ELSEVIER

Ecological Modelling 138 (2001) 255–263

ECOLOGICAL MODELLING

www.elsevier.com/locate/ecolmodel

Spatio-temporal interpolation of climatic variables over large region of complex terrain using neural networks

Oleg Antonić [a,*], Josip Križan [b], Antun Marki [c], Dragan Bukovec [d]

[a] *Rudjer Bošković Institute, Bijenička 54, 10000 Zagreb, Croatia*
[b] *Oikon Ltd., Vlade Prekrata 20, 10000 Zagreb, Croatia*
[c] *Department of Geophysics, Andrija Mohorovičić Geophysical Institute, Faculty of Science, University of Zagreb, Horvatovac bb, 10000 Zagreb, Croatia*
[d] *Croatian Natural History Museum, Demetrova 1, 10000 Zagreb, Croatia*

Abstract

Empirical models for seven climatic variables (monthly mean air temperature, monthly mean daily minimum and maximum air temperature, monthly mean relative humidity, monthly precipitation, monthly mean global solar irradiation and monthly potential evapotranspiration) were built using neural networks. Climatic data from 127 weather stations were used, comprising more than 30000 cases for each variable. Independent estimators were elevation, latitude, longitude, month and time series of respective climatic variable observed at two weather stations (coastal and inland), which have long time-series of climatic variables (from mid last century). Goodness of fit by model was very high for all climatic variables ($R > 0.98$), except for monthly mean relative humidity and monthly precipitation, for which it was somewhat lower ($R = 0.84$ and $R = 0.80$, respectively). Differences in residuals around model were insignificant between months, but significant between weather stations, both for all climatic variables. This was the reason for calculation of mean residuals for all stations, which were spatially interpolated by kriging and used as a model correction. Similarly interpolated standard deviation and standard error of residuals are estimators of the model precision and model error, respectively. Goodness of fit after the averaging of monthly values between years was very high for all climatic variables, which enables construction of spatial distributions of average climate (climatic atlas) for a given period. Presented interpolation models provide reliable, both spatial and temporal estimations of climatic variables, especially useful for dendroecological analysis.

Keywords: Air temperature; Dendroecology; Digital elevation model; Kriging; Potential evapotranspiration; Precipitation; Relative humidity; Solar irradiation

1. Introduction

The explanation of the dendrochronological variability and building of the dendroecological models usually requires time-series of climatic

* Corresponding author. Fax: +385-1-4680084.
E-mail address: oantonic@rudjer.irb.hr (O. Antonić).

PII: S0304-3800(00)00406-3

data (see e.g. Chang and Aguilar, 1980; Friend and Hafley, 1989; Devall et al., 1991). Ideally, a site-specific topoclimatic data (reflecting influence of local relief) are required, which enable both temporal and spatial dendroecological analyses. This kind of climatic data are usually not available. Temporal analysis can be done using the data obtained at the nearest weather station (if it is close enough to the site of interest). If time-series from a nearby station are not long enough, they can be extrapolated using the regression models, with respective time-series from distant weather stations as independent estimators (e.g. Antonić et al., in press). Spatial topoclimatic variability can be ignored (e.g. using the averaging of the tree ring time-series of different trees) or included in the analysis using estimators derived from digital elevation model (DEM, e.g. Antonić, 1996).

When the nearest station is too distant, a spatial interpolation of the climatic variables is necessary to provide reliable data. If the network of weather stations is satisfactorily dense for the existing spatial climatic variability and/or given scale-dependant purpose, an ordinary kriging (Robertson, 1987; Isaaks and Srivastava, 1989) can be used (see e.g. Phillips and Marks, 1996). This approach is unsatisfactory when interpolation over large regions of complex terrain is needed and when the weather station network is too sparse to describe influence of elevation. In this case, the spatial interpolation of the macroclimatic parameters usually include: (1) calculation of the elevation lapse rates using the regression model; (2) standardisation to the sea level using the regression lapse rates to separate regional trends from elevation; (3) spatial interpolation at the sea level to the whole area of interest (using e.g. approach of Franke (1982) or Mitas and Mitasova (1988)); and (4) reprojection of the interpolated values to actual elevation using regression parameters and spatial distribution of elevations (usually DEM). This kind of interpolation technique was applied by, e.g. Hutchinson and Bischof (1983), Mitchell (1991), Daly et al. (1994), Lennon and Turner (1995) and Thornton et al. (1997).

1.1. Aim and hypothesis

The aim of this study was to introduce a spatial interpolation of climatic variables obtained at the numerous meteorological stations over the large region of complex terrain, simultaneously with temporal interpolation during the long period covered only by several stations. Each part of this task (1 — spatial interpolation; 2 — involving the elevations; and 3 — temporal interpolation) can be solved separately using the above mentioned methods. This 'step-by-step' approach is appropriate for the site-specific climatic study such as presented by Antonić et al. (in press). But this approach is hardly applicable when the climate of large area has to be interpolated into the past, especially in the case of complex terrain, due to the irregular number of working stations during the time. Moreover, such separate spatial interpolations for each time unit (e.g. month) of a long period, lead to the output of numerous spatial distributions, which is hardly manageable within the geographic information system (GIS).

In this study, the hypothesis was that spatio-temporal distributions of climatic variables expressed at the level of monthly statistics, could be described as empirical functions of latitude, longitude, elevation and respective climatic time-series obtained at the limited number of weather stations. In other words, the existence of characteristic spatial climatic patterns was assumed. Expected complexity of these functions led to the use of neural networks (NN) as the suitable modelling tool.

2. Material and methods

2.1. Study area and climatic data

The study area was entire territory of the Republic of Croatia (56 538 km^2, Fig. 1). Croatia is a country with large global climatic variability, from the warm and dry Mediterranean climate to the cold and wet mountainous climate (see Fig. 2, compare also Fig. 1). Approximately half of the area belongs to the karst region with extremely rugged karst relief, which strongly affects local climate.

The first weather station in Croatia (Hvar, see Fig. 1) was established in 1858. Until the end of the 19th century another two stations were established (Zagreb in 1861 and Osijek in 1899, see Fig. 1). Numerous weather stations were established during the 20th century (Fig. 3). In this study, climatic data from 127 weather stations were used. Seven climatic variables were included: (1) monthly mean air temperature; (2) monthly mean daily minimum air temperature; (3) monthly mean daily maximum air temperature; (4) monthly mean relative humidity; (5) monthly precipitation; (6) monthly mean global solar irradiation on horizontal surface at ground; and (7) monthly potential evapotranspiration. The first five climatic variables were taken directly from the weather station chronicles. The last two variables were modelled for each weather station as a function of other climatic variables observed at the respective station, using the model of Nikolov and Zeller (1992) for global solar irradiation and the model of Priestly and Taylor (1972) (see also Bonan, 1989) for potential evapotranspiration.

2.2. Data analysis

The basic idea was to estimate values of each climatic variable observed at the majority of stations as an empirical function of latitude, longitude, elevation, and respective values observed at a limited number of selected stations which have long time-series of climatic data. These stations could be named as ‘anchorage stations’, because their time-series are independent estimators in the

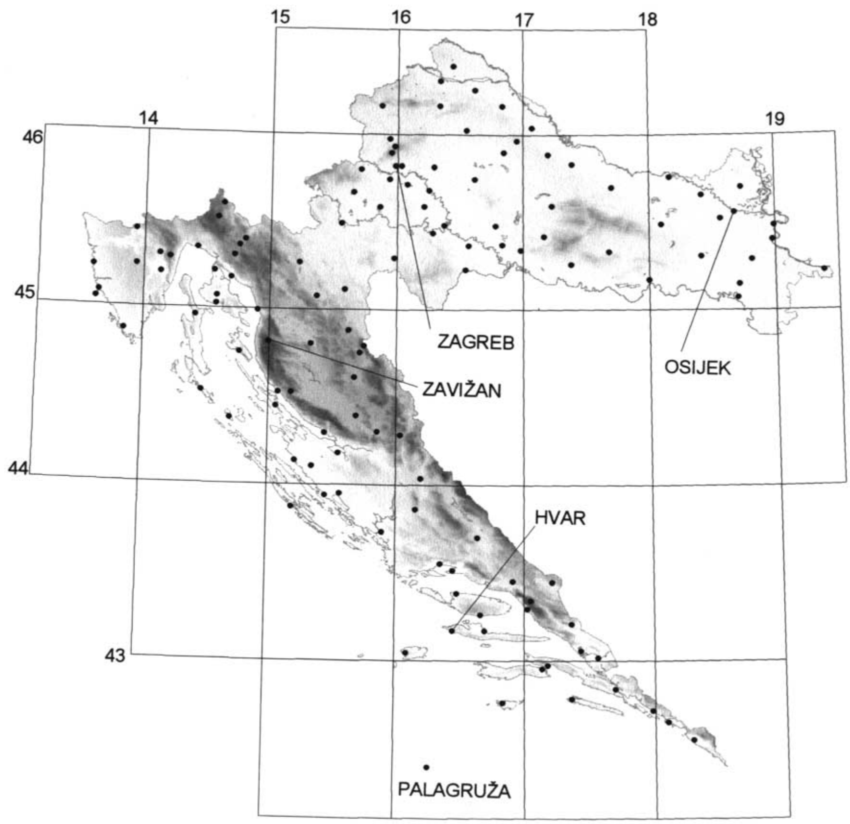

Fig. 1. The Republic of Croatia as a research area. Spatial distribution of altitude is stretched in the grey scale from white (0 m) to dark grey (1762 m). Locations of the weather stations are marked as black points. Stations mentioned in the text are labelled with full names. Latitude and longitude in degrees are superimposed.

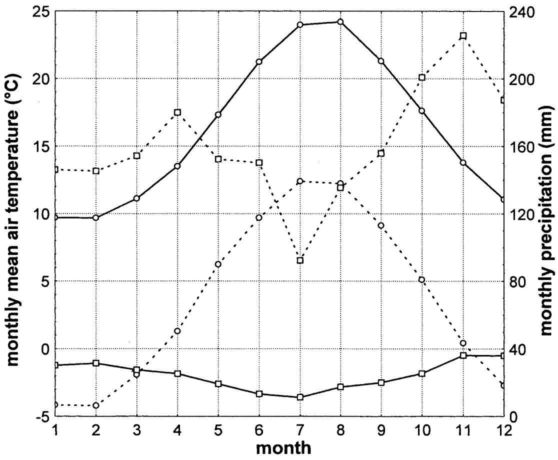

Fig. 2. An illustration of the spatial climatic variability in Croatia. Solid line – Palagruza station, hatched line, Zavizan station. Circles, monthly mean air temperature; squares, monthly precipitation, both averaged for the period of 1956–1995. Compare Fig. 1 for locations.

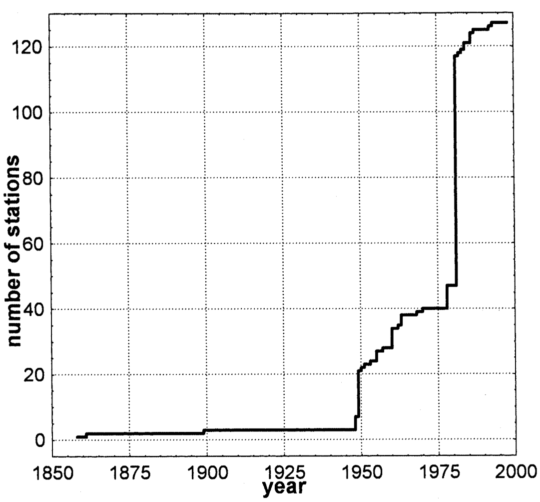

Fig. 3. Number of weather stations in Croatia which started to work up to respective year. Temporary weather stations were not counted.

interpolation model. It could be expected that increase in the number of anchorage stations should increase model reliability but decrease the period which is covered by spatio-temporal interpolation. Initially, three candidates for the anchorage stations were examined: Hvar, Zagreb and Osijek. During the preliminary data analysis, it was found that the use of Hvar (which has the longest period of observations), as a single anchorage station, did not yield satisfactory results for the inland region. Addition of Zagreb, as the second anchorage station, yielded acceptable results for the entire Croatia (see below). Addition of Osijek, as the third anchorage station, did not improve the model significantly. Thus, the final interpolation models were built using the two anchorage stations: Hvar (representing the mediterranean climate) and Zagreb (representing the inland climate).

Interpolation models were derived as empirical functions using the feedforward NN with multilayer perceptrons (MLP) which is appropriate for regression problems (see, e.g. Bishop, 1995; Patterson, 1996). Input layer contained the following independent estimators: (1) month as a nominal variable which was branched into 12 ‘dummy’ variables (Ott, 1993); (2) station latitude; (3) station longitude; (4) station elevation; (5) value of respective climatic variable observed at Hvar in the same month of the same year; and (6) value of respective climatic variable observed at Zagreb in the same month of the same year. The output was the value of respective climatic variable observed at the respective station in the given month and year. During the preliminary data analysis, different NN architectures were tested, regarding the number of hidden layers and their neurons. NN architecture, finally accepted for all climatic parameters, had two hidden layers, each with six neurons (Fig. 4). Each neuron in the network, excluding input layer neurons, calculates its output value using the expression:

$$\beta = act\left(a + \sum_{i=1}^{n} b_i \alpha_i\right) \quad (1)$$

where β is neuron output value, α_i is i-th neuron input value, n is number of input connections, a and b_i are empirical parameters (a is neuron threshold and b_i is i-th input weight), and act is activation function:

$$act(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

The whole model can be described by expression (compare Fig. 4):

$$y = act\left(ot + \sum_{k=1}^{6} ow_k act \left(ht_{2,k} + \sum_{j=1}^{6} hw_{j,k} act \left(ht_{1,j} + \sum_{i=1}^{17} iw_{i,j} in_i \right)\right)\right) \quad (3)$$

where in_i is value of the input layer neuron defined by:

$$in_i = \begin{cases} \begin{cases} 0 & x_1 \neq i \\ 1 & x_1 = i \end{cases} & , i = 1 \ldots 12 \\ x_v & , v = i - 11, i = 13 \ldots 17 \end{cases}, \quad (4)$$

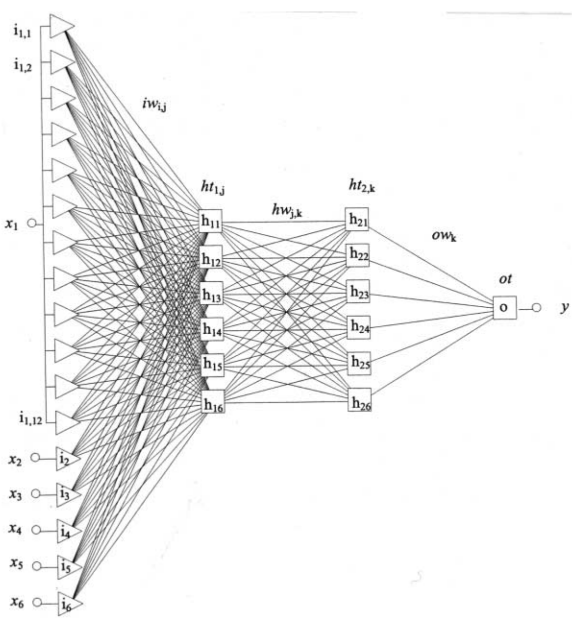

Fig. 4. Architecture of used neural network. x_1 is month, x_2 is station latitude, x_3 is station longitude, x_4 is station elevation, x_5 is value of respective climatic variable observed at Hvar in the same month of the same year, x_6 is value of respective climatic variable observed at Zagreb in the same month of the same year, i_i is input layer neuron, $h_{1,j}$ is the first hidden layer neuron, $h_{2,k}$ is the second hidden layer neuron, o is output neuron, y is output value (estimated value of climatic variable for input values x_v, $v = 1 \ldots 6$), $iw_{i,j}$, $ht_{1,j}$, $hw_{j,k}$, $ht_{2,k}$, ow_k and ot ($i = 1 \ldots 17$, $j = 1 \ldots 6$, $k = 1 \ldots 6$) are empirical parameters (thresholds and weights, compare Eq. (3)).

$ht_{1,j}$ are neuron thresholds in the first hidden layer ($j = 1 \ldots 6$), $ht_{2,k}$ are neuron thresholds in the second hidden layer ($k = 1 \ldots 6$), ot is output neuron threshold, $iw_{i,j}$ are weights of the connections between input layer and the first hidden layer ($i = 1 \ldots 17$, $j = 1 \ldots 6$), $hw_{j,k}$ are weights of the connections between the first and the second hidden layer ($j, k = 1 \ldots 6$), ow_k are weights of the connections between the second hidden layer and output ($k = 1 \ldots 6$), x_v are input values of independent estimators, y is output value (estimated value of climatic variable for input values x_v) and act is logistic function.

In total, over 30000 cases for each climatic variable were included in the model development, namely 10000 randomly selected cases for NN verification set and the rest for the NN training (parameter estimation). Goodness of NN model fit was estimated by correlation coefficient R which was derived directly from observed and estimated values using expression:

$$R = (1 - V_r / V_t)^{1/2} \quad (5)$$

where V_r is residual variance and V_t is total variance. Total R (through all stations, years and months) was calculated separately for the training set and for the verification set, aiming to detect possible overfitting (see e.g. Lawrence et al., 1997). The model testing on the fully independent data set was not undertaken, because the model is basically designed for spatio-temporal interpolation of existing climatic data, and not for spatial extrapolation or temporal prediction of climatic variables. Analysis of variance (ANOVA, Ott, 1993) was used to test the differences in the residuals around model between months and also between weather stations. Separate correlation coefficients for particular stations were also calculated.

Data sets used for the model development were unbalanced, i.e. particular stations were represented with different number of cases, due to the different starts of work. Balancing of the data set would significantly decrease the number of available cases or the number of included weather stations. Consequently, t-test (Ott, 1993) was used to test does the number of cases per station increase goodness of fit. For this purpose, two

Table 1
Correlation between values estimated by the interpolation model and observed values[a]

	METMP	MITMP	MATMP	RLHUM	PRECP	GLRAD	EVAPR
Total data set statistics							
N	34 006	32 774	32 920	33 412	34 092	33 733	31 970
R – training set	0.9936	0.9837	0.9955	0.8497	0.8151	0.9951	0.9920
R – verification set	0.9933	0.9828	0.9955	0.8443	0.8043	0.9949	0.9922
F – ANOVA (month)	1.0310	1.4576	0.6259	1.3027	0.6020	1.2087	1.7280
P (F) – ANOVA (month)	0.4151	0.1399	0.8083	0.2154	0.8288	0.2745	0.0610
F – ANOVA (station)	**297.934**	**469.676**	**54.781**	**105.606**	**27.710**	**73.564**	**77.166**
P (F) – ANOVA (station)	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
Mean statistics through the all stations after correction							
N – median	179	175	175.5	176.5	180	176	174
R – mean left of N – median	0.9960	0.9920	0.9953	0.6939	0.7324	0.9944	0.9917
R – mean right of N – median	0.9959	0.9920	0.9951	0.7508	0.7482	0.9962	0.9929
t	0.2317	−0.0657	0.3101	**−2.0133**	−0.6268	**−3.4119**	−1.2180
P (t)	0.8172	0.9477	0.7570	**0.0463**	0.5320	**0.0009**	0.2255
Mean statistics through the 21 stations with full 40-year data set after averaging of months between years							
R – mean	0.9998	0.9994	0.9993	0.9895	0.9773	0.9995	0.9991

[a] METMP, monthly mean air temperature; MITMP, monthly mean daily minimum air temperature; MATMP, monthly mean daily maximum air temperature; RLHUM, monthly mean relative humidity; PRECP, monthly precipitation; GLRAD, monthly mean global solar irradiation; EVAPR, monthly potential evapotranspiration. N is the number of cases, R is the correlation coefficient, F is the proportion of between-group and within-group variance, P (F) is probability of F-statistics, t is the value of t-statistics, P (t) is probability of t-statistics. Bold values indicate differences significant at the probability level of $P = 0.05$.

distributions of correlation coefficients were compared: (1) obtained at stations with number of cases less than median number of cases for all stations; and (2) obtained at stations with number of cases greater than the median.

Number of 21 weather stations with full 40-year data sets (1956–1995) of all examined climatic parameters were selected to calculate total R after the averaging of the estimated and observed monthly values between the years. This was done to test how the interpolation models, which are primarily spatio-temporal, estimate spatial pattern of interannually averaged climatic parameters only.

3. Results and discussion

Statistical results for all climatic variables are summarised in Table 1. In total, goodness of fit was very high for the monthly mean air temperature, monthly mean daily minimum and maximum air temperature, mean monthly global solar irradiation and monthly potential evapotranspiration, and somewhat lower for the monthly mean relative humidity and monthly precipitation. The last could be explained by the additional impact of topography which cannot be described by the elevation only (see e.g. Antonić, 1996).

Verification set and training set had almost the same goodness of fit for all climatic variables. According to ANOVA, differences in residuals around models between months were insignificant for all climatic variables (Table 1) and they can be ignored.

On the contrary, the differences in residuals between stations were significant for all climatic variables (Table 1), which could be explained by the influence of the local climatic peculiarities. Due to this fact, mean residual for each weather station and for each climatic variable was calculated. Yielded mean residuals were accepted as locally specific corrections of the model and they were spatially interpolated over the entire study area using kriging (e.g. Robertson, 1987; Isaaks and Srivastava, 1989). All kriging interpolations were done with a Gaussian semivariogram model and maximum of eight nearest neighbour supporting points (weather stations) within a radius of 100 km from the subject grid point.

The same procedure was used for spatial interpolation of standard deviation and standard error of residuals calculated for each weather station. Yielded spatial distributions were accepted as estimators of the model precision and model error, respectively, for the entire study area.

Thus, spatio-temporal interpolation model for one climatic variable consists of: (1) neural network; (2) spatial distribution of locally specific corrections; (3) spatial distribution of model precision; and (4) spatial distribution of model error. Using such an interpolation model, a monthly value of a chosen climatic variable can be estimated for the entire study area and for specific year and month, within the period covered by observations at the anchorage stations Hvar and Zagreb (see Fig. 5 for example). This enables reliable estimation of the long time-series of climatic variables, monthly (Fig. 6) or annually based (Fig. 7), for any point in Croatia.

Results of t-test (Table 1) show that mean correlation coefficients obtained at stations with a number of cases greater than the median are not significantly higher than the mean correlation coefficients obtained at stations with a number of cases less than the median, for all climatic vari-

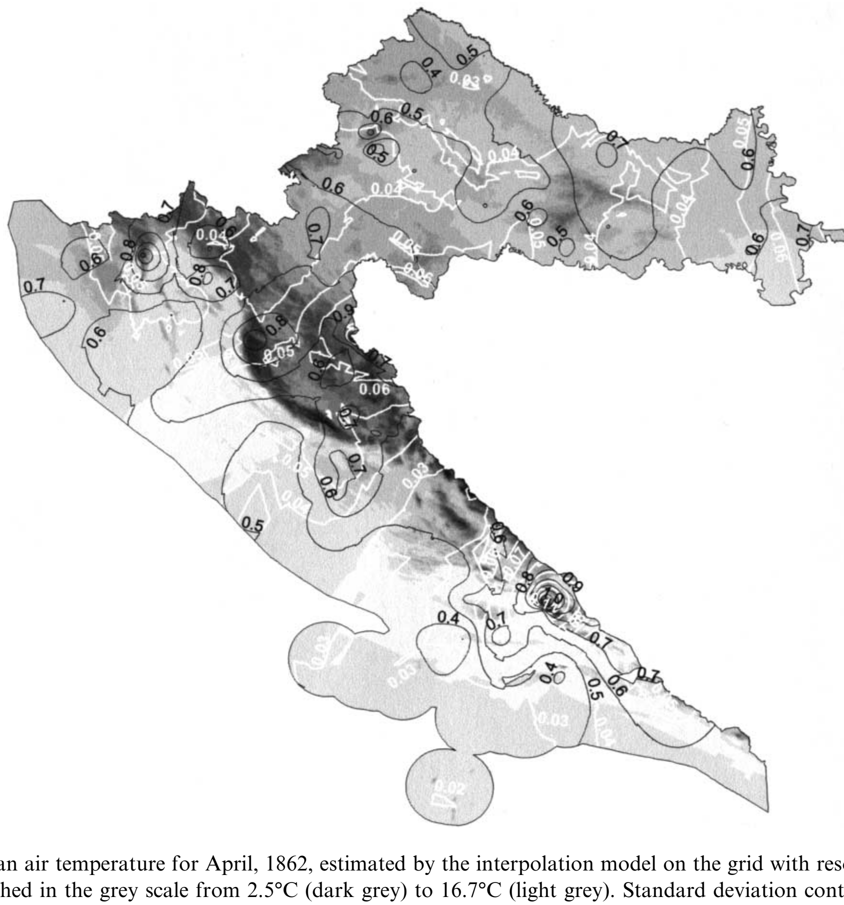

Fig. 5. Monthly mean air temperature for April, 1862, estimated by the interpolation model on the grid with resolution of 300 × 300 m. Values are stretched in the grey scale from 2.5°C (dark grey) to 16.7°C (light grey). Standard deviation contours (black line and labels) and standard error contours (white line and labels) are superimposed.

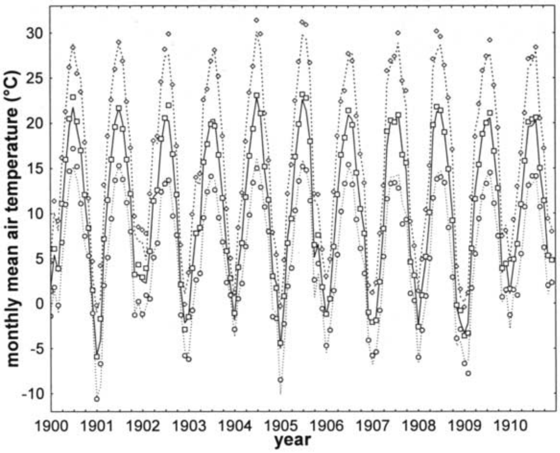

Fig. 6. Correlation between values estimated by the interpolation model (lines) and observed values (symbols) for the monthly mean air temperature (solid line, squares), monthly mean daily minimum air temperature (dotted line, circles) and monthly mean daily maximum air temperature (hatched line, rhombs) for Osijek station and for the first decade in the 20th century.

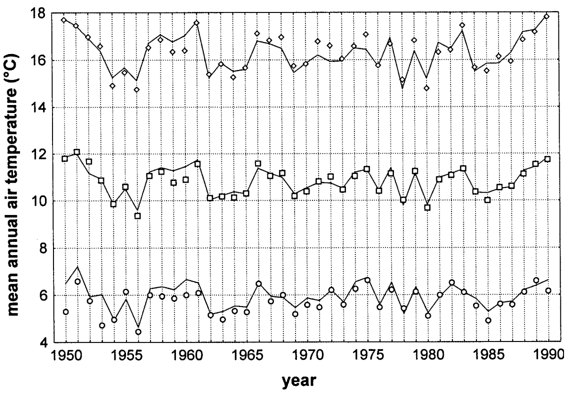

Fig. 7. Correlation between values estimated by the interpolation model (lines) and observed values (symbols) for the mean annual air temperature (squares), mean annual daily minimum air temperature (circles) and mean annual daily maximum air temperature (rhombs) for Osijek station and for the period of 1950–1990.

ables except the relative humidity and global radiation. For these two variables influence of the unbalanced data set could be assumed.

Correlation between values estimated and observed at 21 weather stations with full 40-year data sets (1956–1995), after the averaging of monthly values between the years, is very high for all climatic variables (Table 1). This suggests usability of presented interpolation models for the construction of spatial distributions of average climate (climatic atlas) for the given period covered by observations at the anchorage stations and for the entire Croatian territory.

The conclusion could be made that presented interpolation models provide reliable, both spatial and temporal estimations of climatic variables. Almost all spatial and temporal climatic variability over a large region of complex terrain is explained by the neural network as a function of observations obtained at two weather stations and DEM. Spatial and temporal interpolation errors do not depend on the spatial density of weather stations.

Presented interpolation models could be applied to the particular localities for the purpose of reconstruction of climatic time-series, monthly or annually based, for the long period. This is particularly useful for site-specific dendroecological studies. Additionally, the models could be applied to calculate spatial distribution of climatic variables for the area of interest, for specific month and year or averaged, using needed spatial resolution. This could be combined with more detailed topoclimatic modeling based on DEM (see e.g. Antonić, 1996). For instance, spatial distribution of global solar irradiation on horizontal surface at ground is needed as input variable for the calculation of the topographic solar irradiation (Antonić, 1998, see also Dubayah and Rich, 1995). Furthermore, the models could be used for complex spatio-temporal climatic analysis, e.g. in the research of climatic changes. Usability of the models for spatial extrapolation and temporal prediction of climatic variables have to be tested in the future.

General applicability of NNs as a tool for spatial interpolation has to be also tested in the future in comparison with the other interpolation methods. This has not been the subject of this research, which has been focused on overcoming the lack of data in the past, a problem unsolvable by ordinary spatial interpolation.

Acknowledgements

The authors wish to express a gratitude to Meteorological and Hydrological Service of Croatia (DHMZ), Zagreb, for access to the weather station data. We are also grateful to two anonymous reviewers for their comments. This research has been supported by the Croatian Ministry for Research and Technology and by the OIKON Ltd., Zagreb.

References

Antonić, O., 1996. Application of spatial modelling in the karst bioclimatology. Croat. Met. J. 31, 95–102.

Antonić, O., 1998. Modelling daily topographic solar radiation without site-specific hourly radiation data. Ecol. Model. 113, 31–40.

Antonić, O., Marki, A., Hatić, D., in press. Modelling of the climatic variables as a part of the dendroecological study in the forest Repaš, Prekodravlje, Croatia. Croat. Met. J.

Bishop, C., 1995. Neural Networks for Pattern Recognition. Oxford University Press, Oxford.

Bonan, G.A., 1989. A computer model of the solar radiation, soil moisture and soil thermal regimes in boreal forests. Ecol. Model. 45, 275–306.

Chang, M., Aguilar, G.J.R., 1980. Effects of climate and soil on the radial growth of loblolly pine (*Pinus taeda* L.) in a humid environment of the southeastern U.S.A. For. Ecol. Manage. 3, 141–150.

Daly, C., Neilson, R.P., Phillips, D.L., 1994. A statistical-topographic model for mapping climatological precipitation over mountainous terrain. J. Appl. Meteorol. 33, 140–158.

Devall, M.S., Grender, J.M., Koretz, J., 1991. Dendroecological analysis of a longleaf pine *Pinus palustris* forest in Mississippi. Vegetatio 93, 1–8.

Dubayah, R., Rich, P.M., 1995. Topographic solar radiation models for GIS. Int. J. Geographical Inf. Syst. 9, 405–419.

Franke, R., 1982. Smooth interpolation of scattered data by local thin plate splines. Comp. Maths Appl. 8, 237–281.

Friend, A.L., Hafley, W.L., 1989. Climatic limitations to growth in loblolly and shortleaf pine (*Pinus taeda* and *P. echinata*): a dendroclimatological approach. For. Ecol. Manage. 26, 113–122.

Hutchinson, M.F., Bischof, R.J., 1983. A new method for estimating the spatial distribution of mean seasonal and annual rainfall applied to Hunter Valley, New South Wales. Aust. Meteorol. Mag. 31, 179–184.

Isaaks, E.H., Srivastava, R.M., 1989. An Introduction to Applied Geostatistics. Oxford University Press, New York, pp. 278–322.

Lawrence, S., Giles, C.L., Tsoi, A.C., 1997. Lessons in neural network training: overfitting may be harder than expected. Proceedings of the Fourteenth National Conference on Artificial Intelligence, AAAI-97, AAAI Press, Menlo Park, California, pp. 540–545.

Lennon, J.J., Turner, J.R.G., 1995. Predicting the spatial distribution of climate: temperature in Great Britain. J. Anim. Ecol. 64, 370–392.

Mitas, L., Mitasova, H., 1988. General variational approach to the interpolation problem. Comput. Math Appl. 16, 983–992.

Mitchell, N.D., 1991. The derivation of climate surfaces for New Zealand, and their application to the bioclimatic analysis of the distribution of Kauri (*Agathis australis*). N.Z. J. R. Soc. 21, 13–24.

Nikolov, N.T., Zeller, K.F., 1992. A solar radiation algorithm for ecosystem dynamic models. Ecol. Model. 61, 149–168.

Ott, R.L., 1993. An Introduction to Statistical Methods and Data Analysis. Duxbury Press, Belmont, p. 1051.

Patterson, D., 1996. Artificial Neural Networks. Prentice Hall, Singapore.

Phillips, D.L., Marks, D.G., 1996. Spatial uncertainty analysis: propagation of interpolation errors in spatially distributed models. Ecol. Model. 91, 213–229.

Priestly, C.H.B., Taylor, R.J., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. Mont. Weather Rev. 100, 81–92.

Robertson, G.P., 1987. Geostatistics in ecology: interpolating with known variance. Ecology 68, 744–748.

Thornton, P.E., Running, S.W., White, M.A., 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain. J. Hydrol. 190, 214–251.