental curves versus unmodified and modified UDEL, for the Uspallata, San Juan km 101 and Ñacuñán stations. Year 2014. Source: self-made elaboration.

Table 3. Similarity values (mSSIM) between measured data (I) versus unmodified and modified UDEL, for the three weather stations used in the validation.

Weather station	Time length and period (MM/YYYY)	I vs. UDEL unmodified	I vs. UDEL modified
Uspallata	21.4 years – (08/1993 – 12/2014)	0.8304	0.9678
San Juan km 101	5 years - (01/2010 – 12/2014)	0.9440	0.9834
Ñacuñán	6 years - (01/2009 – 12/2014)	0.9899	0.9871

Source: self-made elaboration.

represented, Table 3 shows the mSSIM values according to the different lengths of the time series available at these stations.

As shown in Table 3, by the three considered cases, the modified UDEL was closer to the instrumental series than the unadjusted curve. This situation is more accentuated on the Uspallata station than on the other two, and is reiterated given the lower general precision of the UDEL on mountainous areas. The slight difference for Ñacuñán unmodified and modified values (0.0028), can be due the application of modified proposal methodology in flatted areas (like Ñacuñán surrounding site). The above can also be seen in Figure 7A, where, on the one hand, it is represented that towards the west of the gridded zone, there are cells with temperature differences greater than 7.5 °C concerning the instrumental values. On the other hand, towards the central and eastern zone of the region, the differences are lower (Ñacuñán station), presumably because it is a site with lower and regular topography.

In contrast, in Table 3, the numerical difference of similarity shows that the precision of the modified UDEL is greater than its original pair, except in the Ñacuñán station, where the original database presents a slight difference over the modified one. According to the above, the greatest difference for Uspallata stands out (for the reasons stated in the previous paragraph). Such evaluation reaffirms the potential of UDEL in terms of its application to cells without measurements and the advantages of using mSSIM as a similarity index.

DISCUSSIONS AND CONCLUSIONS

As part of the better-fitting ICB selection, the mSSIM shows advantages over Pearson's *r* coefficient as a similarity index. Such a situation allowed defining that UDEL data had higher similarity than CRU and GHCN data. It should be noted that although the differences were less than 3 % and the statistical significance was 5 %, the choice of UDEL was given by proximity to the records and about the other ICB ones.

Although several investigations (Rusticucci et al., 2014; Bustos et al., 2016; Casado & Picone 2018; Rivera et al., 2018; Qiu et al., 2019) base the ICB validation by applying Pearson's *r*, the mSSIM index was positioned as an indicator that presents greater precision in terms of similarity measurement. The mSSIM may be considered as a modified correlation coefficient, with potential applications in meteorology and climatology (Mo et al., 2013).

In this way, mSSIM proved to be more effective in recognizing which ICB had higher similarity with instrumental data, unlike the results found with Pearson's *r*, which did not show clarity in establishing differences. Although one of the objectives of ICB was to provide a source of alternative climate data to the instrumental records, the differences with the latter may not be exposed to measures such as Pearson's *r*. While such data sources may under or over-estimate the measurements, they generally maintain the same structural pattern due to their relationship of origin with instrumental data.

Once the UDEL database was chosen, the adjustment was applied through a proposed modification to improve the accuracy. By averaging the time series of each cell, and grouping them into six clusters, the regression equation found for their respective weather stations (modified UDEL versus original UDEL) was applied. In this way, cells without measurement stations could be corrected using the stations within the cluster where they were located as a reference.

It should be noted that unlike the proposed modification made, some authors (Rusticucci et al., 2014; Ferrelli et al., 2016) tried to validate and modify reanalysis databases by applying linear regression equations (instrumental data versus unmodified modeling). In this way, they applied the equations found based on closeness by indicators such as Pearson's r . Thus, the regressions were applied without taking into account the similarity, which ends up underestimating or overestimating the existing instrumental data (depending on the gaps between them and the model used).

Concerned to validations, they indicated differences of less than 1 °C between modified UDEL versus the instrumental data, compared to more than 2 °C difference between the original UDEL and the records measured in the pixels located west and east of the region (mountainous area). Instead, for those cells without instrumental stations, the tests indicated an increase in the mSSIM values, bringing the curve of the measured data closer to the modified UDEL, and therefore, increasing its reliability.

It is important to emphasize that originally, UDEL has greater inconsistencies in mountainous areas, where it is advisable to apply the proposed modification to reduce the differences in magnitude between them. However, since in regions with more regular topography, the differences between UDEL and its modified pair are not significantly large, depending on the site, it is possible to use such ICB without applying modifications.

Although the ICB have their origin in the instrumental data, it is advisable to subject them to a contrast of similarity with the instrumental data. The aim, is to be able to apply the proposal to sites far from weather stations, so it is neces-

sary to modify it and thus increase its similarity with the measured data. In this way, cells without instrumental measurements (as in the center-east of the study area), could have temperature data series with greater reliability, thus reducing the uncertainty of spatial measurement gaps (such as the vast semi-desert and uninhabited plains of northeast Mendoza).

Finally, it is remarkable the role of mSSIM as a similarity index to differentiate between the ICB considered, as well as between the observed and modified UDEL values later on. Although it is advisable to use data from measurements, their lack in certain areas (such as the northeast of the studied region) makes it necessary to resort to alternative sources of information. In these cases, these databases represent a valuable source of climate information, as well as an alternative to the weaknesses of the conventional meteorological network.

The replicability of the applied methodology is recommended for plains regions and not quite accurate for mountainous zones and coastal border lines (last fact mentioned due to the limitations of some ICB itself). Future works could use validated ICB as an important data source for the non-measured regions.

REFERENCES

- Boninsegna, J. & Villalba, R. (2014). Documento marco sobre la oferta hídrica en los oasis de riego de Mendoza y San Juan. Los condicionantes geográficos y climáticos. Retrieved from https://inta.gob.ar/sites/default/files/script-tmpinta_documento_marco_sobre_la_oferta_hidrica_en_los_oa_1.pdf
- Bustos, M. Ferrelli, F. Cisneros, M. Piccolo, M. & Gil, V. (2016). Estudio preliminar del ajuste entre datos meteorológicos in situ y del reanálisis (NCEP/NCAR) en distintos ambientes de la provincia de Buenos Aires, Argentina. *Estudios Geográficos*, 77(280), 335–343. <https://doi.org/10.3989/egoeogr.2016.i280.485>
- Casado, A. & Picone, N. (2018). Aplicabilidad de los datos grillados para el análisis espaciotemporal de las precipitaciones, provincia de Buenos Aires (Argentina). *Párrafos Geográficos*, 17(1), 46–62. Retrieved from http://www.igeopat.org/parrafosgeograficos/images/RevistasPG/2018_17_1/29-3.pdf
- Farr, T. Rosen, P. Caro, E. Crippen, R. Duren, R. Hensley, S. Kobrick, M. Paller, M. Rodriguez, E.

- Roth, L. Seal, D. Shaffer, S. Shimada, J. Umland, J. Werner, M. Oskin, M. Burbank, D. & Alsdorf, D. (2007). *The Shuttle Radar Topography Mission*. Reviews of Geophysics. Volume 45, Issue 2. <https://doi.org/10.1029/2005RG000183>
- Ferrelli, F. Bustos, M. Piccolo, M. Huamantincio, M. & Perillo, G. (2016). Downscaling de variables climáticas a partir del reanálisis NCEP/NCAR en el sudoeste de la provincia de Buenos Aires (Argentina). *Papeles de Geografía*, 2016, pp. 62. <https://doi.org/10.6018/geografia/2016/239051>
- Instituto Argentino de Investigaciones de Zonas Áridas IADIZA. (2019). Red Ambiental IADIZA. Retrieved from https://www.mendoza-conicet.gob.ar/ladyot/red_iadiza/index.htm
- Instituto Nacional de Tecnología Agropecuaria INTA. (2019). Sistema de información y gestión agrometeorológica – SIGA. Retrieved from <http://siga2.inta.gov.ar/#/>
- Lui, A. (2019). “Meteoro”, el científico tuitero que informa sobre el tiempo. *Diario El Sol*, Argentina. Retrieved from <https://www.elsol.com.ar/meteoro-el-cientifico-tuitero-que-informa-sobre-el-tiempo>
- Matsuura, K. & Willmot, C. (2018). Terrestrial Air Temperature: 1900–2017 Gridded Monthly Time Series. Version 5.01. Retrieved from http://climate.geog.udel.edu/~climate/html_pages/Global2017_README_GlobalTsT2017.html
- Mo, R. Ye, C. & Whitfield, P. (2013). Some Similarity Indices with Potential Meteorological Applications. National Laboratory for Coastal and Mountain Meteorology. Technical Report, No. 2013-002. Retrieved from http://www3.telus.net/ruping/lib/reports/2013-002-CMML-TR_Mo_etal.pdf
- Mo, R. Ye, C. & Whitfield, P. (2014). Application Potential of Four Nontraditional Similarity Metrics in Hydrometeorology. *Journal of Hydrometeorology*, 15(5), 1862–1880. <https://doi.org/10.1175/JHM-D-13-0140.1>
- Nashwan, M. Shahid, S. & Chung, E. (2019). Development of high-resolution daily gridded temperature datasets for the central north region of Egypt. *Scientific Data*, 6(1), 1–13. <https://doi.org/10.1038/s41597-019-0144-0>
- National Center for Atmospheric Research Staff NCARS. (2019). The Climate Data Guide: GHCN (Global Historical Climatology Network) Related Gridded Products. Retrieved from <https://climatedataguide.ucar.edu/climate-data/ghcn-global-historical-climatology-network-related-gridded-products>
- National Oceanic and Atmospheric Administration NOAA. (2019). Earth System Research Laboratory. Physical Sciences Division. Gridded Climate Data-sets. Retrieved from <https://www.esrl.noaa.gov/psd/data/gridded/index.html>
- Qiu, M. Liu, B. Liu, Y. Zhang, Y. & Han, S. (2019). Evaluating the accuracy of a gridded near-surface temperature dataset over Mainland China. *Atmosphere*, 10(5). <https://doi.org/10.3390/atmos10050250>
- Rivera, J. Marianetti, G. & Hinrichs, S. (2018). Validation of CHIRPS precipitation dataset along the Central Andes of Argentina. *Atmospheric Research*, 213(Jan), 437–449. <https://doi.org/10.1016/j.atmosres.2018.06.023>
- Rusticucci, M. Zazulie, N. & Raga, G. (2014). Regional winter climate of the southern central Andes: Assessing the performance of ERA-Interim for climate studies. *Journal of Geophysical Research: Atmospheres*, 119(3), 8568–8582. <https://doi.org/10.1002/2013JD021167>
- Santos, J. (2014). En la Argentina hay una deficiencia de meteorólogos grandísima. Artículo periodístico en línea. Retrieved from <http://www.universidad.com.ar/en-la-argentina-hay-una-deficiencia-de-meteorologos-grandisima>
- Secretaría de Infraestructura y Política Hídrica de la Nación SIPHN. (2019). Base de datos hidrológica integrada. Retrieved from <http://bdhi.hidricosar-argentina.gob.ar/>
- Servicio Meteorológico Nacional SMN. (2019). Estadísticas climáticas normales. Retrieved from <http://www.smn.gov.ar/serviciosclimaticos>
- Shelton, M. (2008). *Hydroclimatology: Perspectives and Applications*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511754746>
- Takido, K. Saavedra, O. Ryo, M. Tanuma, K. Ushio, T. & Kubota, T. (2016). Spatiotemporal evaluation of the gauge-adjusted global satellite mapping of precipitation at the basin scale. *Journal of the Meteorological Society of Japan*, 94(2), 185–195. <https://doi.org/10.2151/jmsj.2016-010>
- University of East Anglia Climatic Research Unit. Harris, I. Jones, P. Osborn, T. (2020). CRU TS4.04: Climatic Research Unit (CRU) Time-Series (TS) version 4.04 of high-resolution gridded data of month-by-month variation in climate (Jan. 1901– Dec. 2019). Centre for Environmental Data Analysis, date of citation. <https://catalogue.ceda.ac.uk/uuid/89e1e34ec3554dc98594a5732622bce9>
- University of East Anglia. (2019). High-resolution gridded datasets (and derived products). Retrieved from <https://crudata.uea.ac.uk/cru/data/hrg/>
- Wang, Z. & Bovik, A. (2002). A universal image quality index. *IEEE Signal Processing Letters*, 9(3), 81–84. <https://doi.org/10.1109/97.995823>
- World Meteorological Organization WMO. (2011). Guide to Climatological Practices. WMO-No 100.

- Retrieved from http://www.wmo.int/pages/prog/wcp/ccl/guide/documents/WMO_100_en.pdf
- World Meteorological Organization WMO. (2017). *WMO Guidelines on the Calculation of Climate Normals*. WMO (World Meteorology Organization), (1203), 2929. Retrieved from https://library.wmo.int/doc_num.php?explnum_id=4166.
- Wong, J. Razavi, S. Bonsal, B. Wheeler, H. & Asong, Z. (2017). Inter-comparison of daily precipitation products for large-scale hydro-climatic applications over Canada. *Hydrology and Earth System Sciences*, 21(4), 2163–2185. <https://doi.org/10.5194/hess-21-2163-2017>
- Zhao, C. Wang, Z. Li, H. Wu, X. Qiao, S. & Sun, J. (2019). A new approach for medical image enhancement based on luminance-level modulation and gradient modulation. *Biomedical Signal Processing and Control*, 48, 189–196. <https://doi.org/10.1016/j.bspc.2018.10.008>