Article

Application of Hybrid Prediction Methods in Spatial Assessment of Inland Excess Water Hazard

Annamária Laborczi [1], Csaba Bozán [2], János Körösparti [2], Gábor Szatmári [1,*], Balázs Kajári [2], Norbert Túri [2], György Kerezsi [2] and László Pásztor [1]

[1] Institute for Soil Sciences and Agricultural Chemistry, Centre for Agricultural Research, Herman Ottó út 15, H-1022 Budapest, Hungary; laborczi@rissac.hu (A.L.); pasztor@rissac.hu (L.P.)

[2] Research Institute of Irrigation and Water Management, National Agricultural Research and Innovation Centre, Anna-liget str. 35, H-5540 Szarvas, Hungary; bozan.csaba@ovki.naik.hu (C.B.); korosparti.janos@ovki.naik.hu (J.K.); kajari.balazs@ovki.naik.hu (B.K.); turi.norbert@ovki.naik.hu (N.T.); kerezsi.gyorgy@ovki.naik.hu (G.K.)

* Correspondence: szatmari@rissac.hu

check for updates

Abstract: Inland excess water is temporary water inundation that occurs in flat-lands due to both precipitation and groundwater emerging on the surface as substantial sources. Inland excess water is an interrelated natural and human induced land degradation phenomenon, which causes several problems in the flat-land regions of Hungary covering nearly half of the country. Identification of areas with high risk requires spatial modelling, that is mapping of the specific natural hazard. Various external environmental factors determine the behavior of the occurrence, frequency of inland excess water. Spatial auxiliary information representing inland excess water forming environmental factors were taken into account to support the spatial inference of the locally experienced inland excess water frequency observations. Two hybrid spatial prediction approaches were tested to construct reliable maps, namely Regression Kriging (RK) and Random Forest with Ordinary Kriging (RFK) using spatially exhaustive auxiliary data on soil, geology, topography, land use, and climate. Comparing the results of the two approaches, we did not find significant differences in their accuracy. Although both methods are appropriate for predicting inland excess water hazard, we suggest the usage of RFK, since (i) it is more suitable for revealing non-linear and more complex relations than RK, (ii) it requires less presupposition on and preprocessing of the applied data, (iii) and keeps the range of the reference data, while RK tends more heavily to smooth the estimations, while (iv) it provides a variable rank, providing explicit information on the importance of the used predictors.

Keywords: geostatistics; hybrid spatial prediction model; inundation; land degradation; machine learning; risk mapping

1. Introduction

Inland excess water (IEW) is temporary water inundation, a form of surplus surface water, which occurs in flatlands due to both precipitation and groundwater emerging on the surface as substantial sources. It occurs most frequently in local depressions of large flat areas, irrespective of river floods. A complex interaction of natural (e.g., meteorological, hydrogeological, pedological, topographical), and anthropogenic (e.g., land use, agricultural engineering) factors contribute to the occurrence of IEW [1,2]. It causes several social, economic, and environmental problems in the flat-land regions of Hungary, covering nearly half of the country [3]. Although IEW has received the most extended scientific attention in Hungary, the phenomenon is not confined to this geographic region [4]. It is observed all over the world, where soils are characterized by low water permeability/infiltration, and surface runoff

is limited. Literature frequently uses the phrase 'waterlogging' as IEW, and they mostly examine crop responses for waterlogging. Occurrences were reported from, other European countries (France [5], Romania [6,7], Serbia [8,9]), but also in Africa (Egypt [10], Ethiopia [11,12], Nigeria [13], South Africa [14]), Australia [11,15–18], Asia (Bangladesh [19], China [20,21], India [16,22,23], Russia [24,25], Uzbekistan [25]), and South and North America (Argentina, Chile, and the USA [26–29]). Climate change is having significant impact on the hydrologic cycle, affecting water resource systems [30]. In relation with this impact, the frequency of IEW inundations is likely to be concerned.

IEW inundation data can be achieved from two different sources. Traditionally, in field observations (in Hungary systematically collected since 1935; [31]) coordinated by water management authorities are summarized on (paper) maps. The Hungarian regional Water Management Directorates created IEW maps systematically based on in-situ observations collected in the course of continuous field survey [32] and 1:10,000 and 1:25,000 topographic base maps [33]. These maps are varying in geographical extent, scale, and spatial accuracy. Data collections of the IEW inundations were carried out mainly on county level, which lead to differences in the temporal resolution of the datasets. The majority of the maps are hand-drawn, displaying single inundation events. From the large-scale observations, 1:50,000 and 1:100,000 synergy maps were derived, which are sometimes the only actually available data sources.

With the appearance of publicly available remote sensing data, like aerial and space-borne imagery and the development of image processing techniques, the in situ observations were complemented and IEW could be identified and classified in a more efficient and effective way [34]. There are two drawbacks of space- and/or air-borne data acquisitions. (i) They can provide only a snapshot of the inundation areas, thus most extended phases are not necessarily represented. (ii) Furthermore, the differentiation between natural waters and wetlands or IEW inundation is problematic for image interpretation and processing [35]. Last but not least, datasets compiled by the interpretation of aerial or satellite images originate only from the last two decades. Nevertheless, numerous works were dedicated to the identification and mapping of IEW based on remotely sensed information. Csekő [36] used radar data, Csornai et al. [37] combined radar identification with optical sensors, Rakonczai et al. [38] compared Landsat-based classification of IEW with in situ measurements, Licskó [39] and Szatmári et al. [40,41] used aerial data to identify IEW inundations, Mucsi and Henits [42] used sub-pixel based classification on a Landsat time series, Csendes and Mucsi [43] used hyperspectral imaging combined with the potentials of airborne scanning to monitor environmental processes, Van Leeuwen [44] used a new approach using a combination of an artificial neural network (ANN) and a geographic information system (GIS), Van Leeuwen et al. [45] presented the first results of a system that can monitor IEW over a large area with sufficient detail at a high interval and in a timely matter. The methodology is based on freely available satellite imagery, and a map with known water bodies to train the method to identify inundations.

Due to data quality and reliability of both types of data collections, there have been initiatives on the identification of IEW forming factors together with hazard mapping based on the static and dynamic influential factors. Mezősi et al. [46] investigated potential impacts of climate change on the Great Hungarian Plain based on regional climate models. They found a slight decrease of IEW hazard. However, future prediction had high uncertainty, because IEW is an exceedingly complex phenomenon, and they involved only climatic parameters into the hazard analysis. Barta [47] measured hydro-meteorological and pedological factors that influence the formation of IEW. At his study area, he was able to differentiate the two most frequent types of IEW: (1) the upwelling or vertical type, where groundwater table is increasing, and (2) the accumulative or horizontal type, where the water accumulates under gravity in the lowest areas due to limited infiltration and/or runoff, independent from the groundwater table or communicating by capillary system. Based on his results, in case of the accumulative type, the rate and temporal progress of infiltration, its extreme values in relation to soil saturation were also estimated. The measurements and his evaluation were based on monitoring points, and the results were not spatially assessed. Van Leeuwen et al. developed an approach using a combination of an artificial neural network (ANN) and a geographic information system (GIS) in

order to identify IEW, and investigate its forming factors in a study area in southern Hungary [44]. Nađ et al. [8] carried out spatial assessment of IEW risk in a study area in Serbia. The hazard map was derived from analysis of satellite images from a period of four dates, the vulnerability map is based on a land cover reclassification, the risk map was generated by the combination of the hazard and impact maps.

Spatial modelling and mapping of environmental phenomena and variables are a complex task since many of them are results of complex biological, chemical, and physical interactions between the atmosphere, biosphere, lithosphere, pedosphere, hydrosphere, and anthroposphere that may operate on different scales. As Hengl [48] pointed out, for spatial modelling of environmental variables it is better to use hybrid techniques, which combine geostatistics with classical or more advanced statistical techniques. Hatvani et al. [49] examined ice core-derived water stable isotope records in an Antarctic macro region by using multiple linear regression analysis and ordinary kriging. Fehér and Rakonczai [50] used spatio-temporal sequential Gaussian cosimulation for modelling and analyzing shallow groundwater fluctuation and its effect on Hungarian landscapes. Recently, geostatistics applied together with machine learning has gained much more attention in the spatial modelling of environmental variables. For example, Koch et al. [51,52] used random forest regression kriging and random forest combined with residual Gaussian simulation for modelling the shallow groundwater and the depth of redox interface, respectively. Szabó et al. [53] compared the performance of random-forest-based pedotransfer functions and random forest combined with kriging in deriving 3D soil hydraulic properties. Pásztor et al. [54] mapped risk of IEW of a Hungarian county, Bozán et al. [55] mapped relative frequency of IEW inundation on the Great Hungarian Plain. Both mapping processes were carried out by Regression Kriging method, based on the relationship between the occurrence of IEW inundations and its driving factors. The result map originated from the sum of the Multiple Linear Regression model and the interpolated (Ordinary Kriging) residuals.
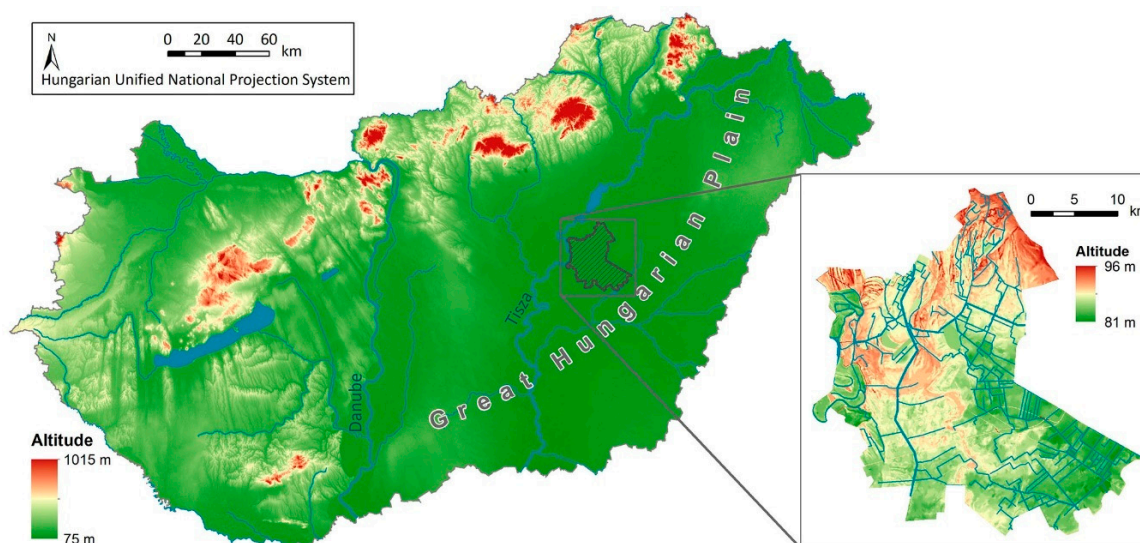
The present paper aimed at spatial modelling of IEW hazard in a Hungarian study area by two hybrid spatial prediction approaches, which combine multivariate statistics and machine learning with geostatistics. We applied Regression Kriging (RK) and Random Forest combined with Ordinary Kriging (RFK) based on locally experienced IEW frequency observations, involving spatially exhaustive auxiliary data representing IEW forming environmental factors. We also investigated the effect of the applied predictors on the results. We ran the predictive models with two combinations of auxiliary variables to test the effect of the introduction of new predictors (linked to at least one of the determining factors).

2. Materials and Methods

2.1. Study Area

The investigated area (788.7 km^2; Figure 1) which is situated in Jász-Nagykun-Szolnok County in the lowland featured Great Hungarian Plain, is entitled "*10.07. Kisújszállás Excess Water Protection Section*" (EWPS). It is supervised by the Middle Tisza District Water Directorate. Motivations for selection of the studied area were as follows:

- majority of the area is hazarded due to excess water inundations;
- poor vertical drainage of its soils due to heavy texture (high amount of expanding clay minerals, low permeability, limited infiltration);
- a significant part of the investigated area is traditionally agricultural land used for productive farming, where arable crop production has dominated since the regulation of the rivers of the Great Hungarian Plain;
- there are meteorological data series available in the necessary length and quality;
- the area represented a pilot for the improvement of integrated management practices for public authorities to mitigate heavy rain risks and excess water hazard in the frame of the RAINMAN project (INTERREG CE968: "RAINMAN").

Figure 1. Overview map with relief of Hungary and the Study Area ('10.07. Kisújszállás Excess Water Protection Section').

The area bordered by the Tisza River has a very diverse geomorphology, which covers the middle part of the Great Hungarian Plain. Climate of the "*10.07. Kisújszállás EWPS*" is moderately warm-dry, the annual precipitation is 500–550 mm. The area is mainly suitable for growing drought-tolerant, long growing, and high heat demanding varieties, where the importance of water retention and irrigation is increasing. In addition to the frequent droughts, the excess waters can cause considerable damage due to the lowland featured area.

A significant part (~70–80%) of the area is covered by Vertisols and Chernozems Reference Soil Groups according to World Reference Base for Soil Resources (WRB, [56]). Solonetz soils also occur in smaller patches. Due to its lowland character, the canal density is well above the national average. These canals generally serve drainage; however, a considerable length of the canals have dual purposes to serve irrigation water. Depth of groundwater varies, generally 0.5–2.5 m with seasonal fluctuation.

Preliminary examination of the *10.07 Kisújszállás EWPS* found that a significant part of this area was endangered by excess waters for some natural reasons. Especially in the eastern border areas, where 26 of the examined 39 years were inundated in a different level by excess water. It was typical for the water coverage to stay in the area for 10 to 15 days. A significant proportion of cultivated crops could be fully destroyed by such inundation. Land use is characterized by a high proportion of agricultural land (91%), and within it a very high proportion of arable land (close to 79%), well above the national average. Grassland (11.08%) and forest (7.77%) are also heavily represented.
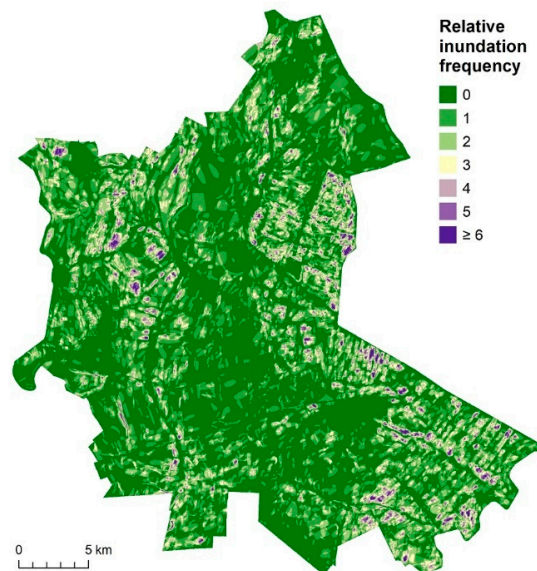
2.2. Reference Data

The responsible Water Management Directorates provided seasonal maps of areas affected by IEW from the period between 1962 and 2014. This legacy information was digitized, vectorized, and then aggregated (Figure 2).

Map of temporally aggregated legacy information on IEW inundation frequency provided reference data as follows. Multiple conditioned random samplings were carried out on the vectorized legacy data. The first condition was to sample each patch of the map by points equal to the square root of the area (in hectares) of the inundation frequency polygons. With the second condition, we made an exception with small, but frequently inundated patches. If the polygon was smaller than 1 ha, but inundation frequency was greater than 5, the polygon got 1 sampling point. The third condition referred to the minimum allowed distance between random points, which was set to 50 m (the pixel size used in the spatial model). With the fourth condition, points lying over settlements and water bodies were omitted. As any map, the used inundation frequency map just models the

reality, thus it cannot be used as absolute reference and consequently its point sampling introduces certain inaccuracies into the applied reference data set. To reduce this effect and in order to have a more balanced sampling, ten random point datasets were constructed. The generated sample sets contain ~13,000 virtual observation sites. The reference data sets were compiled as relative inundation frequency, the values were extracted for each of the 10 random point data sets. In accordance with this, mapping models were run 10 times.

Figure 2. Map of temporally aggregated legacy information on inland excess water (IEW) inundation frequency. The values do not show real frequencies, they can be considered as an indicator (Number of inundation events occurred within the period from where observations are available).

2.3. Environmental Co-Variables

The predictive models were run with two combinations of auxiliary variables (also called environmental co-variables). Basic set (BS) consists of variables formerly used by Bozán et al. (2018). To test the effect of the introduction of new predictors (linked to at least one of the determining factors), an extended set (ES) was also compiled and used.

The effect of **soil** on the occurrence of IEW was modelled and spatially represented by the soil physical property layer and the 'landscape management soil type' of the Digital Kreybig Soil Information System (DKSIS; Pásztor et al., 2012). DKSIS is one of the most important nationwide spatial soil databases of Hungary, it is the reambulated and GIS-developed version of the legacy data of the soil survey lead by Kreybig [57,58]. In the legacy data, soil physical categories were elaborated according to water retention capability, permeability, and infiltration rate. Landscape management soil types were defined from the viewpoint of crop production, aggregated from pH, $CaCO_3$ content, and soil texture [59].

For a more sophisticated characterization of soils, the extended set was completed with hydrophysical soil properties represented by continuous variables. The 3D Soil Hydraulic Database of Europe (EU-SoilHydroGrids ver1.0, [60]) provides information on the most frequently required soil hydraulic properties at 250 m resolution at 7 soil depths up to 2 m with full European coverage. The following layers were used and clipped from EU-SoilHydroGrids as co-variables: saturated water content (pF = 0), water content at field capacity (pF = 2.5), wilting point (pF = 4.2), and saturated hydraulic conductivity. We converted the available information of the 7 available soil depths (0 cm, 5 cm, 15 cm, 30 cm, 60 cm, 100 cm, 200 cm) for 0–30 cm, 30–60 cm, 60–100 cm, 100–200 cm depth intervals.

Climate was represented by four spatial layers provided by the Hungarian Meteorological Service. Average annual precipitation, average annual temperature, average annual evaporation, and

average annual evapotranspiration were compiled using the MISH method elaborated for the spatial interpolation of surface meteorological elements based on 30 year observation with 0.5′ resolution [61]. The layers are available at 100 m resolution.

In addition, humidity index (HUMI) layer was also applied in the prediction. HUMI is used for the characterization of water stress periods. It was calculated by a 10% possibility of occurrence of root square of sum of monthly weighted precipitation and sum of monthly weighted potential evapotranspiration ratio [62]. The Hungarian Meteorological Service and the responsible Water Management Directorates provided precipitation and evapotranspiration data of 68 meteorological observation stations, covering the period of 1961–2014.

The effect of **land use** was characterized and spatially represented by a numeric coefficient based on the National CORINE Land Cover 1:50,000 database (CLC50) [63]. The categories were parameterized with expert-based land use indices characterizing their role in the formation of IEW [1]. According to our method, the lower the values of land use factor (i.e., artificial areas 0.6–1.0; arable lands 0.3–1.0; permanent crops 2.5; pastures 0.6; forest and natural vegetation 1.0–5.0; wetlands 0.1; etc.), the more significant their role in IEW development is.

Topography was taken into account based on the data of the countrywide Hungarian HydroDEM Digital Elevation Model. The database, compiled by the General Directorate of Water Management, is aimed at supporting flood risk mapping and risk management planning processes. The data were available at 50 m per pixel raster resolution. Besides elevation, we applied the following morphometric derivatives in the mapping process: Channel Network Base Level, Channel Network Distance (ES), Closed Depressions, LS-Factor, Mass Balance Index, Multiresolution Index of Ridge Top Flatness (MRRTF), Multiresolution Index of Valley Bottom Flatness (MRVBF), Plan Curvature, Profile Curvature, Relative Slope Position (ES), SAGA Wetness Index, Topographic Position Index, Topographic Wetness Index, Valley Depth (ES), Vertical Distance to Channel Network.

Hydrogeology was spatially represented and was taken into consideration by the depth and the thickness of the uppermost aquitard, for which data was provided by the Geological and Geophysical Institute of Hungary (predecessor of Mining and Geological Survey of Hungary), and the standard depth of groundwater calculated on the average 10 highest values within the last 50 years. The reference groundwater data of well observations were provided by the General Directorate of Water Management. In order to get spatially exhaustive groundwater data, spatial interpolation (co-kriging) had to be carried out, using elevation provided by HydroDEM as a proper spatial co-variable [64].

The movement of groundwater cannot be studied by itself, since the flow systems of the sub-surface waters have a very significant influence on it. In order to clarify the role of sub-surface waters in the formation of excess water inundations, a distinction should be made between recharge and discharge zones. Groundwater flow, i.e., the difference between the amount of water infiltrated into and out of the groundwater for 2×2-km cells was calculated. The map prepared and provided by the Hungarian Mining and Geological Service, shows the recharge and discharge areas in the non-productive state, where the recharge cells can be characterized with positive values and the discharge areas with negative values (mm/year). It was used in the extended predictor variable set.

ES also contained a layer displaying distance from surface water bodies. It was calculated using Euclidean distance metrics in ESRI ArcGIS 10.6 software.

2.4. Preprocessing of Environmental Co-Variables

The auxiliary data set was preprocessed for spatial analysis. Raster layers were transformed to a common grid system, masked to the study area, and resampled to 50-m spatial resolution. Data of Hungarian Meteorological Service and EU-SoilHydroGrids data were converted by cubic convolution method. Since in the case of saturated hydraulic conductivity, cubic convolution produced artifacts, we applied nearest neighbor method. Categorical maps were also converted into the 50-m grid system by maximum area method.

To reduce the number of predictor variables in RK, and to avoid their multicollinearity, principal component analysis (PCA) was carried out on the continuous variables. Before PCA, all of the auxiliary variables were normalized to a 0–255 scale. In further analysis, the first principal components were used, which together explained 99% of the variance. Categorical variables were considered as indicator variables. Every category got a new layer: in case of occurrence, the grid value was set to 255, while out-of-category areas were coded with 0. In Random Forest combined with Ordinary Kriging (RFK) models, categorical co-variables were handled as factors, therefore there was no need to distinguish them in preprocessing.

2.5. The Applied Hybrid Prediction Methods

In this study, two hybrid prediction methods were used for spatial prediction of IEW risk, namely Regression Kriging and Random Forest combined with Ordinary Kriging. Both techniques rest on the same assumption, i.e., the target variable being mapped can be described and modelled in terms of a deterministic component and a stochastic component, which is

$$Z(\mathbf{u}) = m(\mathbf{u}) + \varepsilon(\mathbf{u}),$$

where m(**u**) is the deterministic component describing structural variation, ε(**u**) is the stochastic component consisting of random variation that could be spatially correlated, and **u** is the vector of the geographical coordinates.

From a practical point of view, the main difference between RK and RFK is that how they describe and model the deterministic part of variation. In case of RK, the assumption made on the linear relationship between the target variable and the environmental co-variables may be too rigorous because this relationship could be more complex and this assumption can be valid just for the first approximation (Malone et al., 2018). That was the reason why digital environmental mapping has been interested in machine learning algorithms, which are able to describe and model the deterministic component via predictive models by applying different principles. Among the machine learning algorithms used in digital environmental mapping, random forest plays an accentuated role.

2.5.1. Regression Kriging (RK)

Regression Kriging (RK), also termed as Universal Kriging or Kriging with an External Drift [65], combines regression of the target variable on environmental co-variables with kriging of the regression residuals [65,66]. In this study, a Multiple Linear Regression (MLR) analysis was carried out to describe and model the relationship between the reference data set (with extracted information of IEW inundation frequency, see '2.2. Reference Data' section) and the environmental co-variables listed in '2.3. Environmental Co-Variables' section. In the course of MLR analysis, 0.05 significance level was applied, furthermore, the stepwise method was used for selecting the relevant environmental co-variables. The model obtained by MLR analysis described the deterministic component of IEW risk. The stochastic component was described by geostatistical modelling, namely Ordinary Kriging, using the regression residuals, which represented the variation that could not be explained by the MLR model [67]. The RK prediction can be obtained by summing the prediction of the MLR model and the prediction of Ordinary Kriging.

Finally, ten realizations were generated with both basic and extended variable sets using the ten random point datasets. Mean of the ten realizations provided the final result map of the IEW indundation probability by RK. Standard deviation and median map were also compiled from the ten realizations to test the robustness and indicate the reliability of the aggregated models.

2.5.2. Random Forest Combined with Ordinary Kriging (RFK)

Random Forest combined with Ordinary Kriging (RFK) is a relatively new hybrid method used in digital environmental mapping [53,68,69], which combines predictive model of Random Forest with

kriging of the Random Forest residuals [70]. In RFK, the deterministic component is described by Random Forest (RF) as opposed to RK, where the deterministic component was described by MLR. The RF algorithm generates (depending on its settings and on the type of the dependent variable) a number of regression or classification trees. The model relies on averaging the result of the trees, which are grown independently from each other [71]. In the course of RF modelling, the number of trees was set at 100. As two environmental co-variable packages were applied for the RFK method: for BS mtry was set at 7, for ES mtry was set at 14. The RF part of the RFK models provides a variable rank, reflecting which co-variables play a more important role in the prediction model. The stochastic component was described by geostatistical modelling, namely Ordinary Kriging, using the RF residuals. The RFK prediction can be obtained by summing the prediction of RF and the prediction of Ordinary Kriging.

As in the case of RK, ten realizations were generated with both basic and extended variable sets using the ten random point datasets. Mean of the ten realizations provided the final result map of the IEW indundation probability by RFK. Standard deviation and median map were also compiled from the ten realizations, to test the robustness and indicate the reliability of the aggregated models.

2.6. Validation

In the present research, airborne originated data on inundation frequency were used only in the validation of mapping results. The data originated from relative IEW inundancy layer of Lechner Knowledge Centre Nonprofit Ltd. [72], for the period 1998–2016. The dataset (made available by courtesy) consists of 1451 points (Figure 3), inundation frequency is characterized in percentage categorized by ten (0%–10%, 10%–20%, 20%–30%, etc.).
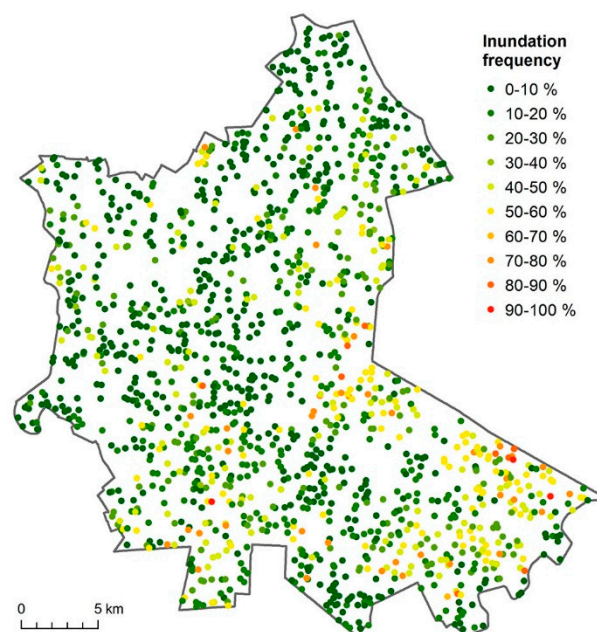

Figure 3. The 1451 locations of the validation dataset.

For validation, the estimated inundation risk value of the four result maps (RK$_{BS}$, RK$_{ES}$, RFK$_{BS}$, RFK$_{ES}$) were extracted at the locations of the 1451 validation points. As comparison, it was revealed if the value of the result map intersected the percentage interval of the validation dataset, or else, how many categories are the difference between the validation and the predicted values. The differences are depicted in a bar chart.

Another kind of, only partly independent, validation was also carried out. The vectorized legacy layer of inundation, which was also the basis of the randomly sampled reference data points, was converted to raster format. Values of the four result maps (RK$_{BS}$, RK$_{ES}$, RFK$_{BS}$, RFK$_{ES}$) were rounded

to integer. Then a pixel by pixel comparison was carried out between the legacy data and the four result layers. The differences are depicted in a bar chart.
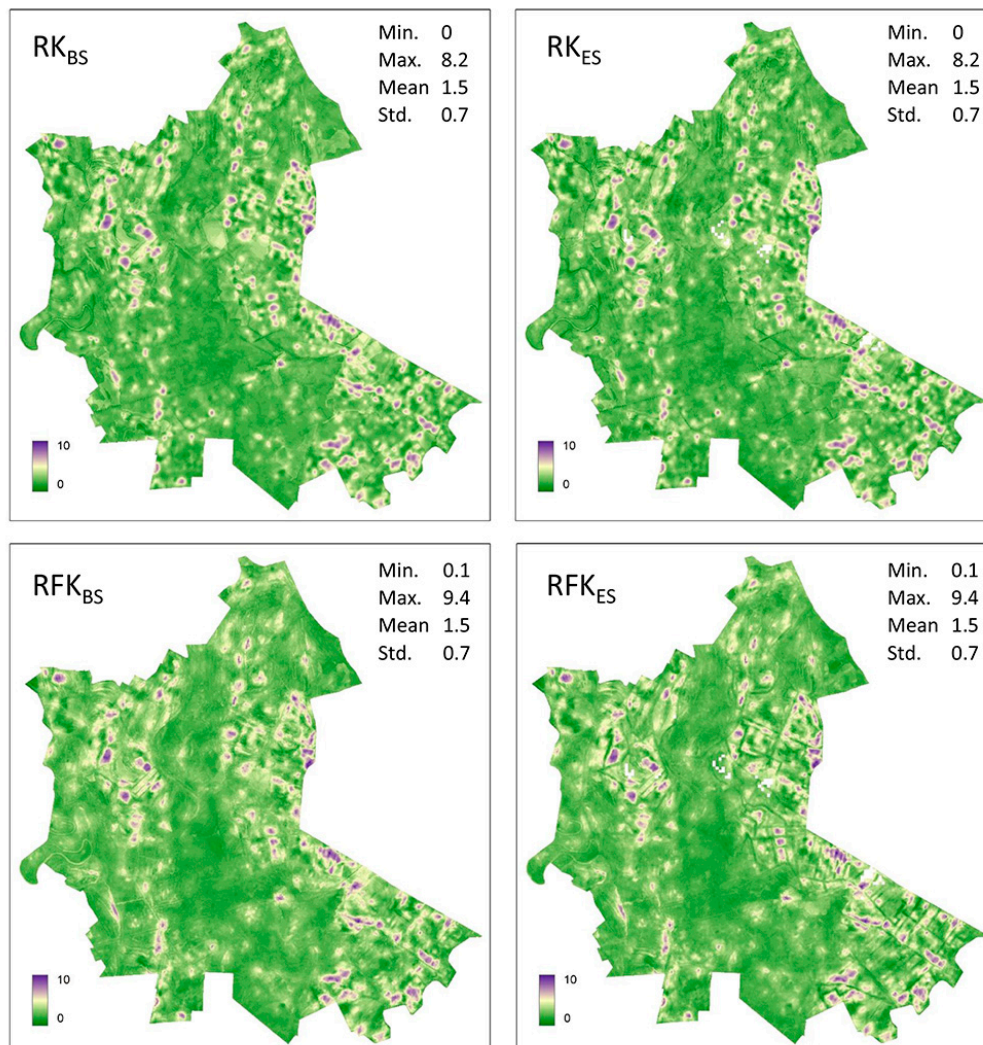
2.7. Software Background

Morphometric derivatives were calculated by SAGA GIS tools (Conrad et al., 2015). RK modelling were run in SAGA GIS environment, while RFK modelling were carried out in R statistical software [73]. Calculation of mean, standard deviation, and median maps, as well as editing the layouts of result maps were compiled in ArcGIS 10.6. Validation queries were run in ArcGIS 10.6, evaluation and depiction of validation results were carried out in MS Excel.

3. Results

3.1. Result Maps

The four final result IEW indundation probability maps (mean of the ten realizations of RK_{BS}, RK_{ES}, RFK_{BS}, RFK_{ES}) are presented in Figure 4. There are a few white pixels in the maps created with ES, which means 'NoData' pixels. They are actually fishponds, and originate from the EU-SoilHydroGrids co-variate layers, since water bodies are masked out in these originally 250-m resolution layers.

Figure 4. IEW indundation probability result maps of the two methods (RK – Regression Kriging, RFK – Regression Forest combined with Ordinary Kriging) by application of two co-variable packages (basic set (BS), extended set (ES)).

As for visible comparison to the inundation map (Figure 4), patterns of the more frequently inundated areas are similarly noticeable in all of the four result maps. Minimum values are ~0, mean values are 1.5, and standard deviations are 0.7 in all of the four result maps; however, maximum values are significantly higher in RFK predictions (9.4) than in RK predictions (8.2). The latter means that RK is capable of narrowing down the range of the reference data in the prediction, while RFK provides more reliable results.

In map RFK$_{ES}$, structure of the 'distance from surface water bodies' layer become remarkably discernible. This appears in the variable importance rank of RF part in the RFK models (Table 1). RF indicated that the 'distance from surface water bodies' co-variable has the most predictive power in the prediction ten times out of ten cases.

Table 1. Occurrences of environmental co-variables in variable importance rank (position 1–5) of Random Forest (RF) in the Random Forest combined with Ordinary Kriging (RFK) models (BS: basic set, ES: extended set of co-variables).

Set	Environmental Co-Variable	RFK BS					RFK ES				
		1.	2.	3.	4.	5.	1.	2.	3.	4.	5.
ES	distance from surface water bodies						10				
	groundwater recharge and discharge areas							8	2		
	saturated water content in 0–30 cm soil depth								2	2	
BS&ES	average annual precipitation	3	2	1	1			2	1	4	
	average annual temperature					1			2		1
	average annual evapotranspiration	1	1		1				1		3
	average annual evaporation		1	1	3	3				1	3
	humidity index (HUMI)					1					
	Channel Network Base Level										1
	Closed Depressions	1	2	1	1	2			1	1	
	Elevation					1					
	SAGA Wetness Index			5	1	1					
	Vertical Distance to Channel Network	5	3	1					1	2	2
	groundwater level		1	1	3	1					

The second most important variables according to the RFK$_{ES}$ models (Table 1) are 'groundwater recharge and discharge areas' (eight times) and 'average annual precipitation' (two times). The third place shows not so consistent image: 'groundwater recharge and discharge areas', 'average annual temperature', and 'saturated water content in 0–30 cm soil depth' occurs in two times, 'Closed Depressions', 'Vertical Distance to Channel Network', 'average annual precipitation', and 'average annual evapotranspiration' in one-one times. The fourth place in the ranking are 'average annual precipitation' (four times), 'saturated water content in 0–30 cm soil depth', and 'Vertical Distance to Channel Network' (two times), 'Closed Depressions', and 'average annual evaporation' (one time).

The ranking in variable importance in RFK$_{BS}$ (Table 1) is formed as follows. 'Vertical Distance to Channel Network' (five times), 'average annual precipitation' (three times), 'Closed Depressions', and 'average annual evapotranspiration' (one time) have the most predictive power. Second place have 'Vertical Distance to Channel Network' (three times), 'Closed Depressions', and 'average annual precipitation' (two times), 'average annual evapotranspiration', 'average annual evaporation', and 'groundwater level' (one time). Third important variable is 'SAGA Wetness Index' (five times), 'Vertical Distance to Channel Network', 'Closed Depressions', 'average annual precipitation', 'average annual evaporation', and 'groundwater level' (one time). Fourth place have 'groundwater level', and 'average annual evaporation' (three times), 'Closed Depressions', 'SAGA Wetness Index', 'average annual evapotranspiration', and 'average annual precipitation' (one time).
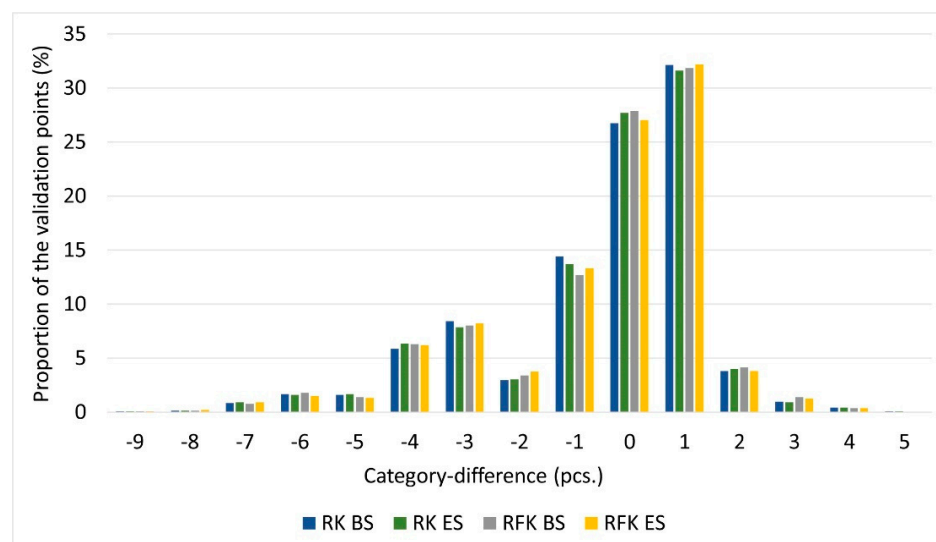
3.2. Validation

Results of validation by independent data are summarized in Table 2 and Figure 5. In all four cases (RK$_{BS}$, RK$_{ES}$, RFK$_{BS}$, RFK$_{ES}$), difference is not more than one category (−1, 0 or 1) in 72%–73% of the samples. Prediction is greater than observation in 37%–38% of the cases, observed values are greater than predicted values in 34%–36% of the points.

Results of the pixel by pixel comparison between the legacy data and the four result layers are summarized in Table 3 and Figure 6. According to the results, predicted values are significantly greater than observed values. In all four cases, 1 category is the difference in more than half of the pixels. However, in 85%–88% of the pixels, the difference is not more than one (−1, 0 or 1).

Table 2. Category-difference between the independent validation dataset, and the categorized predicted values expressed in percentage.
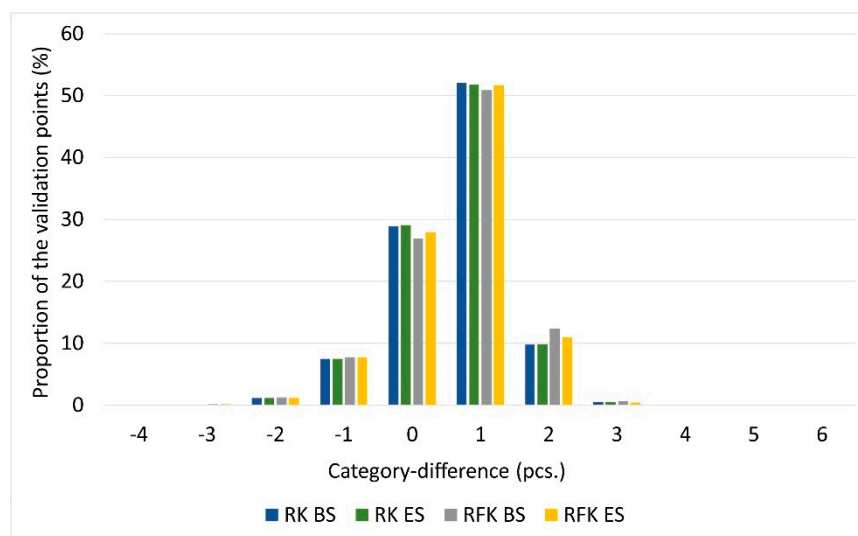
Prediction	−9	−8	−7	−6	−5	−4	−3	−2	−1	0	1	2	3	4	5
RK$_{BS}$	0.07	0.14	0.83	1.65	1.59	5.86	8.41	2.96	14.40	26.74	32.12	3.79	0.96	0.41	0.07
RK$_{ES}$	0.07	0.14	0.90	1.59	1.65	6.34	7.86	3.03	13.71	27.71	31.63	4.00	0.90	0.41	0.07
RFK$_{BS}$	0.07	0.14	0.76	1.79	1.38	6.27	7.99	3.38	12.68	27.84	31.84	4.14	1.38	0.34	
RFK$_{ES}$	0.07	0.21	0.90	1.52	1.31	6.20	8.20	3.72	13.30	27.02	32.18	3.79	1.24	0.34	

Figure 5. Category-difference between the independent validation dataset, and the categorized predicted values.

Table 3. Difference between the legacy map's inundation frequency and the categorized inundation result maps expressed in percentage.

Difference	RK$_{BS}$	RK$_{ES}$	RFK$_{BS}$	RFK$_{ES}$
6			0.0	0.0
5	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0
3	0.5	0.5	0.6	0.4
2	9.8	9.8	12.4	10.9
1	52.1	51.8	50.9	51.7
0	28.9	29.1	26.9	27.9
−1	7.5	7.5	7.7	7.7
−2	1.2	1.1	1.3	1.1
−3	0.1	0.1	0.2	0.2
−4	0.0	0.0	0.0	0.0

Figure 6. Difference between the legacy map's inundation frequency and the categorized inundation result maps.

4. Discussion

Our results showed that there is no significant difference between the accuracy of the two methods used in this study suggesting that both RK and RFK could be appropriate for predicting and mapping IEW hazard. Although both methods performed equally well, we suggest the usage of RFK instead of RK. First of all, numerous studies have demonstrated that random forest commonly outperforms classical statistical techniques (e.g., [69,70,74]). This is because random forest is able to explore, describe, and model complex, non-linear relationships between the response and predictor variables, and in addition, it has been elaborated on a different philosophy aiming at giving the most accurate prediction [71]. Besides, not only does less assumption have to be made on RFK than on RK, but also less preprocessing is needed in RFK [69]. One of the main disadvantages of RK is that it can extrapolate in the feature space, whereas random forest keeps the range of the reference data. Last but not least, RFK can list the rank of the applied environmental co-variables providing explicit information on the importance of them in making prediction.

More co-variables were involved for thematic improvement; however, we did not find that involving them into the modelling would significantly increase the accuracy. On the other hand, according to the importance ranks, the variables added to the ES proved to be the most determining predictors. A possible explanation of similar accuracy of modelling with the two co-variate sets can be the relatively poor spatial resolution of the EU-SoilHydroGrids and layer on recharge and discharge areas, which thematically extended the set of the predictor environmental co-variables.

Our results on prediction importance are in accordance with those of Van Leeuwen et al. [44]. They found that relief had a very important influence in the investigations. We also found, that morphometric derivatives show significant importance in variable importance ranking. The influence of the soil was small in [44], which was assigned to the limited variation of soils on their small study area. In our pilot soil properties, neither seemed essentially important co-variables, as only one soil layer occur in the third and fourth place in the rank. Their results improved by including distance to anthropogenic objects in the training and simulation. Similarly to our findings, distance from surface water bodies proved to be the most important co-variable above all and the distance to channels proved to be the most influential anthropogenic factor.

Our results can be the basis of further investigations, it can support authorities and decision makers in water management projects and issues, improving integrated management practices. Prediction accuracy can be increased by more frequently collected reference data. Van Leeuwen et al. [45] presented a promising method that is capable of continuously identifying IEW over large areas for

operative purposes. However, they concluded that more scientific research is needed to improve the determination of the threshold for the active data processing workflow and to reduce the number of false positives.

We could provide a more specific support, if we could distinguish which type of IEW appeared at the area [47]. Furthermore, it would be useful to involve land use information and agricultural practices into the investigations: not only as co-variable, but as part of an improved risk analysis model. Since inundation of different land use categories implies different economical risks [8]. More precise analyses could be achieved if annual information of land use changes (e.g., crop rotation, agricultural engineering) were involved.

Finally, our results can contribute to climate scenario analyses. Mezősi et al. [46] involved only climatic factors into their investigations, they used neither IEW inundation data, nor other environmental factors. We assume, that involving more data and environmental factors can decrease the uncertainty of future predictions.

According to our former experiences [54,55], the applied hybrid spatial prediction approaches can be suggested to be used not only in relatively small, but in larger study areas as well.

5. Conclusions

To summarize in brief, our aim was spatial modelling of IEW hazard in a Hungarian study area with two hybrid spatial prediction approaches, which combine multivariate statistics and machine learning respectively with geostatistics. We applied Regression Kriging (RK) and Random Forest combined with Ordinary Kriging (RFK) based on locally experienced IEW frequency observations, involving spatially exhaustive auxiliary data representing IEW forming environmental factors. We also investigated the effect of the applied predictors on the results. We ran the predictive models with two combinations of auxiliary variables to test the effect of the introduction of new predictors. According to the results of the two approaches, we did not find significant differences in their accuracy. Although both methods are appropriate for predicting inland excess water hazard, we suggest the usage of RFK, since (i) it is more suitable for revealing non-linear and more complex relations than RK, (ii) it requires less presupposition on and preprocessing of the applied data (iii) keeps the range of the reference data, while RK tends more heavily to smooth the estimations and (iv) it provides a variable rank, providing explicit information on the importance of the used predictors. Involving more co-variables into the mapping process for thematic extension did not prove to be effective according to the accuracy assessment, presumably due to the poor spatial resolution of soil hydrophysical data and the layer on recharge and discharge areas. We attribute this failure to the inference of the expected improvement in thematic extension with the relatively poor spatial representation of the potential key (inundation forming) factors.

Based on our results, we conclude that area-based conditioned random sampling on vectorized legacy data is appropriate as reference data for IEW indundation probability mapping, if modelling is based on multiple datasets. It would be interesting to make further investigations on the accuracy of the results running the models not only with 10, but 20 or even more randomly sampled reference datasets. In the recently occurring nationwide IEW inundation hazard mapping we have increased this number to 20.

Although we did not find significant difference in accuracy provided by the two co-variable packages, we consider the co-variables in the ES package more than useful. We are planning to make further investigations on IEW hazard mapping with the application of more detailed spatial soil hydrophysical data, when it is available for the territory of Hungary.

Significant improvement in prediction accuracy could be expected from more frequently collected reference data. If IEW inundation events were monitored continuously, at unified spatial and temporal resolution, both inundation probability and hazard could be predicted more accurately. The recently developed National Earth Observation Information System [75] and its services could provide and are also expected to make a significant step forward in this field.

References

1. Bozán, C.; Körösparti, J.; Pásztor, L.; Kuti, L.; Kozák, P.; Pálfai, I. *Gis-Based Mapping of Excess Water Inundation Hazard in CsongráD County (Hungary)*; Analele Universiţăţii din Oradea; Universităţii din Oradea: Oradea, Romania, 2009; Volume XIV, pp. 678–684.
2. Pálfai, I. A belvíz definíciói [Definitions of inland excess waters]. *Vízügyi Közlemények* **2001**, *83*, 376–392.
3. Pálfai, I. Az Alföld belvízi veszélyeztetettsége és aszályossága [Map of excess water hazard and drought in the Great Hungarian Plain]. In *A víz szerepe és jelentősége az Alföldön*; Nagyalföld Alapítvány: Nagyalföld, Hungary, 2000; pp. 85–96.
4. Szatmári, J.; Van Leeuwen, B. (Eds.) *Inland Excess Water–Belvíz–Suvišne Unutrašnje Vode*; Szegedi Tudományegyetem–Univerzitet u Novom Sadu: Szeged–Novi Sad, Serbia, 2013; ISBN 978-963-306-263-0.
5. Merot, P.; Ezzahar, B.; Walter, C.; Aurousseau, P. Mapping waterlogging of soils using digital terrain models. *Hydrol. Process.* **1995**, *9*, 27–34. [CrossRef]
6. Romanescu, G.; Stoleriu, C.; Zaharia, C. Territorial Repartition and Ecological Importance of Wetlands in Moldova (Romania). *J. Environ. Sci. Eng.* **2011**, *5*, 1435–1444.
7. Halbac-Cotoara-Zamfir, R.; Gunal, H.; Birkás, M.; Rusu, T.; Brejea, R. Successful and Unsuccessful Stories in Restoring Despoiled and Degraded Lands in Eastern Europe. *Adv. Environ. Biol.* **2015**, *9*, 368–376.
8. Nađ, I.; Marković, V.; Pavlović, M.; Stankov, U.; Vuksanović, G. Assessing inland excess water risk in Kanjiza (Serbia). *Geogr. CGS* **2018**, *123*, 141–158. [CrossRef]
9. Brkic, M.; Dogan, V.; Obradovic, D.; Zivanov, M. Hardware realization of measurement and monitoring system for level of groundwater. In Proceedings of the IX. Symposium Industrial Electronics INDEL, Banja Luka, Bosnia and Herzegovina, 1–3 November 2012; pp. 124–127.
10. Awad, S.R.; El Fakharany, Z.M. Mitigation of waterlogging problem in El-Salhiya area, Egypt. *Water Sci.* **2020**, *34*, 1–12. [CrossRef]
11. Setter, T.L.; Waters, I. Review of prospects for germplasm improvement for waterlogging tolerance in wheat, barley and oats. *Plant Soil* **2003**, *253*, 1–34. [CrossRef]
12. Bulti, M.; Abdulatif, A. Review on Agricultural Problems and Their Management in Ethiopia. *Turkish J. Agric. Food Sci. Technol.* **2019**, *7*, 1189–1202.
13. Bergé-Nguyen, M.; Crétaux, J.-F. Inundations in the Inner Niger Delta: Monitoring and Analysis Using MODIS and Global Precipitation Datasets. *Remote Sens.* **2015**, *7*, 2127–2151. [CrossRef]
14. Ojo, O.I.; Ochieng, G.M.; Otieno, F.O.A. Assessment of water logging and salinity problems in South Africa: An overview of Vaal harts irrigation scheme. In *WIT Transactions on Ecology and the Environment*; WIT Press: Southampton, UK, 2011; pp. 477–484.
15. Manik, S.M.N.; Pengilley, G.; Dean, G.; Field, B.; Shabala, S.; Zhou, M. Soil and crop management practices to minimize the impact of waterlogging on crop productivity. *Front. Plant Sci.* **2019**, *10*, 140.

16. Yaduvanshi, N.P.S.; Setter, T.L.; Sharma, S.K.; Singh, K.N.; Kulshreshtha, N. Influence of waterlogging on yield of wheat (Triticum aestivum), redox potentials, and concentrations of microelements in different soils in India and Australia. *Soil Res.* **2012**, *50*, 489–499. [CrossRef]

17. Bakker, D.M.; Hamilton, G.J.; Houlbrooke, D.J.; Spann, C.; Burgel, A. Van Productivity of crops grown on raised beds on duplex soils prone to waterlogging in Western Australia. *Aust. J. Exp. Agric.* **2007**, *47*, 1368–1376. [CrossRef]

18. Solaiman, Z.; Colmer, T.D.; Loss, S.P.; Thomson, B.D.; Siddique, K.H.M. Growth responses of cool-season grain legumes to transient waterlogging. *Aust. J. Agric. Res.* **2007**, *58*, 406–412. [CrossRef]

19. Hossain, A.; Uddin, S.N. Mechanisms of waterlogging tolerance in wheat: Morphological and metabolic adaptations under hypoxia or anoxia. *Aust. J. Crop Sci.* **2011**, *5*, 1094–1101.

20. Wu, X.; Tang, Y.; Li, C.; McHugh, A.D.; Li, Z.; Wu, C. Individual and combined effects of soil waterlogging and compaction on physiological characteristics of wheat in southwestern China. *Field Crops Res.* **2018**, *215*, 163–172. [CrossRef]

21. Wang, A.J.; Liu, G. Causes for the Formation of Waterlogged Land in the Black Soil Region of Northeastern China. *Adv. Mater. Res.* **2012**, *610–613*, 2925–2930. [CrossRef]

22. Panigrahi, B.; Chandra Paul, J. Managing Drainage Congestion to Increase Crop Production and Productivity in Hirakud Command, India. *J. Agric. Eng. Biotechnol.* **2015**, *3*, 32–40. [CrossRef]

23. Rout, P.K.; Paul, J.C.; Panigrahi, B. Development of land and water management plan based on geoinformation technique for Puincha watershed, Odisha. *J. Soil Water Conserv.* **2017**, *16*, 126–132. [CrossRef]

24. Boiarskii, B.; Hasegawa, H.; Muratov, A.; Sudeykin, V. Application of UAV-derived digital elevation model in agricultural field to determine waterlogged soil areas in Amur region, Russia. *Int. J. Eng. Adv. Technol.* **2019**, *8*, 520–523.

25. FAO and ITPS. *Status of the World's Soil Resources (SWSR)—Main Report*; Food and Agriculture Organization of the United Nations and Intergovernmental Technical Panel on Soils: Rome, Italy, 2015; ISBN 9789251090046.

26. Linkemer, G.; Board, J.E.; Musgrave, M.E. Waterlogging Effects on Growth and Yield Components in Late-Planted Soybean. *Crop Sci.* **1998**, *38*, 1576–1584. [CrossRef]

27. Barickman, T.C.; Simpson, C.R.; Sams, C.E. Waterlogging Causes Early Modification in the Physiological Performance, Carotenoids, Chlorophylls, Proline, and Soluble Sugars of Cucumber Plants. *Plants* **2019**, *8*, 160. [CrossRef]

28. Morales-Olmedo, M.; Ortiz, M.; Sellés, G. Effects of transient soil waterlogging and its importance for rootstock selection. *Chil. J. Agric. Res.* **2015**, *75*, 45–56. [CrossRef]

29. de San Celedonio, R.P.; Abeledo, L.G.; Miralles, D.J. Physiological traits associated with reductions in grain number in wheat and barley under waterlogging. *Plant Soil* **2018**, *429*, 469–481. [CrossRef]

30. Várallyay, G. Láng István, Csete László és Jolánkai Márton (szerk.): A globális klímaváltozás: Hazai hatások és válaszok (A VAHAVA Jelentés) [The global climate change: Effects and answers in Hungary (The VAHAHA report)]. *Agrokémia és Talaj* **2007**, *56*, 199–202. [CrossRef]

31. Somlyódy, L.; Nováky, B.; Simonffy, Z. Éghajlatváltozás, szélsőségek és vízgazdálkodás [Climate change, extremities, water management]. In *"Klíma-21" Füzetek Klímaváltozás—Hatások—Válaszok*; Csete, L., Ed.; MTA KSZI Klímavédelmi Kutatások Koordinációs Iroda: Budapest, Hungary, 2010; pp. 15–32.

32. Pálfai, I. Magyarország belvíz-veszélyeztetettségi térképe [Map of excess water inundation-prone areas of Hungary]. *Vízügyi Közlemények* **2003**, *85*, 510–524.

33. Pálfai, I. A belvizek hidrológiai jellemzése [Hydrology of undrained runoff in Hungary]. *Hidrológiai Közlöny* **1988**, *68*, 320–329.

34. Rakonczai, J.; Csató, S.; Mucsi, L.; Kovács, F.; Szatmári, J. Az 1999. és 2000. évi alföldi belvíz-elöntések kiértékelésének gyakorlati tapasztalatai [Practical experiences with identification of inland excess water in the year 1999 and 2000]. *Vízügyi Közlemények* **2003**, *85*, 317–336.

35. Körösparti, J.; Bozán, C.; Andrási, G.; Túri, N.; Takács, K.; Laborczi, A.; Pásztor, L. Geostatisztikai módszerek alkalmazása a belvíz-veszélyeztetettségi térképezésben [Inland excess water risk mapping by geostatistical methods]. In Proceedings of the MHT XXXIV. Országos Vándorgyűlés, Debrecen, Hungary, 6–8 July 2016; MHT: Debrecen, Hungary, 2016.

36. Csekő, Á. Árvíz- és belvízfelmérés radar felvételekkel [Flood and excess water monitoring with radar images]. *Geodézia és Kartográfia* **2002**, *55*, 16–22.

37. Csornai, G.; Lelkes, M.; Nádor, G.; Wirnhardt, C. Operatív árvíz- és belvíz-monitoring távérzékeléssel [Operative flood and excess water monitoring based on remote sensing]. *Geodézia és Kartográfia* **2000**, *50*, 6–12.

38. Rakonczai, J.; Mucsi, L.; Szatmári, J.; Kovács, F.; Csató, S. A belvizes területek elhatárolásának módszertani lehetőségei [Opportunities for inland excess water mapping]. In Proceedings of the A Magyar Földrajzi Konferencia Tudományos Közleményei, Szeged, Hungary, 25–27 Ocotber 2001; p. 14.

39. Licskó, B. A belvizek légi felmérésének tapasztalatai [Experiences of airborne scanning of excess water inundations]. In Proceedings of the MHT XXVII. Országos Vándorgyűlés; Magyar Hidrológiai Társaság, Baja, Mexico, 1–3 July 2009.

40. Szatmári, J.; Szijj, N.; Mucsi, L.; Tobak, Z.; Van Leeuwen, B.; Lévai, C. A belvízelöntések térképezését és a belvízképződés modellezését megalapozó térbeli adatgyűjtés [Mapping of excess water inundations and modeling of excess water formation with spatial database]. In *Proceedings of the Az Elmélet éS Gyakorlat Találkozása a Térinformatikában II*; Lóki, J., Ed.; Debrecen University Press: Debrecen, Hungary, 2011; pp. 27–35.

41. Szatmári, J.; Tobak, Z.; Van Leeuwen, B.; Dolleschall, J. A belvízelöntések térképezését megalapozó adatgyűjtés és a belvízképződés modellezése neurális hálózattal [Data acquisition for inland excess water mapping and modelling using artificial neural networks]. *Földrajzi Közlemények* **2011**, *135*, 351–363.

42. Mucsi, L.; Henits, L. Creating excess water inundation maps by sub-pixel classification of medium resolution satellite images. *J. Environ. Geogr.* **2010**, *3*, 31–40.

43. Csendes, B.; Mucsi, L. Inland excess water mapping using hyperspectral imagery. *Geogr. Pannonica* **2016**, *20*, 191–196. [CrossRef]

44. Van Leeuwen, B. Identification of inland excess water floodings using an artificial neural network. *Carpathian J. Earth Environ. Sci.* **2012**, *7*, 173–180.

45. Van Leeuwen, B.; Tobak, Z.; Kovács, F.; Sipos, G. Towards a continuous inland excess water flood monitoring system based on remote sensing data. *J. Environ. Geogr.* **2017**, *10*, 9–15. [CrossRef]

46. Mezősi, G.; Bata, T.; Meyer, B.C.; Blanka, V.; Ladányi, Z. Climate Change Impacts on Environmental Hazards on the Great Hungarian Plain, Carpathian Basin. *Int. J. Disaster Risk Sci.* **2014**, *5*, 136–146. [CrossRef]

47. Barta, K. Inland excess water projections based on meteorological and pedological monitoring data on a study area located in the Southern part of the Great Hungarian Plain. *J. Environ. Geogr.* **2013**, *6*, 31–37. [CrossRef]

48. Hengl, T.; Sierdsema, H.; Radović, A.; Dilo, A. Spatial prediction of species' distributions from occurrence-only records: Combining point pattern analysis, ENFA and regression-kriging. *Ecol. Model.* **2009**, *220*, 3499–3511. [CrossRef]

49. Hatvani, I.G.; Leuenberger, M.; Kohán, B.; Kern, Z. Geostatistical analysis and isoscape of ice core derived water stable isotope records in an Antarctic macro region. *Polar Sci.* **2017**, *13*, 23–32. [CrossRef]

50. Fehér, Z.Z.; Rakonczai, J. Analysing the sensitivity of Hungarian landscapes based on climate change induced shallow groundwater fluctuation. *Hungarian Geogr. Bull.* **2019**, *68*, 355–372. [CrossRef]

51. Koch, J.; Stisen, S.; Refsgaard, J.C.; Ernstsen, V.; Jakobsen, P.R.; Højberg, A.L. Modeling depth of the redox interface at high resolution at national scale using random forest and residual gaussian simulation. *Water Resour. Res.* **2019**, *55*, 1451–1469. [CrossRef]

52. Koch, J.; Berger, H.; Henriksen, H.J.; Sonnenborg, T.O. Modelling of the shallow water table at high spatial resolution using random forests. *Hydrol. Earth Syst. Sci.* **2019**, *23*, 4603–4619.

53. Szabó, B.; Szatmári, G.; Takács, K.; Laborczi, A.; Makó, A.; Rajkai, K.; Pásztor, L. Mapping soil hydraulic properties using random-forest-based pedotransfer functions and geostatistics. *Hydrol. Earth Syst. Sci.* **2019**, *23*, 2615–2635.

54. Pásztor, L.; Körösparti, J.; Bozán, C.; Laborczi, A.; Takács, K. Spatial risk assessment of hydrological extremities: Inland excess water hazard, Szabolcs-Szatmár-Bereg County, Hungary. *J. Maps* **2015**, *11*, 636–644.

55. Bozán, C.; Takács, K.; Körösparti, J.; Laborczi, A.; Túri, N.; Pásztor, L. Integrated spatial assessment of inland excess water hazard on the Great Hungarian Plain. *Land Degrad. Dev.* **2018**, *29*, 4373–4386.

56. IUSS Working Group WRB. *World Reference Base for Soil Resources 2014, Update 2015. International Soil Classification System for Naming Soils and Creating Legends for Soil Maps*; FAO: Rome, Italy, 2015; ISBN 978-92-5-108369-7.

57. Pásztor, L.; Szabó, J.; Bakacsi, Z.; Matus, J.; Laborczi, A. Compilation of 1:50,000 scale digital soil maps for Hungary based on the digital Kreybig soil information system. *J. Maps* **2012**, *8*, 215–219.

58. Kreybig, L. Magyar Királyi Földtani Intézet talajfelvételi, vizsgálati és térképezési módszere (The survey, analytical and mapping method of the Hungarian Royal Institute of Geology). *Magy. Királyi Földtani Intézet Évkönyve* **1937**, *31*, 147–244.

59. Pásztor, L.; Laborczi, A.; Takács, K.; Szatmári, G.; Bakacsi, Z.; Szabó, J. Variations for the Implementation of SCORPAN's "S". In *Digital Soil Mapping Across Paradigms, Scales and Boundaries*; Zhang, G.-L., Brus, D.J., Liu, F., Song, X.-D., Lagacherie, P., Eds.; Springer Science+Business Media: Singapore, 2016; pp. 331–342. ISBN 978-981-10-0414-8.

60. Tóth, B.; Weynants, M.; Pásztor, L.; Hengl, T. 3D soil hydraulic database of Europe at 250 m resolution. *Hydrol. Process.* **2017**, *31*, 2662–2666.

61. Szentimrey, T.; Bihari, Z. Mathematical background of the spatial interpolation methods and the software MISH (Meteorological Interpolation based on Surface Homogenized Data Basis). In Proceedings of the Conference on Spatial Interpolation in Climatology and Meteorology, Budapest, Hungary, 3–7 April 2007; pp. 17–27.

62. Bozán, C.; Körösparti, J.; Pásztor, L.; Pálfai, I. Excess water hazard mapping on the South Great Hungarian Plain. In Proceedings of the 13th International Conference on Environmental Science and Technology (CEST), Athens, Greece, 5–7 September 2013; pp. 5–7.

63. Büttner, G.; Maucha, G.; Bíró, M.; Kosztra, B.; Pataki, R.; Petrik, O. National land cover database at scale 1:50000 in Hungary. *EARSeL eProceedings* **2004**, *3*, 323–330.

64. Chung, J.w.; Rogers, J.D. Interpolations of Groundwater Table Elevation in Dissected Uplands. *Ground Water* **2012**, *50*, 598–607.

65. Hengl, T.; Heuvelink, G.B.M.; Stein, A. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma* **2004**, *120*, 75–93. [CrossRef]

66. Hengl, T. *A Practical Guide to Geostatistical Mapping of Environmental Variables*; Institute for Environment and Sustainability: Ispra, Italy, 2007; ISBN 9789279069048.

67. Odeh, I.O.A.; McBratney, A.B.; Chittleborough, D.J. Further results on prediction of soil properties from terrain attributes: Heterotopic cokriging and regression-kriging. *Geoderma* **1995**, *67*, 215–226. [CrossRef]

68. Keskin, H.; Grunwald, S. Regression kriging as a workhorse in the digital soil mapper's toolbox. *Geoderma* **2018**, *326*, 22–41. [CrossRef]

69. Szatmári, G.; Pásztor, L. Comparison of various uncertainty modelling approaches based on geostatistics and machine learning algorithms. *Geoderma* **2019**, *337*, 1329–1340. [CrossRef]

70. Hengl, T.; Heuvelink, G.B.M.; Kempen, B.; Leenaars, J.G.B.; Walsh, M.G.; Shepherd, K.D.; Sila, A.; MacMillan, R.A.; De Jesus, J.M.; Tamene, L.; et al. Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. *PLoS ONE* **2015**, *10*, e0125814.

71. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

72. Lechner, K.C.N.L. Relative Inland Excess Water Inundancy Layer of Lechner Knowledge Centre Nonprofit Ltd. Available online: http://map.fomi.hu/copernicus/ (accessed on 21 May 2019).

73. *R Core Team R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing; R Core Team: Firminy, France, 2019.

74. Veronesi, F.; Schillaci, C. Comparison between geostatistical and machine learning models as predictors of topsoil organic carbon with a focus on local uncertainty estimation. *Ecol. Indic.* **2019**, *101*, 1032–1044. [CrossRef]

75. KIFÜ; NISZ; Lechner, K.C.N.L. Földmegfigyelési Információs Rendszer [National Earth Observation Information System]. Available online: https://kifu.gov.hu/kofop_fir (accessed on 21 May 2019).