Interreg
Danube Transnational Programme
DriDanube

Training manual on drought risk assessment

WP	5
Activity	5.4
Activity leader	OMSZ
Number and name of the deliverable/output	D 5.4.1 Training manual on use of Drought User Service – impacts (WP4) and risk assessment
Participating partners	OMSZ, ALL
Type of the deliverable/output (analysis, report, guideline, workshop, brochure, etc.)	Guideline
Purpose of the deliverable/output	Summary of drought risk assessment
Connection with other deliverables	Guideline on use of Drought User Service
Start date	01.01.2017
End date	30.09.2018

Contents

Theoretical background

Methodology of drought risk assessment

One of the aims of the DriDanube project is the drought risk assessment. However, unfortunately there is no commonly accepted procedure for this purpose. Therefore, it was necessary to review and prepare the methodology that can be applied for drought risk assessment in this project.

Methodology promotes the same approach to drought risk assessment as it is described in European Commission's Risk Assessment and Mapping Guidelines for Disaster Management.

In our development we focused on agricultural drought because we had impact data mainly from agriculture.

Main features of the method are as follows, detailed description can be read in Annex 1:

- Risk is defined as the expected value of the loss function
- If there isn't drought then loss function is 0
- If there is drought then loss function is the difference between the conditional expectation of yield function if there isn't drought and the yield function
- A special case is built in the algorithm: drought is defined by SPI, but any other drought index can be used.
- Relative yield values are used because the yield values are very different.
- Regression model was established for the relative yield function with the temperature mean and the logarithm of precipitation sum (3 or 6 months)

Steps of the algorithm

1. Selection of the relevant meteorological variables.

2. Collecting sample for the crop yield and the meteorological variables.

3. Estimation of the relative crop yield function that is a regression of the relative crop yield on the meteorological variables.

4. On the basis of the results of item 3 developing software for calculation of the following series and estimations in case of given meteorological data series:
   - Estimated relative crop yield functions
   - Several *SPI* series
   - Estimated meteorological drought risk

5. Applying software, risk calculations on station and gridded climate data series. Mapping of risk for CarpatClim, DanubeClim area based on gridded data series.

6. Mapping of risk outside CarpatClim, DanubeClim area based on E-OBS data series (Haylock et al, 2018").

Analysis of extreme rainless periods (droughts) as approach to drought risk representation

By most definitions, drought risk can be calculated if probability of drought occurrence and losses caused by drought are known. Aim of the Zelenhasic and Todorovic (ZT) method application was to calculate probability of rainless periods (droughts) occurrence in the DriDanube partner countries and to identify different stochastic characteristics of these events (rainless periods – droughts). Note that 'rainless period' and 'drought' are used in training manual text as synonyms.

Annex 3 provides mathematical background of the ZT method; here we present only its main features:

- it is a general stochastic model of extreme rainless events (droughts) at certain location;
- drought is defined as at least 20 consecutive days long period with less than 3 mm of daily rainfall;
- droughts are independent events, represented by identically distributed random variables that follow the Poisson probability law;
- method considers all important components of the process - drought duration, time of the occurrence, number of droughts in a given time interval [0,t], and the duration of the longest drought in a given time interval [0,t];
- method provides return periods of the longest droughts, i.e. probability of longest drought occurrence;
- application of the ZT method for the vegetation season - starting on 1st April and ending on 30th September.

For all selected locations in the DriDanube partner countries, methodological steps in analyzing rainless periods were as follows:

<u>At local level</u> (each selected location):

1. From meteorological databases (CarpatClim or its extension DanubeClim, or station data) download year-by-year series of daily rainfalls for historical period of 30 years (1981-2010).
2. Filter downloaded series to extract only rainfall data for vegetation season in each year.
3. Identify all rainless periods in each year longer than 20 days in the given time frame of 30 years for the vegetation season
4. Identify longest recorded (historical) rainless period by year, duration and starting and ending datum.
5. Calculate distribution of the number of rainless periods (Poisson).
6. Calculate distribution of the durations of all rainless periods (exponential).
7. Calculate distribution of the durations of longest (one by year) rainless periods (double exponential).
8. Compute return periods of the longest rainless periods based on related distribution function obtained in step 7.
9. Calculate distribution of the time of occurrence of rainless periods (Gamma).

<u>At regional level</u> (DriDanube countries):

10. Summarize and interpret the results from local level calculations
11. Geostatistical interpolation and mapping of the longest rainless periods for different return periods as obtained in step 8.

Input data

Input data for drought risk assessment

Data requirements from project partners were impact data (crop yield) and meteorological data for the step 3 of Algorithm i.e. sample for regression. Because of the strict data policy in the region the Consortium decided to provide monthly precipitation and temperature data for the same period and locations as the crop data (Figure 1, Figure 2, Table 2)

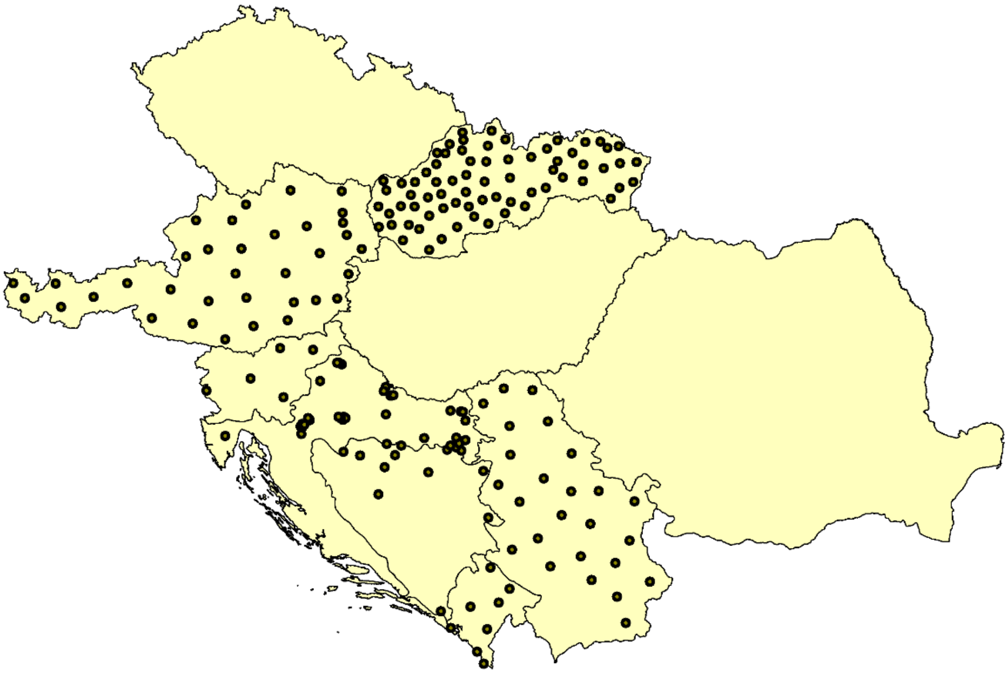


Figure 1 Crop yield data in partner countries

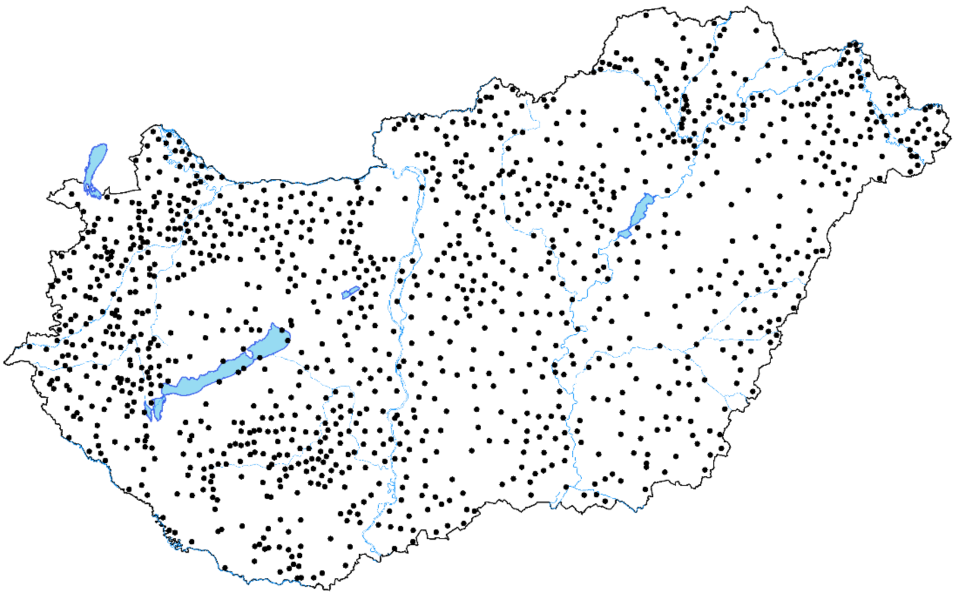


Figure 2 Yield data in Hungary

	Maize	Wheat	Barley	Rape	Oats	Sunflower	Sugar Beet	Potatoes	Pear	Apple	Plum	Citrus	Grape	Olive	Cabbage
Austria	X	X	X	X	X	X	X	X		X			X		
Bosnia and Hercegovina	X	X	X		X										
Czech Republic															
Croatia	X	X	X	X		X									
Hungary	X	X	X	X		X									
Montenegro	X	X						X	X	X	X	X	X	X	
Romania															
Serbia	X	X	X	X		X									
Slovakia	X	X	X	X		X	X	X							
Slovenia	X	X						X		X			X		X

Table 1 Yield data per countries

	Number of data series	Period (years)
Austria	20-30	2003-2014
Bosnia and Hercegovina	10-12	2001-2016
Czech Republic		
Croatia	34	4
Hungary	1000-8000	2001-2016
Montenegro	2-10	2001-2011
Romania		
Serbia	25	2004-2013
Slovakia	71	2000-2016
Slovenia	5	2010-2016

Table 2 Number of yield data series and their length

Input data for rainless periods assessment by the ZT method

Input data for the ZT method are daily rainfall data during vegetation season staring on 1st April and ending on 30th September within historical period 1981-2010 for 170 locations across DriDanube partner countries. Locations were proposed by the FAUNS and partners were asked to check representativeness and to approve appropriateness of spatial distribution of selected locations in each country. Figure below presents spatial distribution of locations used for rainless periods assessment. Input data for the ZT method are daily rainfalls downloaded from databases CarpatClim and DanubeClim as agreed by the Consortium at earlier stage of the project. Smaller part of missing data is obtained by direct communication with partners in Austria, Czech Republic and Slovenia.

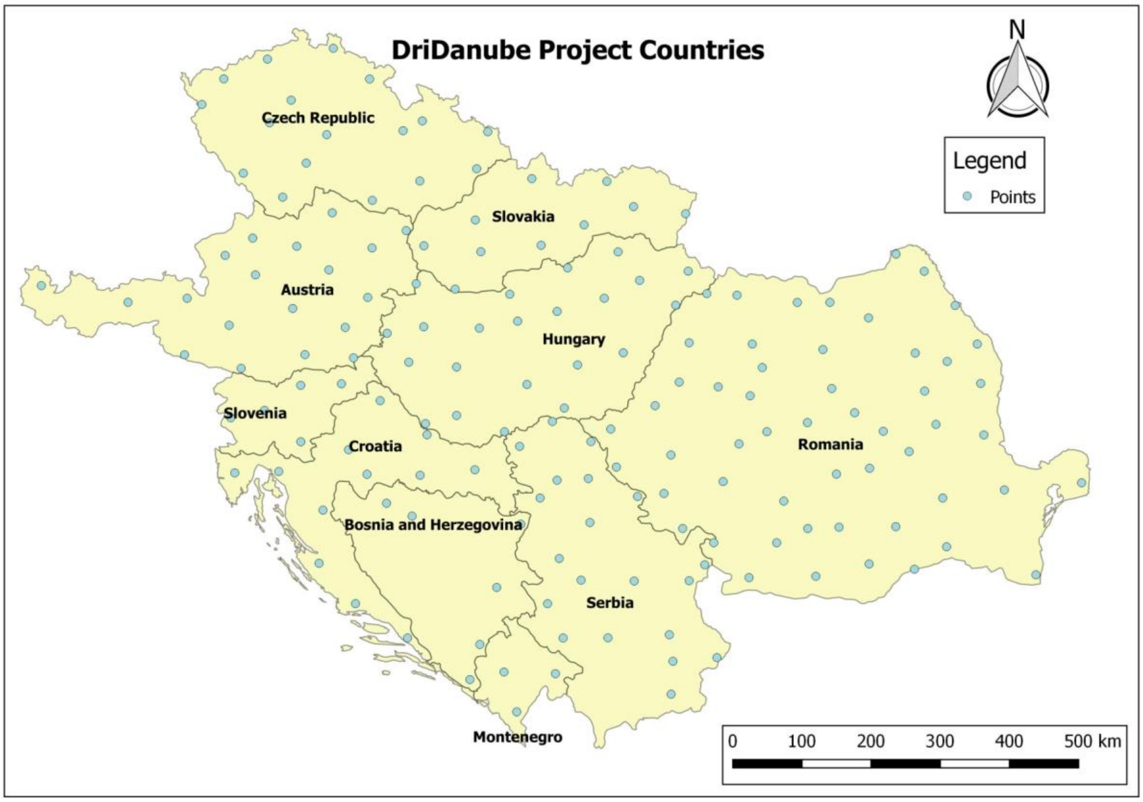


Figure 1 Points of daily rainfall data

Number of points per partner countries and source of daily rainfall data are provided in table given below.

Partner country	Number of points	Source of daily rainfall data
Austria	19	Missing information
Bosnia and Herzegovina	6	DanubeClim
Czech Republic	17	Missing information
Croatia	12	Observed data
Hungary	22	CarpatClim + DanubeClim
Montenegro	4	DanubeClim
Romania	43	CarpatClim
Serbia	20	CarpatClim + DanubeClim
Slovakia	22	CarpatClim
Slovenia	5	Observed data

Table 3 Number of points per partner countries and source of daily rainfall data

Outputs

Outputs of the drought risk assessment

The outputs of the drought risk assessment are:

1. Algorithm of drought risk assessment
2. Software of drought risk assessment
3. Manual for risk assessment
4. Drought risk maps

The outputs will be available in Drought User Service. The software RED (Risk Estimation of Drought) is freely available for project partners and for any other users. In the manual (Annex 2) users can read about the structure of the software.

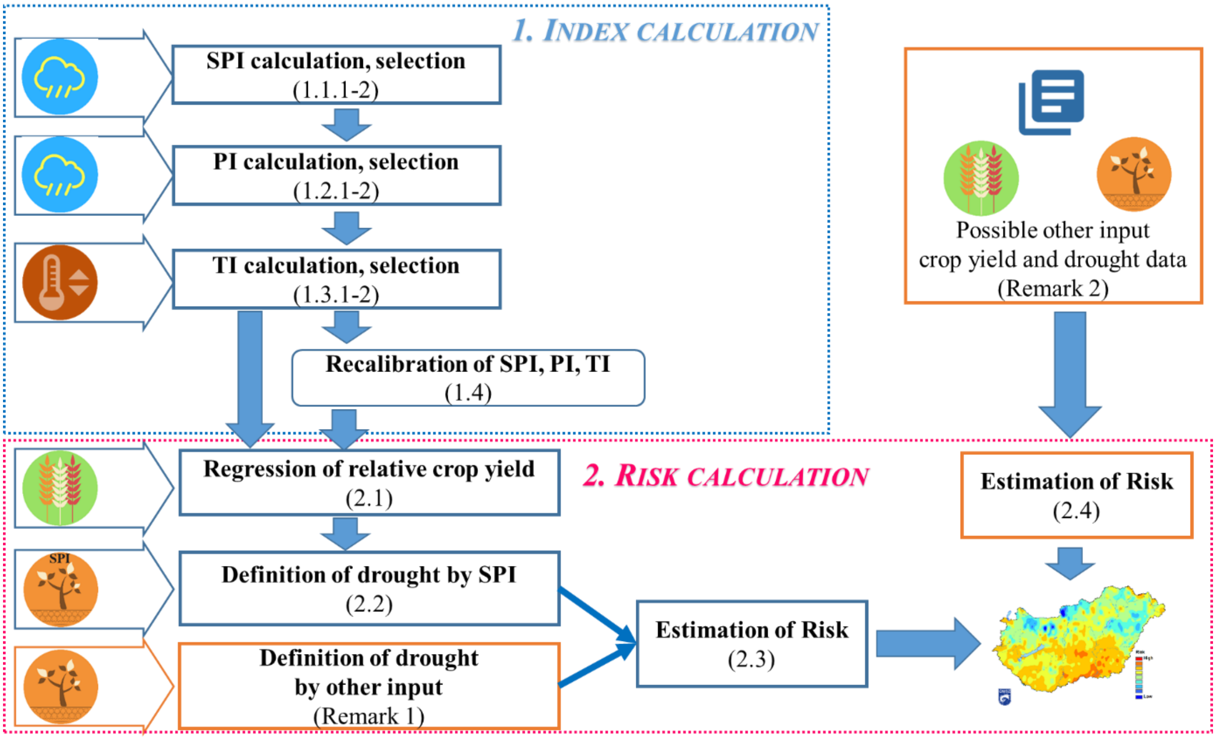


Figure 2. Flow diagram for RED (Risk Estimation of Drought) software

Final outputs of drought risk estimation are drought risk matrices (Figure 3 and 4) and maps (Figure 5-8). The maps for the whole DriDanube region will be available in Drought User Service.

Risk matrices were created on the outputs of the RED software. Regional and country wide averages were calculated from gridded risk values. The risk categories in cells with crop names come from calculated risk values, the risk categories in cells without crop names are estimations from the neigbouring cells. Same colours mean the same risk values both on matrices and maps for every crops and probability values.

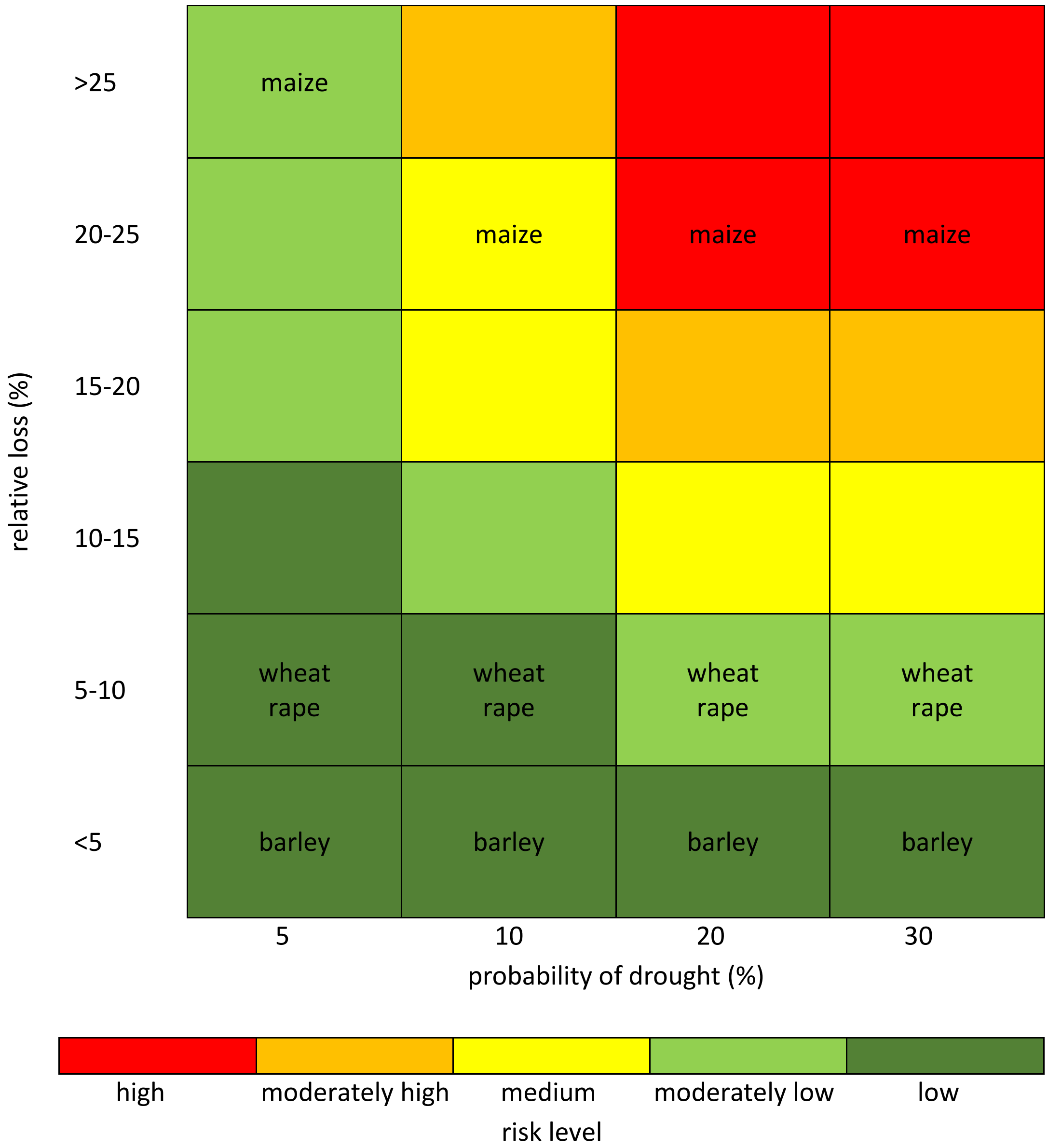
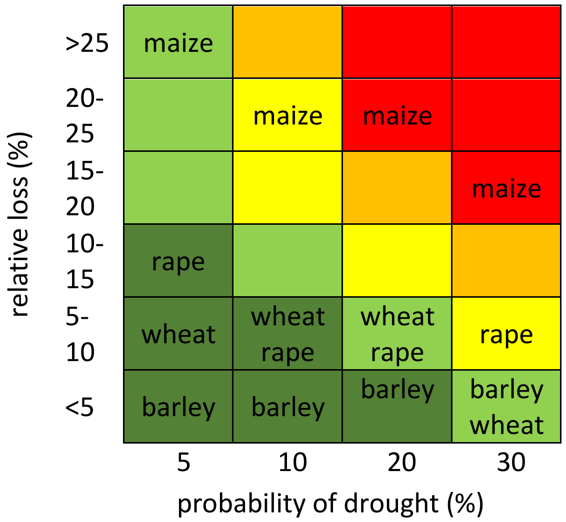
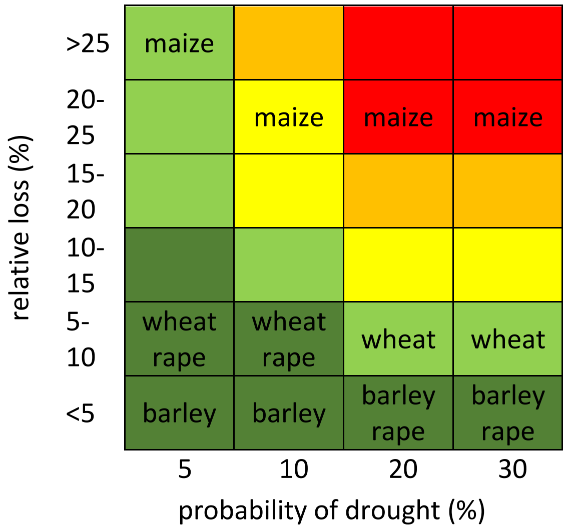


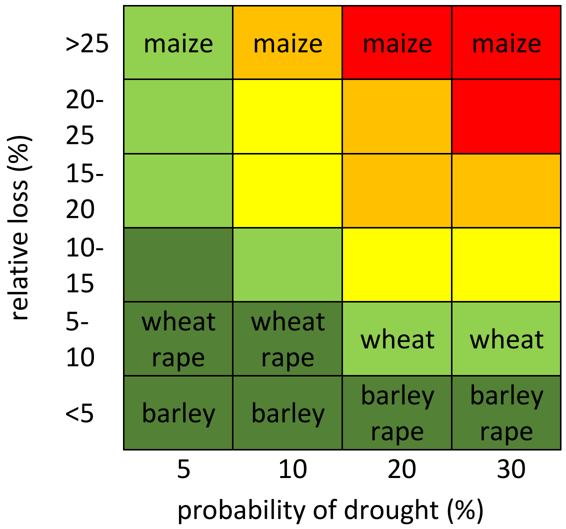
Figure 3. Risk matrix for the DriDanube region



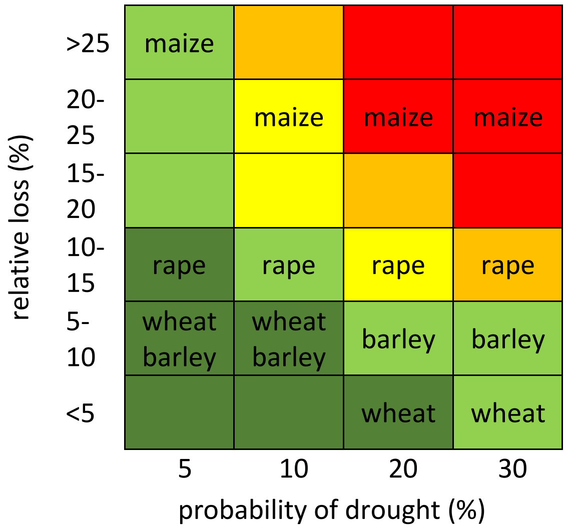
Austria



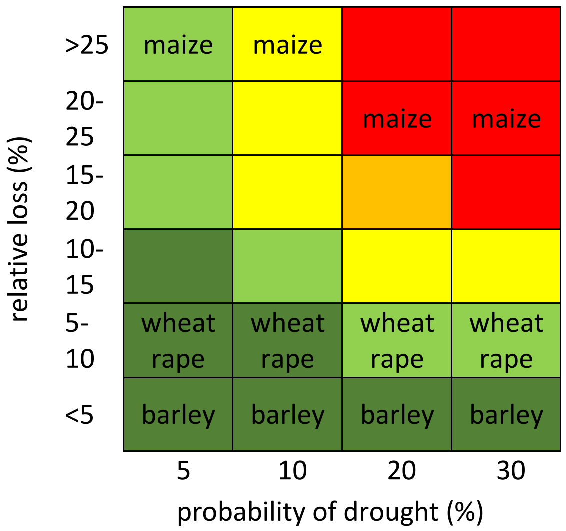
Bosnia and Herzegovina



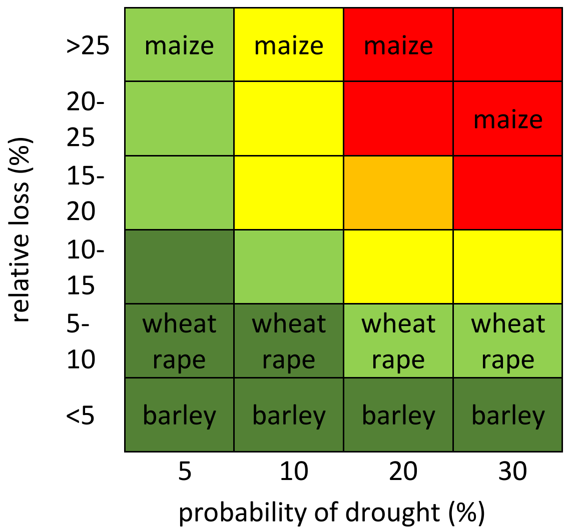
Croatia



Czech Republic



Hungary



Montenegro

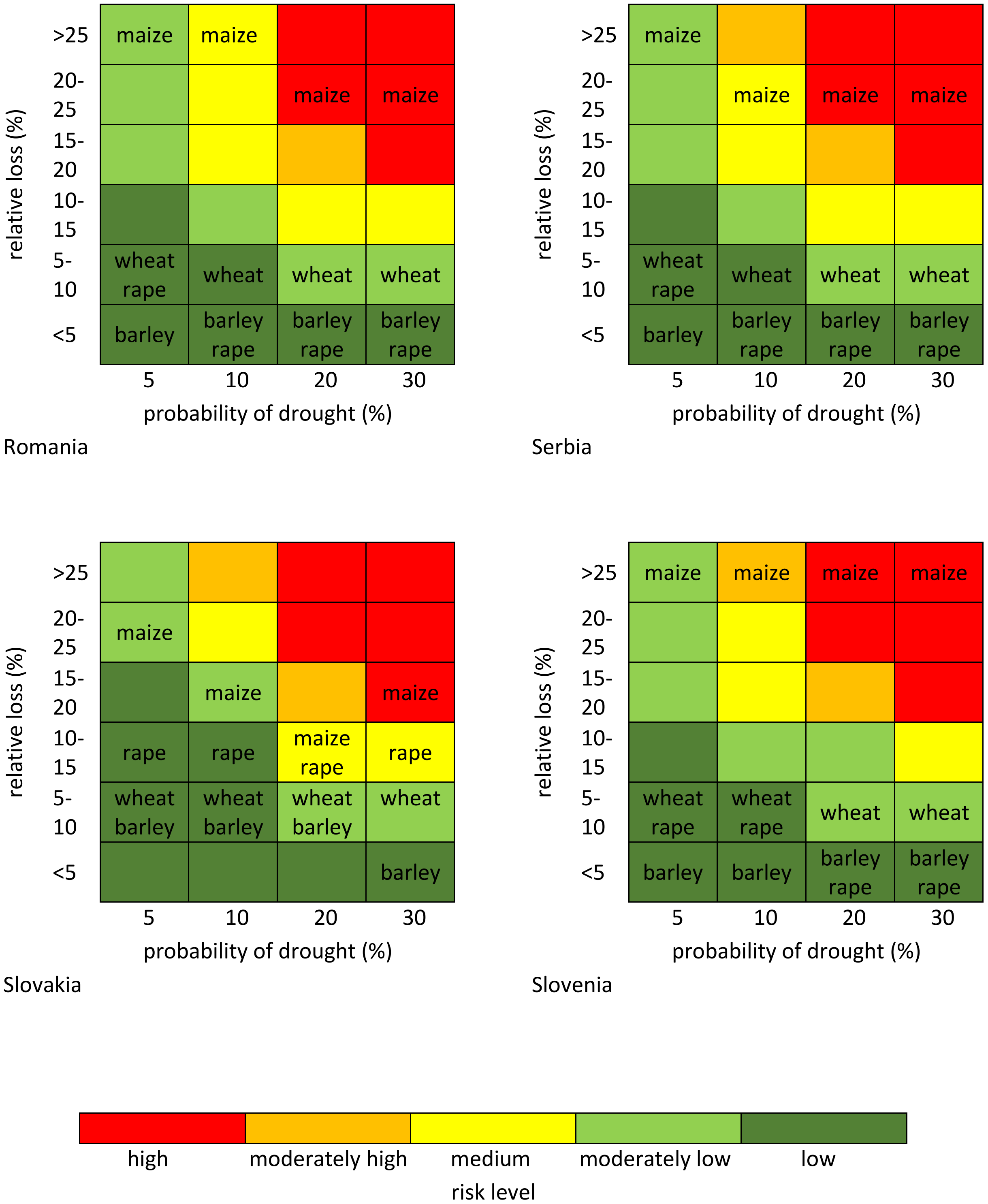


Figure 4. Risk matrices for the project partner countries

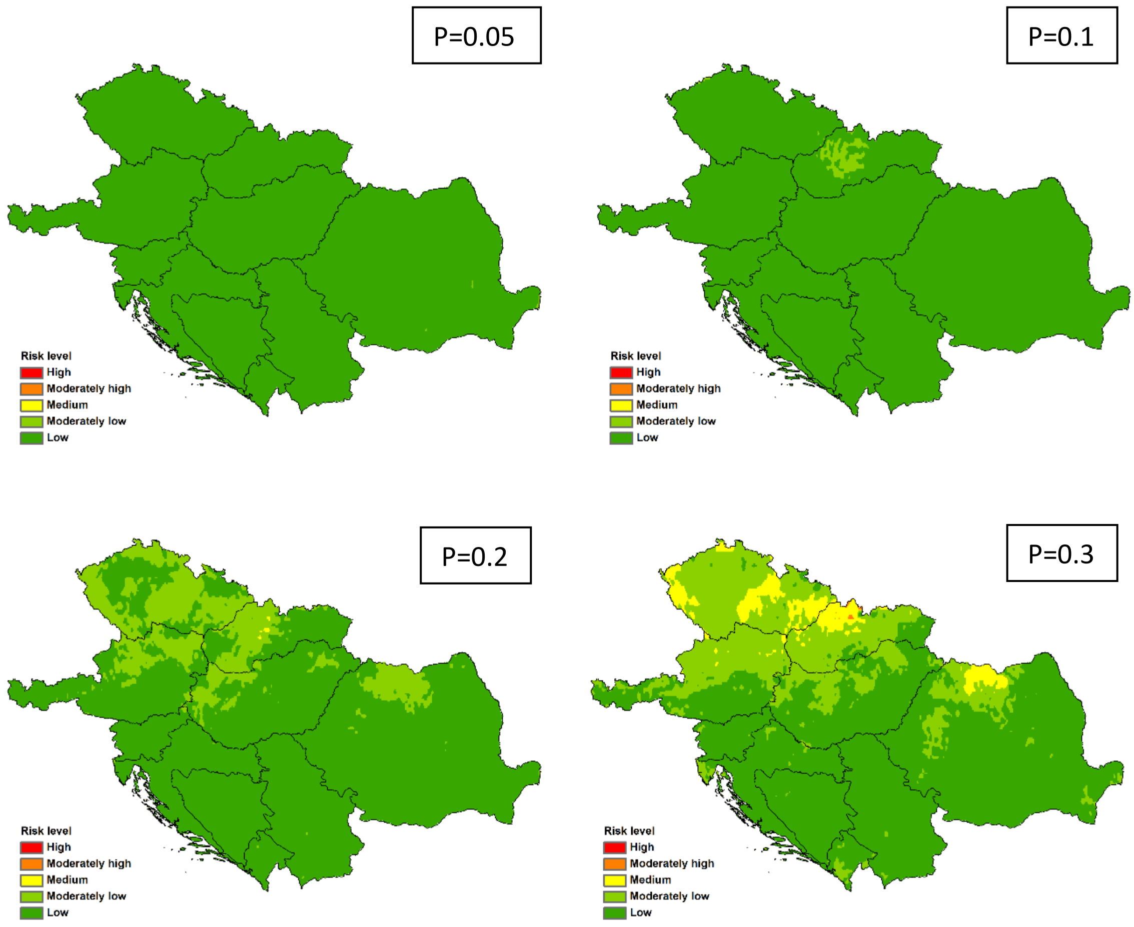


Figure 5. Risk maps for barley on different drought probability levels (P)

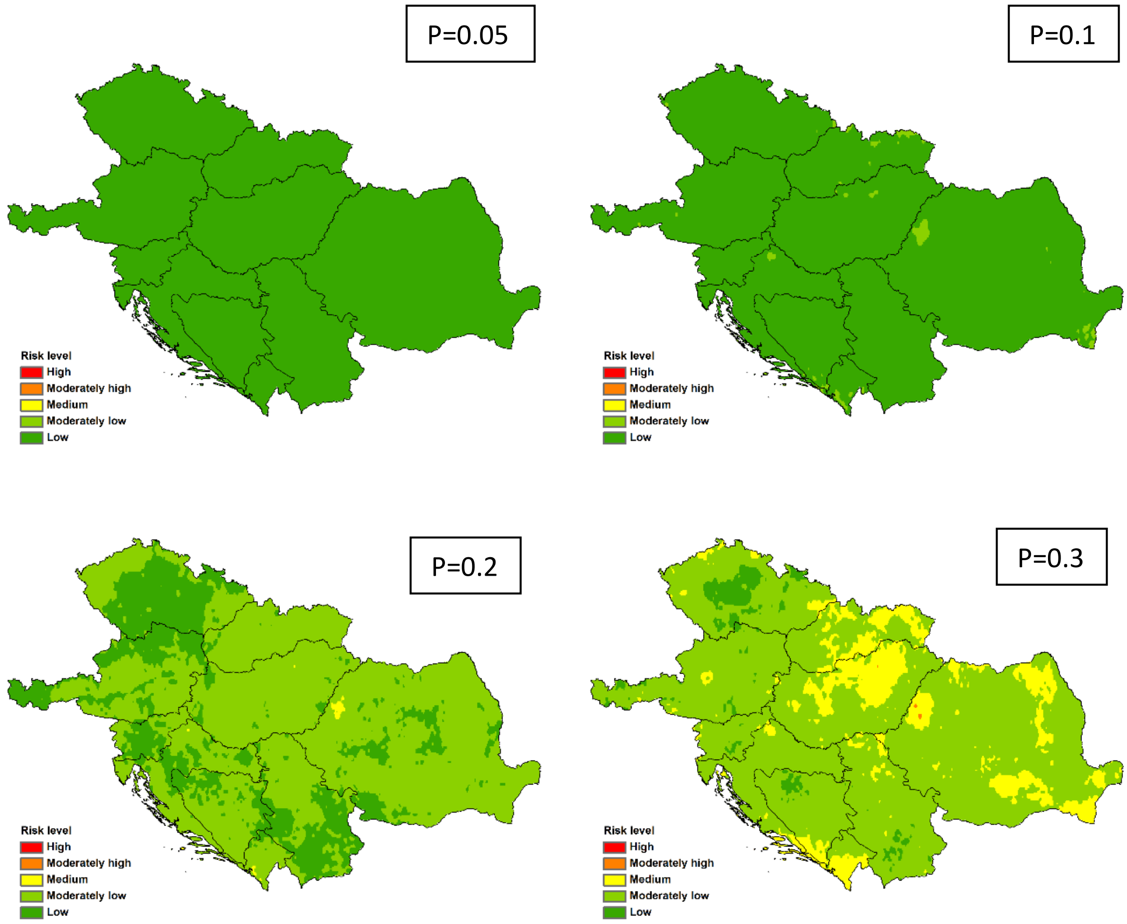


Figure 6. Risk maps for wheat on different drought probability levels (P)

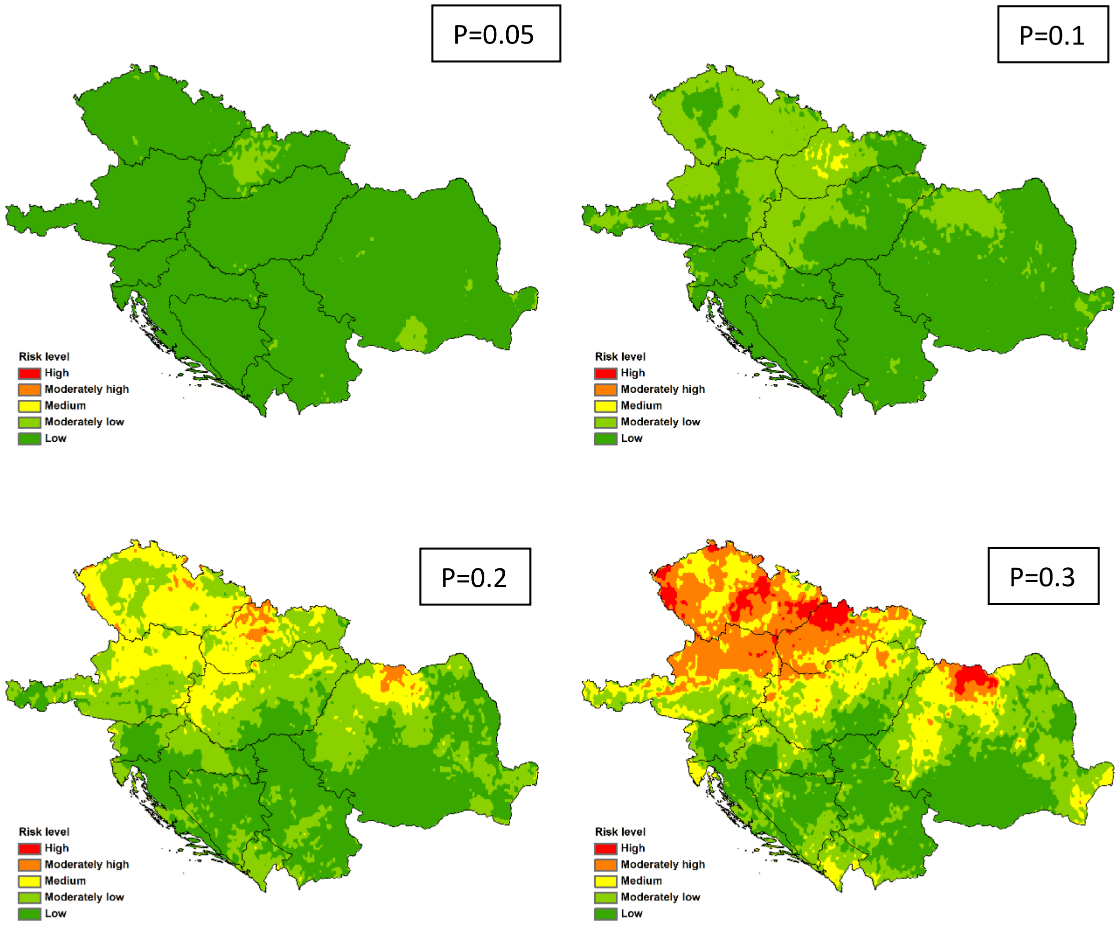


Figure 7. Risk maps for rape on different drought probability levels (P)

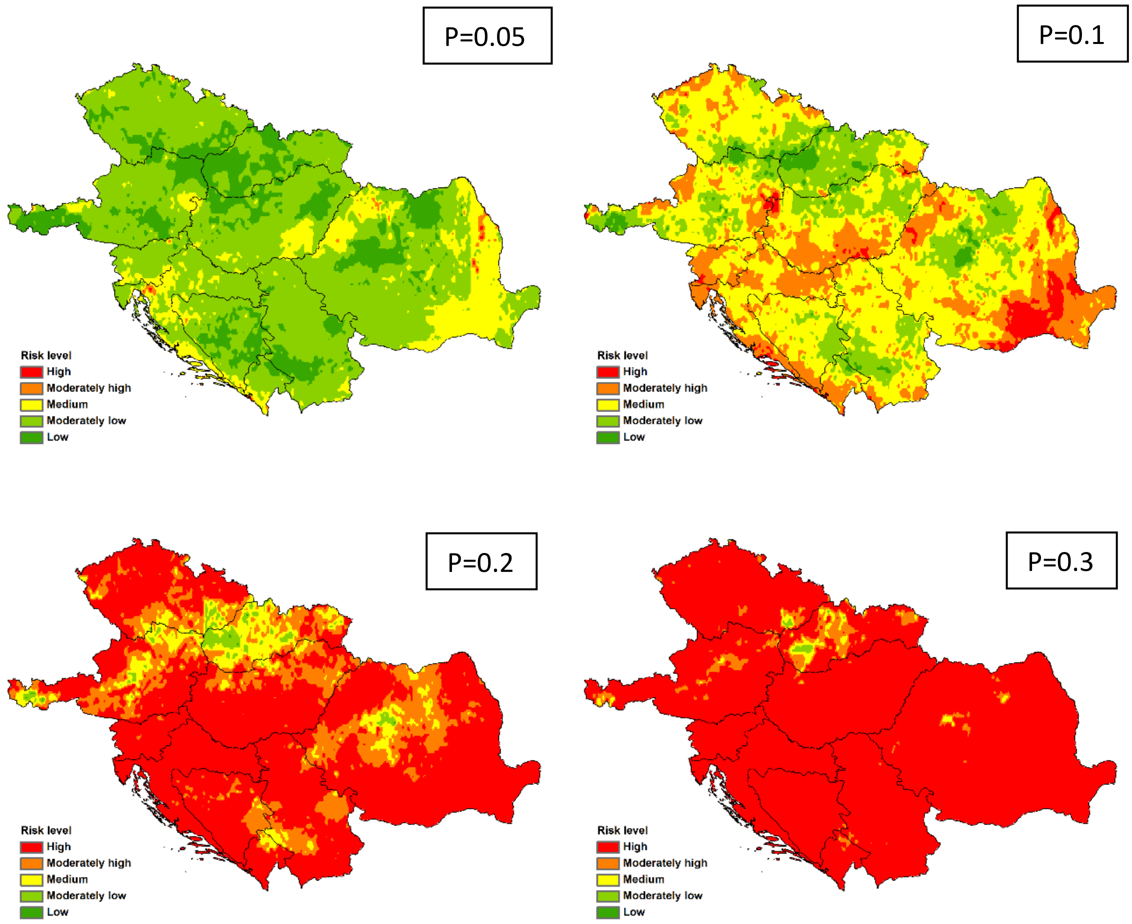


Figure 8. Risk maps for maize on different drought probability levels (P)

Output of rainless periods assessment

Output of the rainless periods assessment generated by the DROUGHTS software consists of wide spectrum of detail information about analyzed historical time series of daily rainfall at 170 locations within the DriDanube countries. Key output information based on performed methodological steps 1-11 are 6 maps for the whole region containing iso-lines of the longest rainless periods for typical return periods of 100, 50, 20, 10, 5 and 2 years. Spatial resolution of maps is about 1 km x 1 km and value of this field (pixel) is representing probability of longest drought duration for the given return period.

The maps are produced using geostatistical interpolation based on results of drought analyses. For better visual representation, different colours of pixel are used, according to values of drought duration in days.

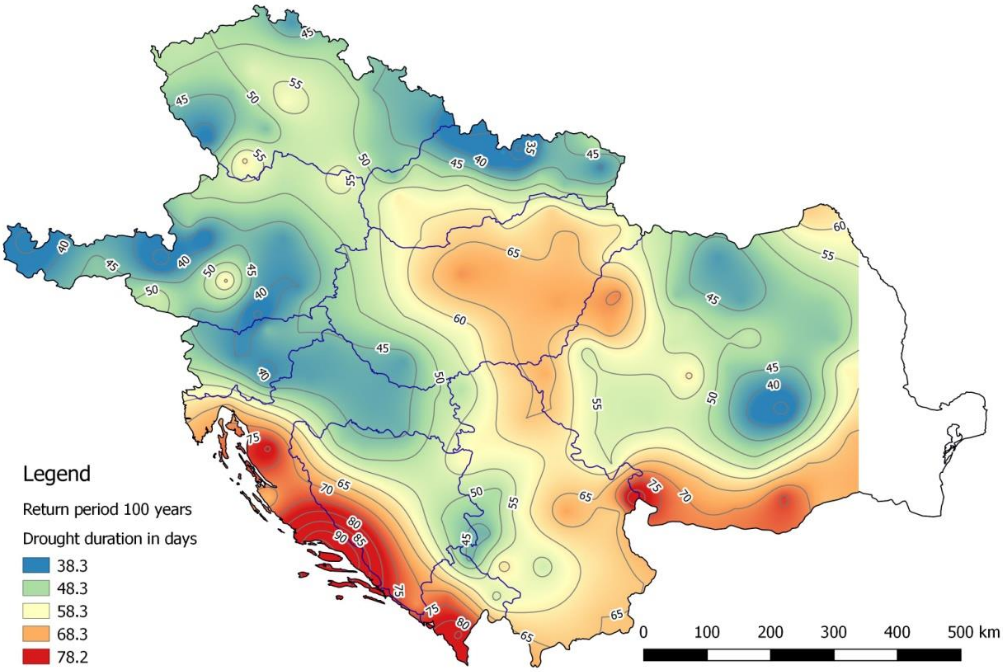


Figure 7 Rainless periods having a 100 year return period (vegetation season April 1 – September 30) for the DriDanube region

Another type of output is a map of locations with a possibility to obtain additional information (number of droughts in the 1981-2010 period; longest drought in days; starting, ending and mid date of the longest drought; average precipitation for the vegetation season in 1981-2010 period) from the data attribute table for each of locations.

Maps are posted on the web platform of the project and will be available to public for any further use, e.g., on-line estimation that on-going drought (rainless period) may last after certain number of days with associated probability given as return period estimate; in turn, this estimate (probability of drought duration) can be directly interpreted as numerical indicator of drought risk with consequence – a loss in agricultural production.

Acknowledgement

We acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (http://ensembles-eu.metoffice.com) and the data providers in the ECA&D project (https://www.ecad.eu)

We acknowledge the CARPATCLIM Database © European Commission - JRC, 2013

Annex

Annex 1. Algorithm of drought Risk Assessment

1. Introduction

One of the aim of the DriDanube project is the drought risk assessment within WP5 and the main responsible partner of this activity is the Hungarian Meteorological Service.

However, quoting from the Applicaion Form (AF) "unfortunately there is no commonly accepted procedure" for this purpose. Therefore it was necessary to review and prepare the methodology that can be applied for drought risk assessment in this project and the methodologies collected in the review WP5.1 may be background for WP5.2. The Activity Title of WP5.2 is, "Preparation of common methodology for drought risk assessment".

There is also a recommendation at Activity Description in AF i.e. "WP5.2 will promote the same approach to drought risk assessment as it is described in European Commission's Risk Assessment and Mapping Guidelines for Disaster Management. Detailed instructions on how to assess drought risk, will be prepared." Looking into this document we can read the following sentences.

"Risk Assessment and Mapping Guidelines for Disaster Management

5. RISK ASSESSMENT METHODS

5.1. Conceptual Framework and Basic Methodology

5.1.1. Risks: combining the consequences of a hazard with the likelihood of its occurrence

According to ISO 31010, risks are the combination of the consequences of an event or hazard and the associated likelihood of its occurrence…, risk can be expressed algebraically as,

Risk = hazard impact * probability of occurrence" (1)

According to the recommended principle, the risk is the product of the hazard impact and the probability of occurrence. Since it is rather a qualitative phrasing than an algebraic expression therefore it is necessary to formulate the problem according to the mathematical statistical conventions. Using this mathematical statistical formulation, the meteorological drought events can be characterized by certain loss functions (hazard impact) that depend on the relevant meteorological variables. Then according to the mathematical statistics, the meteorological drought risk can be defined as the expected value of the loss function. Consequently, the risk of meteorological drought events depends on the loss function and the probability distribution of the meteorological variables describing drought.

However there was another methodological problem i.e. how we can get reliable meteorological drought loss values or hazard impact data to estimate the loss function. This problem was formulated in AF as follows.

"Drought identification is also a process of finding, recognizing and describing risks. The purpose of the risk identification stage is to find and recognize all likely harards and significant consequences. If it is possible, the best solution is to receive hazard impact data from different target groups (e.g. estimation of loss in crop yield). In absence of impact data qualitative methods (e.g. expert opinions, intelligence information checklists, etc.) can also be applied."

The reality is that unfortunately there does not exist reliable meteorological drought loss values or hazard impact data moreover the qualitative methods are not adequate if want to develop quantitave algorithm or procedure. For solving this problem we used the crop yield values and some drought indices (e.g. SPI) for drought identification. Then the drought loss function was derived from the crop yield function using drought index.

According to AF, finally, a statistical method with a software will be established for drought risk assessment and the description of the methodology and software will be available to proceed in the Drought User Service (WP3). The relating Deliverables are as follows.

D5.2.1 Algorithm of drought risk assessment

D5.2.2 Software of drought risk calculation

D5.2.3 Manual for risk assessment

If we want to prepare an algorithm of drought risk assessment and there is no acceptable mathematical methodology for this problem then we have to develop it! The following sections include the description of the developed mathematical methodology. The final result of this methodology can be summarized such as the drought risk is the product of the probability of drought and the differences of the conditional expectations of yield given the events no drought and drought. The estimation procedure for this theoretical drought risk is also presented.

2. Applying the general mathematical methodology for drought risk assessment

The meteorological drought risk assessment can be based on certain meteorological variables.

Meteorological variables

Let us assume we have the following meteorological variables in space (s) and time (t year): $\mathbf{X}(\mathbf{s},t)=\left[X_1(\mathbf{s},t),....,X_N(\mathbf{s},t)\right]^{\mathrm{T}} \quad \mathbf{s}\in S$ (e.g. grid) , $t=1,..,n$

These meteorological variables are written in vector form and they may be e.g. several precipitation and temperature data. Their probability distribution can be defined by the following way:

i, Continuous distribution, joint density function: $f(\mathbf{x};\,\mathbf{s})$ (2)

ii, Discrete distribution: $p_i(\mathbf{s}) = \mathrm{P}(\mathbf{X}(\mathbf{s},t)=\mathbf{a}_i)$, $i=1,2,\ldots$ (3)

Meteorological drought loss (hazard impact)

There are also some meteorological drought loss values that characterize the hazard impact of the drought quantitatively:

$$\mathrm{L}(\mathbf{s},t)=\mathrm{Loss}(\mathbf{s},t)$$

These loss values may be absolute values in weight or relative ones in percent.

Meteorological drought loss (hazard impact) function

The loss value $\mathrm{Loss}(\mathbf{s},t)$ itself is a random quantity because it depends also on the outcome of the meteorological random variables $\mathbf{X}(\mathbf{s},t)$. Therefore we can define the meteorological drought loss (hazard impact) function,

$$L(\mathbf{X}(\mathbf{s},t))=\mathrm{E}\left(\mathrm{Loss}(\mathbf{s},t)\,\middle|\,\mathbf{X}(\mathbf{s},t)\right) \tag{4}$$

that is the conditional expectation of the meteorological loss given the meteorological variables, or with other phrase it is the regression of $\mathrm{Loss}(\mathbf{s},t)$ on $\mathbf{X}(\mathbf{s},t)$.

Meteorological drought risk

According to the mathematical definition the risk is the expected value of the loss i.e.,

$$Risk(\mathbf{s})=\mathrm{E}(\mathrm{Loss}(\mathbf{s},t))=\mathrm{E}(L(\mathbf{X}(\mathbf{s},t))) \tag{5}$$

and it is equal also to the expected value of the meteorological loss function as a consequence of some mathematical theorems. It can be assumed that the risk values depend on the locations only.

The basic cases

Using the notations (2), (3) the basic cases are as follows,

i, Continuous distribution: $Risk(\mathbf{s})=\mathrm{E}(L(\mathbf{X}(\mathbf{s},t)))=\int_{R^N} L(\mathbf{x})\, f(\mathbf{x};\,\mathbf{s})\, d\mathbf{x}$

ii, Discrete distribution: $Risk(\mathbf{s})=\mathrm{E}(L(\mathbf{X}(\mathbf{s},t)))=\sum_{i=1} L(\mathbf{a}_i)\cdot p_i(\mathbf{s})$

Special case

Le us assume the simplest two stage discrete model that means,

$L(\mathbf{a}_1)$=hazard impact, $p_1(\mathbf{s})$=probability of occurrence, $L(\mathbf{a}_2)=0$, $p_2(\mathbf{s})=1-p_1(\mathbf{s})$

Then, $Risk(\mathbf{s}) = \sum_{i=1}^{2} L(\mathbf{a}_i) \cdot p_i(\mathbf{s})$ = hazard impact*probability of occurrence

i.e. the formula (1) has been obtained that is the drought risk assessment as it is described in European Commission's Risk Assessment and Mapping Guidelines for Disaster Management. However this simple model is not adequate for the meteorological drought risk assessment.

Estimation of the meteorological drought risk

The estimation procedure can be implemented according to the statistical conventions.

i, Let $\hat{L}(\mathbf{x})$ be an estimation of the loss function $L(\mathbf{x})$ based on a statistical sample:

$$\mathrm{Loss}(\mathbf{S}_k, T_l),\ \mathbf{X}(\mathbf{S}_k, T_l)\ \ (k = 1,..,K;\ l = 1,..,L)$$

Then a regression procedure is implemented that needs the examination of the schema and shape of the potential loss functions i.e. the regression of the loss on the meteorological variables, sampling for the loss and meteorological variables and estimation of the statistical parameters of the loss function on the basis of the above sample.

ii, Then the estimated meteorological drought risk:

$$\hat{Risk}(\mathbf{s}) = \frac{1}{n}\sum_{t=1}^{n} \hat{L}(\mathbf{X}(\mathbf{s},t)) \qquad (6)$$

since $Risk(\mathbf{s}) = \mathrm{E}(L(\mathbf{X}(\mathbf{s},t)))$.

3. Meteorological drought risk assessment using crop yield and drought identification

In general there is no usable information or sample for the meteorological drought loss values $\mathrm{L}(\mathbf{s},t)=\mathrm{Loss}(\mathbf{s},t)$. Therefore we have developed a method to derive the meteorological drought loss function by using the crop yield values and some drought identification.

Crop yield values

Let us we have some crop yield values:

$$\mathrm{Y}(\mathbf{s},t) = \mathrm{Yield}(\mathbf{s},t)$$

These yield values may be absolute values in weight or relative ones in percent.

Meteorological yield function

Similarly to the definition of loss function $L\left(\mathbf{X}\left(\mathbf{s},t\right)\right)$ (3) the meteorological yield function can be defined as,

$$\mathrm{E}\left(\mathrm{Yield}\left(\mathbf{s},t\right) \mid \mathbf{X}\left(\mathbf{s},t\right)\right) = Y\left(\mathbf{X}\left(\mathbf{s},t\right)\right)$$

that is the conditional expectation of the crop yield given the meteorological variables, or with other phrase it is the regression of $\mathrm{Yield}\left(\mathbf{s},t\right)$ on $\mathbf{X}\left(\mathbf{s},t\right)$.

Drought identification

As it was mentioned at the Introduction the drought identification is also a process of finding, recognizing and describing risks.

The drought identification can be formulated mathematically by a set D as follows:

There is drought if and only if $\mathbf{X}(\mathbf{s},t) \in D$.

Then the probability of drought is, $P_D = \mathrm{P}\left(\mathbf{X}(\mathbf{s},t) \in D\right)$.

Loss function based on yield function and drought identification

Using the yield function $Y\left(\mathbf{X}\left(\mathbf{s},t\right)\right)$ and drought identification set D the meteorological drought loss function can be defined in the following way:

$$L(\mathbf{X}(\mathbf{s},t)) = 0 \quad \text{if} \quad \mathbf{X}(\mathbf{s},t) \notin D \text{ (there is no drought)} \tag{7}$$

$$L(\mathbf{X}(\mathbf{s},t)) = \mathrm{E}\left(Y(\mathbf{X}(\mathbf{s},t)) \mid \mathbf{X}(\mathbf{s},t) \notin D\right) - Y(\mathbf{X}(\mathbf{s},t)) \quad \text{if} \quad \mathbf{X}(\mathbf{s},t) \in D \text{ (there is drought)}$$

where $\mathrm{E}\left(Y(\mathbf{X}(\mathbf{s},t)) \mid \mathbf{X}(\mathbf{s},t) \notin D\right)$ is the conditional expectation of yield if there is no drought.

Meteorological drought risk based on yield function and drought identification

Using the equation (5) and the definition (7) then the meteorological drought risk can be expressed as,

$$Risk\left(\mathbf{s}\right) = \mathrm{E}\left(L\left(\mathbf{X}\left(\mathbf{s},t\right)\right)\right) =$$

$$= P_D \cdot \left(\mathrm{E}\left(Y\left(\mathbf{X}\left(\mathbf{s},t\right)\right) \mid \mathbf{X}\left(\mathbf{s},t\right) \notin D\right) - \mathrm{E}\left(Y\left(\mathbf{X}\left(\mathbf{s},t\right)\right) \mid \mathbf{X}\left(\mathbf{s},t\right) \in D\right)\right) \tag{8}$$

Proof.

According to the equation (5) and the law of total probability,

$$Risk(\mathbf{s}) = \mathrm{E}\left(L\left(\mathbf{X}(\mathbf{s},t)\right)\right) =$$

$$= \mathrm{E}\left(L\left(\mathbf{X}(\mathbf{s},t)\right) \middle| \mathbf{X}(\mathbf{s},t) \notin D\right) \cdot (1 - P_D) + \mathrm{E}\left(L\left(\mathbf{X}(\mathbf{s},t)\right) \middle| \mathbf{X}(\mathbf{s},t) \in D\right) \cdot P_D =$$

and using the definition (7),

$$= \mathrm{E}\left(\mathrm{E}\left(Y\left(\mathbf{X}(\mathbf{s},t)\right) \middle| \mathbf{X}(\mathbf{s},t) \notin D\right) - Y\left(\mathbf{X}(\mathbf{s},t)\right) \middle| \mathbf{X}(\mathbf{s},t) \in D\right) \cdot P_D =$$

$$= P_D \cdot \left(\mathrm{E}\left(Y\left(\mathbf{X}(\mathbf{s},t)\right) \middle| \mathbf{X}(\mathbf{s},t) \notin D\right) - \mathrm{E}\left(Y\left(\mathbf{X}(\mathbf{s},t)\right) \middle| \mathbf{X}(\mathbf{s},t) \in D\right)\right)$$

Estimation of loss function based on estimated yield function and drought identification

Let $\hat{Y}(\mathbf{X}(\mathbf{s},t))$ $(t = 1,..,n)$ be an estimation of the yield function values based on a sample: $\mathrm{Y}(\mathbf{S}_k, T_l) = \mathrm{Yield}(\mathbf{S}_k, T_l)$, $\mathbf{X}(\mathbf{S}_k, T_l)$ $(k = 1,.., K;\, l = 1,.., L)$

Then in accordance with (7) the estimation of loss function using yield function estimation and drought identification set D:

$$\hat{L}\left(\mathbf{X}(\mathbf{s},t)\right) = 0 \quad \text{if} \quad \mathbf{X}(\mathbf{s},t) \notin D \text{ (there is no drought)} \qquad (9)$$

$$\hat{L}(\mathbf{X}(\mathbf{s},t)) = \left(\frac{1}{n - n_D} \sum_{\mathbf{X}(\mathbf{s},t) \notin D} \hat{Y}\left(\mathbf{X}(\mathbf{s},t)\right)\right) - \hat{Y}\left(\mathbf{X}(\mathbf{s},t)\right) \quad \text{if} \quad \mathbf{X}(\mathbf{s},t) \in D \text{ (there is drought)}$$

where $n_D = \sum_{\mathbf{X}(\mathbf{s},t) \in D} 1$ is the frequency of drought.

Estimation of meteorological drought risk based on estimated yield function and drought identification

Using the equation (6) and the estimation (9) the following estimation of the meteorological drought risk can be obtained,

$$\hat{Risk}(\mathbf{s}) = \frac{1}{n} \sum_{t=1}^{n} \hat{L}\left(\mathbf{X}(\mathbf{s},t)\right) = \hat{P}_D \cdot \left(\left(\frac{1}{n - n_D} \sum_{\mathbf{X}(\mathbf{s},t) \notin D} \hat{Y}\left(\mathbf{X}(\mathbf{s},t)\right)\right) - \left(\frac{1}{n_D} \sum_{\mathbf{X}(\mathbf{s},t) \in D} \hat{Y}\left(\mathbf{X}(\mathbf{s},t)\right)\right)\right) \qquad (10)$$

where $\hat{P}_D = \frac{n_D}{n}$ is the estimated probability of drought.

It is an aesthetic formula, thus it must be good!

Proof.

According to the equation (6) and the estimation (9),

$$\hat{R}isk(\mathbf{s})=\frac{1}{n}\sum_{t=1}^{n}\hat{L}(\mathbf{X}(\mathbf{s},t))=\frac{1}{n}\sum_{\mathbf{X}(\mathbf{s},t)\in D}\left(\left(\frac{1}{n-n_D}\sum_{\mathbf{X}(\mathbf{s},t)\notin D}\hat{Y}(\mathbf{X}(\mathbf{s},t))\right)-\hat{Y}(\mathbf{X}(\mathbf{s},t))\right)=$$

$$=\frac{n_D}{n}\left(\frac{1}{n-n_D}\sum_{\mathbf{X}(\mathbf{s},t)\notin D}\hat{Y}(\mathbf{X}(\mathbf{s},t))\right)-\frac{1}{n}\sum_{\mathbf{X}(\mathbf{s},t)\in D}\hat{Y}(\mathbf{X}(\mathbf{s},t))=$$

$$=\hat{P}_D\cdot\left(\left(\frac{1}{n-n_D}\sum_{\mathbf{X}(\mathbf{s},t)\notin D}\hat{Y}(\mathbf{X}(\mathbf{s},t))\right)-\left(\frac{1}{n_D}\sum_{\mathbf{X}(\mathbf{s},t)\in D}\hat{Y}(\mathbf{X}(\mathbf{s},t))\right)\right)$$

4. Meteorological drought risk assessment using crop yield and SPI

Definition of identification set D by an SPI

There may be a special case when the identification set D is defined by a given SPI. According to the definition SPI is such a transformation of the precipitation sum X on a given period of months that SPI has standard normal distribution. Applying the usual notations, that is $SPI(X)\in N(0,1)$ and assuming X is a component of the used meteorological variables $\mathbf{X}$, then the following definition formula can be obtained,

$$D=\{\,\mathbf{X}\mid SPI<C_P\,\}\text{ where }\mathrm{P}(SPI<C_P)=\Phi(C_P)=P_D$$

where $\Phi(x)$ is the standard normal distribution function and the critical value C_P belongs to the probability of drought P_D.

Loss function based on yield function and SPI series

If we have $SPI(\mathbf{s},t)$ $(t=1,..,n)$ series at the locations $\mathbf{s}\in S$ then the loss function based on yield function and SPI can be formulated directly by these series. Using the formula (7),

$$L(\mathbf{X}(\mathbf{s},t))=0\quad\text{if}\quad SPI(\mathbf{s},t)\geq C_P\ \text{(there is no drought)}\tag{11}$$

$$L(\mathbf{X}(\mathbf{s},t))=\mathrm{E}(Y(\mathbf{X}(\mathbf{s},t))\mid SPI(\mathbf{s},t)\geq C_P)-Y(\mathbf{X}(\mathbf{s},t))\quad\text{if}\quad SPI(\mathbf{s},t)<C_P\ \text{(there is drought)}$$

where $\mathrm{E}(Y(\mathbf{X}(\mathbf{s},t))\mid SPI(\mathbf{s},t)\geq C_P)$ is the conditional expectation of yield if there is no drought.

Meteorological drought risk based on yield function and SPI series

Moreover using the equation (8) the meteorological drought risk also can be expressed by the SPI series as,

$$Risk(\mathbf{s}) = \mathrm{E}\left(L\left(\mathbf{X}(\mathbf{s},t)\right)\right) =$$

$$= P_D \cdot \left(\mathrm{E}\left(Y\left(\mathbf{X}(\mathbf{s},t)\right) \middle| SPI(\mathbf{s},t) \geq C_P\right) - \mathrm{E}\left(Y\left(\mathbf{X}(\mathbf{s},t)\right) \middle| SPI(\mathbf{s},t) < C_P\right)\right) \quad (12)$$

Estimation of loss function based on estimated yield function and SPI series

The estimation of the loss function using estimated yield function $\hat{Y}(\mathbf{X}(\mathbf{s},t))$ and $SPI(\mathbf{s},t)$ series by formula (9) is,

$$\hat{L}(\mathbf{X}(\mathbf{s},t)) = 0 \quad \text{if} \quad SPI(\mathbf{s},t) \geq C_P \text{ (there is no drought)} \quad (13)$$

$$\hat{L}(\mathbf{X}(\mathbf{s},t)) = \left(\frac{1}{n-n_D} \sum_{SPI(\mathbf{s},t)\geq C_P} \hat{Y}(\mathbf{X}(\mathbf{s},t))\right) - \hat{Y}(\mathbf{X}(\mathbf{s},t)) \quad \text{if} \quad SPI(\mathbf{s},t) < C_P \text{ (there is drought)}$$

where $n_D = \sum_{SPI(\mathbf{s},t)<C_P} 1$ is the frequency of drought.

Estimation of drought risk based on estimated yield function and SPI series

The estimation of the meteorological drought risk using estimated yield function $\hat{Y}(\mathbf{X}(\mathbf{s},t))$ and $SPI(\mathbf{s},t)$ series applying the formula (10) can be written as,

$$\hat{Risk}(\mathbf{s}) = \frac{1}{n}\sum_{t=1}^{n} \hat{L}\left(\mathbf{X}(\mathbf{s},t)\right) = \hat{P}_D \cdot \left(\left(\frac{1}{n-n_D} \sum_{SPI(\mathbf{s},t)\geq C_P} \hat{Y}(\mathbf{X}(\mathbf{s},t))\right) - \left(\frac{1}{n_D} \sum_{SPI(\mathbf{s},t)<C_P} \hat{Y}(\mathbf{X}(\mathbf{s},t))\right)\right) \quad (14)$$

where $\hat{P}_D = \frac{n_D}{n}$ is the estimated probability of drought.

Remark
Instead of SPI optional drought index or drought definition can be used!

5. Regression model applied for estimation of the yield function

The estimation of the yield function $Y\left(\mathbf{X}(\mathbf{s},t)\right)$ values should be based on some sample: $\mathrm{Y}(\mathbf{S}_k, T_l) = \mathrm{Yield}(\mathbf{S}_k, T_l)$, $\mathbf{X}(\mathbf{S}_k, T_l)$ $(k = 1,., K; l = 1,., L)$.

We had sample for the absolute yield values $\mathrm{Y}(\mathbf{s},t)$ in weight however there is a distribution problem of such yield sample $\mathrm{Y}(\mathbf{S}_k, T_l)$ in general i.e. the expected values $\mathrm{E}\left(\mathrm{Y}(\mathbf{S}_k, T_l)\right)$ may be very different.

Therefore we used the relative yield in percent:

$$\mathrm{Y}_{rel}\left(\mathbf{s},t\right)=\frac{\mathrm{Y}\left(\mathbf{s},t\right)-\mathrm{E}\left(\mathrm{Y}\left(\mathbf{s},t\right)\right)}{\mathrm{E}\left(\mathrm{Y}\left(\mathbf{s},t\right)\right)}\cdot 100\% \qquad (15)$$

assuming about the expected values that $\mathrm{E}\left(\mathrm{Y}\left(\mathbf{s},t\right)\right)=E(\mathbf{s})$.

Regression model for the relative yield function

Using the equality $\mathrm{E}\left(\mathrm{Y}_{rel}\left(\mathbf{s},t\right)\right)\equiv 0$ we applied the following linear regression model for the relative yield function,

$$\hat{Y}_{rel}\left(\mathbf{X}(\mathbf{s},t)\right)=\alpha_1\cdot\left(\ln X_1\left(\mathbf{s},t\right)-E_{\ln,1}\left(\mathbf{s}\right)\right)+\alpha_2\cdot\left(X_2\left(\mathbf{s},t\right)-E_2\left(\mathbf{s}\right)\right) \qquad (16)$$

where $X_1\left(\mathbf{s},t\right)$ is precipitation sum and $X_2\left(\mathbf{s},t\right)$ is temperature mean for a given 3 or 6 months period. Furthermore there are the expected values varying in space i.e.,

$E_{\ln,1}(\mathbf{s})=\mathrm{E}\left(\ln X_1\left(\mathbf{s},t\right)\right)$ and $E_2(\mathbf{s})=\mathrm{E}\left(X_2\left(\mathbf{s},t\right)\right)$

The common coefficients α_1,α_2 can be estimated by the method of least squares.

Remark 1

At these longer periods (3 or 6 months) the independent variables are near normally distributed i.e. $\ln X_1\left(\mathbf{s},t\right)\tilde{\in} N\left(E_{\ln,1}(\mathbf{s}),D_{\ln,1}(\mathbf{s})\right)$, $X_2\left(\mathbf{s},t\right)\tilde{\in} N\left(E_2(\mathbf{s}),D_2(\mathbf{s})\right)$, where the standard deviations are $D_{\ln,1}(\mathbf{s})=\mathrm{D}\left(\ln X_1\left(\mathbf{s},t\right)\right)$ and $D_2(\mathbf{s})=\mathrm{D}\left(X_2\left(\mathbf{s},t\right)\right)$.

Consequently there is a near linear connection with the SPI belonging to the same period i.e.,

$$SPI\left(X_1\left(\mathbf{s},t\right)\right)\approx\frac{\ln X_1\left(\mathbf{s},t\right)-E_{\ln,1}\left(\mathbf{s}\right)}{D_{\ln,1}(\mathbf{s})} \qquad (17)$$

At the risk estimation formula (14) this $SPI\left(X_1\left(\mathbf{s},t\right)\right)$ can be applied. Then we can evaluate the connection and signification of the different type SPI series with the real drought loss, by calculation of $SPI\left(X_1\left(\mathbf{s},t\right)\right)$ (17), $\hat{Y}_{rel}\left(X_1\left(\mathbf{s},t\right),X_2\left(\mathbf{s},t\right)\right)$ (16) and $\hat{R}isk\left(\mathbf{s}\right)$ (14). At the software there will be possibility to examine several SPI series with different periods.

Remark 2

We did not intend to develop a general crop-weather model for the project region that would have been beyond our possibilities of course. Our intention was to estimate an expected value only i.e. the risk instead of a more precise crop yield estimation. Therefore a not to complicated regression model was also acceptable for our aim. However in case of having crop-weather model with good quality then the estimated crop yield values can be used according to the equations (10), (14). Similarly, we repeat again that instead of SPI other optional drought index or drought definition can be used!

6. Algorithm for drought risk assessment applied in the project

On the basis of the developed methodology the following algorithm is planned to be applied for the drought risk assessment in the project. A software is also developed in order to implement the steps.

Steps of the algorithm

1. Selection of the relevant meteorological variables $\mathbf{X}(\mathbf{s},t)$.

Monthly precipitation and temperature series can be selected during the procedure.

Homogenized station and gridded climate data series are used.

Gridded databases: CarpatClim, DanubeClim

2. Collecting sample for the crop yield $\text{Yield}(\mathbf{s},t)$ and the meteorological variables $\mathbf{X}(\mathbf{s},t)$.

3. Estimation of the relative crop yield function $\hat{Y}_{rel}\left(\mathbf{X}(\mathbf{s},t)\right)$ that is a regression of the relative crop yield on the meteorological variables according to the Section 5.

4. On the basis of the results of item 3 developing software for calculation of the following series and estimations in case of given meteorological data series $\mathbf{X}(\mathbf{s},t)$ $(t=1,..,n)$:

- Estimated relative crop yield functions $\hat{Y}_{rel}\left(\mathbf{X}(\mathbf{s},t)\right)$ $(t=1,..,n)$ according to the equation (16) at Section 5.

- Several $SPI(\mathbf{s},t)$ $(t=1,..,n)$ series.

- Estimated meteorological drought risk $\hat{Risk}(\mathbf{s})$ according to the equation (14) at Section 4.

The software will be sent to the partners.

5. Applying software, risk calculations on station and gridded climate data series.

Mapping of risk for CarpatClim, DanubeClim area based on gridded data series.

6. Mapping of risk outside CarpatClim, DanubeClim area based on other gridded data series (E-OBS).

7. Data requirement

Some quotations from the Application Form: "All partners will prepare information and datasets for drought risk evaluation. They will have to collect hazard impacts and meteorological data in their countries and apply the developed software of drought risk calculation on their own data." "In CARPATCLIM project daily gridded meteorological database was established that can be the base of drought risk calculation in this project."

Data for sample

Data requirements from project partners were impact data (crop yield and/or crop losses) and meteorological data for the step 3 of Algorithm i.e. sample for regression. In Hungary we had access to detailed yield database (2001-2016) on about 1900 pilot sites in the Farm Accountancy Data Network (FADN) for the four main plants in Hungary: maize, wheat, rape and barley. As regards the reliable meteorological variables the precipitation and temperature series homogenized by method MASH (*Szentimrey*) can be interpolated by method MISH (*Szentimrey and Bihari*) for the pilot sites.

Gridded databases CarpatClim, DanubeClim and E-OBS for Mapping

According to the step 5 of the Algorithm mapping of risk for CarpatClim, DanubeClim areas is based on gridded data series.

Main properties of CarpatClim database (Szalai et al):

Daily gridded data series for basic meteorological variables in Carpathian Region (1961-2010) Spatial resolution: 0.1°

Project of JRC (2011-2013) (10 participants)

Methodology: MASH (*Szentimrey*) for homogenization, MISH (*Szentimrey and Bihari*) for gridding

DanubeClim: extension of the Carpatclim database

New regions: Bosnia and Herzegovina (Republika Srpska), Montenegro, South part of Serbia, Western part of Hungary

Same methods: MISH-MASH

Bilateral contracts with JRC

Figure 1. Map for the CarpatClim and DanubeClim regions

According to the step 6 of the Algorithm mapping of risk outside CarpatClim, DanubeClim area is based on E-OBS data series (Haylock et al, 2018").

Main properties of E-OBS database:

The ensemble version is available on a 0.1 and 0.25 degree regular grid for the elements (daily mean temperature, daily minimum temperature, daily maximum temperature, daily precipitation sum and daily averaged sea level pressure). They cover the area: 25N-71.5N x 25W-45E.
The ensemble dataset is constructed through a conditional simulation procedure. For each of the 100 members of the ensemble a spatially correlated random field is produced using a pre-calculated spatial correlation function. The mean across the 100 members is calculated and is provided as the "best-guess" fields. The spread is calculated as the difference between the 5th and 95th percentiles over the ensemble to provide a measure indicate of the 90% uncertainty range.

8. References

CarpatClim Project: http://www.carpatclim-eu.org/pages/home/

Drought Management Centre for South- East Europe – DMCSEE: Summary of project results https://www.met.hu/doc/DMCSEE/DMCSEE_final_publication.pdf

Drought Risk Danube Region Project (DriDanube): Application Form (AF)

EUROPEAN COMMISSION: COMMISSION STAFF WORKING PAPER, Risk Assessment and Mapping Guidelines for Disaster Management, Brussels, 21.12.2010 SEC (2010) 1626 final

Haylock, M.R., N. Hofstra, A.M.G. Klein Tank, E.J. Klok, P.D. Jones, M. New. 2008: A European daily high-resolution gridded dataset of surface temperature and precipitation. J. Geophys. Res (Atmospheres), 113, D20119, doi:10.1029/2008JD10201

Szalai, S., Auer, I., Hiebl, J., Milkovich, J., Radim, T. Stepanek, P., Zahradnicek, P., Bihari, Z., Lakatos, M., Szentimrey, T., Limanowka, D., Kilar, P., Cheval, S., Deak, Gy., Mihic, D., Antolovic, I., Mihajlovic, V., Nejedlik, P., Stastny, P., Mikulova, K., Nabyvanets, I., Skyryk, O., Krakovskaya, S.,Vogt, J., Antofie, T., Spinoni, J.: Climate of the Greater Carpathian Region. Final Technical Report.www.carpatclim-eu.org.

Szentimrey, T., 2014: Manual of homogenization software MASHv3.03, Hungarian Meteorological Service, p. 69.

Szentimrey, T., Bihari, Z., 2014: Manual of interpolation software MISHv1.03, Hungarian Meteorological Service, p. 60

WMO: Standardized Precipitation Index, User Guide
http://www.wamis.org/agm/pubs/SPI/WMO_1090_EN.pdf

Annex 2. Manual of the *Risk Estimation of Drought* (REDv1.01) software

(R E D v1.01)

Tamás Szentimrey

Varimax Limited Partnership
Budapest, Hungary
e-mail: szentimrey.t@gmail.com

I. INTRODUCTION

One of the aim of the DriDanube project is the drought risk assessment within WP5 and the main responsible partner of this activity is the Hungarian Meteorological Service.

However unfortunately there is no commonly accepted procedure for this purpose therefore it was necessary to review and prepare the mathematical methodology that can be applied for drought risk assessment in this project.

Varimax Limited Partnership as a subcontractor of the Hungarian Meteorological Service undertook to develop the mathematical methodology and the software.

The elaborated mathematical methodology is described in the deliverable D5.2.1: Algorithm of drought risk assessment.

This D5.2.1 is the mathematical background of this software RED (Risk Estimation of Drought) that is developed for the deliverable D5.2.2: Software of drought risk calculation.

II. MATHEMATICAL BACKGROUND

1. Estimation of drought risk based on estimated yield function and SPI series

According to the formula (14) at Section 4 in D5.2.1, the estimation of the meteorological drought risk using estimated yield function $\hat{Y}(\mathbf{X}(\mathbf{s},t))$ and $SPI(\mathbf{s},t)$ series can be written as,

$$\hat{R}isk(\mathbf{s}) = \hat{P}_D \cdot \left(\left(\frac{1}{n-n_D} \sum_{SPI(\mathbf{s},t)\geq C_P} \hat{Y}(\mathbf{X}(\mathbf{s},t)) \right) - \left(\frac{1}{n_D} \sum_{SPI(\mathbf{s},t)<C_P} \hat{Y}(\mathbf{X}(\mathbf{s},t)) \right) \right) \quad (1)$$

where $\hat{P}_D$ is the estimated probability of drought and C_P is the critical value belonging to this probability. The final result of this methodology can be summarized such as the estimated drought risk is the product of the probability of drought and the estimated conditional expectation of the loss given drought. We remark that instead of SPI optional drought index or drought definition can be used!

2. Regression model for the relative yield function

According to the formula (16) at Section 5 in D5.2.1 the regression model is,

$$\hat{Y}_{rel}(\mathbf{X}(\mathbf{s},t)) = \alpha_1 \cdot (\ln X_1(\mathbf{s},t) - E_{\ln,1}(\mathbf{s})) + \alpha_2 \cdot (X_2(\mathbf{s},t) - E_2(\mathbf{s})) \quad (2)$$

where $\hat{Y}_{rel}(\mathbf{X}(\mathbf{s},t))$ is the relative yield function in percent (D5.2.1,(15)), $X_1(\mathbf{s},t)$ is precipitation sum and $X_2(\mathbf{s},t)$ is temperature mean for a given 3 or 6 months period. Furthermore there are the appropriate expected values $E_{\ln,1}(\mathbf{s})$, $E_2(\mathbf{s})$ varying in space.
The common coefficients α_1, α_2 can be estimated by the method of least squares.
At this software the variables are, YieldPercent= $\hat{Y}_{rel}(\mathbf{X}(\mathbf{s},t))$, precipitation index PI= $\ln X_1(\mathbf{s},t) - E_{\ln,1}(\mathbf{s})$, temperature index TI= $X_2(\mathbf{s},t) - E_2(\mathbf{s})$ and standardized precipitation index SPI= $SPI(X_1(\mathbf{s},t))$. At the risk estimation formula (1) this $SPI(X_1(\mathbf{s},t))$ can be applied in accordance with Remark1 at Section 5 in D5.2.1. Then we can evaluate the connection and signification of the different type SPI series with the real drought loss.
At this software there is possibility to examine several SPI series with various periods. These examinations can be implemented for several 3 or 6 months periods and the estimated regression coefficients α_1, α_2 are included by the parameter files Reg{.}.par in the subdirectory RED\RiskCalculation\RegPar. These parameter files are given for four plants: maize, wheat, rape and barley. Detailed information can be found at Section VI.

III. THE STRUCTURE OF PROGRAM SYSTEM

Main Directory: **RED**

- Subdirectory **IndexCalculation**
    - Subdirectory **SPI** (for Standard Precipitation Index SPI)
        - Subdirectory **SPIcalc**
            - Main Program Files
            - Subdirectory sub (do not use it including subroutines)
            - Subdirectory work (do not use it)
        - Subdirectory **SPIselect**
            - Main Program Files
            - Subdirectory sub (do not use it including subroutines)
    - Subdirectory **RegressionIndeces**
        - Subdirectory **PI** (for Precipitation Index PI)
            - Subdirectory **PIcalc**
                - Main Program Files
                - Subdirectory sub (do not use it including subroutines)
                - Subdirectory work (do not use it)
            - Subdirectory **PIselect**
                - Main Program Files
                - Subdirectory sub (do not use it including subroutines)
        - Subdirectory **TI** (for Temperature Index TI)
            - Subdirectory **TIcalc**
                - Main Program Files
                - Subdirectory sub (do not use it including subroutines)
                - Subdirectory work (do not use it)
            - Subdirectory **TIselect**
                - Main Program Files
                - Subdirectory sub (do not use it including subroutines)
    - Subdirectory **IndexReCalibration**(for SPI, PI, TI together)
        - Main program files
        - Subdirectory sub (do not use it including subroutines)

- Subdirectory **RiskCalculation**
    - Main Program Files
    - Subdirectory **RegPar**
    - Subdirectory **RiskYD**

Directory **Example**

- Example for the main Input/Output files

IV. THE MAIN STEPS OF RISK ESTIMATION

1. INDEX CALCULATION

1.1.1 SPI calculation for 12 months

Input: PrecMonthly.ser
(station or gridded) monthly precipitation series for 12 months in one file,
max: 200 year, max: 10000 stations or grid points;
Calibration period; Run lengths (3 and 6 are suggested)
Output: series SPI{J}.ser (J run length) for 12 months; SPI.par
Run1: IndexCalculation\SPI\SPIcalc\StartSPI.bat
Run2: IndexCalculation\SPI\SPIcalc\MonthlySPI.bat
(by using SPI_monthly_CP.exe from DMCSEE for one location)
Run3: IndexCalculation\SPI\SPIcalc\CopyCalcSelect.bat

1.1.2 Selection of SPI for given month

Input: SPI{J}.ser, SPI.par (copied from 1.1.1 automatically); M (index of month)
Output: series SPI.ser for run length J and month M; LastSPI.txt
Run: IndexCalculation\SPI\SPIselect\SelectSPI.bat

1.2.1 PI calculation for 12 months

Input: PrecMonthly.ser (same as at 1.1.1); SPI.par (copied from 1.1.1 automatically)
Calibration period, Run lengths (automatic 1.1.1 SPI parametrization is suggested)
Output: series PI{J}.ser (J run length) for 12 months, PI.par
Run1: IndexCalculation\RegressionIndeces\PI\PIcalc\MonthlyPI.bat
Run2: IndexCalculation\RegressionIndeces\PI\PIcalc\CopyCalcSelect.bat

1.2.2 Selection of PI for given month

Input: PI{J}.ser, PI.par (copied from 1.2.1 automatically); M (index of month)
Output: series PI.ser for run length J and month M; LastPI.txt
Run: IndexCalculation\RegressionIndeces\PI\PIselect\SelectPI.bat

1.3.1 TI calculation for 12 months

Input: TempMonthly.ser
Monthly temperature series for 12 months in one file, for the same locations and time as PrecMonthly.ser at 1.1.1.; SPI.par (copied from 1.1.1 automatically)
Calibration period, Run lengths (automatic 1.1.1 SPI parametrization is suggested)
Output: series TI{J}.ser (J run length) for 12 months, TI.par
Run1: IndexCalculation\RegressionIndeces\TI\TIcalc\MonthlyTI.bat
Run2: IndexCalculation\RegressionIndeces\TI\TIcalc\CopyCalcSelect.bat

1.3.2 Selection of TI for given month

Input: TI{J}.ser, TI.par (copied from 1.3.1 automatically); M (index of month)
Output: series TI.ser for run length J and month M; LastTI.txt
Run: IndexCalculation\RegressionIndeces\TI\TIselect\SelectTI.bat

1.4 Recalibration of indices SPI, PI, TI

Input: SPI.ser, PI.ser, TI.ser with same parametrization (run length, month);
new calibration period; SPI.par (copied from 1.1.1 automatically)
Output: SPI.ser, PI.ser, TI.ser with the new calibration period;
Calibration.stat (some control statistics)
Run: IndexCalculation\IndexReCalibration\Calibration.bat

2. RISK CALCULATION

2.1 Regression of relative crop yield

Input: SPI.ser, PI.ser, TI.ser (copied from 1.1.2, 1.2.2, 1.3.2 automatically);
SPI.par (copied from 1.1.1 automatically); Reg.par (to be copied from subdir RegPar);
Index of regression on Reg.par; time interval of estimation
Output: YieldPercent.ser (series of estimated relative crop yield (%));
Risk.par, work.ser
Run: RiskCalculation\YieldRegression.exe

2.2 Definition of drought by SPI

Input: SPI.ser, Risk.par; probability of drought (0.3, 0.25, 0.2, 0.15, 0.1, 0.05)
Output: DroughtSPI.ser (indicator series of drought defined by SPI and probability:
drought: 1, no drought: 0), Risk.par
Run: RiskCalculation\DroughtSPI.exe

2.3 Estimation of risk on the basis of regression (2.1) and drought (2.2)

Input: YieldPercent.ser, DroughtSPI.ser, Risk.par
Output: Risk.res (estimated expected loss (%) and risk values)
Run: RiskCalculation\RiskCalc.exe

Remark 1
We had the remark at Section II.1 that instead of SPI optional drought index or drought definition can be used! In this case at step 2.3 the input file DroughtSPI.ser has to be replaced with other indicator series (0-1) of drought. The format of this input file can be seen on Fig.7 at Section V. The probality of drought must be given on the parameter file Risk.par whose format in the first row: value 1, number of locations (I6); value 2, first year (i6); value 3, last year (i6); value 3, probality of drought (f6.2).
Another possibility is to run: RiskCalculation\RiskDP.exe
Input: YieldPercent.ser, Drought.ser (same as 2.4);
Output: RiskDP.res (similar as RiskYDP.res at 2.4)

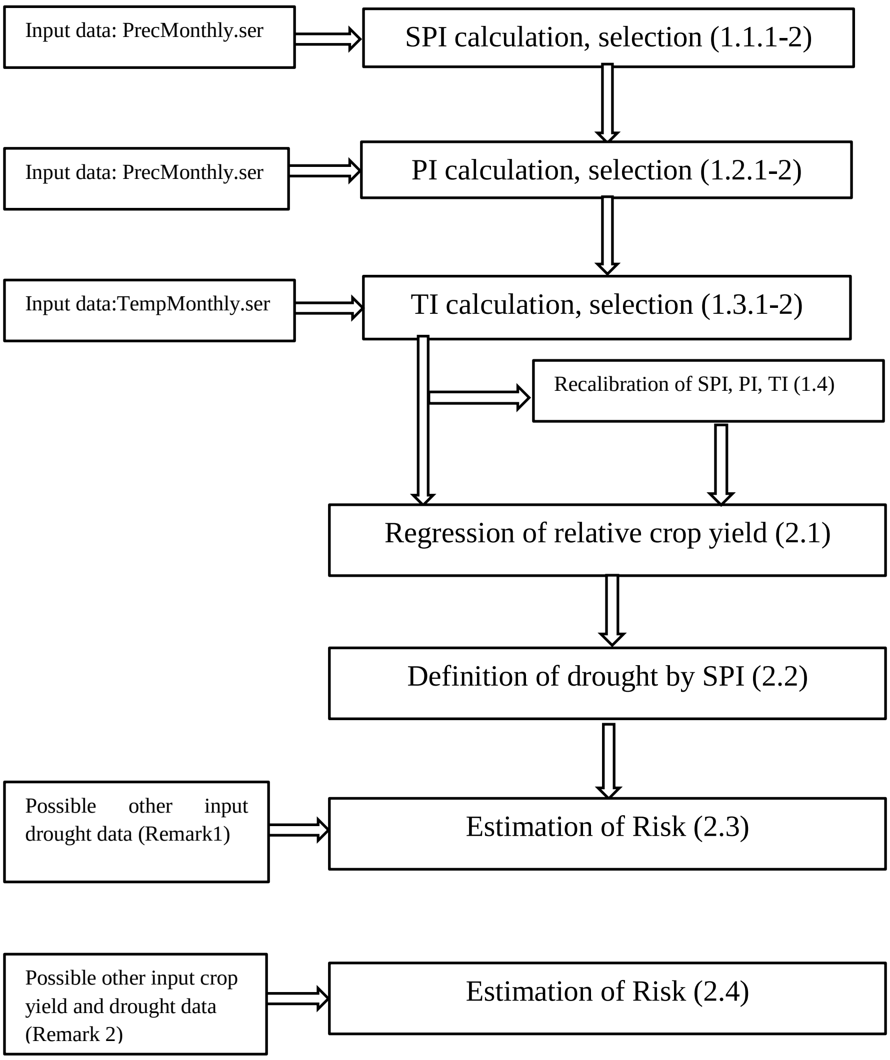


Figure 1. Flow diagram for steps 1.1-2.3 or 2.4

Remark 2
We did not intend to develop a general crop-weather model for the project region that would have been beyond our possibilities of course. Our intention was to estimate an expected value only i.e. the risk instead of a more precise crop yield estimation. Therefore, a not to complicated regression model was acceptable for our aim. However, in case of having crop-weather model and identification of drought with good quality then these values can be used according to the formula (10) at Section 3 in D5.2.1 by the following program.

2.4 Estimation of risk on the basis of given yield and drought

Input: Yield.ser (crop yield series), Drought.ser (indicator series of drought)
Output: RiskYDP.res
(estimated expected loss and risk values with probabilities of drought (PD))
Run: RiskCalculation\RiskYD\RiskYDP.exe

V. THE MAIN INPUT/OUTPUT FILES

Input Files: PrecMonthly.ser, TempMonthly.ser, Reg.par (see Section V.)

Output Files:
SPI.ser, PI.ser, TI.ser: index series
YieldPercent.ser: series of estimated relative crop yield (%)
DroughtSPI.ser: indicator series of drought defined by SPI and probability
Risk.res: estimated expected loss (%) and risk values

Format of PrecMonthly.ser, TempMonthly.ser:
Many years' monthly data series without missing values (max. number of series: 10000, max. number of years: 200)
row 1: station number or index, or grid index (obligatory!), Format: I8
column 1: date of year (I4)
column 2: month (I3)
column i+2: series i.
Data Format: F8.2 (See the Data Files of EXAMPLE)

Example: monthly data series (1961-2010) for 10 locations

		1	2	3	4	5	6	7	8	9	10
1961	1	29.67	27.39	24.43	22.51	22.33	19.49	20.14	20.47	20.08	18.07
1961	2	25.64	22.19	21.51	23.08	22.71	19.45	18.16	20.03	17.98	16.28
1961	3	9.64	9.21	11.01	14.75	15.31	14.88	13.31	10.11	10.16	9.62
......											
2010	9	144.19	131.40	123.81	104.35	94.68	91.29	86.79	96.30	89.50	81.86
2010	10	25.65	18.82	20.80	21.14	22.78	25.82	24.03	20.57	21.40	18.84
2010	11	89.75	81.23	77.79	71.30	68.30	64.54	61.42	73.56	76.09	70.87
2010	12	91.66	87.52	78.19	74.91	73.52	69.04	75.50	89.41	87.48	79.60

Figure 1. Format of input file PrecMonthly.ser

		1	2	3	4	5	6	7	8	9	10
1961	1	-3.34	-3.95	-4.15	-3.92	-3.72	-3.60	-2.98	-3.04	-4.44	-4.76
1961	2	-0.62	-0.95	-1.24	-0.86	-0.77	-0.49	0.08	-0.08	-1.27	-1.39
1961	3	5.95	6.17	5.71	6.02	5.99	6.24	6.77	6.37	5.56	5.28
....											
2010	9	13.20	12.90	13.26	13.69	13.64	14.26	13.92	13.08	12.74	13.78
2010	10	6.78	6.31	6.49	6.89	6.91	7.37	7.31	6.85	6.03	6.79
2010	11	6.72	6.79	6.52	6.82	6.78	7.08	7.40	7.22	6.53	6.50
2010	12	-3.64	-3.93	-4.13	-3.65	-3.35	-2.90	-2.26	-2.45	-3.74	-3.74

Figure 2. Format of input file TempMonthly.ser

Calibration period: 1961-1990 Month: 8 spi6.ser

	1	2	3	4	5	6	7	8	9	10
1961	-1.66	-1.60	-1.88	-1.73	-1.76	-1.86	-2.00	-2.57	-1.94	-2.10
1962	-1.28	-1.15	-1.13	-1.06	-0.92	-1.26	-1.34	-1.12	-1.23	-1.21
1963	-0.06	-0.32	-0.41	-0.50	-0.66	-1.04	-1.20	-1.20	-1.42	-1.27
......										
2008	-0.04	-0.29	-0.43	-0.32	-0.48	0.13	0.28	0.65	0.95	1.00
2009	-0.58	-0.51	-0.45	-0.35	-0.33	-0.42	-0.58	-0.45	-1.26	-1.14
2010	2.58	3.10	2.83	2.78	2.50	2.52	2.62	2.47	2.87	3.27

Figure 3. Format of output file SPI.ser

```
Calibration period: 1961-1990   Month: 8   pi6.ser
          1      2      3      4      5      6      7      8      9     10
1961  -0.44  -0.41  -0.50  -0.46  -0.47  -0.51  -0.55  -0.81  -0.50  -0.54
1962  -0.33  -0.29  -0.28  -0.27  -0.23  -0.33  -0.36  -0.32  -0.30  -0.29
1963   0.00  -0.07  -0.09  -0.12  -0.16  -0.27  -0.31  -0.34  -0.35  -0.31
........................
2008   0.00  -0.06  -0.10  -0.07  -0.11   0.04   0.08   0.19   0.23   0.24
2009  -0.14  -0.12  -0.10  -0.08  -0.07  -0.10  -0.14  -0.12  -0.31  -0.27
2010   0.59   0.69   0.64   0.64   0.58   0.59   0.62   0.64   0.63   0.70
```

Figure 4. Format of output file PI.ser

```
Calibration period: 1961-1990   Month: 8    ti6.ser
          1      2      3      4      5      6      7      8      9     10
1961   0.81   0.81   0.80   0.75   0.71   0.71   0.67   0.71   0.67   0.66
1962  -0.56  -0.54  -0.55  -0.59  -0.62  -0.65  -0.66  -0.64  -0.70  -0.71
1963   0.42   0.42   0.42   0.38   0.38   0.35   0.34   0.32   0.33   0.33
.................................
2008   1.03   0.99   1.02   0.98   1.00   0.98   0.99   0.99   0.97   0.96
2009   1.44   1.47   1.50   1.51   1.53   1.55   1.57   1.59   1.58   1.58
2010   0.87   0.84   0.89   0.93   0.93   1.00   1.02   0.99   1.02   1.03
```

Figure 5. Format of output file TI.ser

```
MAIZE              Period of Months:  3 - 8           Yield(%)
            1       2       3       4       5       6       7       8       9      10
1961   -17.85  -17.19  -19.46  -17.30  -16.99  -18.74  -19.46  -27.04  -17.62  -18.66
1962    -0.02    0.72    1.35    2.54    4.19    1.02    0.26    1.61    2.82    3.23
1963    -0.90   -3.15   -3.49   -3.47   -4.46   -7.92   -8.97   -9.22   -9.63   -8.46
..................
2008    -7.42   -8.94  -10.19   -8.43   -9.64   -5.65   -4.58   -0.99    0.37    0.78
2009   -15.86  -15.81  -15.32  -14.38  -14.14  -15.81  -17.17  -16.40  -21.83  -20.65
2010    11.43   14.44   12.69   12.73   11.15   10.11   10.78   12.08   11.46   13.40
```

Figure 6. Format of output file YieldPercent.ser

```
Calibration period: 1961-1990   Month: 8    spi6.ser        Probability of Drought:   0.10
         1     2     3     4     5     6     7     8     9    10
1961     1     1     1     1     1     1     1     1     1     1
1962     1     0     0     0     0     0     1     0     0     0
1963     0     0     0     0     0     0     0     0     1     0
….........
2008     0     0     0     0     0     0     0     0     0     0
2009     0     0     0     0     0     0     0     0     0     0
2010     0     0     0     0     0     0     0     0     0     0
```

Figure 7. Format of output file DroughtSPI.ser (drought: 1, no drought: 0)

```
MAIZE              Period of Months:  3 - 8
Time Interval: 1961-2010
Probability of Drought:  0.10
Location    Loss(%)      Risk
      1      13.40      1.340
      2      17.36      1.736
      3      20.15      2.015
      4      24.67      2.467
      5      25.96      2.596
      6      22.29      2.229
      7      16.46      1.646
      8      26.73      2.673
      9      22.30      2.230
     10      24.13      2.413
```

Figure 8. Format of output file Risk.res

(Remark: Loss(%) is the estimated conditional expectation of the loss given drought.)

VI. MAPPING OF RISK

Final output, risk.res contains estimated expected loss (%) and risk values for the intput (station or grid) points given in PrecMonthly.ser and TempMonthly.ser.

For creating maps, the risk values can be integrated in any GIS software after giving the coordinates of the input points in a text or excel file (Riskmap.dat).

Location	fi	la	Loss(%)	Risk
1	45.9	17.9	13.40	1.340
2	45.7	17.8	17.36	1.736
3	45.3	17.5	20.15	2.015
4	45.4	18.2	24.67	2.467
5	45.3	18.3	25.96	2.596
6	46.7	21.3	22.29	2.229
7	47.7	19.5	16.46	1.646
8	47.2	20.8	26.73	2.673
9	46.9	21.0	22.30	2.230
10	45.8	19.9	24.13	2.413

Figure 9. Format of output file Riskmap.dat

An interpolation method included in the GIS software can be used to make map from station values or smooth the grid point data.

VII. REGRESSION PARAMETER FILES (REG.PAR)

One of the key issues of the developed risk estimation methodology is the regression of the relative crop yield on the meteorological variables according to the formula I.(1). The type of the regression formula seems acceptable so we intended to estimate the unknown parameters α_1, α_2 by the method of least squares.

Then the first step is collecting sample for the crop yield $\text{Yield}(\mathbf{s},t)$ and the meteorological variables $X_1(\mathbf{s},t)$, $X_2(\mathbf{s},t)$. Data requirements from project partners were impact data (crop yield) and meteorological data for sample of regression. For this purpose, we used also detailed yield database (2001-2016) on about 1900 pilot sites in the Farm Accountancy Data Network (FADN) for the four main plants in Hungary: maize, wheat, rape and barley. As regards the reliable meteorological variables the precipitation and temperature series homogenized by method MASH (*Szentimrey*) were interpolated by method MISH (*Szentimrey and Bihari*) for the pilot sites.
The regression on the sample were implemented for the most important 3- or 6-months periods and the estimated regression coefficients α_1, α_2 are included by the parameter files in the subdirectory RED\RiskCalculation\RegPar. These parameter files are given for four plants: maize, wheat, rape and barley. For example, the RegMaize.par is the following.

```
MAIZE
ind  nv   period      PI        TI        corr
  1   2    9 - 2    32.275     0.030     0.285
  2   2   10 - 3    25.697     0.949     0.246
  3   2   11 - 4    32.281     3.209     0.298
  4   2   12 - 5    14.468     0.486     0.136
  5   2    1 - 6    23.351    -0.691     0.251
  6   2    2 - 7    45.632    -0.864     0.429
  7   2    3 - 8    29.044   -10.683     0.488
  8   2    4 - 9    20.211   -12.006     0.427
  9   2    9 -11    21.537    -5.143     0.374
 10   2   10 -12    25.736     0.082     0.305
 11   2   11 - 1    13.115    -1.469     0.162
 12   2   12 - 2     0.004     1.284     0.067
 13   2    1 - 3     7.177     0.766     0.114
 14   2    2 - 4    18.144     3.139     0.303
 15   2    3 - 5    11.509    -4.227     0.227
 16   2    4 - 6    10.632   -14.004     0.399
 17   2    5 - 7    28.579   -15.428     0.572
 18   2    6 - 8    18.590   -14.386     0.578
 19   2    7 - 9    15.508    -7.286     0.359
```

Figure 1. Regression parameter file RegMaize.par

Notations of the parameter file:
ind: index of regression (input of RiskCalculation\YieldRegression.exe)
nv: number of independent variables
period: period of months (with length 3 or 6)
PI: coefficient α_1 of PI= $\ln X_1(\mathbf{s},t) - E_{\ln,1}(\mathbf{s})$ at regression formula I.(2)
TI: coefficient α_2 of TI= $X_2(\mathbf{s},t) - E_2(\mathbf{s})$ at regression formula I.(2)
corr: multiple correlation of the regression

Four parameter files (RegMaize.par, RegWheat.par, RegRape.par, RegBarley.par) can be found in the subdirectory RiskCalculation\RegPar. Before application please to copy and rename the actual parameter file into the directory RiskCalculation as Reg.par and during running of RiskCalculation\YieldRegression.exe choose the index of regression.

WHEAT

ind	nv	period	PI	TI	corr
1	2	9 - 2	23.417	3.895	0.311
2	2	10 - 3	22.189	3.660	0.315
3	2	11 - 4	18.230	5.015	0.284
4	2	12 - 5	12.447	3.512	0.191
5	2	1 - 6	16.513	2.597	0.221
6	2	2 - 7	20.626	1.603	0.239
7	2	3 - 8	6.941	-7.539	0.290
8	2	4 - 9	-5.554	-12.447	0.285
9	2	9 -11	16.329	-1.430	0.330
10	2	10 -12	15.496	1.372	0.237
11	2	11 - 1	0.440	1.827	0.087
12	2	12 - 2	2.827	3.760	0.248
13	2	1 - 3	9.491	2.673	0.269
14	2	2 - 4	11.944	3.789	0.320
15	2	3 - 5	9.135	-2.852	0.218
16	2	4 - 6	-7.854	-15.451	0.369
17	2	5 - 7	-1.828	-12.801	0.409
18	2	6 - 8	-5.774	-10.448	0.316
19	2	7 - 9	4.262	-3.066	0.160

Figure 2. Regression parameter file RegWheat.par

RAPE

ind	nv	period	PI	TI	corr
1	2	9 - 2	9.692	6.461	0.275
2	2	10 - 3	15.843	6.143	0.326
3	2	11 - 4	11.889	7.756	0.330
4	2	12 - 5	14.089	7.465	0.307
5	2	1 - 6	16.934	6.272	0.273
6	2	2 - 7	19.139	6.093	0.244
7	2	3 - 8	9.821	0.049	0.100
8	2	4 - 9	1.675	-4.271	0.110
9	2	9 -11	7.129	1.430	0.122
10	2	10 -12	6.400	4.107	0.176
11	2	11 - 1	-3.255	4.928	0.242
12	2	12 - 2	7.422	5.840	0.351
13	2	1 - 3	11.678	4.036	0.349
14	2	2 - 4	11.842	5.782	0.347
15	2	3 - 5	10.116	4.041	0.135
16	2	4 - 6	-8.130	-7.027	0.135
17	2	5 - 7	1.992	-6.853	0.195
18	2	6 - 8	-4.410	-6.811	0.168
19	2	7 - 9	6.560	-0.826	0.114

Figure 3. Regression parameter file RegRape.par

BARLEY

ind	nv	period	PI	TI	corr
1	2	9 - 2	15.432	4.227	0.243
2	2	10 - 3	12.461	4.394	0.247
3	2	11 - 4	6.710	5.150	0.223
4	2	12 - 5	5.280	4.829	0.196
5	2	1 - 6	8.072	4.370	0.182
6	2	2 - 7	10.642	3.859	0.153
7	2	3 - 8	4.410	-2.322	0.110
8	2	4 - 9	-2.641	-5.558	0.120
9	2	9 -11	13.108	-1.447	0.252
10	2	10 -12	9.004	0.884	0.133
11	2	11 - 1	-3.471	2.485	0.137
12	2	12 - 2	-0.488	4.280	0.271
13	2	1 - 3	4.996	3.506	0.263
14	2	2 - 4	4.856	4.116	0.243
15	2	3 - 5	5.086	-0.021	0.081
16	2	4 - 6	-11.084	-10.521	0.220
17	2	5 - 7	-2.452	-7.090	0.206
18	2	6 - 8	-3.060	-4.797	0.134
19	2	7 - 9	7.214	0.937	0.104

Figure 4. Regression parameter file RegBarley.par

VIII. EXAMPLE: MAPPING OF RISK FOR HUNGARY

Some figures are presented below for illustration of mapping risk.
These risk maps were based on gridded data series. The risk calculation was implemented for a grid with spatial resolution 0.1° for Hungary and the output risk.res values are presented.

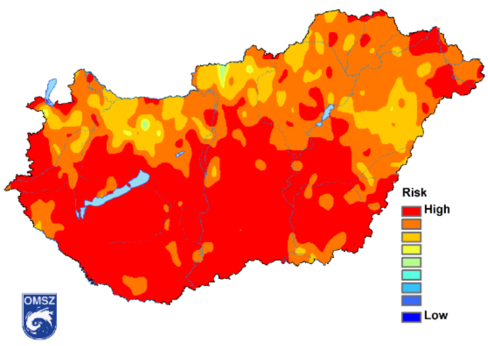


Figure 1. Risk Map for Maize (period: 5-7, drought probability: 0.2)

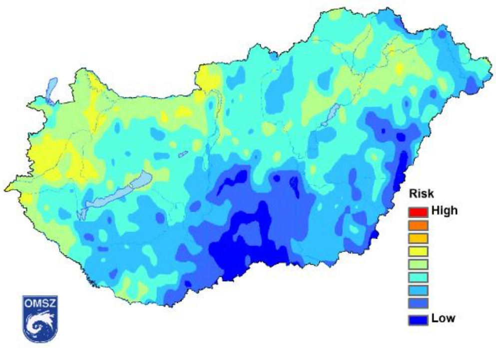


Figure 2. Risk Map for Wheat (period: 1-6, drought probability: 0.2)

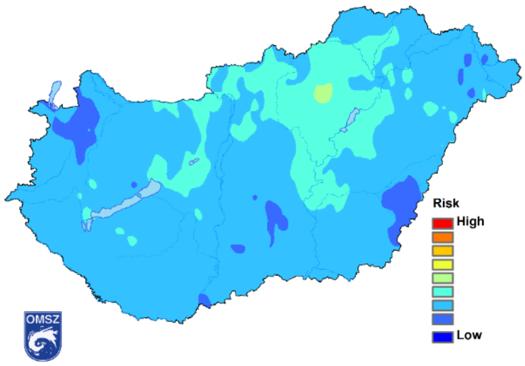


Figure 3. Risk Map for Rape (period: 2-4, drought probability: 0.2)

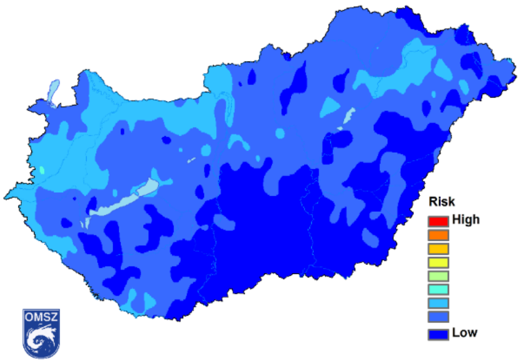


Figure 4. Risk Map for Barley (period: 2-4, drought probability: 0.2)

References

CarpatClim Project: http://www.carpatclim-eu.org/pages/home/

Drought Management Centre for South- East Europe – DMCSEE: Summary of project results https://www.met.hu/doc/DMCSEE/DMCSEE_final_publication.pdf

Drought Risk Danube Region Project (DriDanube): Application Form (AF)

DriDanube, Deliverable D5.2.1: Algorithm of drought risk assessment

EUROPEAN COMMISSION: COMMISSION STAFF WORKING PAPER, Risk Assessment and Mapping Guidelines for Disaster Management, Brussels, 21.12.2010 SEC(2010) 1626 final

Szentimrey, T., 2014: Manual of homogenization software MASHv3.03, Hungarian Meteorological Service, p. 69.

Szentimrey, T., Bihari, Z., 2014: Manual of interpolation software MISHv1.03, Hungarian Meteorological Service, p. 60

WMO: Standardized Precipitation Index, User Guide http://www.wamis.org/agm/pubs/SPI/WMO_1090_EN.pdf

Annex 3. Rainless periods (droughts) assessment by the Zelenhasic-Todorovic method

1. Mathematical background

General notice: terms 'rainless period' and 'drought' are used as synonyms.

The ZT method (after Zelenhasic and Todorovic, CSU in Fort Collins - USA) is a general stochastic model of extremes, here rainless periods. The method uses daily data of rainfalls, and firstly identifies rainless periods above a given reference value, Y_r. All important components of the process, such as number of droughts, drought duration, time of occurrence, in a given time interval [0,*t*], the longest drought in a given time interval [0,*t*], and its time of occurrence, are then taken into consideration, determination of distribution functions and interpretation of the method's output.

Droughts are defined as the upper extremes of dry weather intervals and are treated as a random number of random variables in an interval of time [0,*t*]. By assumption, droughts are independent events, represented by identically distributed random variables which follow the Poisson probability law.

An application of the ZT method can be conducted for the part of calendar year, for instance, the vegetation season starting on 1st April and ending on 30th September. This period is of the prime importance for agriculture, and the project. Therefore, it will be used hereafter for theoretical exposition of the ZT method as closed time interval [0,*t*]. Notice, however, that the ZT method can be applied for the entire calendar year or any selected part of it.

Drought can be defined in many ways, depending on local conditions (such as soil, planting, regional climate, etc.) and sectoral characterization such as river transport, urban water supply, water management etc. In particular, from the agricultural standpoint, drought is considered as undesired event described by consecutive rainless days with a precipitation of less than 3 mm in any day, and lasting for at least 20 days. Experts in agronomy indicated that values 3mm and 20 consecutive days is sufficiently reliable benchmark for recognizing hazard rainless events as key descriptor of agro-meteorological droughts.

Regarding the reference value Y_r, two groups of rainless events *m* exist:

a) $Y_m > Y_r$, with $(Y_m - Y_r) > 0$ (when the rain depth during event *m* is above reference value of 20 days)
b) $Y_m \leq Y_r$. (otherwise)

Events $Y_m > Y_r$ can be considered as the extreme rainless periods, and treated as a stochastic (random) variable. In turn, droughts can be designated as X_v, where v = 1,2, ..., m.

Each drought event is discrete event with discretization step of one day. It is composed of the following defining descriptive parameters:

(1) drought duration (in days), X_v
(2) time of the beginning of a drought (calendar datum), $\tau_b(v)$
(3) time of the end of a drought (calendar datum), $\tau_e(v)$
(4) time of a drought occurrence, $\tau(v)$, defined as midpoint between starting and ending datum

$$\tau(\nu)=\frac{1}{2}\left[\tau_b(\nu)+\tau_e(\nu)\right]$$

(5) order number of a drought, ν, for a given time interval [0,t] for a particular vegetation season, where ν = 1, 2, ...

Considering the entire process of droughts, three additional magnitudes enter the analysis:

(6) total number of droughts, k, within the time interval [0,t], where k=0, 1, 2, ...
(7) the longest (largest) drought within a time interval [0,t], $X(t)=\sup X_\nu$, and $\tau(\nu)\leq t$

(8) time of the occurrence, $\tau(t)$, of the longest (largest) drought within time interval [0,t].

The ZT method also involves random variable Z_ν for time interval [0,t] defined as $Z_\nu = X_\nu - Y_r$. For adopted reference value of 20 days, it follows $Z_\nu = X_\nu - 20$, where Z_ν and X_ν are measured in days.

According to the nature of drought phenomenon, it is obvious that their total number in time interval [0,t], as well as their duration and times of occurrence, are random variables. This way the stochastic process of extreme rainless periods can be considered as completely described stochastic process and each component of the process can be analyzed by use of some associated distribution functions. The statistical tests (Pearson χ^2-test and Kolmogorov-Smirnov test) are useful to apply to check an agreement between theoretical and empirical distribution functions for all analyzed components of the process.

1.1 Distribution of the number of droughts

The distribution of the number of droughts at a given location is an important component of the method and many earlier works (e.g., Zelenhasić, 1970; Berić et al., 1990) indicated that it is expected that the number of droughts in [0,t] at given location will also be distributed according to the Poisson probability law (PPL). For instance, at different meteorological stations in Vojvodina Province (Serbia) it was shown that PPL is valid for each of six-monthly periods in the vegetation season (periods starting with 1st of April and lasting after 30, 61, 91, 122, 153 and 183 days). In order to estimate function $\lambda(t)$ for period of interest, that is vegetation season (long 183 days) in DriDanube countries, the distribution of the number of droughts is defined as stated by the Poisson probability law

$$P\left(E_k^t\right)=\frac{\left[\lambda_1(t)\right]^k}{k!}\mathrm{e}^{-\lambda_1(t)} \qquad (1)$$

In Eq. (1) E_k^t is the event that exactly k droughts occurred in interval [0,t] (k = 0,1,2, …) and $\lambda_1(t)$ is the mean number of droughts in a time interval [0,t], computed across all years in analyzed series of historical daily rainfalls.

Notice: Hereafter continuous interval [0,t] corresponds to discrete domain [1,183] determined by starting and ending date of vegetation season: 1st April (1) - 30th September (183).

1.2 Distribution of droughts duration

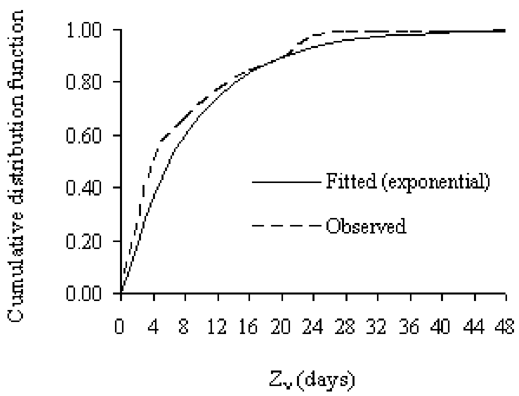
To determine the distribution of the random variable Z_v and analyze drought durations as stochastic variable X_v , the time interval [0,*t*] should be set (and it is) equal to the vegetation season so that corresponding theoretical exponential distribution function is determined as:

$$B(Z)=1-e^{-\lambda_2 \cdot Z}; \quad Z \geq 0 \qquad (2)$$

where λ_2 is a parameter estimated as $\lambda_2 = \frac{1}{\bar{Z}}$.

For any given location (e.g. meteorological station or grid point in CarpatClim or DanubeClim data base), it is easy to identify from the data record the maximum observed values of Z_v (max Z_{rec}), and to compute mean and standard deviation of the random variable Z_v.

Computer program ZTDM computes all required parameters for drawing both observed and corresponding theoretical distribution functions of the random variable Z_v as shown below.



The Kolmogorov-Smirnov test is used to confirm the good agreement between observed and theoretical distribution functions, or to indicate if there is not satisfactory agreement. In later case data input has to be checked and if there are not errors in it, theoretical distribution function should not be used and analyst should stick to empirical distribution function only. In most of applications Kolmogorov-Smirnov test indicate good agreements of empirical and theoretical distribution functions.

1.3 Distribution of the longest droughts duration

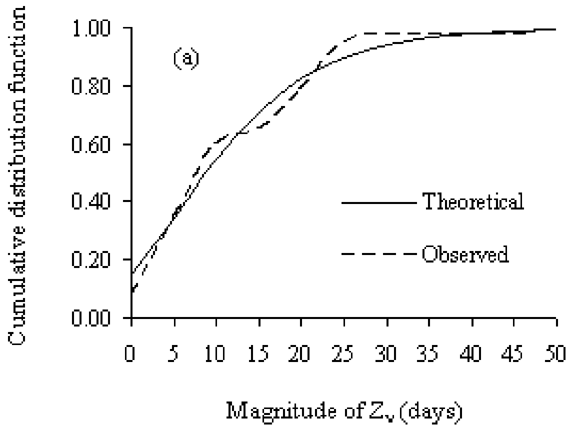
The theoretical distribution function of the longest drought is determined as double exponential distribution function:

$$F(Z/t)=e^{-\lambda_1(t)\cdot e^{-\lambda_2 \cdot Z}} \quad Z \geq 0 \qquad (3)$$

where $\lambda_1(t)$ is the mean number of droughts during the vegetation season; and λ_2 - parameter of distribution of drought durations during the vegetation season.

Again, as in previous case (all droughts assessment), both the Kolmogorov-Smirnov and the chi-square goodness of fit tests are used in the DRZTM computer program to check goodness of agreement between the two distribution functions. Diagram presented below illustrates this case.

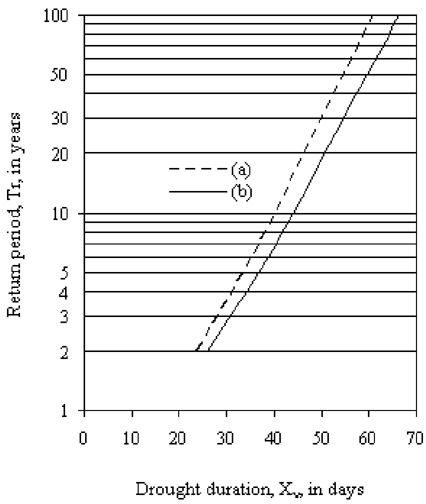
Otherwise, empirical function should be used instead, like in previous case (Section 1.2).



Practical value of the theoretical distribution function of the longest droughts is possibility to estimate numerical values X_v of droughts for different return periods. Based on equation (4) , with stochastic variable and with assumption of 20 days as reference rainless point ($Z_v = X_v - Y_r = X_v - 20$), the return period T_r of stochastic variable Z_v is easy to compute as:

$$T_r = \frac{1}{1 - F(Z/t)} \tag{5}$$

Typical relationship $T_r = f(X_v)$ is shown in diagram below.



This diagram helps to determine the return period for longest drought at given location (here grid point) and is of the prime interest for the project as well as for users from different sectors such as agriculture, water management, climate dependent projects, insurance, etc.

1.4 Distribution of the time of occurrence of the longest drought

It is important to know which part of the vegetation season has the highest probability of the longest drought occurring in it. Recall that the longest dry weather intervals in the vegetation season are plotted as impulse functions in the middle of their time intervals with the magnitude equal to their durations. For the time when the supremum of a random variables occurs, a function can be defined as:

$$F\left\{\tau(t)\le u\right\}=e^{-\lambda_1(t)}+\frac{\lambda(u)}{\lambda_1(t)}\left[1-e^{-\lambda_1(t)}\right] \quad (6)$$

where $\lambda_1(t)$ is the mean number of droughts in the vegetation season, *u* (in days) is numerical value which is taken by random variable $\tau(t)$, and where $0 \le u \le t$. $\lambda(u)$ is a value taken from function $\lambda(t)$ and $\tau(t)$ is a moment in time interval [0,t] when the longest extreme rainless period occurred.

An example final version of equation (6) is

$$F\left\{\tau(t)\le u\right\}=0.1508+0.4489\lambda(u) \quad (7)$$

and again, the Kolmogorov-Smirnov goodness of fit test is used in the DRZTM program to check for a good agreement between observed data and values from the above distribution function.

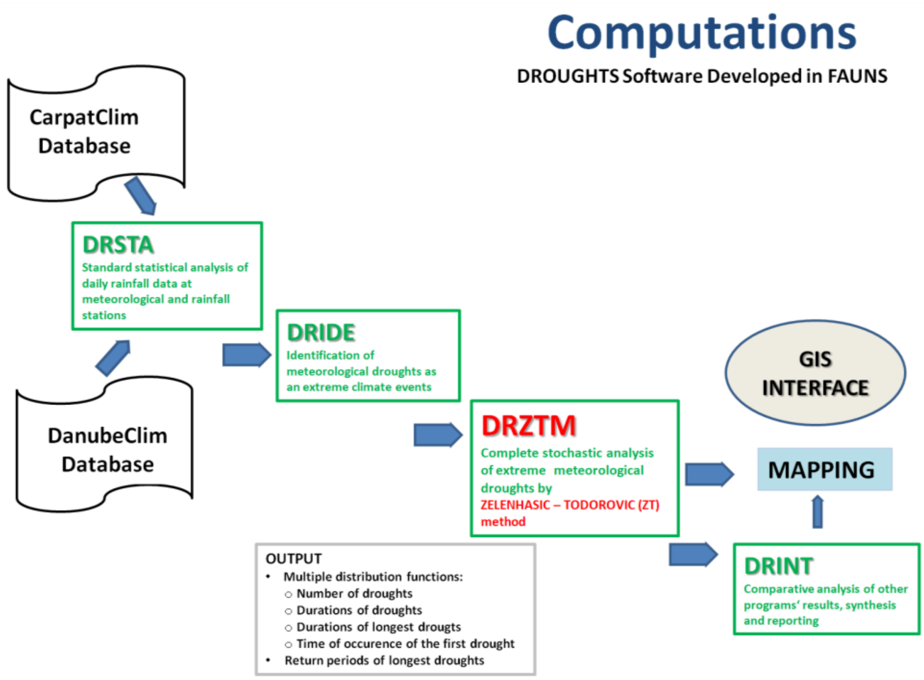
Use of equations (6) and (7) and observed data enables user to identify the part of the vegetation season with the highest probability of having the longest drought. In case of Serbia, for example, it was found that in a part of vegetation season from second half of August and first half of September it is most probable to expect longest droughts.

2. DROUGHTS Software

Based on collected daily rainfall data for representative rainfall/meteorological stations in the region (covered by the project), the programming system DROUGHTS is developed in the FAUNS. Its architecture is shown in the figure below. The major part of the system is developed for project purposes only and run on standard PC platforms under MS Windows 7+. Computer programs are written in Fortran programming language.

The DROUGHTS is used to perform comprehensive analysis of extreme rainless periods of different return periods and enable mapping of risky areas. It is only briefly described in the following subsections.

Note that the core of the system is the DRZTM program which realizes major steps described in Section 1 (Mathematical background). The DRZTM is already used for detail stochastic analysis of rainless periods as drought events at all 170 selected grid points in a region covered by the project. The most important parts of its output are used as a data set for mapping.



2.1 Description of programs within methodological framework

Only base input/output solutions will be presented. Additional features of all programs can be used for extended statistical analyses and are not described here.

DRSTA – Standard statistical analysis of daily rainfall data at grid points

Input: Daily yearly rainfall data in millimeters for multi-year period 1981-2010 at given grid point (in general for given meteorological/rainfall station), for complete calendar year. Data input adjusted from a file in .prn format after it is created by extracting data from the .xls file. Program offers other input data preparation solutions.

Output: Computed basic statistics for vegetation season only consists of monthly (Apr-Sep) and yearly sums, averages, max/min values, standard deviations, and variation coefficients. Trend for yearly sums of rainfalls in vegetation season. Weibull empirical distribution of Apr-Sep sums of rainfalls. Output is also the data file with rainfalls for vegetation season only. File is created for given grid point, one at time.

DRIDE – Identification of meteorological droughts as an extreme climate events

Input : Output file generated by the DRSTA program containing daily rainfall data in millimeters for multi-year period at given grid point (for vegetation season only).

Output: Identified extreme drought events in each year at a given station, in vegetation season, as rainless periods longer than 20 days. Days with less than 3 mm of rainfall are considered as rainless.

Other features: Reference values of 3 mm (to recognize rainless days) and 20 (consecutive rainless days) are changeable to distinguish even more extreme drought events; Additional criteria for determining rainless periods (e.g. internal sums of consecutive days during drought above certain reference value; not fully implemented yet).

DRZTM – Complete stochastic analysis of extreme rainless extreme periods by the ZT method

Input: Parameters (grid point identification, period of analysis, number of droughts, steps of ZT method to be executed); Data about all identified rainless periods during vegetation season above reference level of 20 days) for multi-year period at given grid point. Data set includes order number of drought, year, middle datum of drought, duration of drought.

Output:

Step 1: Print all droughts data (starting, ending, mid-point) and dates of occurrence of all historical droughts
Step 2: Distribution of the number of droughts for different time periods
Step 3: Distribution of drought lengths
Step 4: Distribution of the longest yearly droughts
Step 6: Distribution of the time of occurrence of the largest drought

Program generates both empirical and theoretical distributions in steps 2 through 5. Output of this program will be is used for integration and synthesis analyses within the program DRINT whose execution should enable automatic communication with GIS sub-system for interpolation and mapping return periods of longest droughts in the DriDanube region. DRINT is still in the development phase, but this is of no significant importance for the AF required output.

3. Application (Testing and Verification Phase)

Based on applications of a software system DROUGHTS (programs DRSTA, DRIDE and DRZTM), and interpolation and mapping in GIS environment, a sample mapping presented in the figure in the Section *Output of rainless periods* assessment is given for illustrative purposes. The implementation of methodology for stochastic analysis of rainless periods in DriDanube region is in its final stage. Testing and verification phase is underway. It is expected that necessary adjustments will be made to enable synergy of FAUNS algorithm results with the results of the algorithm developed by OMSZ. Estimates of risks due to the droughts will obviously be subject to interpretation of probabilities of dry periods, actual weather conditions and predicted losses in agricultural production in each of the partner countries. A feedback from all the partners in the project will help to make step forward in complete risk assessments caused by the droughts.

4. References

[1] Berić M., Zelenhasić E., Srđević B. (1990). Extreme dry weather intervals of the growing season in Backa, Yugoslavia. Water Resources Management 4: 79–95.
[2] CarpatClim Project: http://www.carpatclim-eu.org/pages/home/
[3] Drought Risk Danube Region Project (DriDanube): Application Form (AF)
[4] Rajić M., Bezdan A. (2012). Contribution to research of droughts in Vojvodina Province. Carpathian Journal of Earth and Environmental Sciences 7(1): 101–107.

[5] Srdjevic B. (1989): ZTDRS - Computer program for streamflow drought analysis by the ZT method, Proc. and Program Catalogue, 2nd International Software Exibition for Environmental Science and Engineering, 73-77, Como, Italy, 1989.
[6] Todorović P., Zelenhasić E. (1970) A Stochastic Model for Flood Analysis. Water Resources Research 6(6): 1641–1648.
[7] Zelenhasić E. (2002) On the extreme streamflow drought analysis. Water Resources Management 16(2): 105–132.
[8] Zelenhasić E. and Salvai A. (1987) A Method of Streamflow Drought Analysis. Water Resources research 23(1): 156 – 168.
[9] Zelenhasić E. (1970) A stohastic model for flood analysis. Ph.D. dissertation. Colorado State Univeristy, Fort Collins, USA.

Authors

OMSZ: Zita Bihari, Tamás Szentimrey, Lilla Hoffmann

FAUNS: Zorica Srđević, Bojan Srđević, Pavel Benka, Milica Rajić, Jasna Grabić

11. November 2018.

.....