Accurate classification of bioclimatic data: spatial analysis

Giuseppe Passarella, Emanuele Barca, Delia Bruno, Sabino Maggi and Rita Masciale
CNR-IRSA Water Research Institute,
Via F. De Blasio 5, 70132 Bari, Italy
Email: sabino.maggi@cnr.it

Aimé Lay-Ekuakille
Department of Innovation Engineering,
Università del Salento,
Via Monteroni, 73100 Lecce, Italy

Abstract— This paper presents a general methodology for processing bioclimatic data in the spatial domain whose main goal is to derive indications related to the moisture/dryness level of a region and provide water management authorities with information about its irrigation requirements. The methodology uses point-scale measurements of weather related data to perform a detailed analysis of the spatial behavior of the corresponding bioclimatic indicators at the continuous regional scale. The proposed methodology, although more demanding in terms of computation resources, gives more accurate results than standard approximate approaches available in current GIS packages. This methodology has been applied to a particular case study using the well known De Martonne index as a bioclimatic indicator.

Index Terms—Bioclimatic indices, time-series, kriging, spatial analysis.

I. Introduction

The bioclimatic characterization of a region is usually based on some well established and easy-to-calculate bioclimatic indices that synthesize the moisture/dryness level of the region. Bioclimatic indices provide precious information to water delivery and management authorities about the irrigation requirements of the area under consideration. Bioclimatic indices also allow to estimate the amount of recharge from infiltration occurring in the area [1]–[3], an important indicator to prevent the lowering of the groundwater level and the resulting salinization of coastal aquifers [4].

Bioclimatic indices are very useful tools also for land management, where they may contribute to assess the degree of hydrologic stress of a given area [5], [6]. Such assessment is particularly important in Mediterranean regions, where climate scenarios predict a trend towards a warmer and more arid climate, with the expansion of the arid and semi-arid areas.

Bioclimatic indices are calculated from weather related measurements taken at local meteorological stations spread more or less randomly across a given area. For the reasons given given above, it is important to be able to extend these point-scale data over the whole region, using some interpolation method that transforms the irregularly located measurements into continuous (or almost continuous) maps.

Geostatistics is widely applied in several environmental fields [7]–[11]), because of its robust theoretical background and the wide availability of software applications and tools.

Approximate interpolation methods of the point-scale data over an extended region, such as the Thiessen polygon or inverse distance methods, are available in current GIS packages. These methods are fast and very easy to use, but often they cannot produce accurate maps of the bioclimatic indexes, unless the number of local measurement stations is exceedingly high. Also, approximate interpolation methods do not provide any information about the uncertainty associated with the interpolated values.

Under the assumption that the bioclimatic indices are spatially auto-correlated random variables [12], it is possible to estimate their spatial behavior over the whole area using standard geostatistical estimators, such as ordinary kriging.

In this paper we present a general methodology based on this estimator for processing bioclimatic data in the spatial domain. Starting from time-series of weather related variables recorded at local meteorological stations, such as the average temperature and the total annual precipitation, the methodology proceeds through the validation of the observed data [13], [14], the choice of the "most representative" bioclimatic index for the environment under consideration [15], and its spatial interpolation over the whole study area, to precisely define the climatic zones into which the region is divided and to infer proper indications of their possible evolution and trend over time, as shown in a companion paper that analyses the same bioclimatic data in the temporal domain [16].

Kriging is a complex and computationally expensive task, in particular when it is necessary to interpolate large time-series, since it involves the calculation of a variogram for each measured variable at each time step [17]. As in this paper we are interested only in the spatial behavior of the bioclimatic indices, to speed up calculations we have removed the temporal dependence of the bioclimatic indices by summarizing the entire time-series data at each monitoring point with its median value over the whole observation time.

The proposed methodology has been applied to a case study involving the Apulia region in Italy, based on thermo-pluviometric data registered at 82 stations of the weather network managed by the Regional Civil Protection. A validation

978-92-990075-8-7

TABLE I
CLIMATIC ZONES CLASSIFIED ACCORDING TO THE DE MARTONNE INDEX.

Classification	Index range	Description
Hyper-arid	$I_{DM} < 5$	Desertic
Arid	$I_{DM} < 15$	Needs continuous irrigation
Semi-arid	$I_{DM} < 20$	Needs irrigation
Dry subhumid	$I_{DM} < 30$	Needs supplementary irrigation
Humid	$I_{DM} \leq 60$	Needs occasional irrigation
Perhumid	$I_{DM} > 60$	Water self-sufficiency

of the methodology has been performed by comparing its results with those obtained with the Thiessen polygon method. The detailed maps defining the climatic zones of Apulia and their evolution in time [16], could be useful to study the effects of global warming in a typical Mediterranean environment.

II. BIOCLIMATIC INDICES

The dryness of a given zone is classified according to so-called bioclimatic indices (or aridity indices), which correlate total precipitations and average temperatures, often used as a proxy for evapotranspiration data. The values of these two weather related variables are recorded at least on a daily basis and are aggregated with respect to the same time step, chosen according to the available data and the goals of the analysis (monthly, each season, yearly).

Among several bioclimatic indices available in the scientific literature, the De Martonne [18] index has been selected to illustrate the methodology used here. The annual De Martonne index I_{DM} is defined as

$$I_{DM} = \frac{P_a}{T_a + 10}, \quad (1)$$

where P_a (mm) is the annual amount of precipitation and T_a (°C) is the mean annual air temperature.

The De Martonne index allows to specify the various degrees of moisture and dryness and to classify the extreme environmental or weather conditions experienced by crops present in a given zone. Therefore, this index is of great importance not only from the environmental or climatological point of view but also for land management and agriculture. The association between the values of the De Martonne index and the degree of moisture/dryness of an area can be performed according to the classification scheme of Table I.

III. SPATIAL ANALYSIS

A. Missing data

The analysis of spatio-temporal data series must usually face the missing data problem. Missed observations may occur due to instrumentation or human error, to problems when transmitting data from the measurement stations to the central data collection server, or even to database mismatches due to the different data formats used in the measurement stations.

Missing data can cause serious problems. Most statistical procedures automatically remove missing data and may give misleading results if the amount of remaining data is not sufficient to perform the analysis.

Hence the need to rely on reliable methods to rebuild the time-series data and fill the data gaps. The reconstruction of missing data is a complex and delicate process that, if applied lightly, can introduce further distortions in the data series [13], [14], [19]. Any method for estimating missing data should: (1) be able to make reliable assumptions about the underlying causes of missing data; (2) leave internal data relations unchanged (e.g. mean, standard deviation, etc.); (3) leave the overall uncertainty associated with the data unchanged.

In this work, missing data have been reconstructed using the Multiple Imputations Chained Equations (MICE) iterative method, one of the most advanced and effective methods to handle missing data [20], [21].

B. Interpolation

Once the missing data have been fixed and the data matrices containing the time-series of measurements are completed, it is possible to calculate the values of the selected bioclimatic indices at each site at the desired time aggregation level. By definition, the bioclimatic indices reflect a localized situation and taken individually do not give information about processes occurring at a larger scale. In order to gain information about these processes, it is essential to transform the point-scale data to a regional scale, by making use of proper geostatistical tools.

The interpolation of point-scale observations over the whole study area is usually done by superimposing a grid mesh and using an appropriate method to predict the values of each measured variable (such as precipitation and temperature here) at each node of the grid. A map algebra process combines each grid into a new grid containing the values of the desired bioclimatic index. The last step is to classify the values of the index at each grid point according to Table I, producing a spatial map of the categorized bioclimatic index over the area under consideration.

If the method used to interpolate the original measurements is too coarse, the derived map of the bioclimatic index will be affected by an even higher approximation error, making it unable to predict bioclimatic data at a wider regional scale.

In this paper, we have used an alternative approach based on the consideration that also the bioclimatic indices can be considered spatially auto-correlated variables, since they are simple combinations of spatial variables. The assumption can be tested by means of the so-called Moran I index [22], a well-known statistical test whose theoretical values range from $+1$ for positive spatial autocorrelation, to -1 when values at neighboring points are more dissimilar than what can be expected by random chance (negative spatial autocorrelation).

This approach allows to first calculate the bioclimatic index at each measurement site and then to apply the desired interpolation method only on this spatial variable, reducing the total computation time by a factor approximately equal to the number of weather related variables used for the calculation of each bioclimatic index.

Two complementary methods for the calculation of the spatial distribution of the point-scale index values over the whole study area have been used here. The first method is based on Thiessen polygons and assigns the point-value of each observation to the whole polygon built up around the measurement station, while the second method, ordinary kriging, uses the spatial auto-correlation of the observations, evaluated through the empirical variogram, to estimate the values of the spatial variable over the whole study area [12], [23], [24].

The two methods differ for their basic assumptions, the complexity of calculation and the significance of results. The method based on Thiessen polygons is a very simple and fast approximate method and, for this reason, it is implemented in the main Geographical Information Systems (GIS). Ordinary kriging requires more restrictive hypotheses on the statistical distribution and spatial continuity of the data, is much more demanding from a computational point of view, yet provides more reliable results and produces an accurate spatial distribution pattern of the bioclimatic indices over the study area.

C. Thiessen polygons

The most straightforward interpolation method is the Thiessen (or Voronoj) polygon methods, that defines individual areas of influence around each point of a discrete data set.

The generation of Thiessen polygons is based on the following algorithm: (1) the study area is divided into triangles (Delauney triangles) by connecting the location of each data point to two of its nearest neighbors so that each triangle has the smallest possible area (or the largest possible angles); (2) each triangle is circumscribed by a circle whose center is the intersection of the medians to each side of the circle; (3) the Thiessen polygons are built by connecting all the centers of the circles. A characteristic property of Thiessen polygons is that each polygon contains one and only one measurement site. Within the polygon the value of the variable remains unchanged.

More formally, if $z(\mathbf{x}_\alpha),\ \alpha = 1, 2, ..., N$ is the set of spatial observations at the N sites $\mathbf{x}_\alpha$, every grid node $\mathbf{x}_j$ located within the Thiessen polygon containing $\mathbf{x}_\alpha$ satisfies the relation $|\mathbf{x}_j - \mathbf{x}_\alpha| < |\mathbf{x}_j - \mathbf{x}_{\alpha'}|$, in other words it is closer to $\mathbf{x}_\alpha$ than to any other site $\mathbf{x}_{\alpha'}$ located outside the polygon. At these nodes, the value of the spatial observation is set to

$$\hat{z}(\mathbf{x}_j) = z(\mathbf{x}_\alpha), \quad j = 1, 2, ...W, \tag{2}$$

where $z(\mathbf{x}_\alpha)$ is the value measured at the network location $\mathbf{x}_\alpha$ contained within the polygon.

D. Ordinary kriging

Kriging is a geostatistical estimator that models a dependent variable by considering the structure of its spatial variability, and uses the empirical variogram as a measure of the dissimilarity of measurements taken at the same distance [25]. The empirical semi-variogram $\gamma(\mathbf{h})$ is defined as the average of the squared sum difference between all spatial observations $z(\mathbf{x}_\alpha)$ separated by the same lag distance $\mathbf{h}$,

$$\gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} \left[z(\mathbf{x}_\alpha) - z(\mathbf{x}_\alpha + \mathbf{h})\right]^2, \tag{3}$$

where $N(\mathbf{h})$ is the number of pairs of data locations that are separated by $\mathbf{h}$ (plus or minus a predefined tolerance). The variogram is a function of both distance and direction and can thus account for any spatial anisotropy in the measured data. Actual variogram calculations are performed by pooling data pairs into separate lag bins. Larger lag bins allow more data pairs for estimation but reduce the amount of detail of the variogram.

Fitting a curve to an empirical variogram allows to define a calculated variogram model that can be used to model the spatial variation of the data. The actual functional form of the variogram model depends on the data series and on the lag tolerance. Ideally the variogram starts from a a non-zero value (nugget) at zero lag and increases with $\mathbf{h}$, reaching a constant value (sill) at a well defined lag distance (range), where the sill approximates the variance s^2 of the whole data set. Beyond the range, spatial autocorrelation among the data is essentially zero.

Ordinary kriging is a generalized least-squares regression technique that predicts the values of a variable $\hat{z}(\mathbf{x}_j)$ at each node $\mathbf{x}_j$ of a grid covering the area under consideration from a weighted linear combination of n neighboring observations $z(\mathbf{x}_\alpha)$. The basic expression of ordinary kriging is

$$\hat{z}(\mathbf{x}_j) = \sum_{i=1}^{n} \omega_\alpha(\mathbf{x}_j) z(\mathbf{x}_\alpha), \quad j = 1, 2, ...W, \tag{4}$$

where W is the number of grid nodes and $\omega_\alpha(\mathbf{x}_j)$ is the weight assigned to each observation, so that $\sum_{i=1}^{n} \omega_\alpha(\mathbf{x}_j) = 1$. The weights $\omega_\alpha(\mathbf{x}_j)$ are determined from the calculated variogram of the data set, so that observations closer to the node $\mathbf{x}_j$ are weighted more than those located farther away.

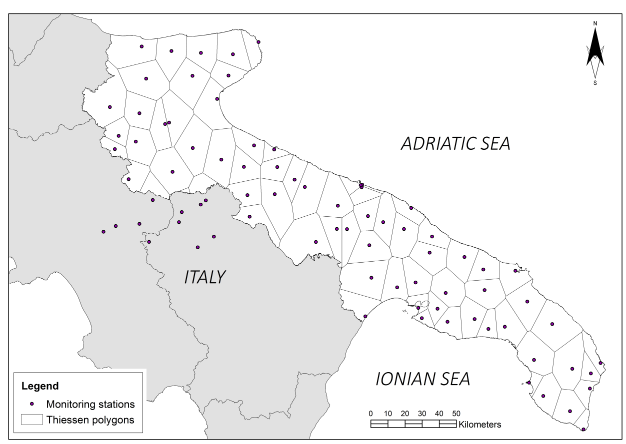

Fig. 1. Map of the Apulia region with the 82 weather monitoring stations used for this case study and the Thiessen polygons built around each site.

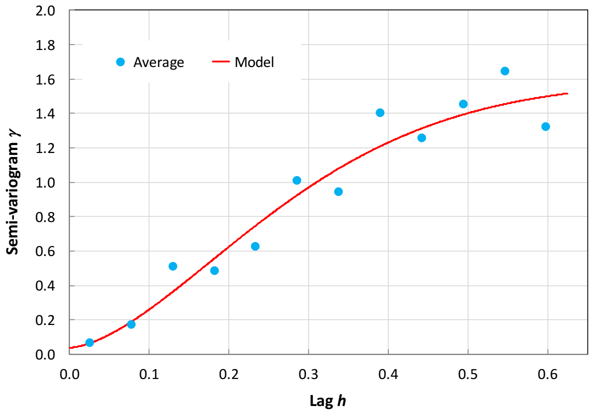

Fig. 2. Variogram of the median of the De Martonne index in the 1931–2010 time frame: (circles) empirical variogram, (solid line) best variogram model.

IV. Case study

The annual De Martonne index (1) has been calculated for the 82 weather monitoring stations of the Civil Protection of Apulia, Italy where precipitation and temperature data from 1931 to 2010 are available. The monitoring stations are spread more or less uniformly across the area of the region, with a density of around 1 station per 250 km^2 (Fig. 1). The Figure also shows the Thiessen polygons built around each measurement site. Some monitoring stations located outside the area of the region are shown in the Figure but are not considered in this case study.

Missing data have been reconstructed with the MICE method, using the `mice` package implemented in the R environment [26], [27].

Since we are interested here only in the spatial behavior of the De Martonne index, the annual values of the index have been aggregated by computing, at each monitoring point, the median of the De Martonne index over the whole considered temporal interval. The choice of the median instead of other statistical indices is based on continuous nature of the data and on the consideration that the median is a more stable measure of centrality than the mean since it is less sensitive to the presence of outliers in the data.

The degree of spatial autocorrelation of the median values has been evaluated with the Moran index, obtaining for the aggregated data set an index $I = 0.1471$ that suggests the presence of a positive autocorrelation and, without considering the sign, is about 10 times larger than the expected value of -0.0125 (slight negative spatial autocorrelation). The p-value of 1.92×10^{-12} is also several orders of magnitude less than the given $P = 0.05$, confirming the assumption that, at least in the 1931–2010 time frame considered here, the median values of the De Martonne index are spatially autocorrelated over the study area.

Once the assumption of spatial autocorrelation has been verified, it has been possible to calculate the empirical variogram of the data and to determine from it the best variogram model that best fits the available data (Figure 2). Next, the median values of the De Martonne index have been interpolated over the whole study area, using the two spatialization methods described above. Figure 2 shows the empirical variogram (circles) and the best calculated variogram model (solid line) of the median values of the De Martonne index for the 82 monitoring stations considered in this case study.

The corresponding variogram model has been determined by iteratively changing the model function and its characteristic parameters until convergence is reached. The best calculated variogram model equation is

$$2\gamma(h) = 0.0379 \text{ nugget} + 1.556 \text{ stable}(0.625, 1.622),$$

where the $\text{stable}(\cdot)$ function is one of the standard variogram model functions used in geostatistical analysis [28].

The last step of the analysis is the interpolation of point-scale median index values over the whole study area. Figures 3 and 4 show the regional maps of the De Martonne median index for the period 1931–2010, calculated with the Thiessen polygon and ordinary kriging methods, respectively.

The comparison of the two maps clearly shows that, although the results of the two methods are in overall agreement, the map calculated with the approximate Thiessen polygon method is patchy and quite unrealistic, and the abrupt change of the index values at the border of each polygon makes it impossible to evaluate the bioclimatic characteristics of the region at a spatial scale better than the resolution of the original point measurements. On the other hand, the map calculated by ordinary kriging is much more detailed than the crude polygon approximation and naturally incorporate the spatial continuity of the observations, allowing to define the bioclimatic characteristics of the different areas of the region in a much more accurate way.

These maps can be used to define the bioclimatic zones of the region, following the classification scheme of Table I. The area of each zone, calculated with the two interpolation methods used in this work, is listed in Table II. Both methods

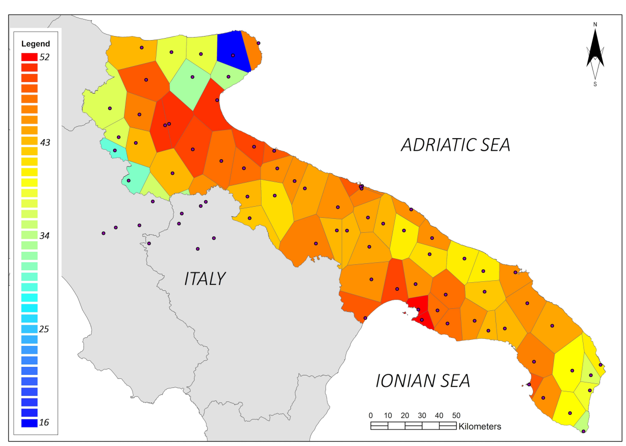

Fig. 3. Median of the De Martonne index interpolated over the study area using the Thiessen polygon method.

TABLE II
AREA OF THE CLIMATE ZONES OF APULIA, CALCULATED WITH THE THIESSEN POLYGON AND ORDINARY KRIGING METHODS.

Classification	Index range	Thiessen polygons		Ordinary kriging	
		Area (km^2)	Area (%)	Area (km^2)	Area (%)
Hyper-arid	$I_{DM} < 5$	0	0	0	0
Arid	$I_{DM} < 15$	2179	11	2956	15
Semi-arid	$I_{DM} < 20$	14834	77	13384	69
Dry subhumid	$I_{DM} < 30$	2345	12	3018	16
Humid	$I_{DM} \leq 60$	0	0	0	0
Perhumid	$I_{DM} >, 60$	0	0	0	0

show that the largest part of the region is characterized by a Mediterranean semi-arid climate, while the remaining parts of the region are equally divided into either arid or dry subhumid zones. The extreme climate conditions, hyper-arid and humid/perhumid are both totally absent in Apulia. From the Table it is also apparent that the approximate Thiessen method overestimates the area of the semi-arid climate zone, while the kriging method allows a more accurate definition of the climate zones into which in region is divided.

V. CONCLUSION

This paper presents a general methodology to accurately characterize a regional territory from a bioclimatic point of view. Starting from time-series measurements of temperature and precipitation, the bioclimatic De Martonne index is calculated for each monitoring station. The point-scale data is interpolated it to a continuous regional scale with ordinary kriging. Missing data in the measured time-series are automatically reconstructed using the well-known MICE method.

The proposed methodology, although more complex and demanding in terms of computation time, is theoretically rigorous and gives more accurate results than standard approximate methods, such as those based on Thiessen polygons.

The methodology has been applied to a case study involving the whole Apulia region, in Italy. The analysis has been carried out on time-series of thermo-pluviometric data measured at 82 stations of the weather monitoring network managed by the Regional Civil Protection from 1931 to 2010, using the De Martonne index as a bioclimatic indicator. Detailed maps and tables of the De Martonne index of Apulia have been calculated, defining the climatic zones into which the region is divided.

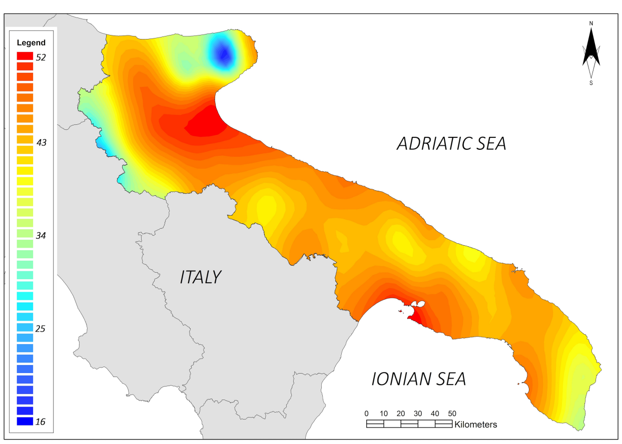

Fig. 4. Median of the De Martonne index interpolated over the study area using the ordinary kriging method.

Together with a companion paper focused on a statistical methods to evaluate the temporal evolution of the bioclimatic indicator, the spatial and temporal analysis can provide water management authorities detailed information concerning the level of aridity of the region and its irrigation needs, at the sub-regional and local scale. Crossing this information with related regional information, such land use or water availability, it is be used to find the best compromise between water requirements and availability, avoiding potential imbalances and optimizing distribution costs.

The detailed maps defining the climatic zones of Apulia and their evolution in time [16], could also be useful tools to study the effects of global warming in a typical Mediterranean region.

ACKNOWLEDGMENT

The data used in the case study has been collected within the European Regional Development Fund POR-FESR Programme 2007-2013, “Analysis of desertification processes in Apulia: causes, effects, mitigation and drought prevention”, carried out in collaboration with the Department of Agricultural and Environmental Science of the Università di Bari and the Department of Civil Engineering of the Politecnico di Bari.

REFERENCES

[1] D. Zaccaria, N. Lamaddalena, C. M. Neale, G. P. Merkley, N. Palmisano, and G. Passarella, “Simulation of peak-demand hydrographs in pressurized irrigation delivery systems using a deterministic–stochastic combined model. part i: model development,” *Irrigation Science*, vol. 31, no. 3, pp. 209–224, 2013. [Online]. Available: http://dx.doi.org/10.1007/s00271-011-0317-x

[2] E. Barca, M. C. Calzolari, G. Passarella, and F. Ungaro, “Predicting shallow water table depth at regional scale: optimizing monitoring network in space and time,” *Water Resources Management*, vol. 27, no. 15, pp. 5171–5190, 2013. [Online]. Available: https://link.springer.com/article/10.1007/s11269-013-0461-6

[3] G. Passarella, E. Barca, D. Sollitto, R. Masciale, and D. E. Bruno, “Cross-calibration of two independent groundwater balance models and evaluation of unknown terms: Th e case of the shallow aquifer of “tavoliere di puglia” (south italy),” *Water Resources Management*, vol. 31, no. 1, pp. 327–340, 2017. [Online]. Available: http://dx.doi.org/10.1007/s11269-016-1527-z

[4] D. Zaccaria, G. Passarella, D. D'Agostino, R. Giordano, and S. S. Solis, "Risk assessment of aquifer salinization in a large-scale coastal irrigation scheme, italy," *CLEAN Soil, Air, Water*, vol. 44, no. 4, pp. 371–382, 2016. [Online]. Available: http://dx.doi.org/10.1002/clen.201400396

[5] F. Attorre, M. Alfò, M. D. Sanctis, F. Francesconi, and F. Bruno, "Comparison of interpolation methods for mapping climatic and bioclimatic variables at regional scale," *International Journal of Climatology*, vol. 127, pp. 1825–1843, 2007. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1002/joc.1495/full

[6] C. Paltineanu, N. Tanasescu, E. Chitu, and I. F. Mihailescu, "Relationships between the de martonne aridity index and water requirements of some representative crops: A case study from romania," *International Agrophysics*, vol. 21, no. 1, pp. 81–93, 2007.

[7] N. Ungaro, G. Marano, A. Vlora, and G. Passarella, "Distribution patterns of two species of swimming crabs (portunidae), liocarcinus depurator and macropipus tuberculatus, in the southwestern adriatic sea (mediterranean sea)," *Crustacean Issues*, vol. 12, pp. 131–138, 2000.

[8] G. Passarella, M. Vurro, V. D'Agostino, G. Giuliano, and M. J. Barcelona, "A probabilistic methodology to assess the risk of groundwater quality degradation," *Environmental Monitoring and Assessment*, vol. 79, no. 1, pp. 57–74, 2002. [Online]. Available: http://dx.doi.org/10.1023/A:1020033808025

[9] G. Passarella and M. C. Caputo, "A methodology for space-time classification of groundwater quality," *Environmental Monitoring and Assessment*, vol. 115, no. 1, pp. 95–117, 2006. [Online]. Available: http://dx.doi.org/10.1007/s10661-006-6547-3

[10] E. Barca and G. Passarella, "Spatial evaluation of the risk of groundwater quality degradation. a comparison between disjunctive kriging and geostatistical simulation," *Environmental Monitoring and Assessment*, vol. 137, no. 1, pp. 261–273, 2008. [Online]. Available: https://link.springer.com/article/10.1007/s10661-007-9758-3

[11] E. Barca, D. E. Bruno, and G. Passarella, "Similarity indices of meteo-climatic gauging stations: definition and comparison," *Environmental Monitoring and Assessment*, vol. 188, no. 7, pp. 403/1–12, 2016. [Online]. Available: https://link.springer.com/article/10.1007/s10661-016-5407-z

[12] G. Matheron, *The theory of regionalised variables and its applications*. Les Cahiers du Centre de Morphologie Math ematique de Fontainebleau, 1971, vol. 5.

[13] R. L. Presti, E. Barca, and G. Passarella, "A methodology for treating missing data applied to daily rainfall data in the candelaro river basin (italy)," *Environmental Monitoring and Assessment,*, vol. 160, no. 1, pp. 1–22, 2010. [Online]. Available: https://link.springer.com/article/10.1007%2Fs10661-008-0653-3

[14] E. Barca, D. E. Bruno, and G. Passarella, "Optimal redesign of environmental monitoring networks by using software msanos," *Environmental Earth Sciences*, vol. 75, no. 14, pp. 1082/1–14, 2016. [Online]. Available: https://link.springer.com/article/10.1007/s12665-016-5885-x

[15] E. Baltas, "Spatial distribution of climatic indices in northern greece," *Meteorological Applications*, vol. 14, pp. 69–78, 2007. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1002/met.7/abstract

[16] G. Passarella, E. Barca, D. E. Bruno, S. Maggi, and A. Lay-Ekuakille, "Accurate classification of bioclimatic data: temporal analysis," *these Proceedings*, 2017.

[17] E. Barca, E. Porcu, D. Bruno, and G. Passarella, "An automated decision support system for aided assessment of variogram models," *Environmental Modelling & Software*, vol. 87, pp. 72–83, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364815216309021

[18] E. de Martonne, *Trait de Gographie Physique*. Librarie Arman Colin, Paris, 1925-1927.

[19] E. Barca, L. Berardi, D. B. Laucelli, G. Passarella, and O. Giustolisi, "Evolutionary polynomial regression application for missing data handling in meteo-climatic gauging stations," in *Proceedings of GRASPA 2015*, 2015.

[20] D. B. Rubin, "Multiple imputation after 18+ years," *Journal of the American Statistical Association*, vol. 91, no. 434, pp. 473–489, 1996. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/01621459.1996.10476908

[21] F. Scheuren, "Multiple imputation: How it began and continues," *The American Statistician*, vol. 59, no. 4, pp. 315–319, 2005. [Online]. Available: https://www.jstor.org/stable/27643702

[22] P. A. P. Moran, "Notes on continuous stochastic phenomena," *Biometrika*, vol. 37, no. 1/2, pp. 17–23, 1950. [Online]. Available: http://www.jstor.org/stable/2332142

[23] A. G. Journel and C. J. Huijbrechts, *Mining Geostatistics*. London, UK: Academic Press, 1978.

[24] E. H. Isaaks and R. M. Srivastava, *An Introduction to Applied Geostatistics*. New York, NY: Oxford University Press, 1989.

[25] J.-P. Chiles and P. Delfiner, *Geostatistics, Modeling Spatial Uncertainty*. Wiley, 1999. [Online]. Available: http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470183152.html

[26] S. van Buuren and K. Groothuis-Oudshoorn, "mice: Multivariate imputation by chained equations in r," *Journal of Statistical Software*, vol. 45, no. 3, pp. 1–67, 2011. [Online]. Available: https://www.jstatsoft.org/v045/i03

[27] N. F. A. Radi, R. Zakaria, and M. A. Azman, "Estimation of missing rainfall data using spatial interpolation and imputation methods," *AIP Conference Proceedings*, vol. 1643, no. 1, pp. 42–48, 2015. [Online]. Available: http://aip.scitation.org/doi/abs/10.1063/1.4907423

[28] K. Johnston, J. M. V. Hoef, K. Krivoruchko, and N. Lucas, *ArcGIS 9: Using ArcGIS Geostatistical Analyst*. ESRI, USA, 2003. [Online]. Available: http://dusk2.geo.orst.edu/gis/geostat_analyst.pdf