

MODELLING THE IMPACT OF CLIMATE CHANGE ON THE HUNGARIAN WINE REGIONS USING RANDOM FOREST

GAÁL, M.^{1*} – MORIONDO, M.² – BINDI, M.³

¹*Corvinus University of Budapest, Department of Mathematics and Informatics
1118 Budapest, Villányi út 29-43., Hungary
(phone: +36-1-482-6261; fax: +36-1-466-9273)*

²*National Research Council of Italy, Institute of Biometeorology
Via G. Caproni 8, 50145 Florence, Italy*

³*University of Florence, Department of Plant, Soil and Environmental Science
Piazzale delle Cascine 18, 50144 Florence, Italy
(phone: +39-0553288257; fax: +39-055332472)*

**Corresponding author
e-mail: marta.gaal@uni-corvinus.hu*

(Received 20th March 2012; accepted 8th May 2012)

Abstract. This paper aims to simulate and analyse the impact of climate change on the Hungarian wine regions using spatial layers of temperature-based bioclimatic indices. Random forest classification was used to analyse the similarities between the present and future climate of the wine regions. The model was firstly calibrated for the present period then applied for the expected future climatic conditions simulated by the RegCM3 model with A1B scenario. Results show that in the near future (2021-2050) the grapevine regions typical of the southern may expand in greater part of the country, while at the end of the century (2071-2100) only the northern part of the country shows some similarities with the present climate. Despite these results, Hungary is expected to remain amongst the regions with good quality grapevine growing conditions, but the structure of the cultivation and/or varieties should be changed.

Keywords: *climate change, viticulture, wine regions, Random Forest classification*

Introduction

Climate change has the potential to greatly impact nearly every sector of agriculture, including viticulture. Grapevine is one of the oldest cultivated plants, along with the process of making wine. The cultivated varieties and the overall wine style that a region produces is a result of the average climatic conditions, while climate variability determines vintage quality differences. Climatic changes, which influence both variability and average conditions, therefore have the potential to impact on growth, grape composition, wine style and spatial distribution of grapevines (Hunter and Bonnardot, 2011). Today's viticultural regions for quality wine production are located in relatively narrow geographical and therefore climatic niches that means high sensitivity and risk both in short-term climate variability and long-term climate change (Jones, 2007; Holland and Smit, 2010). Additionally, grapevine is perennial and during the expected productive life of the vineyards the climate is projected to change significantly.

Jones et al. (2005) found that in the last decades the majority of the U.S. and European wine regions experienced now statistically significant warming trends in the growing season. This tendency was observed also in Hungary (Kocsis et al., 2010). As a consequence, phenological changes are expected to occur, too (Hlaszny et al., 2012;

Ladányi et al., 2010). According to Moisselin et al. (2002) the increase of the mean temperature at a rate of 1°C can cause a geographical shift of the border of the suitable wine regions with 180 km to north. At present Hungary lies at near the northern border of the growing areas, and expected to remain amongst the regions with good quality grapevine growing conditions, but growers must prepare for the expected changes. Analysing the climatic risk factors of Central Hungarian grape growing regions (Szenteleki et al., 2012) it was found that the significant increase of the number of warm and hot days in the future involves the risk of production.

While there are many individual weather and climate factors that can affect grape growth and wine quality (e.g. solar radiation, heat accumulation, temperature extremes, precipitation, wind, and extreme weather events such as hail), growing season length and temperatures are critical aspects because of their major influence on the ability to ripen grapes to optimum levels of sugar, acid and flavour (Jones et al., 2005; Zanathy, 2008, Ladányi, 2010). To evaluate viticulture and wine production in the context of climate suitability and the potential impacts from climate change, various temperature-based bioclimatic indices (e.g. degree-days, temperature of the warmest month, cool night index, average growing season temperatures, Huglin and Winkler indices) can be used (Tonietto and Carbonneau, 2004; Zorer, 2008).

The suitability of an area for vine growing depends also on other ecological factors, e.g. soil, elevation, slope and aspect. In Hungary a production area cadastre with maximum of 400 points was elaborated in the Central Research Institute of Viticulture and Oenology (Table 1).

Table 1. Ecological characteristics for vine areas (source: Szenteleki et al., 2007)

Factors	Maximum points
Climatic factors (winter and vegetative spring-fall frequency of frost)	95
Soil factors (soil type, subsoil, physical characteristics, water supply, humus content, homogeneity, water content of the soil, danger of erosion/blow-out depression)	112
Geographic terrain factors (direction of slope and exposure, height above sea level, discharge of the cold, need to prepare the site/necessity for terracing)	175
Surrounding area (forest, building structures, accessibility)	18

The aim of the present study was to evaluate the expected future climatic conditions in the Hungarian wine regions and compare the possible changes with the present situation. Soil factor was partly considered, but we assume that the other factors (elevation, slope and aspect) will not change, so these factors have not been involved in the analysis.

Materials and methods

Meteorological data

WorldClim is a set of global climate layers (climate grids) with a spatial resolution of 30 arc seconds, often referred to as 1 km spatial resolution. The climate elements considered are monthly precipitation and mean, minimum and maximum temperature. Input data were gathered from various sources and many meteorological stations from the 1950-2000 period, and were interpolated using the thin-plate smoothing spline algorithm (Hijmans et al., 2005). The high resolution of this database does not imply

that the quality of the data is necessarily high in all places. It depends on the local climate variability and on the quality and density of the observations of a given area.

For the validation, observed monthly mean temperature and precipitation data of 32 Hungarian meteorological stations were used, from the period of 1961-1990. Paired *t*-tests were carried out for seasonal data. We found that the precipitation values of the WorldClim database and the Hungarian observations can be considered the same, but there are differences in the monthly mean temperatures. In two stations (Martonvásár and Szabadbattyán) the overestimations of the WorldClim database sometimes are greater than 1°C. However there are strong correlations between the datasets, around 0.88 considering all the 32 meteorological stations, and over 0.9 without the two problematic stations. Taking into account that the periods of the two datasets are not exactly the same, and there is no information about the quality of the observed data, WorldClim database was accepted as a good characterisation of the present climate in Hungary.

Expected regional climate change focused to the Carpathian Basin is modelled by four different RCMs, run by the Department of Meteorology, Eötvös Loránd University, Hungary and by the Hungarian Meteorological Service (Bartholy et al., 2009). Climate scenarios applied in the present study were provided by the Department of Meteorology of the Eötvös Loránd University. They applied the high-resolution version of the Regional Climate Model (RegCM3) over the Carpathian basin using the A1B scenario (Bartholy et al., 2009; Torma et al., 2011). The horizontal grid spacing of this dataset is 10 km – the highest reached by RegCM3 model – and the database contains daily data of several climatic elements for the baseline period (1961-1990) and two time-slices in the future (2021-2050 and 2071-2100). From the daily data monthly precipitation and mean, minimum and maximum temperature data were calculated.

Using the climate scenarios the changes (differences in case of temperature and ratio in case of precipitation) were calculated compared to the baseline, then added (multiplying in case of precipitation) to the higher resolution WorldClim dataset.

Grapevine growth is basically determined by the climatic potential of a region, described with different thermal indices. From the present and the expected future climate data biological effective degree days between 10°C and 19°C (Gladstone index, *BEDD*), as well Winkler (*WIN*) and Huglin (*HI*) indices were calculated. Minimum temperature of the coldest (January, *T_Jan*) and the warmest month (July, *T_Jul*) were considered as limiting factors in grapevine cultivation. The cool night index (minimum temperature in September, *CI*), the average growing season temperature (from April to October, *T_avg*), the yearly average temperature (*T_year*) and the seasonal water balance, which profoundly influence grape and wine quality (Jones et al., 2005; Tonietto and Carbonneau, 2004) were additionally considered in the classification process (see Appendix for a comprehensive description of indexes calculations).

Soil data

Besides climate, soil is also an important environmental factor to which the grapevine is subjected. The Harmonized World Soil Database (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2009) was used to determine the soil water storage capacity values (*s_AWC*). The database distinguishes seven AWC classes. Water deficit values (*WAT_def*) were calculated for the present and for the two future time-slices based on a monthly water-balance model (McCabe and Markstrom, 2007; Gray and McCabe, 2010).

Wine regions and land use data

There are 22 wine regions in Hungary, which have their well defined and widely requested characters based on their typical varieties grown there. The map of the wine regions – provided by the Institute of Geodesy, Cartography and Remote Sensing – is available only in picture (*.jpg) format. Therefore, the first step was the digitalization and creation of a georeferenced GIS database of the regions.

The wine regions are potential areas for growing quality wine. The pixels defining the current grapevine cultivated area over each wine region were derived from CORINE Land Cover database (CORINE 2006), at a spatial resolution of 100x100 m. Considering the mismatch of spatial scale between CORINE and WorldClim (i.e 100 m and 1 Km), before downgrading the CORINE information at a lower spatial resolution, losing much information, 1 Km buffers were created around the CORINE vineyard pixels.

CORINE dataset was additionally used to define the present agricultural land use types in the predicted potential grapevine areas.

Extension of the dataset

The whole area examined was $861 \times 407 = 350427$ pixels, from which

- 158924 pixels are within the borders of Hungary
- 41105 pixels belong to the Hungarian wine regions
- 9081 pixels are Hungarian vineyards based on the CORINE dataset (*Fig. 1*).

According to the original datasets, all maps were created using the WGS-84 coordinate system.

Climate data for the present period are available for the whole area, but the climate scenarios characterise only Hungary.

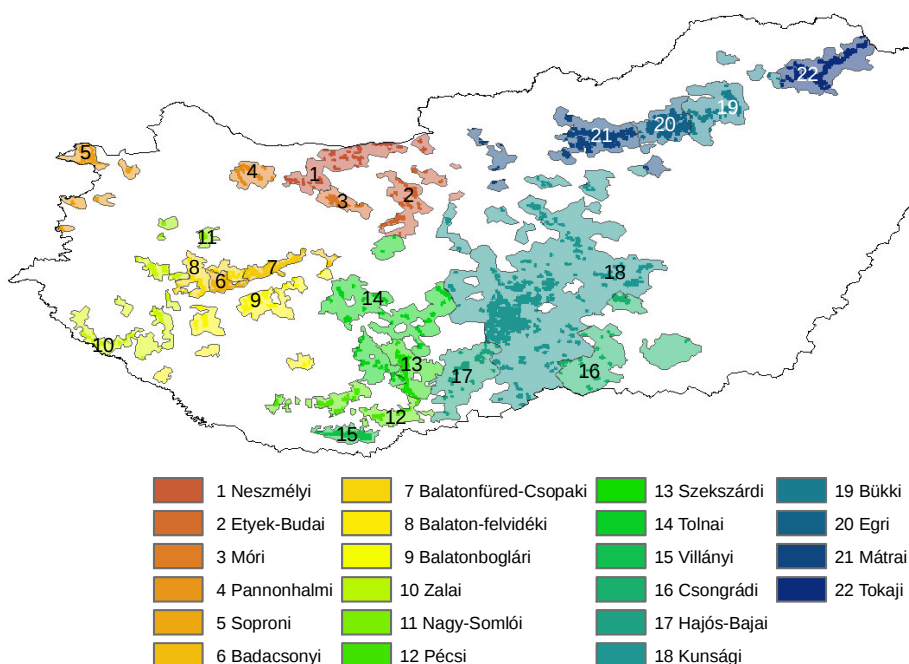


Figure 1. Location, numbers and names of the Hungarian wine regions
the small raster cells indicate the vineyards based on CORINE 2006

Random forest classification

Random forest (RF) classification has become a widely-used predictive model in many scientific disciplines within the past few years. The method is receiving much attention in ecological studies, especially for predicting the effect of climate change on species distribution (e.g. Cutler et al., 2007; Benito Garzón et al., 2008; Attore et al., 2011; Evans et al., 2011; Vaca et al., 2011). The unique advantage of the machine learning algorithm is that complex relationships and spatial patterns can be discovered readily (Evans et al., 2011). It can also be applied when predictor variables are highly correlated (Strobl and Zeileis, 2008).

The RF is a classification method that basically consists of a combination of decision trees where each classifier is generated using a bootstrap sample. The bootstrap sample is randomly split into two subsets, which are used for training (66%) and for internal testing (33%, out-of-bag sample, OOB). A classification tree is fit to each bootstrap sample where each node within a tree is constructed by selecting a random subset of the environmental variables (for this parameter *mtry* was set to 4). This feature reduces the problem of correlated variables because these may be extracted in turn, thus contributing independently to the aggregated tree model.

Each tree is fully grown until a final node is reached and then it is used to predict the classes of OOB observations. This procedure is repeated until the desired number of trees has reached. The algorithm includes the computation of the OOB error estimate, which is calculated for each tree over the data split out of the corresponding bootstrap sample, and then averaged. Because the OOB observations are not used in the training of the trees, these are essentially cross-validated accuracy estimates. According to Evans et al. (2011) model selection was based on a minimization of both “out-of-bag” (OOB) error and largest “within-class” error estimates of several runs.

In the prediction mode, a calibrated RF model consists of an ensemble of classification trees, each of which is allowed one vote for the model prediction. The most voted prediction from all of the trees in the random forest becomes the final model prediction.

The RF algorithm also provides measures of variable importance. The most often used measure is based on the decrease of classification accuracy. The mean decrease in accuracy for a variable is the normalized difference of the classification accuracy for the out-of-bag data when the data for that variable is included as observed, and the classification accuracy for the out-of-bag data when the values of the variable in the out-of-bag data have been randomly permuted. The higher mean decrease accuracy (also called permutation importance) value means the higher importance of a given variable. According to Strobl et al. (2009) all variables whose importance is negative, zero or has a small positive value that lies in the same range as the negative values, can be excluded from further exploration, but during the analyses only positive importance values were found. The interpretation or comparison of the importance measures should rely only on a descriptive ranking of the predictor variables, not on the absolute values.

Calculations were done using the ‘randomForest’ package available in R environment. RF classification was firstly calibrated for the present period and used to derive information on the relative importance of the considered parameters. The calibrated RF model was then applied to predict the possible impact of climatic change on wine regions.

Since the RF algorithm should be preferably trained using a similar amount of test cases and considering that our dataset is highly unbalanced towards areas not covered

with grapevine (94% of the country) we trained RF algorithm in two sequential steps. In the first stage RF was trained to identify the areas suitable for grapevine cultivation, using the vineyards layer as binary response variable (1=presence, 0=absence). As a second step, RF model was trained to discriminate amongst different wine regions and applied to those pixels which were classified as part of grapevine cultivated area.

In the first case, a preliminary test was initially performed to identify the errors associated to different type of training dataset. Selecting the presence-absence values from the whole country – not considering the unbalanced dataset –, the OOB error is low (around 4.3%), but the classification error of the vineyards is very high (more than 50%). Performing a calibration/validation process in this case (80% for calibration and 20% for validation), only 0.2% misclassification can be seen, but the ratio of the vineyard pixels (due to the unbalanced sample) is very low and this can lead to a false evaluation. As the most commonly used classification algorithms, in fact, RF aims to minimize the overall error rate rather than paying special attention to the minority class. In other terms, RF may result in a very good prediction accuracy even misclassifying all the test cases of the minority class.

Using a dataset with almost the same number of pixels belonging to cultivated and not cultivated grapevine areas, the OOB error is slightly increasing, but the classification error decreases.

As a final solution grapevine areas were sampled with all of the CORINE vineyard pixels (9081 pixels), while the non-cultivated areas were sampled with almost the same number of pixels randomly selected within the boundaries of the country, but outside of the wine regions (using the *sample* function the probability of selecting the pixels was set to 7.7% of the 117822 pixels, resulting samples between 8937 and 9211 pixels).

As the different runs of the model – based on randomly selected samples – give slightly different results, five runs were iterated in both present (1950-2000) and future time slices (2021-2050 and 2071-2100), where the most frequently predicted value was accepted.

Results

Model calibration and validation

Grapevine cultivated area

RF was firstly trained to identify the areas suitable for grapevine cultivation on national scale by selecting almost the same number of pixels for grapevine presence and absence (vineyards and no-vineyard areas). Pixels where grapevine is cultivated were sampled with the CORINE vineyards, pixels where grapevine is not cultivated were selected within the boundaries of Hungary, but outside of the wine region areas. Performing a calibration/validation process for this sample set, the validation showed total agreement with the original presence-absence values.

The RF internal validation resulted into a rather satisfying OOB error (9.4%) where the classification error was always higher in case of the no-vineyards (around 11%). Owing these results, the calibrated RF model was considered robust and coherent and applied to the relevant predictor variables calculated for the 2021-2050 and 2071-2100 time slices to derive the grapevine cultivated areas.

WAT_def, *T_Jan*, *T_year*, *HI* and *CI* were the parameters having the major importance to predict the presence of grapevine, while *T_Jul*, *BEDD*, *WIN* and *T_avg* were those having the lowest impact for the prediction (Fig. 2).

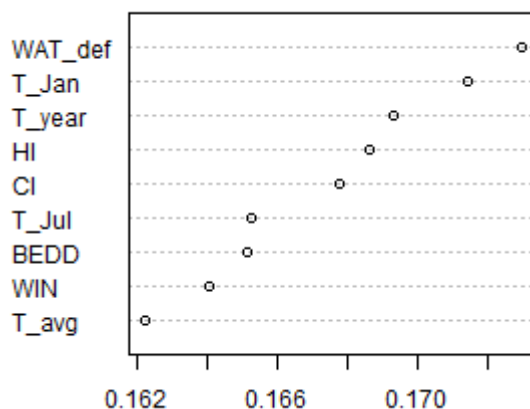


Figure 2. Variable importance based on the mean decrease accuracy in case of vineyard and no-vineyard areas

Analysing the 22 wine regions

A second RF model was calibrated to discriminate the 22 Hungarian wine regions.

The result showed nearly complete classification accuracy with an OOB of 6.36% (average of the five runs) during the calibration stage. The highest misclassification can be found in Badacsonyi and Balaton-felvidéki wine regions, which are neighbours, and in case of the Nagy-Somló wine region, which has the fewest pixels (class errors averaged for the five runs were 22, 23 and 21%, respectively). In the validation test, dataset was randomly split into 80% and 20% of the total cases, which were used to calibrate and validate RF, respectively. The prediction showed almost total agreement with the original regions.

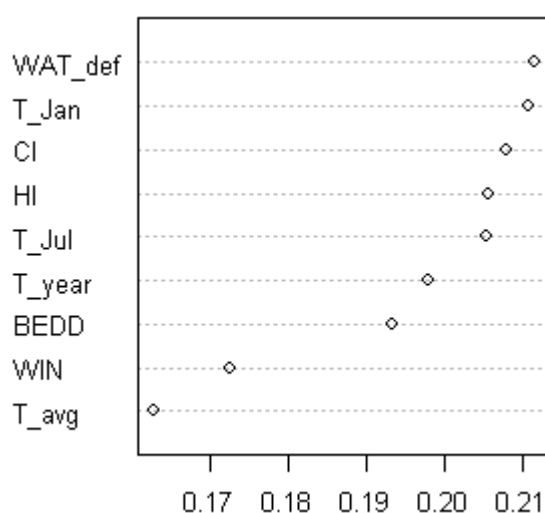


Figure 3. Variable importance based on the mean decrease accuracy in case of the 22 wine regions

WAT_def and T_Jan proved to be the most important. The cool night index (CI), Huglin index (HI) and maximum temperature of July (T_Jul) have similar importance values, and their relative rank varied depending on the random samples (Fig. 3).

The box-plots of the first three important variables may be used to represent the main climatic differences among the wine regions. The whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box.

The water deficit values (Fig. 4) show high variability among the wine regions, but also within the regions. The lowest values can be observed in the western part of the country, in Zalai (10) and Soproni (5) wine regions, while the highest deficit is in Móri (3) wine region.

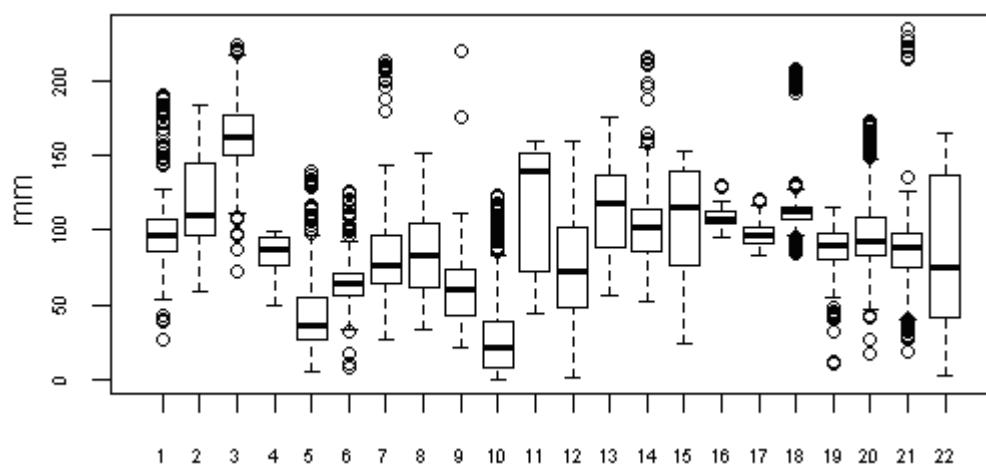


Figure 4. Box-plots of the water deficit (WAT_def) values in the 22 wine regions
1 Neszmélyi, 2 Etyek-Budai, 3 Móri, 4 Pannonhalmi, 5 Soproni, 6 Badacsonyi, 7 Balatonfüred-Csepaki, 8 Balaton-felvidéki, 9 Balatonboglári, 10 Zalai, 11 Nagy-Somlói, 12 Pécsi, 13 Szekszárdi, 14 Tolnai, 15 Villányi, 16 Csongrádi, 17 Hajós-Bajai, 18 Kunsági, 19 Bükki, 20 Egri, 21 Mátrai, 22 Tokaji

Based on the minimum temperature in January (Fig. 5) the regions in the northern part of the country (Bükki (19), Egri (20), Mátrai (21) and Tokaji (22)) are very different, the regions in the middle and southern part of the country can form another group, and the regions in the western part can be separated, too.

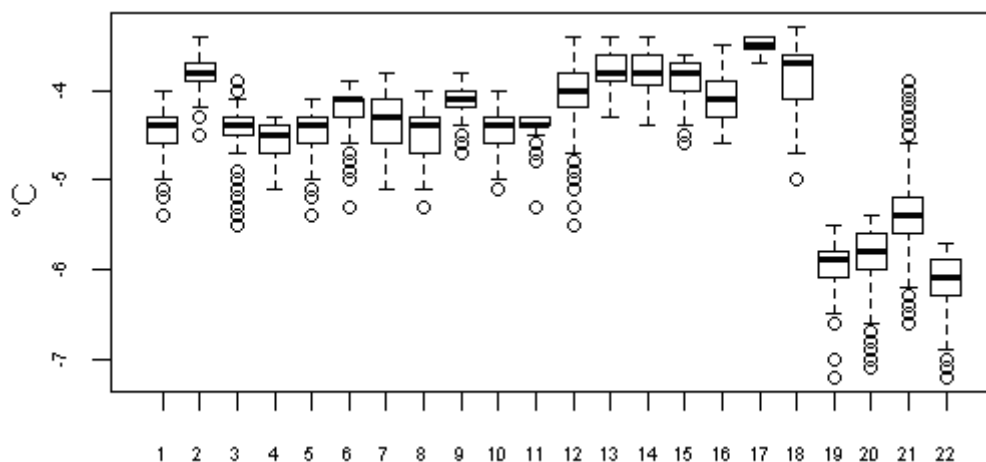


Figure 5. Box-plots of the min. temperature in January (T_Jan) in the 22 wine regions
(region names see at Fig. 4)

The regions in the northern part of the country are very different from the others also based on the cool night index (*Fig. 6*), but in this case Soproni (5) wine region is very similar to them. The regions in the middle and southern part of the country show the highest temperature, while among the remaining regions in the western part of the country there is still variability.

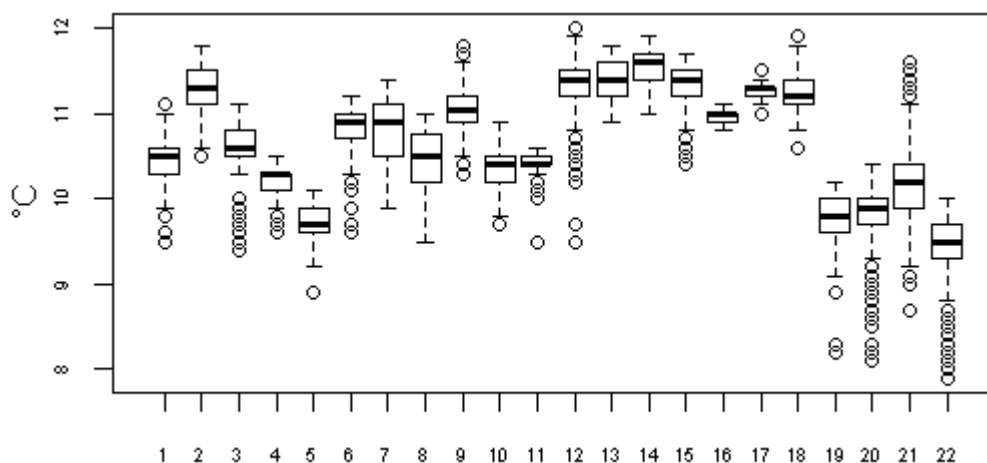


Figure 6. Box-plots of cool night index (CI) values in the 22 wine regions (region names see at *Fig. 4*)

To reveal the climatic structure of the wine regions Principal Component Analysis (PCA) was performed over the indexes averaged per wine regions. The first two components explain the 92.3% of the entire variability.

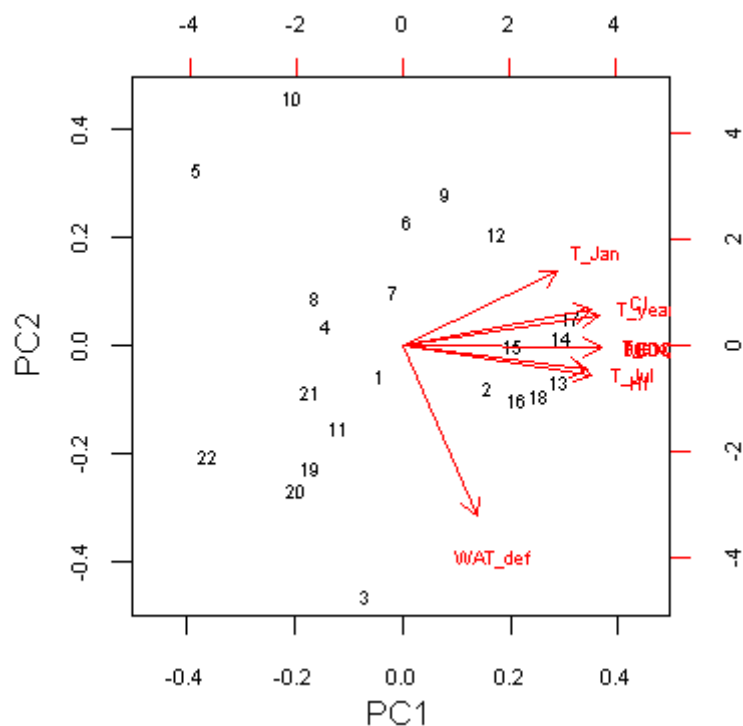


Figure 7. Principal Component Analysis for the wine regions (region names see at *Fig. 4*)

The plot of the factor scores of PC1 and PC2 (Fig. 7) shows that the first component (PC1) is positively correlated to all the predictor variables, but the most important ones are the temperature related indices of the growing season (T_{avg} , WIN and $BEDD$).

The warmer regions in the southern and middle part of the country have positive values, while the cooler regions located in the western and east part of the country have negative values.

The second component is determined by the negative effect of water deficit (WAT_{def}) and is positively correlated with the minimum temperature in January (T_{Jan}). Along this axis the outliers are Zalai (10) and Soproni (5) wine regions, where the water deficit is the lowest (compare with Fig. 4) as well as Móri (3) wine region with the highest deficit. The effect of T_{Jan} appears as a more or less diagonal arrangement in the figure, therefore regions with the lowest minimum temperature values in January can be found in the lower left part.

Prediction

Grapevine areas

The presence of grapevine cultivated areas was correctly predicted on national scale, considering that the model simulates a potential area (i.e. may include areas that are viable for cultivation) and that the predicted area is in any case included within the limits of wine regions. Additionally, the RF model correctly predicted the presence of grapevine areas outside the Hungarian boundaries (Fig. 8b), even though these regions were not considered in the calibration process.

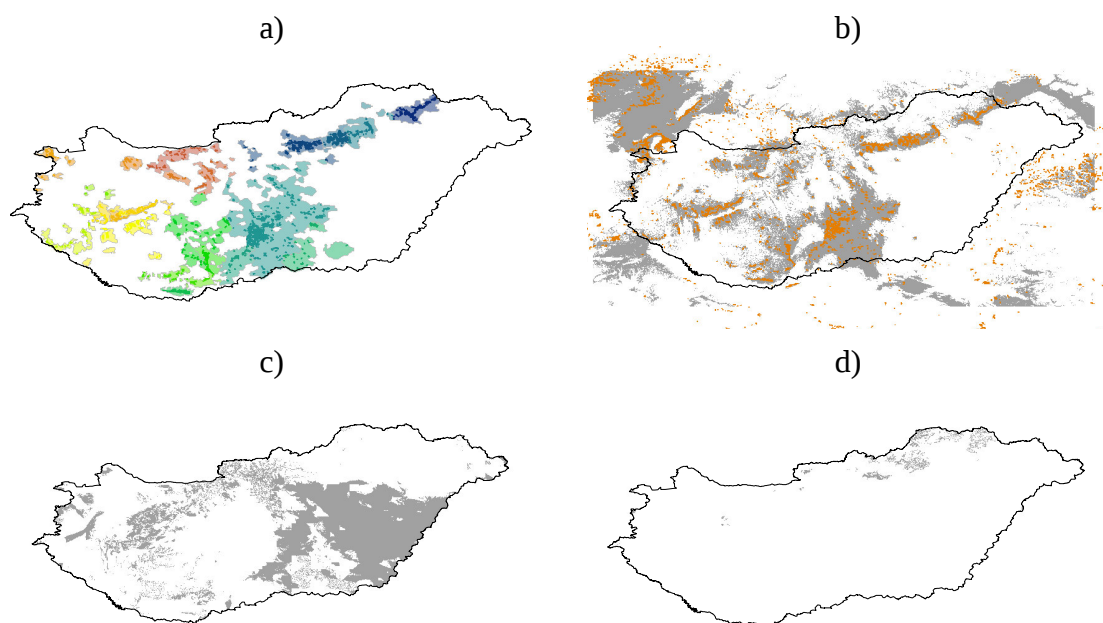


Figure 8. The present Hungarian wine regions and vineyards (a – for more details see Fig. 1), and the similar grapevine areas (with grey) based on the RF classification in the present (b), 2021-2050 (c) and 2071-2100 (d). In the prediction for the present the orange pixels show the CORINE vineyard areas

Prediction for the near future (2021-2050) shows a general shift towards the Great Plain (north-east). Surprising result is that in this time slice the simulation does not show any similarity with the present conditions in the north part, e.g. in the famous Tokaji wine region.

At the end of the century (2071-2100) we can hardly find areas similar to the present ones. These areas can be found only in the north part of the country, in accordance with the general geographical shift of the growing areas do to warming (*Fig. 8c and d*).

Predicted wine region types

In the second step, RF was used to predict the presence of specific wine regions within the predicted grapevine cultivated areas (*Fig. 9*).

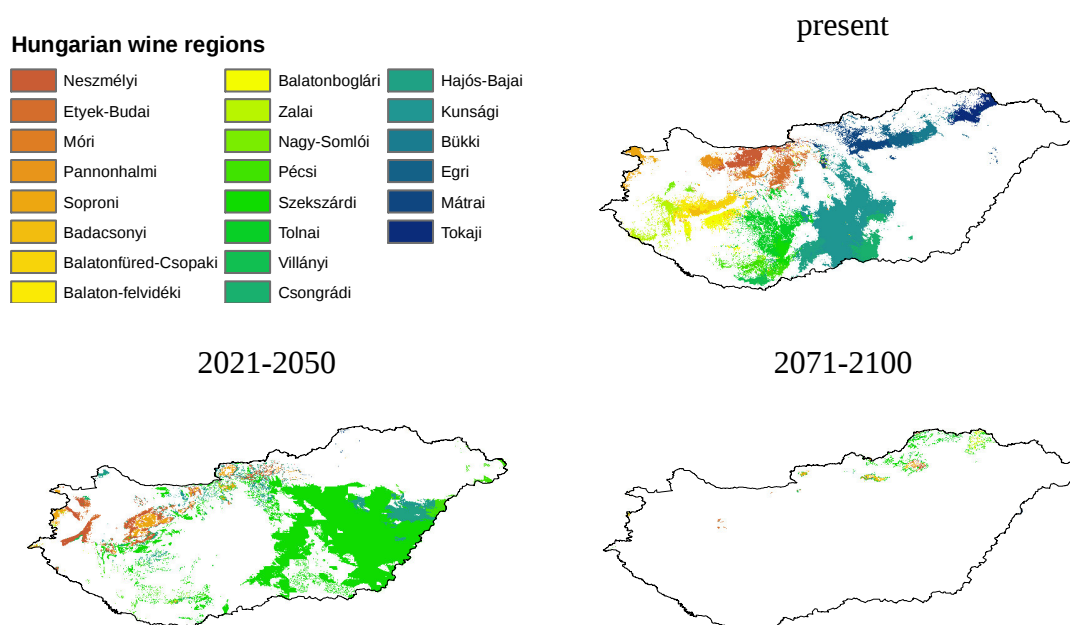


Figure 9. Predicted wine region types for the potential areas

For the present period, as expected, RF correctly simulated the placement of each wine region. In the period of 2021-2050 most of the predicted regions become similar to Szekszárdi and Hajós-Bajai wine regions, which are the warmest regions at present. In the Trans-Danubian area similarities with Neszmélyi and Soproni wine regions can be found.

In the period of 2071-2100 there would be a very little area similar to the present ones, and also the correspondence is not really defined, there is a mixed situation. There are similarities with Szekszárdi, Badacsony, Neszmélyi, Soproni, Balatonboglári and Zalai wine regions – at present located in the south and western part of the country – so a north-east shift can be observed.

Possible land use changes

At the moment, only a few areas where grapevine is predicted to shift are invested in viticulture. Therefore the possible changes in agricultural land use were examined considering the new areas viable for viticulture in future time slices.

Fig. 10 shows the present agricultural land use types of the predicted grapevine areas, and *Table 2* summarizes it. It can be seen, that the ratio of the vineyards in the predicted areas is only 2.4% and 3.4% in the two periods, respectively. The greatest part of the predicted areas – especially in the Great Plain in the near future – is out of the wine regions and the present land use type is non-irrigated arable land or pastures.

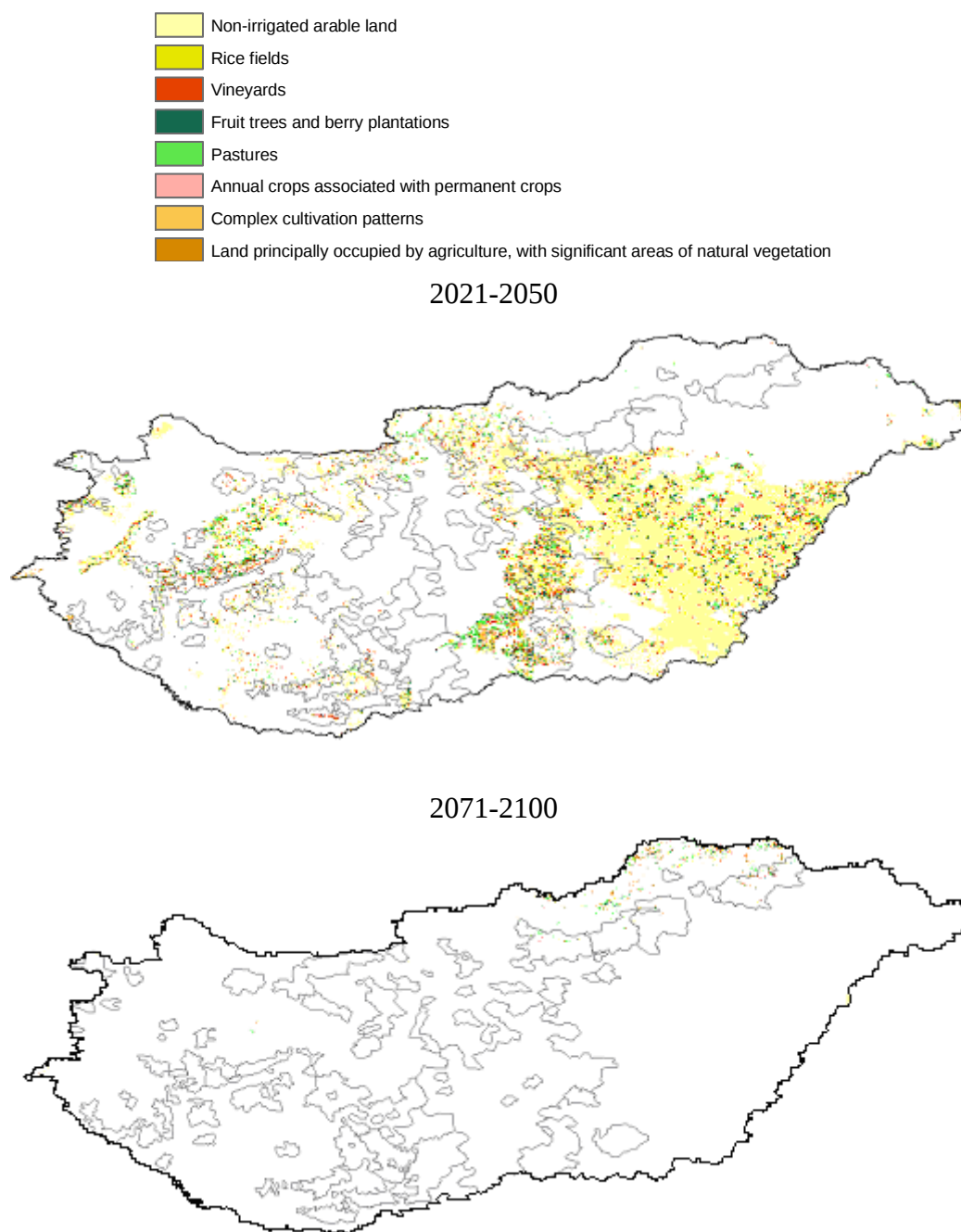


Figure 10. The present agricultural land use types of the predicted grapevine areas

Table 2. *The present agricultural land use types in the predicted grapevine areas*

Present land use type	% in the predicted area 2021-2050	% in the predicted area 2071-2100
Non-irrigated arable land	74.6	67.4
Rice fields	0.3	0
Vineyards	2.4	3.4
Fruit trees and berry plantations	0.8	1.0
Pastures	10.3	10.8
Annual crops associated with permanent crops	-	-
Complex cultivation patterns	6.5	9.1
Land principally occupied by agriculture, with significant areas of natural vegetation	5.1	8.4

The borders of the present wine regions are defined by the Hungarian wine law. One of the adaptation strategies should be the modification of the borders, the other one the modification of the land use structure within the borders of the regions. *Table 3* shows the ratio of the agricultural areas predicted for viticulture compared to the area of the wine regions, as well as the present ratio of the vineyards. According to the data in some regions the expansion of the grapevine areas should be expected in the near future (2021-2050), e.g. Móri, Soproni, Badacsonyi, Balaton-felvidéki, Csongrádi, Kunsági and Mátrai. Some regions, e.g. Szekszárdi and Villányi (which are among the warmest regions), can lose grapevine areas or have to cultivate varieties that are different from the ones at present.

As it was presented earlier, in the period of 2071-2100 there would be a very little area similar to the present ones. From that we can find some parts in the northern wine regions (Bükk, Egri, Mátrai and Tokaji), but the agricultural land use type in the predicted areas is negligible compared to the present vineyards area.

Table 3. *The area of the present vineyards and the agricultural lands predicted viable for viticulture compared to the area of the wine regions (based on the CORINE 2006 database)*

Wine region	Present (%)	2021-2050 (%)	2071-2100 (%)
Neszmélyi	3.93	11.25	0.00
Etyek-Budai	5.09	3.60	0.00
Móri	5.10	21.46	0.00
Pannonhalmi	5.53	5.00	0.00
Soproni	6.96	15.07	0.00
Badacsonyi	19.08	29.95	0.00
Balatonfüred-Csopak	17.23	21.79	0.00
Balaton-felvidéki	6.57	12.30	0.00
Balatonboglári	5.70	6.46	0.00
Zalai	4.73	4.14	0.00
Nagy-Somlói	5.74	1.22	0.00
Pécsi	4.98	6.13	0.00
Szekszárdi	6.20	3.25	0.00
Tolnai	2.64	1.30	0.00
Villányi	13.05	3.39	0.00
Csongrádi	2.12	29.32	0.00
Hajós-Bajai	2.91	4.62	0.00
Kunsági	5.10	26.3	0.00
Bükk	4.67	1.30	0.03
Egri	13.96	2.93	0.01
Mátrai	8.86	26.85	0.05
Tokaji	9.57	0.00	0.23

Expected changes based on the climatic indices

The climatic classification of the grapevine areas is usually done based on the average growing season temperature (*Fig. 11* and *Table 4*) and the Huglin index (*Fig. 12* and *Table 5*). Both parameters indicate that the wine regions are expected to be warmer with usually one category for each period.

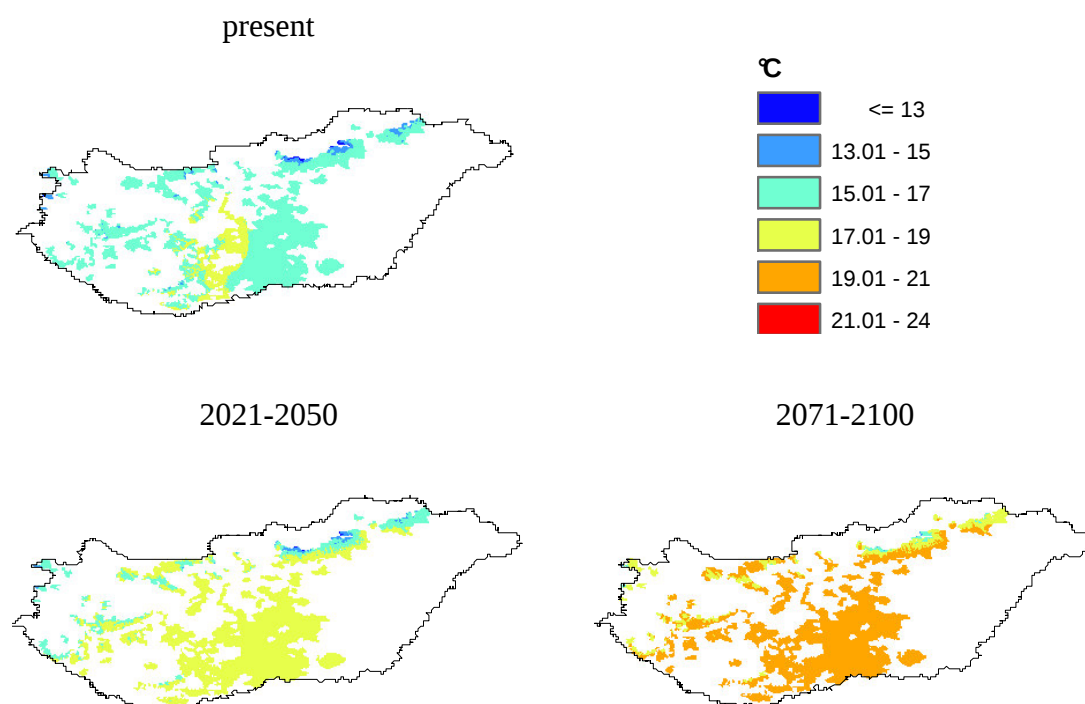


Figure 11. Average growing season temperature in the periods examined

Table 4. Average growing season temperature (Apr-Oct) regarding grapevine maturity (source: Jones et al., 2010)

Temperature (°C)	Climate group
< 13	too cool
13 - 15	cool
15 - 17	intermediate
17 - 19	warm
19 - 21	hot
21 - 24	very hot
24 <	too hot

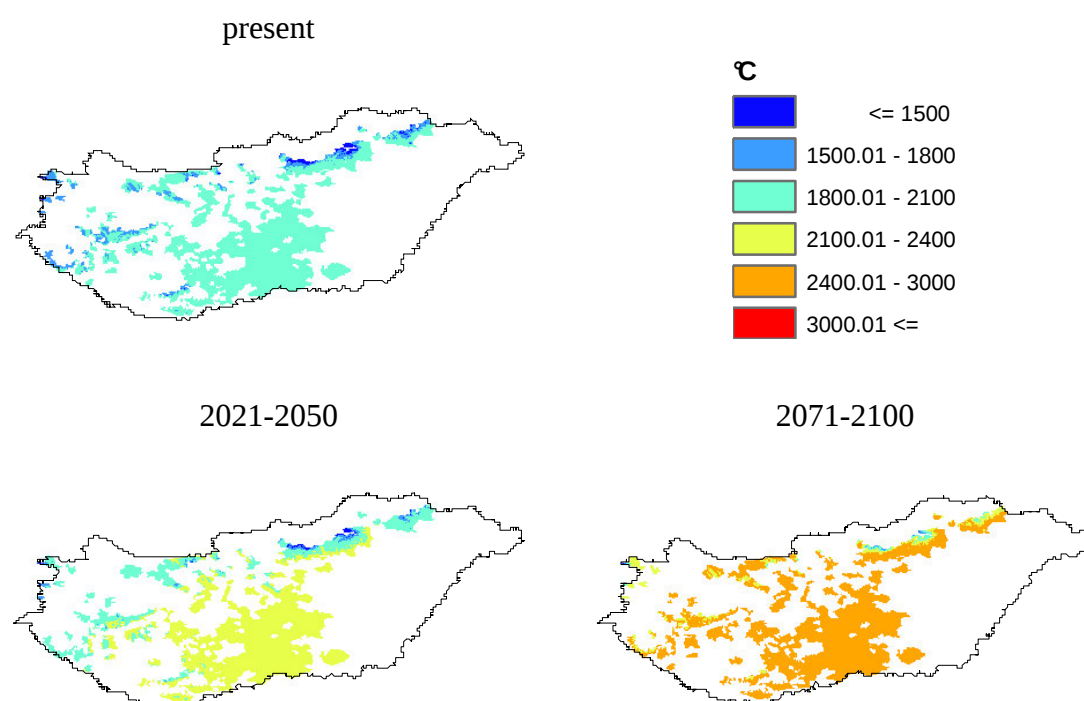


Figure 12. Huglin Index values in the periods examined

Table 5. Groups of site types according to their Huglin Index
(source: Tonietto and Carbonneau, 2004)

Huglin Index (°C)	Class name
$HI \leq 1500$	very cool
$1500 < HI \leq 1800$	cool
$1800 < HI \leq 2100$	temperate
$2100 < HI \leq 2400$	warm temperate
$2400 < HI \leq 3000$	warm to very warm
$3000 < HI$	hot

These maps are in accordance with the formerly presented results of the RF predictions, and confirm that in the period of 2021-2050 the greatest similarity would be with the present southern regions (e.g. Szekszárdi wine region), and in the period of 2071-2100 only in the northern part of the country can be found areas similar to the present climatic conditions.

In the areas – first of all in the south part of the country – where the RF classification does not predict areas similar to the present ones, growers must prepare to apply new varieties and/or agrotechnics in the future. However, these regions remain suitable for quality grapevine growing. The “too hot” category is not expected to appear in the country, even at the end of the century.

Discussion

In this paper a set of widely used viticultural climatic indices were used to analyse the spatial distribution of the Hungarian wine regions. This process was performed by Random Forest, a machine learning approach, which has been demonstrated to be one of the most promising techniques for ecological classification (Cutler et al., 2007; Moriondo et al., 2008, Evans et al., 2011). This approach allowed identifying the potential shift/contraction of the wine regions according to the expected climate change.

In the first step RF was trained to identify the areas suitable for grapevine cultivation, using the present vineyards data as binary response variable (1=presence, 0=absence). This analysis resulted into a rather satisfying OOB error (9.4%), therefore the calibrated RF model was considered robust and coherent and applied to the relevant predictor variables calculated for the 2021-2050 and 2071-2100 time slices to derive the potential grapevine cultivated areas.

In the second step, RF was trained to discriminate amongst the different wine regions. The result showed nearly complete classification accuracy with an OOB of 6.36%. The highest misclassification can be found in some neighbouring regions, and in case of the Nagy-Somló wine region, which has the fewest pixels. This problem of the imbalanced response variables is mentioned by Evans et al. (2011), too.

The analysis of the variable importance indicated that water deficit (*WAT_def*) was the most important variable in the classification process, and this may be related to the high degree of information of this index, which depends on day length, average temperature and cumulated rainfall during the growing season, soil depth and texture. Minimum temperature of January (*T_Jan*) proved to be the second more important. The reason of it should be the significantly cooler values of the regions in the north part of the country (regions 19-22), and the warmer areas in the south and middle part (regions 12-18) compared to the others (*Fig. 5*). The cool night index (*CI*), Huglin index (*HI*) and maximum temperature of July (*T_Jul*) have similar importance values. In case of the *CI* the regions in the northern part, while in case of the *HI* and *T_Jul* the regions in the south and middle part (the warmer parts of the country) show great differences compared to the others.

Huglin Index proved to be a better predictor than the other similar climatic indexes (*WIN*, *BEDD*, *T_avg*) calculated for the growing season. This may be related to the additional information included in this index such as the mean day length in relation to the latitude, second to the fact that the calculation of the thermal component is estimated over the mean day length when most of metabolisms are active. The same result was found by Moriondo et al. (in progress) analysing the European wine regions.

Results show that in the near future (2021-2050) the present climatic conditions of the southern regions can be expected in greater part of the country. Some regions may benefit from the changes, as in their area there are predicted grapevine areas with other agricultural land use types, which may be changed. However, at the end of the century (2071-2100) only in the northern part of the country shows some similarities with the present climate. Despite it, Hungary is expected to remain amongst the regions with good quality grapevine growing conditions, but the structure of the cultivation and/or varieties should be changed. These changes were demonstrated also with the average growing season temperature and Huglin Index values (*Fig. 11 and 12*). Both parameters indicate that the wine regions are expected to be warmer with usually one category for each period.

Acknowledgements. The research was carried out in the frame of the Corvinus Visiting Scholar program supported by the TÁMOP-4.2.1/B-09/1/KMR-2010-0005 project. Many thanks to Prof. Marco Bindi (University of Florence) for hosting at his research group.

REFERENCES

- [1] Attore, F., Alfó, M., De Sanctis, M., Francesconi, F., Valenti, M.V., Bruno, F. (2011): Evaluating the effects of climate change on tree species abundance and distribution in the Italian peninsula – *Applied Vegetation Science* 14: 242-255.
- [2] Bartholy J., Pongrácz R., Torma Cs., Pieczka I., Hunyady A. (2009): Regional climate model experiments for the Carpathian basin. – 89th AMS Annual Meeting / 21st Conference on Climate Variability and Change.
<http://ams.confex.com/ams/pdfpapers/147084.pdf>. Phoenix, AZ. 5p.
- [3] Benito Garzón, M., Sánchez de Dios, R., Sainz Ollero, H. (2008): Effects of climate change on the distribution of Iberian tree species – *Applied Vegetation Science* doi: 10.3170/2008-7-18348
- [4] Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawer, J.J. (2007): Random forests for classification in ecology – *Ecology* 88(11): 2783-2792.
- [5] Evans, J.S., Murphy, M.A., Holden, Z.A., Cushman, S.A. (2011): Modeling species distribution and change using Random Forest – In: Draw, C.A. et al. (eds) *Predictive species and habitat modeling in landscape ecology: Concepts and applications*, DOI 10.1007/978-1-4419-7390-0_8, Springer Science+Business Media
- [6] FAO/IIASA/ISRIC/ISS-CAS/JRC (2009): Harmonized World Soil Database (version 1.1). – FAO, Rome, Italy and IIASA, Laxenburg, Austria.
- [7] Gladstones, J. (1992): *Viticulture and Environment*. – Winetitles, Adelaide.
- [8] Gray, S.T., McCabe, G.J. (2010): A combined water balance and tree ring approach to understanding the potential hydrologic effects of climate change in the central Rocky Mountain region. – *Water resources research* 46, W05513, doi:10.1029/2008WR007650
- [9] Hamon, W.R. (1961): Estimating potential evapotranspiration. – *Journal of the Hydraulics Division, Proc. of the American Society of Civil Engineers* 87: 107-120.
- [10] Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A. (2005): Very high resolution interpolated climate surfaces for global land areas. – *International Journal of Climatology* 25: 1965-1978.
- [11] Hlaszny, E., Hajdu, E., Bisztray, Gy., Ladanyi, M. (2012): Comparison of budburst models predictions for Kékfrankos. – *Applied Ecology and Environmental Research*, 10(1): 75-86.
- [12] Holland, T., Smit, B. (2010): Climate change and the wine industry: current research themes and new directions. – *Journal of Wine Research* 21(2-3): 125-136.
- [13] Huglin, P. (1978): Nouveau mode d'évaluation des possibilités héliothermiques d'un milieu viticole. – *C. R. Acad. Agr. France*, 1117-1126.
- [14] Hunter, J.J., Bonnardot, V. (2011): Suitability of some climatic parameters for grapevine cultivation in South Sfrica, with focus on key physiological processes. – *S. Afr. J. Enol. Vitic.* 32(1): 137-154.
- [15] Jones, G.V., White, M.A., Cooper, O.R., Storchmann, K. (2005): Climate change and global wine quality – *Climatic Change* 73: 319-343.
- [16] Jones, G.V. (2007): Climate Change: Observations, Projections, and General Implications for Viticulture and Wine Production. – Working paper No. 7., Economics Department, Whitman College, USA
- [17] Jones, G.V., Duff, A.A., Hall, A., Myers, J.W. (2010): Spatial analysis of climate in winegrape growing regions in the Western United States – *Am. J. Enol. Vitic.* 61(3): 313-326.

- [18] Kocsis, L., Taksonyi, P., Tarczal, E. (2010): The impact of climate change on grape production – Hungarian Agricultural Research 2: 15-18.
- [19] Ladányi, M. (2010): Climate indicator analysis specialized for viticultural researches. – In: Mihailovic, D.T., Lalic, B. (eds): *Advances in Environmental Modeling and Measurements*, Nova Science Publishers, Chapter 23. pp. 255-264.
- [20] Ladányi, M., Hlaszny, E. Pernes, Gy., Bisztray, Gy. (2010): Climate change impact study based on grapevine phenology modelling. – VIII. International Terroir Congress 2010 Soave (Vr) Italy 14-18 June 2010, 3: 65-71.
- [21] McCabe, G.J., Markstrom, S.L. (2007): A monthly water-balance model driven by a graphical user interface. – U. S. Geological Survey Open-File report 2007-1088, 6 p.
- [22] Moisselin, J.M., Schneider, M., Canellas, C., Mestre, O. (2002): Les changements climatiques en France au XX siècle: étude des longues séries homogénéisées de température et de précipitations. – *La Météorologie* 38: 45-56.
- [23] Moriondo M, Stefanini, F.M, Bindi, M. (2008): Reproduction of olive tree habitat suitability for global change impact assessment. – *Ecological Modelling* 218: 95-109.
- [24] Moriondo, M., Jones, G.V, Bois, B., Dibari, C., Ferrise, R, Trombi, G, Bindi, M (in progress): Wine regions shift in response to climate change
- [25] Strobl, C., Zeileis, A. (2008): Danger: High Power! – Exploring the statistical properties of a test for Random Forest variable importance – Technical Report 17: 1-8.
- [26] Strobl, C., Malley, J., Tutz, G. (2009): An introduction to recursive partitioning – Technical Report 55: 1-42.
- [27] Szenteleki, K., Botos, E.P., Szabo, A., Horvath, Cs., Martinovich, L., Katona, Z. (2007): Definition of the ecological facilities, ecological indicators and quality of products in the Hungarian vine and wine sector using updated GIS – EFITA Conference 2-5 July, 2007, Glasgow, CD-ROM
- [28] Szenteleki, K., Ladányi, M., Gaál, M., Zanathy, G., Bisztray, Gy. (2012): Climatic risk factors of Central Hungarian grape growing regions – *Applied Ecology and Environmental Research*, 10(1): 87-105.
- [29] Tonietto, J., Carbonneau, A. (2004): A multicriteria climatic classification system for grape-growing regions worldwide. – *Agricultural and Forest Meteorology* 124: 81-97.
- [30] Torma, Cs., Coppola, E., Giorgi, F., Bartholy J., Pongrácz R. (2011): Validation of a high resolution version of the regional climate model RegCM3 over the Carpathian Basin. – *Journal of Hydrometeorology* 12: 84-100.
- [31] Vaca, R.A., Golicher, D.J., Cayuela, L. (2011): Using climatically based random forests to downscale coarse-grained potential natural vegetation maps in tropical Mexico – *Applied Vegetation Science* 14: 388-401.
- [32] Winkler, A.J. (1938): The effect of climatic regions – *Wine Rev* 6: 14-16.
- [33] Zanathy, G. 2008: Gondolatok a klímaváltozás szőlőtermesztésre gyakorolt hatásáról – *Agro napló* 12(2): 92-94.
- [34] Zorer, R. (2008): Impacts of Climate Variability and Change on Wine Grape Quality. – *Klímaváltozás hatása a szőlő- és bortermelésre Konferencia*. 2008. november 24. Budapest.

Appendix

Calculation of the indices

Winkler Index (WIN)

The Winkler Index (Winkler, 1938) was introduced to identify the most suitable cultivars for a region. Winkler index is calculated as Growing Degree Days (GDD) between April and October, where daily degree days is given by the difference between mean daily temperature and 10°C when $T_{\text{mean}} > 10^\circ\text{C}$.

Biologically Effective Day Degrees / Gladstone index (BEDD)

The Biologically Effective Day Degrees (Gladstones, 1992) is a bio-climatic index commonly used for the assessment of the climatic suitability of certain areas for the grapevine cultivation. The BEDD is the same as the GDD process described above, only with the additional constraint that there is an upper limit to the accumulation of degree days. It is based on daily average temperature with a 19°C upper cut-off.

Huglin Index (HI)

The Huglin Index, also called Heliothermal Index (Huglin, 1978) was introduced to classify the different viticultural regions of the world in relation to the heliothermal conditions during the grape growth period. The index is calculated as the sum of daily mean and maximum temperature above +10°C from 1st April to 30th September (1st October to 31st March in the Southern Hemisphere). A coefficient takes into account for the mean day length in relation to the latitude ranging from 1.02 to 1.06 between 40° and 50°.

Cool Night Index (CI)

The Cool Night Index (CI) aims at improving the assessment of the qualitative potential of a suitable region by providing a measure of the coolness of the nights during the ripening period. It is based on the observation that warm night conditions are detrimental from a qualitative point of view (e.g. loss of aromas, lesser coloration in the red varieties) (Tonietto and Carbonneau, 2004). The index is the monthly average minimum temperature in September (Northern Hemisphere) or March (Southern Hemisphere).

Water deficit (WAT_def)

The water-balance model used in this work uses a monthly accounting procedure to analyses the allocation of water among various components of the hydrologic system and it is fully described in McCabe and Markstrom (2007).

The spatial mean monthly temperature (T , °C), monthly total precipitation (P , mm), and latitude ($^\circ$, used for the computation of day length), were used for the computation of potential evapotranspiration (PET). Actual evapotranspiration (AET) was then derived from PET, P , moisture storage in the soil (ST , mm), and soil-moisture storage withdrawal (STW , mm). Monthly PET was estimated from T according to the Hamon equation (Hamon, 1961):

$$PET = 13.97 \times d \times D^2 \times W_t$$

where PET is expressed in millimeters per month, d is the number of days in a month, D is the mean monthly hours of daylight in units of 12 hrs, and Wt is a saturated water vapor density term, in grams per cubic meter, calculated by:

$$Wt = (4.95 \times e^{0.062 \times T}) / 100$$

When P for a month is less than PET, then AET is equal to P plus the amount of soil moisture that can be withdrawn from storage in the soil. Soil-moisture storage withdrawal linearly decreases with decreasing ST such that as the soil becomes drier, water becomes more difficult to remove from the soil and less is available for AET. STW was computed as follows:

$$STW = ST_{i-1} - [\text{abs}(P_{\text{total}} - PET) \times ST_{i-1} / STC]$$

where ST_{i-1} is the soil-moisture storage for the previous month and STC is the soil-moisture storage capacity, which was derived on a spatial resolution coherent to the framework (i.e. 1km x 1 km) from the Harmonized World Soil Database (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2009) as a function of soil texture and depth. Monthly water deficit was then calculated as $PET - AET$ when the sum of P and STW was less than PET and summed up over the entire year (WAT_{def}).