

**MAPPING PROBABILITIES OF PRECIPITATION OCCURRENCE ON THE
TERRITORY OF THE REPUBLIC OF SERBIA BY THE METHOD OF
INDICATOR KRIGING**

JELENA PANDŽIĆ¹, BRANISLAV BAJAT¹, JELENA LUKOVIĆ²

¹*University of Belgrade – Faculty of Civil Engineering, Department of Geodesy and Geoinformatics, Bulevar kralja Aleksandra 73, 11000 Belgrade*

²*University of Belgrade – Faculty of Geography, Studentski trg 3/3, 11000 Belgrade*

Abstract: This paper presents the application of indicator kriging as a geostatistical method for the purpose of creating maps of precipitation occurrence probabilities on the territory of the Republic of Serbia for distinctive months during 2009. The difference between this approach to mapping and standard isohyetal maps, which describe precipitation intensity, lies in the fact that this approach points to the potential of the occurrence of a certain amount of precipitation at a specific location for a given time period.

Key words: indikator kriging, precipitation, probability mapping

Introduction

Spatial variations of a continuous feature, such as the concentration of particular chemical elements in soil and precipitation amounts, are phenomena that cannot be described by simple mathematical functions. Instead, such variations are described by stochastic surfaces that are the essence of the application of geostatistical interpolation methods (Burrough and McDonnell, 2006).

The purpose of geostatistics is to provide a prediction of a possible spatial distribution of a certain variable. The prediction usually exists in the form of a created map or series of maps. It occurs in two basic forms: (1) as an estimate or (2) as a simulation (Zhang, 2011). The result of a kriging application is the best linear unbiased estimate (BLUE) of a variable value at a particular location. This estimate is based on the available data sample and the resulting variogram model that depicts the spatial correlation of the data most faithfully. Unlike the estimate, simulations provide a huge number (usually hundreds) of maps of the equally probable variable distribution. The collection of simulated maps allows researchers to determine the uncertainty of a certain variable estimate at a particular location (Pejović et al., 2012).

Indicator kriging is a viable geostatistical interpolation method for describing spatial variations of precipitation as one of the most important climate elements (Bajat et al.,

¹ jelenapandzic@gmail.com

This work was supported by the Ministry of Education, Science and Technological Development of Serbia (Contracts No. III 47014, TR 36009 and III 43007)

2013). The aim of this paper is to carry out a prediction of the probabilities of the occurrence of a certain amount of precipitation on the territory of the Republic of Serbia in distinctive months based on data from the year 2009 by the means of the indicator kriging method. The prediction would ultimately lead to verification (or refutation) of current knowledge on the spatial distribution of rainfall on the aforementioned territory. It is also possible to generate probability maps using the method of Monte Carlo simulation for the purpose of assessing the occurrence probability of a variable (Joksić and Bajat, 2005).

The territory of the Republic of Serbia is characterized by two precipitation regimes: continental and Mediterranean. The greater part of Serbia belongs to the continental regime, which means the greatest amount of precipitation occurs in May and June and the least occurs in February and October. On the other hand, the southwestern part of Serbia is characterized by the Mediterranean regime and experiences a rainier period in November, December and January and drier period in August (RHMS, 2013). Within the paper, the prediction of probabilities of the occurrence of a certain amount of precipitation in February, June, August and October 2009 was carried out for the territory of the Republic of Serbia by using indicator kriging.

Methodology

The mathematical description of the spatial variation of any variable, which is the essence of kriging applications, is performed using the sum of three main components (Burrough and McDonnell, 2006):

$$Z(x) = m(x) + \varepsilon'(x) + \varepsilon'' \quad (1)$$

with:

$Z(x)$ - being the value of a random function,

$m(x)$ - being a deterministic function that describes the so-called structural component, i.e. trend,

$\varepsilon'(x)$ - being a stochastic (random) component that is spatially correlated and represents the remainder of the structural component, also known as a regionalised variable,

ε'' - being the residual error, i.e. spatially uncorrelated noise.

Considering different approaches used for treating some of the components of spatial variation of the variable, especially the trend, as well as the fact whether additional variables are used as predictors, one can distinguish between kriging variants: simple, ordinary, universal, regression, indicator kriging, cokriging, etc. All of the aforementioned kriging types, except for indicator kriging, give the estimated values of the variable of interest, i.e. of the spatial attribute at a certain location as a result. In contrast, indicator kriging doesn't give the information on a particular estimated variable value at a certain location; however, it does provide information on the probability of whether a value at a certain location exceeds or not a set limiting value that was input in advance, i.e. threshold.

Indicator kriging

Ever since Journel (1983) introduced indicator kriging into the area of geostatistics, many scientists, experts, enthusiasts have been working on developing the algorithm of indicator kriging and methods derived from it. Although originally used for assessing mineral resources, as well as for the purpose of mapping many other natural resources, indicator kriging as a technique is today successfully used in the area of water quality

assessment (Tolosana-Delgado, 2007), precipitation mapping (Atkinson and Lloyd, 1998; Sun et al., 2003), and also for drawing conclusions about the prevalence of certain diseases, like schistosomiasis which is one of the most common parasitic diseases in humans (Guimarães et al., 2012).

Indicator kriging is in fact ordinary kriging applied to binarized data or so-called indicators. The essence of indicator kriging is transforming a selected (continuous) variable into a binary variable, whereby every original variable value is replaced with a value of 1 or 0 depending on whether the observed value is below or above a defined threshold (Isaaks and Srivastava, 1989). Mathematically, this nonlinear transformation can be represented by the formula:

$$i(x, z_k) = \begin{cases} 1, & z(x) \leq z_k \\ 0, & z(x) > z_k \end{cases} \quad (2)$$

with:

$z(x)$ - being a measured variable value at the point x ,

z_k - being a boundary value, i.e. threshold,

$i(x, z_k)$ - being a transformed variable value (indicator) at the point x , for the given threshold value z_k .

Indicator kriging is an especially efficient way to limit the effect of extremely big values or outliers on the results of prediction due to the fact that the variable values are assigned the same indicator as other values that are above the set threshold regardless of the absolute difference (Glacken and Blackney, 1998). The outcome of indicator kriging is a conditional cumulative distribution function of the observed variable, i.e. the estimate of the percentage of values in the neighbourhood of every observed point that (do not) exceed the value of the previously defined threshold. In other words, unlike the other forms of kriging, the results of indicator kriging aren't concrete values of the original variable (e.g. 5.2 mg, 106 mm, etc.) at the observed points as a consequence of performed transformations. Those results rather present the probabilities that the values of original variable at those points (do not) exceed the value of a previously defined threshold.

Variogram modelling

Transformation of the original variable values into indicators is just the first step of the prediction procedure using indicator kriging. What follows is variogram modelling. Namely, if one takes into account the effect of the structural component (trend), the remaining variations of the variable become homogeneous, i.e. differences of the variable values at different locations are mostly a function of the distances between those locations (Burrough and McDonnell, 2006). This fact allows the formation of the experimental variogram, which is in fact a graphical representation of the dependence of the semivariance of the measured variable value differences at different points on a distance. Semivariance is calculated as:

$$\hat{\gamma}(h) = \frac{1}{2n} \sum_{i=1}^n [z(x_i) - z(x_i + h)]^2 \quad (3)$$

with n being the number of point pairs at which the variable z was measured and which are at distance h . The experimental variogram obtained this way is modelled by a

mathematical function, i.e. a theoretical model that describes the spatial variation of the variable as faithfully as possible and as such is used in the prediction procedure is fitted to the experimental variogram.

The most commonly used variogram models are spherical, exponential, Gaussian and linear. The first three models are known as transitive variograms because their spatial correlation structure changes with distance. The linear model is a nontransitive variogram considering it doesn't have a sill inside the area in which the measurements exist (Burrough and McDonnell, 2006).

The situation in which the Gaussian model fits the experimental variogram better than any other variogram model indicates that the data is slightly variable, whilst the spherical model has a clear transitive point which indicates that the single variogram model is dominant. Choosing an exponential model suggests that variations show gradual transition inside a range or that the variogram model is nested, i.e. it represents a linear combination of several basic models (Burrough and McDonnell, 2006).

(An)isotropy of the variations should also be considered when modelling a variogram. In case the variogram is calculated by ignoring the direction in which point pairs are searched for, i.e. if the search area is a full circle, the resulting variogram becomes an isotropic variogram and represents a variogram that is averaged over all directions. However, spatial variations can change with direction of propagation, which is often the case for phenomena such as air pollution that show greater continuity in the direction in which winds blow predominantly than in the direction perpendicular to the one previously mentioned. In those cases, several variograms for different directions are calculated and the resulting variograms that have different sills and ranges are called anisotropic. Depending on whether the sill or range stays constant, one can differentiate between geometric and zonal anisotropy (Isaaks and Srivastava, 1989). Of course, it can often happen that both aspects of the anisotropy of spatial variations are present.

Errors in the procedure of prediction using indicator kriging

Bearing in mind that the result of indicator kriging is a conditional cumulative distribution function, i.e. estimated probabilities that the values of the original variable at the observed points (don't) exceed the value of previously defined threshold, all of the estimated values should be in the interval $[0,1]$. Also, estimated values obtained from indicator kriging conducted with several different boundary values, i.e. thresholds, should preserve the order of the thresholds themselves (e.g. if a greater threshold value is used in every consecutive prediction, the estimated values at every observed point should progressively increase as well). However, because of the errors that inevitably occur during prediction, "impossible" situations do happen: for example probabilities greater than 100% are reported and negative probabilities and illogicalities in probability order occur. All of the above are collectively referred to as order relation violations that require the introduction of certain corrections.

The easiest way to overcome enumerated illogicalities is to replace all negative probability values with zero and all values greater than 1 with 1. This would ensure that all estimated values are in the permitted interval $[0,1]$. As for preserving the order of estimated values obtained by predictions in which different thresholds were used (for the purpose of transforming the original variable values into indicators), it is usually ensured by simple averaging. This means that the values at the observed point obtained from predictions conducted with two adjacent thresholds, and in whose case the previously mentioned order condition wasn't met, are replaced with their simple arithmetic mean (Isaaks and Srivastava, 1989).

Examination of the variable distribution

Although indicator kriging doesn't require that the original variable values comply to the normal distribution due to the fact that normality dependence disappears with their binarization (transformation into indicators), it is sometimes useful to examine the available data sample before the prediction itself in order to gain insight into its characteristics. In that sense, one can examine a variable distribution and its (non)correlation with other variable, etc.

Histogram appearance (skewness, kurtosis, etc.) is the simplest way of examining a variable distribution. QQ-plots compare the quantiles of the distribution of the observed variable to the quantiles of the desired distribution (usually normal), whereby the coordinates of the points on the graphs are nothing but the corresponding quantiles. The points on the corresponding qq-plot must lie on one line in order to conclude that a variable is normally distributed.

Prediction of precipitation occurrence probabilities on the territory of the Republic of Serbia

Prediction of probabilities of the occurrence of a certain amount of precipitation on the territory of the Republic of Serbia for characteristic months in 2009 using the method of indicator kriging was performed within the paper. The prediction was conducted for February, June and October bearing in mind that most of Serbia has a continental precipitation regime and the greatest amount of precipitation occurs in May and June, and the least amount occurs in February and October (RHMS, 2013). Prediction was also done for August which is characterized by the minimum amount of precipitation in the areas with a Mediterranean precipitation regime, which is the case with the southwestern part of the territory of the Republic of Serbia.

Data

As an input, data from relatively uniformly distributed weather stations throughout the territory of the Republic of Serbia was used. Geographic coordinates, elevation and cumulative monthly precipitation amount in February, June, August and October 2009 were provided for each station. A total of 197 stations were originally available, but in the end due to identified data errors, a total of 191 stations were used for prediction. Besides the data mentioned, a digital elevation model of the Republic of Serbia with resolution $1 \text{ km} \times 1 \text{ km}$ was also used (estimation was done at grid nodes).

Although it was previously highlighted that in case of applying indicator kriging for prediction it is not necessary for the original variable values to comply to the normal distribution, the distribution of the variables was examined in order to gain insight into the nature of the data. A combination of histograms, qq-plots and Shapiro-Wilk test was used for the purpose of examination. It was affirmed that none of the variables (precipitation amounts in February, June, August and October 2009) can be claimed to be normally distributed.

Generally, by observing the histograms of precipitation sums for the four month periods of 2009 (Figure 1) it can be noticed that all of them are positively skewed, which points to the fact that extremely high precipitation sums in Serbia are present in relatively small spaces.

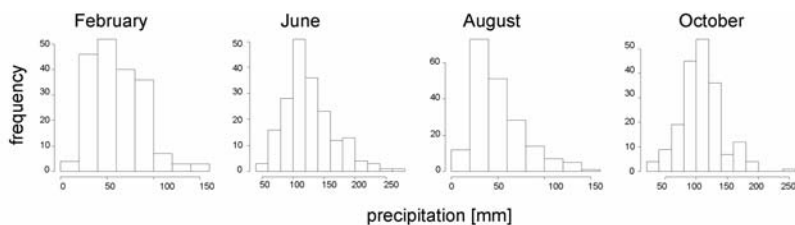


Figure 1. Histograms of precipitation amount for February, June, August and October 2009

Results and discussion

It was previously mentioned in the paper that it is necessary to binarize the original variable values, i.e. to transform them into indicators - values 1 and 0 in order to apply indicator kriging for prediction. As boundary values used for transformation, the quantiles of a corresponding variable distribution were chosen. Three quartiles were defined for each of the observed four months periods of 2009, based on which the transformation, and later prediction, was performed. In the first case the first quartile of the observed variable distribution denoted by q_{25} (the value below which 25% of the variable values, specifically precipitation amount in February, June, August or October are) was used, in the second the median denoted by q_{50} (the value below which 50% of the variable values are), and in the third case the third quartile denoted by q_{75} (the value below which 75% of the variable values are) (Table 1).

Each prediction was independent in the sense that each time a variogram that best fits the experimental variogram was modelled. It was chosen between spherical, exponential and Gaussian model, whereby the final decision was made based on the results of cross-validation. Cross-validation is a procedure that involves prediction of the variable values at points at which the measurements exist. In this particular case, cross-validation was used to obtain numerical indicators for the (un)justified use of a certain variogram model in the prediction procedure. Root mean squared error (RMSE) was used as a parameter based on which the choice was made (Bajat and Štrbac, 2003).

Table 1. Ranges of precipitation values at stations for each of the considered months and threshold values by quartiles

Month	min [mm]	max [mm]	q_{25} [mm]	q_{50} [mm]	q_{75} [mm]
February	5.40	152.50	39.10	57.80	80.15
June	47.30	279.10	100.70	119.50	145.60
August	8.10	145.70	30.75	43.80	62.70
October	21.80	253.00	89.30	105.00	127.00

The summary of chosen variogram models for individual prediction cases is given in Table 2, all based on the minimal RMSE value of the estimated values.

Table 2. Chosen variogram models

Chosen variogram model				
Quartile	February	June	August	October
the first	exponential	spherical	spherical	exponential
the median	exponential	spherical	spherical	spherical
the third	exponential	spherical	exponential	Gaussian

Figure 2 shows the experimental variograms for all conducted predictions together with the corresponding (chosen) variogram models. What is evident is that the so-called hole effect, which points to a form of periodicity in the data, is present in almost every experimental variogram.

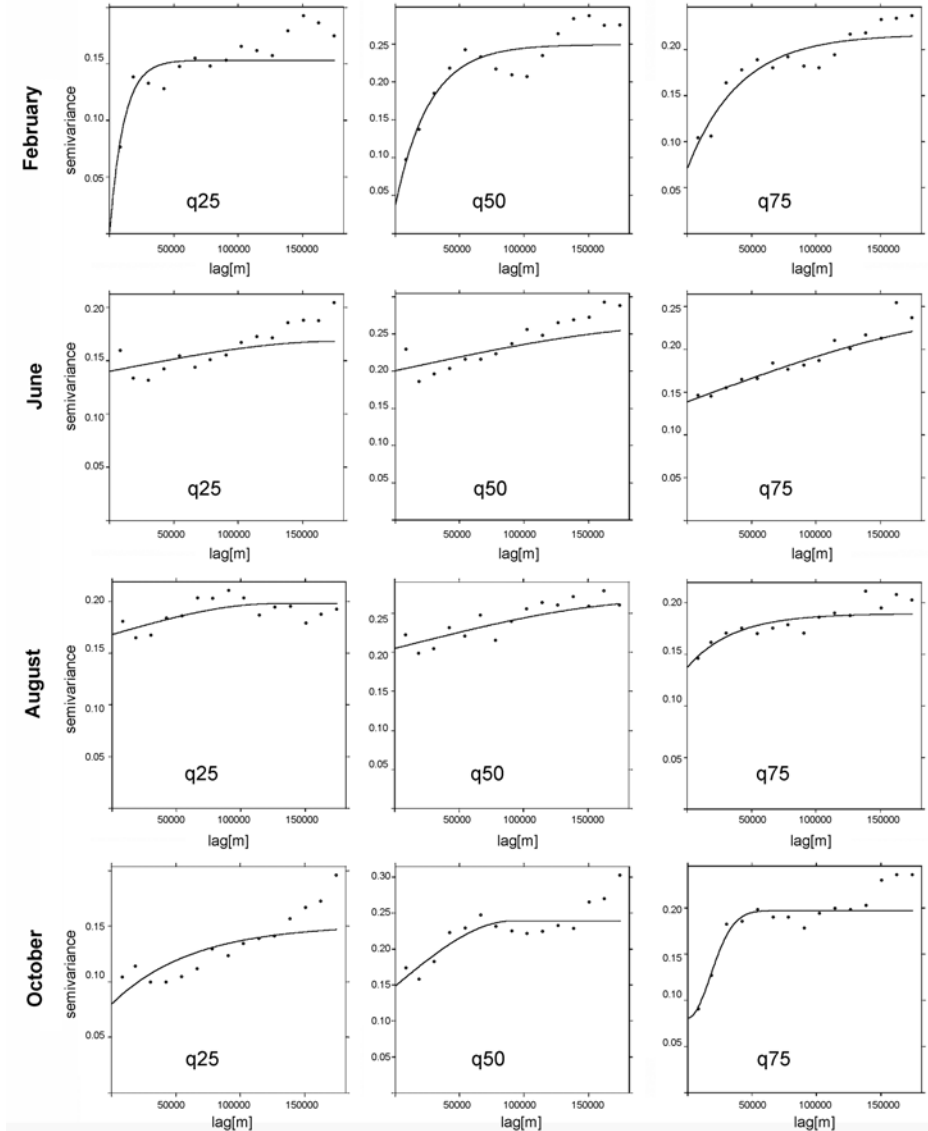


Figure 2. Experimental variograms with chosen variogram models for February, June, August and October 2009

It is important to note that the (an)isotropy of the variations was examined as well, and the result was achieved by calculating variograms in certain directions (north-south, east-west, northeast-southwest and northwest-southeast). Inspection of the obtained experimental variograms showed no real justification for using several directional variograms, therefore variogram models were fitted to the averaged experimental

variograms for whose calculation direction parameter was neglected (search area was a full circle).

The chosen variogram models were used in a prediction based on the indicator kriging method. Taking into consideration all of the above, one can draw a conclusion that the result of the application of this method aren't the desired maps of precipitation amount on the territory of the Republic of Serbia, but are rather probability maps of the occurrence of certain precipitation amount on the territory mentioned, for February, June, August and October 2009. If errors did not occur during the prediction procedure, all estimated probabilities would be in the range of $[0,1]$ and would preserve the order of the thresholds used in prediction. However, errors inevitably occur, and therefore the estimated probabilities were corrected according to the procedure explained earlier in the paper. Firstly, all negative probability values were replaced with zero and all values greater than 1 with one. This ensured that all estimated values are within the permitted range of $[0,1]$. As far as preserving the order of probabilities is concerned, simple averaging of corresponding values was performed: e.g. in case that the estimated probability for a month at an observed location obtained from prediction in which the first quartile (q_{25}) was used was greater than the estimated probability obtained from prediction in which the median (q_{50}) was used, both values were substituted with their average value.

All calculations, as well as preparation of the final graphics, i.e. maps, was performed in the R open source software environment. R software solution offers lots of statistical and graphic tools within available R packages, together with a possibility of expanding and adjusting the environment to one's needs by creating new packages.

Figure 3 shows the maps of the estimated probabilities of the occurrence of a certain precipitation amount on the territory of the Republic of Serbia for distinctive months in 2009.

Three maps are given for each month: a left one uses the corresponding first quartile in the estimation procedure, a middle one for which the median was used and a right one in whose case the third quartile was used for the estimation procedure. The dark-grey colour corresponds to the interval $[0.8,1]$, i.e. $[80\%,100\%]$ and suggests that there is a great possibility that the monthly precipitation amount at an observed location doesn't exceed the threshold (the first quartile, the median or the third quartile). On the other hand, the white colour that corresponds to the interval $[0,0.2]$ points to the fact that it's pretty unlikely the precipitation amount is within given limits, in other words it suggests that there is a huge possibility that the precipitation amount at an observed location exceeds the threshold value. By comparing three belonging maps for every month it is obvious that the area coloured with darker colours increases when going to the right. This means that as the threshold value increases (from the first to the third quartile) so does the probability of precipitation amount at a certain location being within the limits.

Obtained maps correspond to the precipitation distribution in Serbia for analyzed months. Regions with relatively low precipitation amounts are located at the north of the country in the Morava valley and at the territory of Metohija (Ducić and Radovanović, 2005), which is confirmed by Figure 3 (middle map for February, right maps for June and October). On the other hand, parts of eastern Serbia feature considerably larger precipitation amounts (right maps for February and October in Figure 3). However, the rainiest regions of Serbia are located in the west and southwest, which is obvious from most of the maps since those areas are coloured with the brightest colour. Even the relatively bright area that points to a great possibility of a precipitation amount at that location which exceeds a threshold equal to the third quartile (the biggest threshold value used) is present in every map on the right. Considerably higher precipitation amounts in the western part of the country are caused by a cold front and showers brought by cold air masses coming from the Atlantic and western Europe (Unkašević and Tošić, 2011). It is the reason why the western parts of the

country receive more precipitation than the eastern ones, although they are located at the same latitude. The maps for February and October also suggest larger precipitation amounts in the central parts of Serbia.

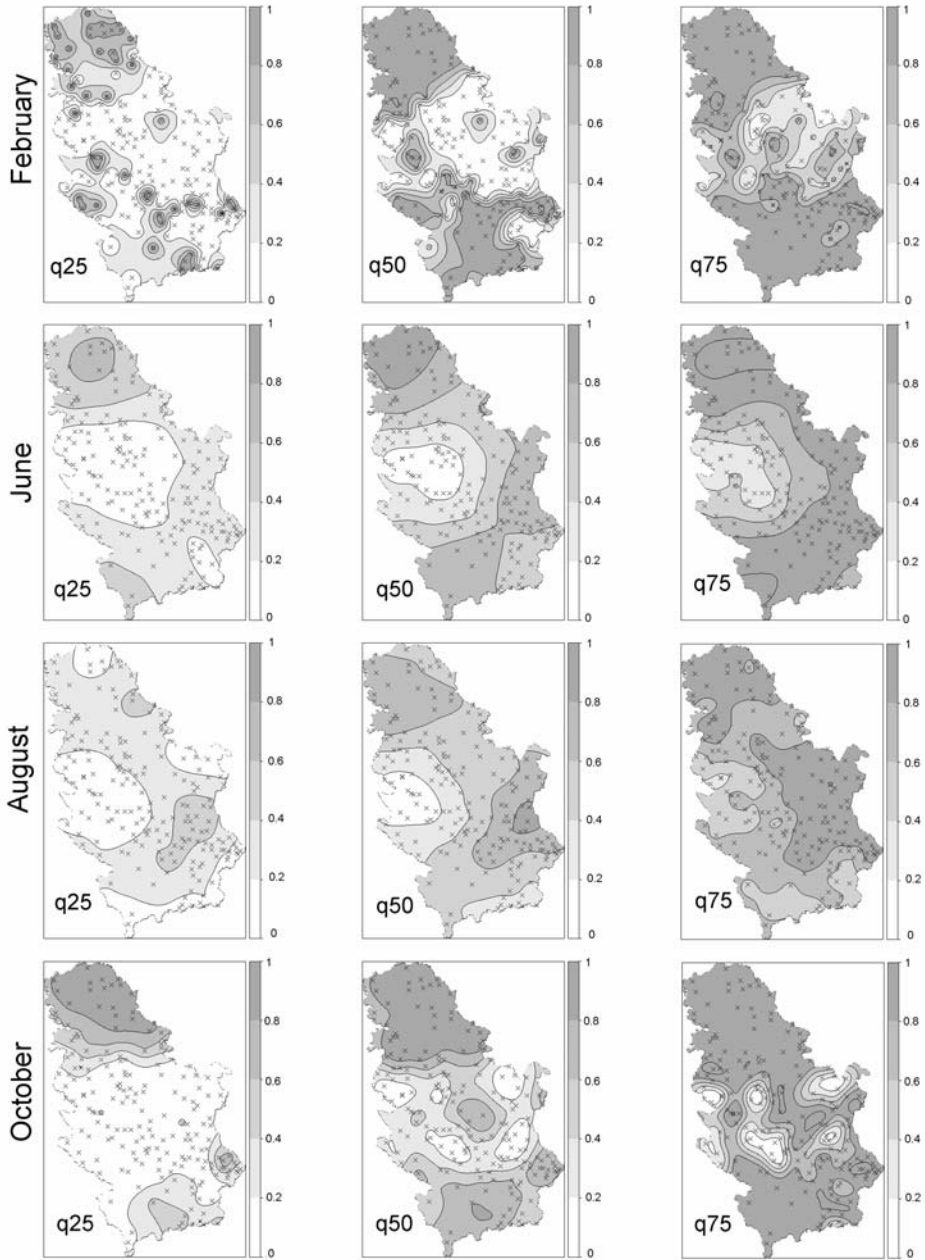


Figure 3. Maps of the estimated probabilities of the occurrence of a certain precipitation amount on the territory of the Republic of Serbia for February, June, August and October 2009

Although the southwestern parts of Serbia generally tend to receive more precipitation, there are some areas like the Pešter plateau that feature unfavourable conditions for precipitation formation (left map and map in the middle for February in Figure 3). Anticyclones with cold and stable weather prevail in this region in winter, whereas descending air currents are common in summer. That is the reason why weather stations at Pešter receive less precipitation (700-800 mm) considering the elevation at which they are located.

Conclusion

Precipitation, as one of the most important climate variables, varies in space (and time). Simple mathematical functions cannot adequately describe those irregular variations and therefore more complex methods, including geostatistical interpolation methods, are required. These methods provide stochastic surfaces as a result and are used for that purpose. Within this paper, indicator kriging was used for describing spatial variations of precipitation on the territory of the Republic of Serbia, i.e. for their prediction.

The result of indicator kriging is a map of probabilities that depict variable values at observed locations. These variable values do not exceed a set boundary value given in advance, which was as such used during the transformation of the original data.

Maps of the probabilities of the occurrence of certain cumulative monthly precipitation amounts on the territory of the Republic of Serbia in February, June, August and October 2009 were created within the paper. Three maps were created for each month, using different threshold values (the first quartile, the median and the third quartile) for transformation of the original data values into indicators. The obtained maps correspond to the spatial distribution of precipitation in Serbia, thereby identifying the northern parts of the country, the Morava valley and Metohija as regions in which the occurrence of a smaller precipitation amount (the amount within the limits of the defined threshold) is more likely to happen. The western parts of Serbia feature greater probabilities of abundant precipitation occurrence, i.e. precipitation which by amount exceeds the limits defined when conducting prediction using the method of indicator kriging.

Maps of probabilities of precipitation occurrence have certain advantages over standard isohyetal maps because they point to the potential of the occurrence of a certain precipitation amount at some location for a given time period. They can be used in agrotechnical projects, especially in countries like Serbia, where yields depend on external meteorological conditions and are primarily limited to precipitation amounts in particular months. These maps are also used as an input for creating spatial risk models (floods, landslides, etc.)

References

- Atkinson, P. M. and Lloyd, C. D. (1998). Mapping Precipitation in Switzerland with Ordinary and Indicator Kriging. *Journal of Geographic Information and Decision Analysis*, Vol. 2, No. 2, p. 65-76.
- Bajat, B. and Štrbac, D. (2003). Quality analysis of digital terrain models for the test area Zlatibor, *Glasnik Srpskog geografskog društva*, Vol. 83, No. 1, p. 31-42. (in Serbian)
- Bajat, B. et al. (2013). Mapping average annual precipitation in Serbia (1961–1990) by using regression kriging. *Theor Appl Climatol*, 112(1), 1-13.
- Bivand, R. S., Pebesma, E. J. and Gómez-Rubio, V. (2008). *Applied Spatial Data Analysis with R*. New York: Springer Science+Business Media, LLC.
- Burrough, P. A. and McDonnell, R. A. (2006). *Principles of Geographic Information Systems*. Serbian translation by Bajat, B. and Blagojević, D., Beograd: Građevinski fakultet.
- Ducić, V. and Radovanović, M. (2005). *Klima Srbije*, Beograd: Zavod za udžbenike i nastavna sredstva, str. 212.
- Glacken, I. and Blackney, P. (1998). *A practitioners implementation of Indicator Kriging*. Vann, J., ed. Proceedings of the Symposium "Beyond Ordinary Kriging: Non-Linear Geostatistical Methods in Practice". Perth: The Geostatistical Association of Australasia, p. 26-39.
- Guimarães, R. J. P. S. et al. (2012). Use of Indicator Kriging to Investigate Schistosomiasis in Minas Gerais State, Brazil. *Journal of Tropical Medicine* [online], Vol. 2012. Available at: <http://www.readcube.com/articles/10.1155/2012/837428?locale=en> [29.08.2013].
- Isaaks, E. H. and Srivastava, R. M. (1989). *Applied Geostatistics*. New York: Oxford University Press.
- Joksić, D. and Bajat, B. (2005). Probability maps as a measure of reliability for intervisibility analysis, *SPATIUM*, No. 12, p. 22-27.
- Journel, A. G. (1983). Nonparametric Estimation of Spatial Distributions. *Mathematical Geology*, Vol. 15, No. 3, p. 445-468.
- Pejović, M., Bajat, B. and Luković, J. (2012). Spatial distribution of interpolation uncertainty: case study of isotherm map of Serbia (1991-2009). *Glasnik Srpskog geografskog društva*, Vol. 92, No. 4, p. 31-50.
- Republic Hydrometeorological Service of Serbia - RHMS (2013). Padavinski režim u Srbiji [online]. Available at: http://www.hidmet.gov.rs/podaci/meteorologija/latin/Pdavinski_rezim_u_Srbiji.pdf [10.08.2013].
- Sun, X., Manton, M. J. and Ebert, E. E. (2003). Regional rainfall estimation using double-kriging of raingauge and satellite observations. BMRC Research Report No. 94. Melbourne: Australian Government, Bureau of Meteorology.
- Tolosana-Delgado, R. (2007): Simplicial Indicator Kriging: presentation [online]. Wuhan: China University of Geosciences. Available at: <http://www.sediment.uni-goettingen.de/staff/tolosana/extra/Wuhan-talk-4.pdf> [21.08.2013].
- Unkašević, M. and Tošić, I. (2011). A statistical analysis of the daily precipitation over Serbia: trends and indices. *Theoretical and Applied Climatology*, 106:69–78.
- Zhang, Y. (2011): Introduction to Geostatistics - Course Notes [online]. Laramie: University of Wyoming, Department of Geology and Geophysics. Available at: <http://geofaculty.uwyo.edu/yzhang/files/Geosta1.pdf> [12.07.2013].

КАРТИРАЊЕ ВЕРОВАТНОЋА ПОЈАВЕ КИШНИХ ПАДАВИНА НА ТЕРИТОРИЈИ РЕПУБЛИКЕ СРБИЈЕ МЕТОДОМ ИНДИКАТОРСКОГ КРИГИНГА

ЈЕЛЕНА ПАНЦИЋ², БРАНИСЛАВ БАЈАТ¹, ЈЕЛЕНА ЛУКОВИЋ²

¹Универзитет у Београду – Грађевински факултет, Одсек за геодезију и геоинформатику, Булевар краља
Александра 73, 11000 Београд

²Универзитет у Београду – Географски факултет, Студентски трг 3/3, 11000 Београд

Апстракт: У раду је дат приказ примене геостатистичке методе индикаторског кригинга за потребе израде карата вероватноћа појаве кишних падавина на територији Републике Србије за карактеристичне месеце у току 2009. године. Разлика између оваквог приступа картирању и стандардних изохијетних карата које описују интензитет падавина је у томе што овај приступ указује на потенцијал појаве одређене количине падавина на некој локацији за дати временски интервал.

Кључне речи: индикаторски кригинг, падавине, картирање вероватноћа

Увод

Просторне варијације неког непрекидног обележја, било да се ради о концентрацији одређеног хемијског елемента у земљишту, количини падавина или нечем трећем, су у основи варијабилне појаве за чије описивање није могуће користити једноставне математичке функције. Због тога се те варијације описују стохастичким површима које су суштина примене геостатистичких метода интерполације (Burgough and McDonnell, 2006).

Циљ геостатистике је управо предикција могуће просторне расподеле неке променљиве. Поменута предикција најчешће егзистира у форми креиране карте или низа карата. Јавља се у два основна облика: као оцена и као симулација (Zhang, 2011). Резултат примене кригинга је најбоља линеарна непомерена оцена (*енгл.* Best Linear Unbiased Estimate - BLUE) вредности променљиве на одређеној локацији. Ова оцена се базира како на доступном узорку података, тако и на моделу вариограма који најверније описује просторну корелацију поменутих података. За разлику од оцене, симулација као резултат даје велики број (најчешће стотинак) карата једнако вероватне расподеле променљиве, на основу којих је могуће утврдити и несигурност оцене одређене променљиве на одређеној локацији (Pejović et al., 2012).

За описивање просторних варијација падавина, као једног од најважнијих климатских елемената, погодна је користити поменуте геостатистичке методе интерполације, тј. кригинг (Bajat et al., 2013). Циљ рада био је да се, применом методе индикаторског кригинга, изврши предикција вероватноћа појаве одређене количине падавина на територији Републике Србије у карактеристичним месецима на основу података из 2009. године, чиме би се потврдила (или оповргла) досадашња сазнања о просторној расподели кишних падавина на поменутој територији. Поред примене индикаторског кригинга у циљу одређивања вероватноће неке појаве, могуће је генерисати карте вероватноћа и методом Монте Карло симулација (Joksić and Bajat, 2005).

Територију Републике Србије одликују два режима падавина: континентални и медитерански. Већи део Србије припада континенталном режиму, који подразумева појаву највеће количине кишних падавина у мају и јуну, а најмање у фебруару и октобру. С друге стране, медитерански режим падавина којим се одликује југозападни део Србије подразумева кишовитији период у новембру, децембру и јануару, а сушнији у августу месецу (RHMS, 2013). У оквиру рада, применом методе индикаторског кригинга, извршена је предикција вероватноћа

² jelenapandzic@gmail.com

Рад представља резултате истраживања у оквиру пројеката ИИИ 47014, ТР 36009 и ИИИ43007 које финансира Министарство просвете, науке и технолошког развоја Републике Србије

појаве одређене количине падавина на територији Републике Србије у фебруару, јуну, августу и октобру 2009. године.

Методологија

Математичко описивање просторне варијације било које променљиве величине, које је суштина примене кригинга, врши се помоћу суме три главне компоненте (Burgough and McDonnell, 2006):

$$Z(x) = m(x) + \varepsilon'(x) + \varepsilon'' \quad (1)$$

где је:

$Z(x)$ - вредност случајне функције,

$m(x)$ - детерминистичка функција која описује тзв. структурну компоненту, тј. тренд,

$\varepsilon'(x)$ - стохастичка (случајна) компонента која је просторно корелисана и представља остатак структурне компоненте, позната и као регионализована променљива,

ε'' - грешка остатка, тј. просторно некорелисани случајни шум.

У зависности од приступа третирању појединих компоненти просторне варијације променљиве, првенствено тренда, као и од тога да ли се нека додатна (допунска) променљива користи током предикције, разликују се варијанте кригинга: прости, обични, универзални, регресиони, индикаторски кригинг, кокригинг, итд. Сви поменути типови кригинга изузев индикаторског за резултат имају процењене вредности посматране променљиве, тј. просторног атрибута на неком простору. Насупрот томе, индикаторски кригинг не пружа информацију о конкретној, оцењеној вредности променљиве на некој локацији, већ о вероватноћи да та вредност на посматраној локацији (не) прелази неку унапред задату вредност, тј. праг (*енгл.* threshold).

Индикаторски кригинг

Од тренутка када је Журнел (*фр.* Journel) 1983. године увео индикаторски кригинг у област геостатистике многи научници, стручњаци, ентузијастички радили су на развоју алгоритма индикаторског кригинга и метода проистеклих из њега. Иако је првенствено коришћен за процену минералних ресурса, као и за потребе картирања многих других природних ресурса, индикаторски кригинг се као техника у данашње време успешно користи и у области процене квалитета вода (Tolosana-Delgado, 2007), картирања падавина (Atkinson and Lloyd, 1998; Sun et al., 2003), али и за доношење закључака о распрострањености појединих обољења, попут шистозомијазе (*лат.* schistosomiasis) која је једна од најчешћих паразитских болести људи (Guimarães et al., 2012).

Индикаторски кригинг је заправо обични кригинг примењен над бинаризованим подацима, тзв. индикаторима. Суштина индикаторског кригинга је да се одабрана (континуална) променљива трансформише у бинарну променљиву, при чему се свака оригинална вредност променљиве замењује вредношћу 1 или 0, у зависности од тога да ли се посматрана вредност налази испод или изнад дефинисаног прага (Isaaks and Srivastava, 1989). Математички, ова нелинеарна трансформација се може представити формулом:

$$i(x, z_k) = \begin{cases} 1, & z(x) \leq z_k \\ 0, & z(x) > z_k \end{cases} \quad (2)$$

где је:

$z(x)$ - мерена вредност променљиве у тачки x ,

z_k - гранична вредност, тј. праг,

$i(x, z_k)$ - трансформисана вредност променљиве (индикатор) у тачки x , за задату вредност прага z_k .

Имајући у виду да се у поступку трансформације вредностима променљиве које су значајно веће од прага додељује исти индикатор као и онима које су тек нешто веће од истог прага, индикаторски кригинг је ефикасан начин за ограничавање утицаја веома великих вредности на резултате предикције (Glacken and Blackney, 1998). Исход спроведеног индикаторског кригинга јесте условна кумулативна функција расподеле посматране променљиве (*енгл.* conditional cumulative distribution function - ccdf), тј. оцена процента вредности у околини сваке посматране тачке које (не) прелазе вредност претходно дефинисаног прага. Другим речима, за разлику од осталих облика кригинга, као последица извршене трансформације, резултат индикаторског кригинга нису конкретне вредности полазне променљиве (нпр. 5.2 mg, 106 mm и сл.) у посматраним тачкама, већ вероватноће са којом вредности полазне променљиве у тим тачкама (не) прелазе вредност претходно дефинисаног прага.

Моделирање вариограма

Трансформација оригиналних вредности променљиве у индикаторе само је први корак у поступку предикције применом индикаторског кригинга. Оно што следи је моделирање вариограма. Наиме, уколико се узме у обзир утицај структурне компоненте (тренда), преостале варијације променљиве постају хомогене, тј. разлике вредности променљиве на различитим локацијама су углавном у функцији растојања између тих локација (Burrough and McDonnell, 2006). Ова чињеница омогућава формирање експерименталног вариограма који је заправо графички приказ зависности полуваријансе разлика мерених вредности променљиве у различитим тачкама од растојања. Полуваријанса се рачуна као:

$$\hat{\gamma}(h) = \frac{1}{2n} \sum_{i=1}^n [z(x_i) - z(x_i + h)]^2 \quad (3)$$

где је n број парова тачака у којима је мерена променљива Z , а које се налазе на међусобном растојању h . Овако добијен експериментални вариограм моделира се неком математичком функцијом, тј. апроксимира се (*енгл.* fit) теоријским моделом који што је могуће верније описује просторну варијацију променљиве и као такав се користи у поступку предикције.

Најчешће коришћени модели вариограма су сферни, експоненцијални, Гаусов и линеарни. Прва три модела позната су као транзитивни вариограми јер се њихова просторна корелациона структура мења са растојањем. Линеарни модел је нетранзитивни вариограм с обзиром да нема праг (*енгл.* sill) унутар подручја у којем постоје мерења (Burrough and McDonnell, 2006).

Уколико се експерименталном вариограму најбоље прилагођава Гаусов модел, то указује на благо варијабилне податке, док сферни модел има изражену прелазну тачку, што упућује на то да је доминантан један модел вариограма. Одабир експоненцијалног модела за апроксимацију експерименталног вариограма сугерише да варијације показују постепени прелаз у оквиру домета или да је модел вариограма сложен, „груписан“ (*енгл.* nested), тј. да представља линеарну комбинацију више основних модела (Burrough and McDonnell, 2006).

(Ан)изотропија варијација је такође нешто о чему треба водити рачуна приликом моделирања вариограма. У случају да се вариограм рачуна игноришући правац у којем се траже парови тачака, тј. уколико је подручје претраживања пун круг, резултујући вариограм, тзв. изотропни вариограм, представља осредњени вариограм по свим правцима. Међутим, просторне варијације се могу мењати са правцем простирања, што је нпр. случај са загађењем ваздуха које показује већи континуитет у правцу у којем претежно дувају ветрови, него у правцу који је управан на претходно поменути. У таквим ситуацијама се рачуна више вариограма, по различитим правцима, а резултујући вариограми, који имају различит праг и домет, називају се анизотропним. У зависности да ли праг или домет остају константни, разликује се геометријска и зонална анизотропија (Isaaks and Srivastava, 1989). Наравно, често се може десити да су истовремено присутна оба вида анизотропије просторних варијација.

Грешке у поступку предикције применом индикаторског кригинга

Имајући у виду раније поменути чињеницу да је резултат индикаторског кригинга условна кумулативна функција расподеле, тј. оцењене вероватноће са којом вредности полазне променљиве у посматраним тачкама (не) прелазе вредност претходно дефинисаног прага, све оцењене вредности требале би бити у интервалу $[0,1]$. Такође, оцењене вредности добијене из индикаторског кригинга спроведеног са више различитих граничних вредности, тј. прагова, морале би чувати поредак самих тих прагова (нпр. уколико је у свакој наредној предикцији коришћена већа вредност прага, и оцењене вредности у свакој посматраној тачки морале би се прогресивно повећавати). Међутим, због грешака до којих неизбежно долази у поступку предикције, дешавају се и „немогуће“ ситуације: појављују се негативне вероватноће, тврди се да се нешто догађа са вероватноћом већом од 100%, појављују се нелогичности у поретку вероватноћа. Све то се једним именом назива кршењем поретка (*енгл.* order relation violations) које изискује увођење одређених поправки.

Најједноставнији начин да се побројане нелогичности превазиђу јесте да се све негативне вредности вероватноћа замене нулом, а све вредности веће од 1 јединицом. Тиме би се обезбедило да се све оцењене вредности налазе у дозвољеном интервалу $[0,1]$. Што се тиче чувања поретка оцењених вредности добијених предикцијама приликом којих су коришћене различите граничне вредности (за потребе трансформације оригиналних вредности променљиве у индикаторе), оно се најчешће обезбеђује простим осредњавањем. То значи да се вредности у посматраној тачки добијене предикцијом извршеном коришћењем два (по поретку) суседна прага, а у чијем случају није задовољен претходни услов, замењују њиховом простом аритметичком средином (Isaaks and Srivastava, 1989).

Испитивање расподеле променљиве

Иако индикаторски кригинг не захтева да се оригиналне вредности променљиве покоравају нормалној расподели управо из разлога што се њиховим бинаризовањем (трансформацијом у индикаторе) губи та зависност од нормалности (Atkinson and Lloyd, 1998), корисно је некад пре саме предикције испитати доступни узорак података, како би се стекао увид у његове особине. У том смислу, може се испитивати расподела неке променљиве, њена (не)корелисаност са другом променљивом и сл.

Изглед хистограма (искошеност, постојање екстремних врхова, тј. „пикова“ и сл.) је најједноставнији начин испитивања расподеле неке променљиве. QQ-plot-ови пореде квантиле расподеле посматране променљиве са квантилима жељене расподеле (најчешће нормалне), при чему су координате тачака на графицима ништа друго до одговарајући квантили. Да би се за неку променљиву могло рећи да је нормално распоређена, тачке на одговарајућем qq-plot-у морају лежати на једној правој.

Предикција вероватноћа појаве кишних падавина на територији Републике Србије

У оквиру рада извршена је предикција вероватноћа појаве одређене количине падавина на територији Републике Србије у карактеристичним месецима 2009. године применом методе индикаторског кригинга. Имајући у виду да највећи део Србије има континентални режим падавина који подразумева највећу количину кишних падавина у мају и јуну, а најмању у фебруару и октобру (RHMS, 2013), управо је предикција вршена за фебруар, јун и октобар. Такође, предикција је вршена и за август који подразумева минималну количину падавина у подручјима са медитеранским режимом падавина, што је случај са југозападним делом територије Републике Србије.

Подаци

Као улазни подаци коришћени су подаци са метеоролошких станица релативно равномерно распоређених на целој територији Републике Србије. За сваку станицу биле су познате њене географске координате, надморска висина, као и збирна месечна количина падавина у фебруару, јуну, августу и октобру месецу 2009. године на тој станици. На располагању је изворно било 197 станица, али су због уочених грешака у подацима за

предикцију коришћени подаци са 191 станице. Осим поменутих података, коришћен је и дигитални модел висина Републике Србије резолуције 1 km x 1 km (оцена је вршена у тачкама грида).

Иако је раније указано на чињеницу да у случају примене индикаторског кригинга није неопходно да се оригиналне вредности променљиве померавају нормалној расподели, ради стицања увида у природу података испитана је расподела променљивих. За потребе испитивања коришћена је комбинација хистограма, qq-plot-ова и Шапиро-Вилк тестова. Утврђено је да се ни за једну променљиву (количина падавина у фебруару, јуну, августу, односно октобру 2009. године) не може тврдити да је нормално распоређена.

Генерално, посматрајући хистограме суме падавина за четири месеца 2009. године (Слика 1) може се уочити да су сви позитивно искошени (*енгл.* positively skewed), што указује на то да су изразито високе суме падавина у Србији присутне на релативно малом простору.

Слика 1. Хистограми количина падавина за фебруар, јун, август и октобар 2009. године Резултати и дискусија

Раније у раду је наведено да је за примену индикаторског кригинга неопходно оригиналне вредности променљиве бинаризовати, тј. трансформисати их у индикаторе - вредности 1 и 0. Као граничне вредности које су коришћене приликом трансформације одабрани су квантили расподеле одговарајуће променљиве. За сваки од четири посматрана месеца 2009. године дефинисана су по три квантила, на основу којих је вршена трансформација и, касније, предикција. У првом случају је коришћен први квантил расподеле посматране променљиве, у ознаци q_{25} (вредност испод које се налази 25% вредности променљиве, конкретно, количине падавина у фебруару, јуну, августу или октобру), у другом медијана у ознаци q_{50} (вредност испод које се налази 50% вредности променљиве), а у трећем случају трећи квантил у ознаци q_{75} (вредност испод које се налази 75% вредности променљиве) (Табела 1).

Свака предикција је била независна, у смислу да је сваки пут изнова моделиран вариограм који је био најбоље прилагођен експерименталном вариограму. Бирало се између сферног, експоненцијалног и Гаусовог модела, при чему је коначан избор вршен на основу резултата крос-валидације. Крос-валидација је поступак који подразумева предикцију вредности променљиве у тачкама у којима постоје мерења. У конкретном случају, крос-валидација је послужила за добијање нумеричких показатеља (не)оправданости коришћења одређеног модела вариограма у поступку предикције. Као параметар на основу којег је извршен избор коришћен је корен средње квадратне грешке (*енгл.* Root Mean Squared Error - RMSE) (Вајат and Штрбац, 2003).

Табела 1. Распони вредности падавина на станицама за сваки од разматраних месеци и вредности прагова по квантилима

У Табели 2 дат је сумарни приказ одабраних модела вариограма за поједине случајеве предикције, а све на основу минималне вредности корена средње квадратне грешке оцењених вредности.

Табела 2. Одабрани модели вариограма

На Слици 2 приказани су експериментални вариограми за све вршене предикције са одговарајућим (одабраним) моделима вариограма. Оно што је евидентно јесте да је у готово свим експерименталним вариограмима присутан тзв. ефекат шупљине (*енгл.* hole effect) који указује на неки облик периодичности у подацима.

Слика 2. Експериментални вариограми са одабраним моделима вариограма за фебруар, јун, август и октобар 2009. године

Важно је напоменути да је испитана и (ан)изотропија варијација, тако што су рачунати вариограми по одређеним правцима (север-југ, исток-запад, североисток-југозапад и северозапад-југоисток). Увидом у добијене експерименталне вариограме утврђено је да не

постоји реална оправданост за коришћење различитих вариограма за различите карактеристичне правце, већ су уместо тога модели вариограма прилагођавани осредњеним експерименталним вариограмима срачунатим не узимајући у обзир параметар правца у којем се траже парови тачака (подручје претраживања било је пун круг).

Одабрани модели вариограма коришћени су у поступку предикције применом методе индикаторског кригинга. Као што се може закључити из свега до сада изложеног, резултат примене ове методе нису креиране карте количина падавина на територији Републике Србије, већ карте вероватноћа појаве одређене количине падавина на територији Републике Србије, у фебруару, јуну, августу и октобру месецу 2009. године. Кад се у поступку предикције не би јављале грешке, све оцењене вероватноће налазиле би се у интервалу $[0,1]$ и чувале би поредак прагова који су коришћени током предикције. Међутим, грешке се неизоставно јављају, па су оцењене вероватноће кориговане у складу са раније објашњеним поступком. Најпре су све негативне вредности вероватноћа замењене нулом, а све вредности веће од 1 јединицом. Тиме је обезбеђено да се све оцењене вредности налазе у дозвољеном интервалу $[0,1]$. Што се тиче очувања поретка вероватноћа, вршено је просто осредњавање одговарајућих вредности: нпр. у случајевима где је за исти месец у посматраној тачки оцењена вероватноћа из предикције у којој је коришћен први квантил (q_{25}) била већа од оцењене вероватноће из предикције у којој је коришћена медијана (q_{50}), обе вредности замењене су њиховом средњом вредношћу.

Сва рачунања, као и припрема коначних графичких прилога, тј. карата, извршена су у оквиру *R open source* програмског окружења. *R* програмско решење потенцијалним корисницима нуди велики број статистичких и графичких алата у оквиру постојећих *R* пакета (*енгл.* packages), уз могућност проширивања и прилагођавања окружења сопственим потребама путем креирања нових пакета.

На Слици 3 приказане су карте оцењених вероватноћа појаве одређене количине падавина на територији Републике Србије у карактеристичним месецима 2009. године.

За сваки месец дате су три карте: лева у случају које је у поступку оцене коришћен одговарајући први квантил, средња за коју је коришћена медијана и десна у чијем случају је коришћен трећи квантил. Тамносива боја која одговара интервалу $[0.8,1]$, тј. $[80\%,100\%]$, указује на велику вероватноћу да на посматраном подручју количина падавина на месечном нивоу не прелази граничну вредност (први квантил, медијану или трећи квантил). Насупрот томе, бела боја која одговара интервалу $[0,0.2]$ указује на малу вероватноћу да је претходно поменуто случај, другим речима, упућује на велику вероватноћу да на посматраном подручју количина падавина прелази граничну вредност. Поредиши три карте за било који месец, евидентно је да се идући надесно повећава површина обојена тамнијим бојама. То значи да се са повећањем вредности прага (од првог до трећег квантила) повећава и вероватноћа да се количина падавина на посматраном подручју налази у границама дозвољене вредности.

Добијене карте одговарају распореду падавина у Србији у анализираним месецима. Области са релативно ниском количином падавина налазе се на северу земље, у долини Мораве и на подручју Метохије (Ducić and Radovanović, 2005), што је и потврђено на Слици 3 (карта у средини за фебруар, десне карте за јун и октобар). Северне делове Србије генерално карактеришу мале количине падавина. Са друге стране, делови источне Србије се одликују знатно већом количином падавина (десне карте за фебруар и октобар на Слици 3). Међутим, најкишовитије области Србије налазе се на западу и југозападу, што је евидентно на већини карата будући да су ове области обојене најсветлијом бојом. Чак се на свакој десној карти задржало то релативно светло подручје које указује на постојање велике вероватноће да у тој области количина падавина прелази граничну вредност која је једнака трећем квантилу (највећој коришћеној вредности прага). Знатно виша количина падавина у западном делу земље узрокована је хладним фронтом и плусковима које доносе хладне ваздушне масе долазећи са Атлантика и из западне Европе (Unkašević and Tošić, 2011). То је разлог због којег западни делови земље примају више талога од источних, иако се налазе на истој географској ширини. Карте за фебруар и октобар такође указују на већу количину падавина у централним деловима Србије.

Слика 3. Карте оцењених вероватноћа појаве одређене количине падавина на територији Републике Србије за фебруар, јун, август и октобар 2009. године

Иако југозападни делови Србије генерално примају више падавина, постоје области које се одликују неповољним условима за образовање падавина, као што је Пештерска висораван (лева карта и карта у средини за фебруар на Слици 3). Зими у овој области преовлађују антициклони са хладним и стабилним временом, док су лети честа силазна ваздушна струјања. То је разлог због којег станице на Пештеру примају мање падавина (700-800 mm) с обзиром на надморску висину на којој се налазе.

Закључак

Кишне падавине, као један од најважнијих климатских елемената, варирају у простору (и времену). Једноставне математичке функције не могу на адекватан начин описати поменуте варијације које су неправилног облика, па се у ту сврху користе комплексније методе, попут геостатистичких метода интерполације које за резултат дају стохастичке површи. У оквиру овог рада, за описивање просторних варијација падавина на територији Републике Србије, тј. за предикцију, примењен је индикаторски кригинг.

Резултат индикаторског кригинга је карта вероватноћа са којом вредности променљиве у посматраним тачкама (не) прелазе унапред дефинисану граничну вредност коришћену приликом трансформације оригиналних података.

У раду су креиране карте вероватноћа појаве одређене збирне месечне количине падавина на територији Републике Србије у фебруару, јуну, августу и октобру 2009. године. За сваки месец креиране су по три карте, користећи различите граничне вредности (први квантил, медијану и трећи квантил) приликом трансформације оригиналних вредности података у индикаторе. Добијене карте одговарају просторном распореду падавина у Србији, издвајајући северне делове земље, долину Мораве и Метохију као области у којима је вероватнија појава мање количине падавина (количине у границама дефинисаног прага). Западни делови Србије одликују се великом вероватноћом за појаве обилнијих падавина, тј. падавина које по количини прелазе граничне вредности дефинисане приликом спровођења предикције применом методе индикаторског кригинга.

Карте вероватноћа појаве падавина имају предности у односу на стандардне изохијетне карте јер указују на потенцијал појаве одређене количине падавина на некој локацији за дати временски интервал. Могу се примењивати у агротехничким пројектима, посебно у земљама као што је Србија, где приноси зависе од спољних метеоролошких услова, пре свега од количине падавина у одређеним месецима. Ове карте се такође примењују као улазни подаци у изради просторних модела ризика (поплаве, клизишта, итд.)

Литература

Литературу видети на 33 страни.