PHD THESIS BOOKLET

Vinicius Deganutti de Barros

Gödöllő Campus

2025

HUNGARIAN UNIVERSITY OF
AGRICULTURE AND LIFE SCIENCES

Hungarian University of Agriculture and Life Sciences

Gödöllő Campus

Spatial AquaCrop. A new tool for utilizing AquaCrop in a raster based environment

Vinicius Deganutti de Barros

Gödöllő

2025

Name of the Doctoral School: Environmental Sciences Doctoral School

Discipline: Environmental Science

Head: Csákiné Dr. Michéli Erika

full professor, DSc, corresponding member of the Hungarian Academy of Sciences, head of insitute

MATE, Institute of Environmental Sciences

Supervisor(s): Waltner István, PhD, habil.

associate professor, head of department

MATE, Insitute of Environmental Sciences

Department of Water Management and Climate Adaptation

....

Approval of the Head of Doctoral School Approval of the Supervisor(s)

1. BACKGROUND OF THE WORK AND ITS AIMS

Water has been a key resource for many areas in human society, specially for agriculture, in which water is a limiting resource and at the same time agriculture is one of the biggest utilizers of water. For that Water management is a key point for agriculture and there are many ways for that, one in specifically is crop models, which can give different information on how to properly manage crops, specially for needed irrigation. One of these models is AquaCrop, which has been created by FAO with an intent to simplify crop modeling. SpatialAquaCrop is a package that has been developed in this PhD to utilize the AquaCrop model in a raster environment and not just as a point based simulation. Based on this necessity of improving water management and reducing its waste, this research focused on utilizing the package SpatialAquaCrop to simulate the AquaCrop model in a spatial way, demonstrating the capability of the package and analyzing 5 different case studies in Hungary, with the focus on water management and soil moisture.

The aims for this research are:

1. Development of a methodology for the raster-based spatial application of the AquaCrop model.
2. Application of the developed methodology in an R-based (open-source) environment.
3. Testing and evaluating the resulting R package through 5 study cases.
4. Using the developed package, evaluating the available spatial soil datasets for Hungary, with focus on their potential use for the spatial extension of the Hungarian Drought Monitoring System.

2. MATERIALS AND METHODS

2.1. SpatialAquaCrop

SpatialAquaCrop (SpatialAquaCrop, 2022) is the name that was given for the used R based methodology and package that was the main tool for the research of the different study cases in this doctoral thesis. This methodology has been designed to be a user-friendly method in R to read spatial datasets and utilize the AquaCrop plug-in to run the AquaCrop model and then output the results as raster files. In its current version of its script, it can read TIF files, and output the results as a TIF file. The output can be any of the different outputs which the AquaCrop model can give.

The primary approach of the package is to run the external AquaCrop plugin software in a way that it will run the AquaCrop model in the specified parameters for each of the points/pixels of the study area, which is represented by a raster files. As mentioned, the package runs the AquaCrop model for each of the different pixel cells and saves the output of each one to be put together in the end again in a raster format or as a table. The AquaCrop plugin utilizes text files to storage all the necessary input data and output data as well, so one of the main functions of the package is to be able to properly input and extract all the information to/from the necessary text files. One important point to take into consideration is that the AquaCrop model is not built in to consider affects from lateral movement from different pixels, which might influence spatial patterns that might have been there, due to mainly topography effects. Because of that even though the package is built in to create spatial datasets, in the future topography effects and lateral movements will have to take into account to more precisely assert the modeled results.

All the parameters that will be used as inputs for the modeling run have to be formatted in a specific formatting in which AquaCrop plugin can read them. SpatialAquaCrop makes sure that all the necessary text files are in the proper formatting, or if the user would want, it is possible to create the text files utilizing the AquaCrop software and SpatialAquaCrop can read the information in those files as well, having the ability to choose where to prepare the input files is a good option, as users who have more experience with the AquaCrop model might prefer to prepare all files with the AquaCrop software.

As of this version of the package there are three scripts which represent the three main functions that aim to gather all the necessary information for AquaCrop plug-in to run and produce the different outputs, which will later be presented in a raster format. Besides being a simpler approach for running the AquaCrop model for spatial data, the package aims to be user friendly as well, following on one of the aims of the AquaCrop model (Raes et al, 2018), which is to simplify the process of crop modeling. This package has been developed in an R environment. It is beneficial if the user has a bit of programing knowledge in R but not necessary, as the package tries to guide any user in an easy/understandable way. An overview of what each of the three functions represent can be seen in Table 1.

The Initial function is called 'Initial_AQC ()', it aims to give the user options on how the run will be conducted, which in this case means, which parameters will be used from the ones that AquaCrop can accept. This is done by utilizing the package "svDialogs" (Grosjean, 2022), which gives the user different questions in which depending on the answer the function will know how to proceed and which parameters to take into consideration for the modeling run. In this part the function gives the option as well that if the user would like to use files created from the AquaCrop software, now would be the time to create them and put them into the indicated folder, which it is mentioned during this initial function. At the end of this initial function an external table is created in which the user will have to manually fill in the different information presented there, this will guide how and where the next function will obtain the data necessary for the run. Besides this table, several others might be created as well if the user would like to consider groundwater, field management, irrigation or initial conditions.

After the table has been filled the next function can be initiated, which is called "Control_AQC ()". This second function's main objective is to create some of the several unique text files in which the main function, the third one, will be used to run the model. In the beginning this function checks of the model will consider groundwater, field management, irrigation or initial conditions and let the user know that with messages as well, so that the user has a better grasp if the model is considering the right selected parameters. Next a blank crop file will be created in case the user would like to manually fill it, or if it is set that an already filled crop file will be used, a message will be shown mentioning it. The final of these initial files is the CO_2 concentration file, which the standard file that AquaCrop provides has been used.

The next important step of this function is the creation of the project file, which acts as main information holder for the AquaCrop plugin to know the paths of the unique text files, so it can read all the different data inputs necessary for the model calculation. Besides this this file has some standard values that the model will consider, like: Beginning and end date of the simulation run, starting depth of root zone expansion curve, thickness of top soil in which soil water depletion has to be determined and others.

The final part of the "Control_AQC ()" function is to extract all the input data that are in raster format and export it to different CSV files that will be read in the next function for the modeling calculation.

The goal is to run the model simulation for each of the pixels in order, store the different outputs and in the end put them back into a raster format and a table format as a CSV file for easier access of the result values. The way all the information from the raster files can be extracted and exported into a proper and accurate format is to vectorize the different raster, making larger data sets lighter and easier to read. For future versions of the package an option to just store internally all the values will have to be implemented to circumvent this issue. In the end the CSV files that will be created are soil parameters, precipitation, base evapotranspiration, minimum temperature and maximum temperature.

The last function, and the main one, is called "Spatial_AQC ()", its aims to read all the CSV files which were created from the spatial input data in the last function, run the AquaCrop model calculation, with the AquaCrop plugin, gather all the created outputs for each of the different pixels of the analyzed area and compile then in different CSV tables and raster files. Some of the outputs that the model can give are crop yield, evapotranspiration, runoff, drainage, biomass, irrigation needed and others. There are as well two different types of outputs that the model can provide, the seasonal ones that consider a value for the whole length of the simulation and daily values.

The way in which AquaCrop plugin runs the model calculation is that it reads all the information that are on the specific text files for the different parameters, the path for the files is set in the project file, and with that information it can create the seasonal and daily output file for that specific set of input parameters. The way in which SpatialAquaCrop approaches this is that it considers each of these runs as one of the pixels present in the raster, and it calculates everything necessary for that pixel and saves the output in a vector. Utilizing a for loop, this calculation is done for each of the different pixels of the raster and all their output data is saved into different vectors and in the end all those vectors are transformed back into the same raster format of the input data.

The seasonal outputs and daily outputs that are currently set in the final script are focused on water management/transportation and some crop related outputs. There are other outputs in which AquaCrop plugin can output, in which in the final version of the package, in which will be submitted to R to be implemented to its CRAM, will have the other outputs added as well. Besides the outputs in which the model can give, one last function was added to the last script, which is to calculate the green water footprint for the simulation, this was added due to the importance of this concept for different types of research that focus on the relationship between crop and water.

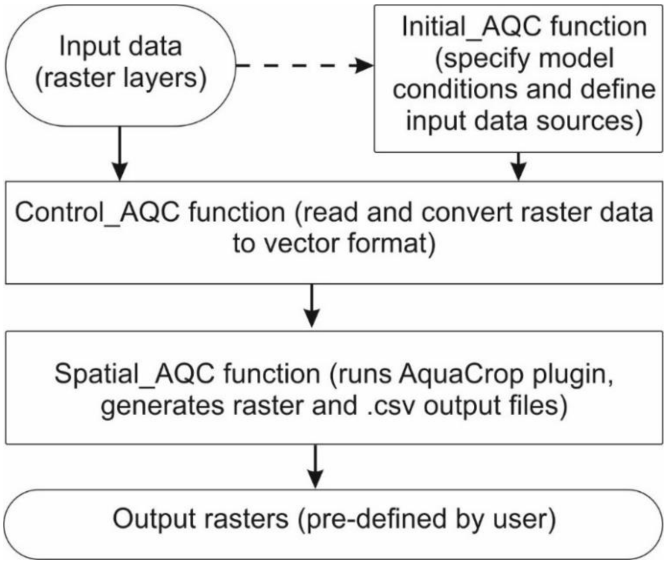

Fig 1 - Basic overview of the SpatialAquaCrop simulation

2.2. First Case study

During the length of this PhD research different approaches with different aims were done while utilizing AquaCrop model and more specific the SpaticalAquaCrop as a methodology for simulating different crops in different scenarios (with a focus on water management, more specifically soil moisture). This thesis will be divided into five different study cases, with explaining their objectives and methodology first, with later their results and discussing those results afterwards. The methodology applied for each study case has a striking similarity to one another, which is to use the SpatialAquaCrop package as its main tool for simulation and analysis.

The first case study aim was to verify how well the SpatialAquaCrop package performed while analyzing the output for two different crops, maize and sunflower, in a small catchment area in Hungary. There was no focus on what was analyzed for this simulation, but still some important results were seen in this study case, which helped to cement the analyses and study cases that came after this one.

For this case study the study area was the Rákos stream watershed, in Pest County, Hungary. Rákos stream is the main river in this study area, it mainly passes through agricultural areas and some urbanized areas as it flows in the direction of the Danube.

As mentioned before this study focused on running the SpatialAquaCrop plugin for two different crops for the whole of the study area, these two crops were maize and sunflower as they are crops which are commonly used in the region. For the input parameters that were chosen, the soil parameters were obtained from the CORINE dataset with 6 data for 6 different depths of soil (0, 5,10, 15, 30 and 60 cm depth). The crop parameters were taken from the database that the AquaCrop software provides; for this analysis the climate data used was considered from just a particular point, the climate station at the research field at MATE university in Godollo, this way the variation in the results will mainly come from the soil parameters; no irrigation was considered, so the simulation ran only considering rainfall, no specific groundwater input was used so in this case the simulation considers that there is no shallow groundwater present in the system; no initial conditions were set; no field management practice was set and the simulation ran for the entirety of the 2020 year. These settings are the most basic in which SpatialAquaCrop can run its simulation, emphasizing again that the goal for this case study is to be just a demonstration of the capabilities of the SpatialAquaCrop package.

2.2. Second Case study

For the second case study, the same area of the Rakos catchment was selected for the study, but for this one both soil and climate parameters had a spatial variation. The main goal of this study was to access first how the AquaCrop model can perform if not considering that the simulation area is a crop field but a grassland and second to access the soil moisture dynamics in this grassland scenario and then compare the results with point-based data that was gotten from the climate station at the research field for the MATE university on Godollo.

The first main point of discussion for this study case is the change of a crop field, in which the AquaCrop model was initially built for, for a grassland field. For this a change in the type of crop and its parameters are needed, for that some research was done, and it was found that there are some studies that have already accomplished this type of “crop” in AquaCrop.

As discussed before, the parameters that are set in the crop file, which the AquaCrop plugin utilizes, dictate what are the crop characteristics and how it will be affected by the climate and soil parameters, so a custom crop file must be created and for that the studies of Allen (Allen et al., 1998) were followed, and a custom alfalfa crop was created. To simulate a grassland alfalfa was selected and its parameters were changed so that it could resemble the best to a grass in a grassland, for example the canopy cover growth of the crop was changed to resemble the closest to one of grass. A set growing cycle of 275 days (starting in in March) was set as well.

The input data for this simulation was obtained from different sources. Soil data was accessed from two sources, field capacity (FC), saturation (SAT) and permanent wilting point (PWP) data were downloaded from EU-SoilHydroGrids ver1.0 (with 250 x 250 m spatial resolution), while soil texture data was derived from the DO-SoReMI.hu initiative (with 100 x 100 m spatial resolution). Even though these two sources have different resolutions, the one with 250 x 250 m resolution was resampled to match the 100 x 100 m spatial resolution. Different depths were considered as well, for this study the different soil depths that were taken in consideration for the simulation were: 0, 5, 15, 30 and 60 cm. Figure 2 presents the spatial variability of soil texture of the top 30 cm layer within the study area.

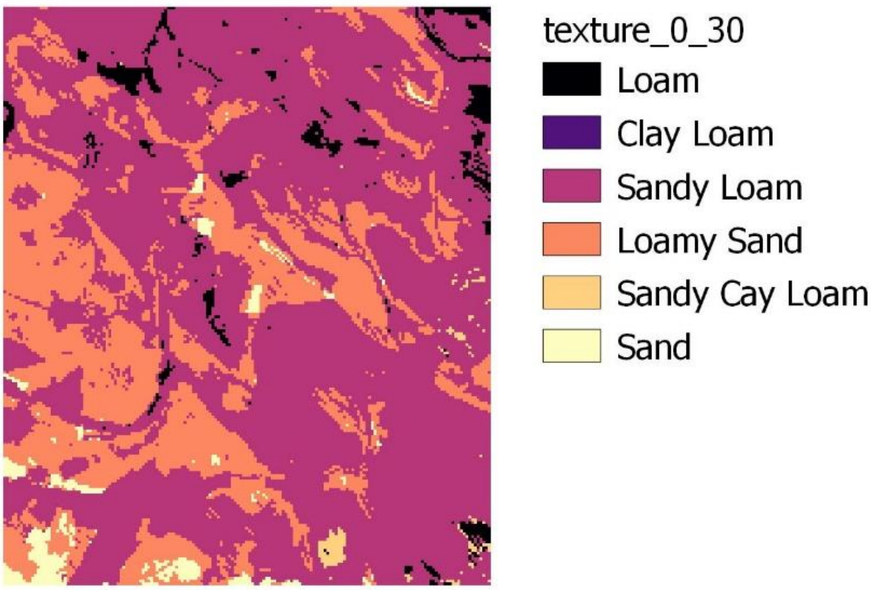

Figure 2 - Texture map up to 30 cm soil depth.

The climate data (daily precipitation, maximum and minimum temperature) used in the simulation was accessed from the Meteorological Data Repository of the

Hungarian Meteorological Service (OMSZ). Daily potential evapotranspiration has been calculated using the Pennman-Monteith equation (Allen et al., 1998). One software which facilitates the calculation of potential evapotranspiration is the ETO Calculator which is available to be downloaded at the FAO website. One advantage of utilizing this software is that the end results already output the result in the proper text file format which is used by the AquaCrop Plugin and the SpatialAquaCrop package. Climatic data was available at a 0.1˚ x 0.1˚ spatial resolution and was interpolated and resampled to the target 100 x 100 m grid.

2.3. Third Case study

After analyzing the results for the second case study it was seen that the soil moisture results for AquaCrop seemed to have a higher variation than what was expected, even though it is know that results from simulated models suffer from a higher variation when compared to field data. There are different options in which could be chosen to approach this issue, for this small study case a comparasion between the results of soil moisture for Aqucrop and Hydrus model has been done and for the final study case the quality of the input parameters was chosen as a focus to try to improve the quality of the soil moisture results.

For this small study the main objective was to compare the soil moisture values obtained from the AquaCrop and Hydrus model and analyze how they behave in comparison to each other. Both simulations ran for the year of 2020 and 2021 and the selected location for the analysis was the meteorological station at MATE university in Godollo. This simulation was done while taking in consideration the same input data for both models, soil parameters were based on the 2018 CORINE land cover dataset and the climate parameters were taken from different sources for each of the years, for 2020 it was from OMS and for 2021 from the data captured by the meteorological station located at the research field. The crop chosen was wheat, as it is a common crop grown in the region. The simulation ran for the entirety of both years.

2.4. Forth Case study

Further down on the research timeline for this PhD thesis, while having a better understanding of the AquaCrop model and reformulated the SpatialAquaCrop package into what would be the final stable version that would be used in this PhD research a new study case was proposed to keep analysing how well results from AquaCrop model can be used in different scenarios and keep presenting the efficacy of the ability of SpatialAquaCrop to process spatial datasets.

As one of the main goals of this thesis is to show the application of AquaCrop model in a raster format applied within the R environment (SpatialAquaCrop package), for this study case first a point-based validation was carried out for soil moisture for a maize field in Martonvasar, Hungary, for 2020. It was possible to gain information on the performance of the model under initial settings having no site-specific parametrization, as the area in Martonvasar can be considered a data scarce area. Besides this validation a comparison between modelled biomass and green canopy crop cover against NDVI was done for winter wheat at an experimental site in Gödöllő for the year of 2020 and 2021, this comparison was done for the growth period until around its senescence. As field scale yield information is considered to be sensitive data it is difficult to obtain from farmers, hence NDVI was used as a proxy for biomass in the validation years. Following the validation efforts the developed SpatialAquaCrop package was used to simulate wheat growth for the year of 2020 and 2021 in the Rákos watershed region for the comparison with NDVI.

Point-based evaluation of the AquaCrop model was carried out in two different sites, on a maize field in Martonvásár for surface soil moisture, and in the experimental field in Gödöllő (which is located inside the Rákos stream catchment) an NDVI comparison between modeled biomass and green canopy crop cover (CC) for winter wheat was done. Specific soil, climate and crop data were taken in consideration for each of the sites for a better parametrization of the model.

For biomass and CC comparison, soil parameters locally analyzed and meteorological data from the local meteorological station (situated at the experimental field for the MATE university, Gödöllő, Hungary) has been utilized (precipitation, temperature) and from that reference evapotranspiration was derived utilizing the Pennman-Monteith equation.

Crop parameters for winter wheat and maize (for the Martonvásár site) were mostly kept the same as the standard ones provided in the AquaCrop software (for the modeling), just the length of the days between sowing to emergence, maximum rooting depth, senescence, flowering and maturity have been changed, in accordance to Szász and winter wheat at the site (Gödöllő) was sown on 1st of December and harvested on 23rd of July .

After NDVI, biomass and CC results were calculated, they were plotted against one another to check for correlation, for that coefficient of determination (R^2) and correlation coefficient were calculated. Statistical significance for the correlation coefficient was checked afterwards as well with the Shapiro-Welch t-test. This comparison was not made for the whole length of the simulation (sowing until harvesting), just until the crop's senescence, which for winter wheat in Hungary is around the beginning of June. This length was chosen because chlorophyl concentration diminishes during senescence lowering the NDVI value while biomass and CC still have a growing trend.

As for the validation at the Martonvasar site, the necessary soil and meteorological data for running the AquaCrop model has been provided by the work of Sándor. The maize field trial was established at Martonvásár, under ploughing and minimal tillage managements in 2020 aiming at the effect of cover crops sawn for the winter period. The plot size is 35 m x 17.5 m for each treatment. The treatments are set up in two replicates. The used maize (Zea mais L.) on the field of the trial was sowed under conventional ploughing without cover crop (i.e. the control treatment of the trial) as it represents the most typical management in the region. The chernozem soil of the experiment is non acidic loam with deep A horizon with 1.96-2.26 m% humus content. Maize was sown on 16th of April and harvested in 21st of October. Soil parameters for the modeling were obtained using field data on soil physical properties and water retention.

Even though these validations and analysis were done in a point-based approach, SpatialAquaCrop package was still used for it, as at this point with the addition of the possibility of extracting daily outputs from any dates and as well extracting the whole timeseries results for selected parameters, the package presents as a good alternative for the AquaCrop software for simulating point based data, while not having the proper visual features that the AquaCrop software has, as it fully built in an R environment, it can be quite beneficial for the researcher utilizing this package, as it can expedite the process of analysing the different outputs that the model creates.

2.5. Fifth case study

For the last study case of this PhD, different points in the South part of Hungary were chosen (29 different meteorological stations in the middle of agricultural fields) in a point-based approach to compare the surface soil moisture near different meteorological stations with the soil moisture results while utilizing the AquaCrop model and the SpatialAquaCrop package to process the points. The meteorological stations are located in different agricultural fields, but the stations themselves are usually separated from the crops and most of the time surrounded by grass. The image bellow show where the stations are located in Hungary (Figure 3). OVF provided data for several meteorological stations throughout all of Hungary, but for focused research a cluster station at the south of Hungary were chosen for this study.

The stations for this case study are mainly located in the Bács-Kiskun County and the Danube-Tisza Interfleuve, which is one of the most drought-prone regions in Hungary. Some characteristics main characteristics of the Danube-Tisza Interfleuve is that there has been a decrease in the groundwater level for the past years, due to different effects such as change in climate patterns (droughts) and channelization of rivers, and a change and degradation of the natural vegetation in the area.

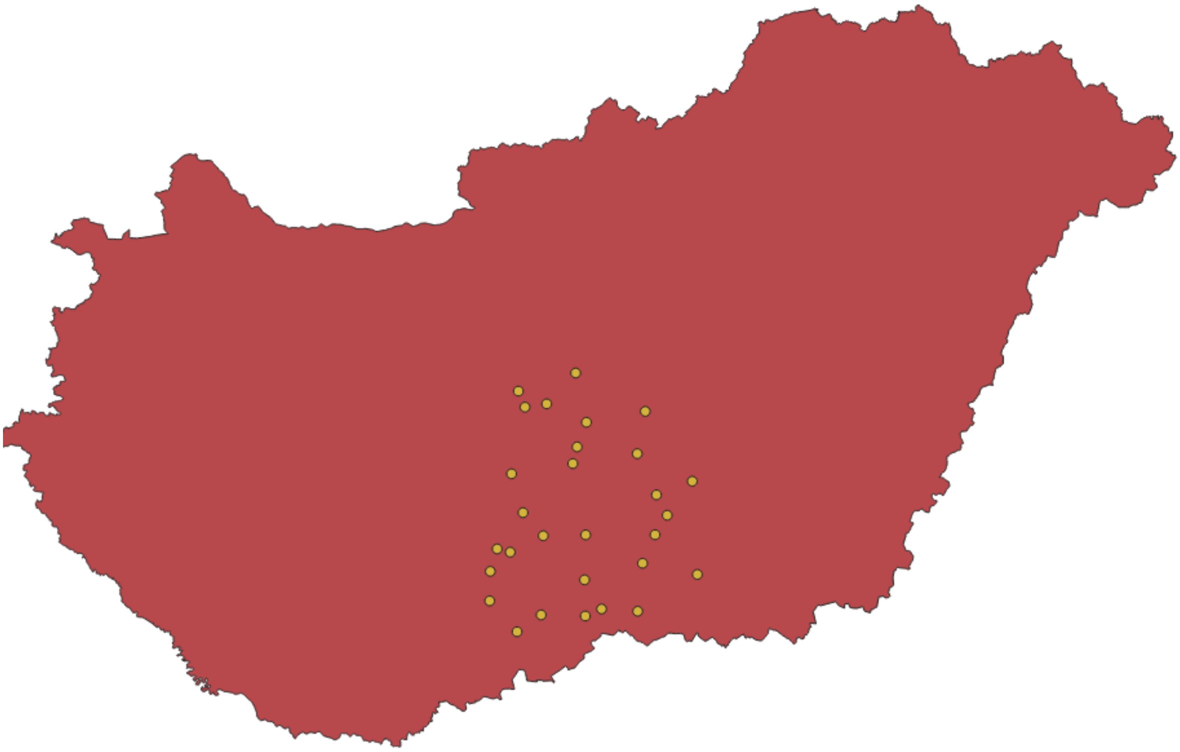

Figure 3 - Location of the meteorological stations within Hungary

The input data for the simulation run were obtained from different sources, the meteorological data was obtained from the different meteorological stations provided by OVF, this data included daily precipitation, maximum and minimum temperature from 2020 to 2022. Evapotranspiration was later calculated utilizing the ETo calculator, which can be downloaded from the FAO website and as prior mentioned the result from the software is already in the format in which both AquaCrop software and SpatialAquaCrop package can utilize.

Soil parameters were taken from different raster sources, there were three sets of crop parameters that were used in this study case. The first set of parameters (Field capacity, permanent wilting point and soil saturation) were taken from EU-SoilHydroGrids ver1.0 (with 250 x 250 m spatial resolution) and the texture was derived from Corine dataset and this texture was compared to the standard Ksat values that AquaCrop provides (figure 42) and the Ksat values taken from it. The other set of values were obtained from which provided all soil parameters needed for the simulation. The third set of parameters is a mix from the Ksat from

AquaCrop and other soil parameters from HUN-REN, as HUN-REN dataset does have a different texture classification, following USDA texture reference, so this texture was as well compared to the standard Ksat values from Aquarop and the related values were used for each of the analyzed points. For the initial runs 100% penetrability and 0% of gravel was considered (the same way for the prior simulations for the other study cases). The exact soil parameters for each of the stations were not available at the time of the study and processing of the data, as it was mentioned prior as the different soil data grid sets were generated with mathematical models, but at the moment OVF has released at https://vizhiany.vizugy.hu/ a new dataset with a in-depth soil description of the different areas where the station are located, providing a different, and in theory, a more accurate dataset that could be used in future researches.

As the meteorological stations are located in closed areas with a high possibility of grass covering the soil, it was chosen again to run the simulation while utilizing a crop file that tries to simulate grass. The base values for the simulation were taken from the work of Terán-Chaves for Ryegrass and of Raes and Kim for alfalfa, taking in consideration these works a new set of parameters was created for this study case. Even though the crop parameters which have been utilized in these studies can simulate the growth of two different types of grass quite well, for this study some further modifications were needed for the crop parameters in the crop file, so that the main goal of comparing the soil moisture in the soil close to the meteorological stations with the simulated on in AquaCrop could be achieved.

As AquaCrop utilizes a base evapotranspiration for its calculations, which is the evapotranspiration from a grassy field, this way it can be assumed that the soil moisture that has been obtained from a grassy field can be considered as base soil moisture for the region, or in other words a soil moisture values that has not been influenced by the different crops or vegetation in the region. Being able to simulate such value is quite beneficial as it can help to understand better the water movement in the region and help with flood prevention.

No optional parameters were chosen at first for the simulation, but after the first results were checked, some of these parameters were added to try to achieve the best simulation results.

The first goal for this analysis was to compare how well AquaCrop would model surface soil moisture in these specific points, while validating the model results with surface soil moisture results that were surveyed in these points, while utilizing raster based data as an input. The simulation ran from 2020 to 2022 and considered that the “New crop” started its growth cycle in March of 2020.

3.RESULTS AND DISCUSSION

3.1 First study case results and discussion

The results in raster format that were obtained in this simulation were: drainage, runoff, evapotranspiration in which R could create utilizing the raster output (.tif format) that was produced by the package, meaning that these results could be read by any GIS software.

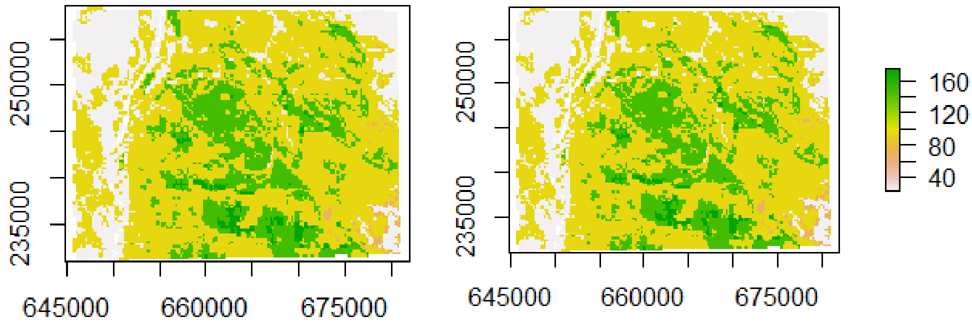

Figure 4 - Drainage (mm) map for maize (left) and sunflower (right)

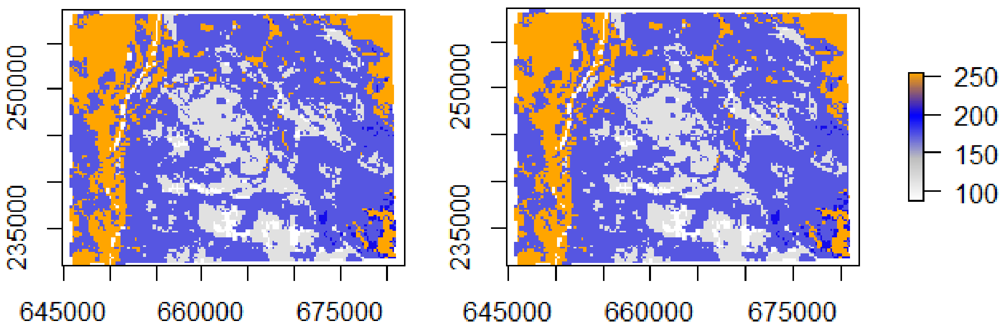

Figure 5 Runoff (mm) map of maize (left) and sunflower (right)

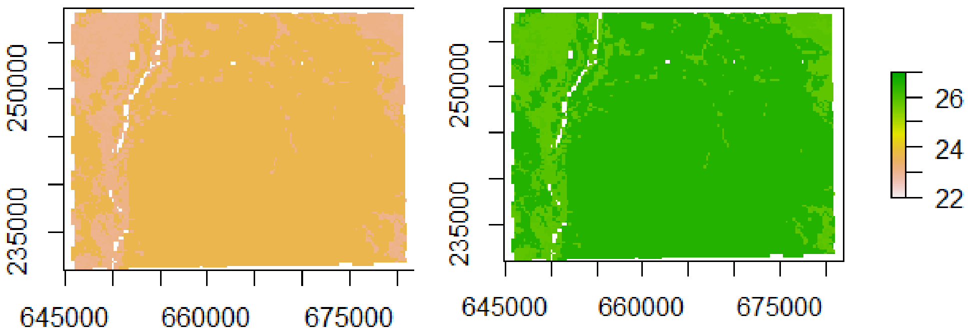

Figure 6 - Evapotranspiration (mm) map of maize (left) and sunflower (right)

Even though the only spatial results that were used in the simulation were the soil parameters, the first results that were obtained in this simulation are pertinent. First it is clear that all the functions in the package worked as the simulation could run without any issue from the beginning to the end and produce the different outputs that it was set to produce.

The results themselves from this study case, even though a simple one, show some relations with other results from prior studies, water footprint and yield (He et al., 2024,water footprint network, 2024). This is important as it means that the AquaCrop plugin was properly used but the SpatialAquaCrop package and the input data used was properly read by the package. Water Footprint is not something that AquaCrop model outputs but adding it to the package as an extra function for users to run seemed like a good decision, as calculating, green water footprint in this case, is quite relevant to modern studies for water management related to crop fields.

Even though not many new things were observed in this first study case, the main point of it was to have a proper first run of the SpatialAquaCrop package and check how well it performs and where some improvements could be made. One important point was to add outputs not just in a spatial manner but in a table format, which would facilitate the analysis of the results, another one was to add the daily results as well, not just the seasonal ones. These two features were added to the next version of the packaged and utilized for the next study case.

3.2 Results second case and discussion

The first results to check are the results from the seasonal outputs in which AquaCrop Plugin generated and SpatialAquaCrop transformed into raster format files. These seasonal files take in consideration the whole simulation period and not the result of only one day

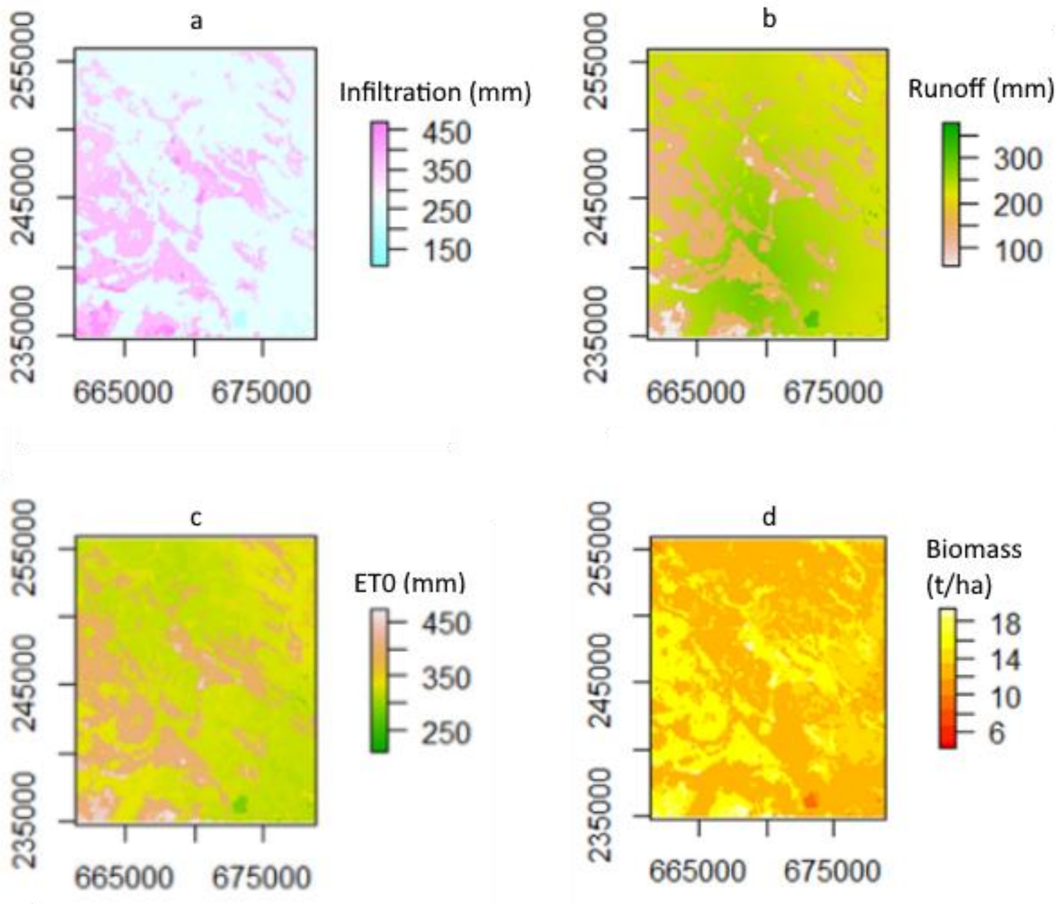

Figure 7 - Spatial variation maps of seasonal Infiltration (4.a.), Runoff (4.b.), Evapotranspiration (4.c.) and Biomass (4.d.).

The ability to output daily results in a spatial format greatly enhances the capability of the SpatialAquaCrop package, resulting in a great addition to the package. Due to this it is possible to analyse the results of the simulation as a time series, which was done for this study case. One of the goals of this study case was to compare simulated soil moisture from the AquaCrop model with field soil moisture values, in a way to validate the results obtained with the simulation.

Figure 8 presents the comparison of precipitation input data (based on an interpolated country level dataset) with the rain gauge data collected at the meteorological station, presenting a clear alignment of precipitation events, with a significant correlation value of 0.86, meaning that the precipitation data used in the simulation was on par with the precipitation data gathered at the meteorological station.

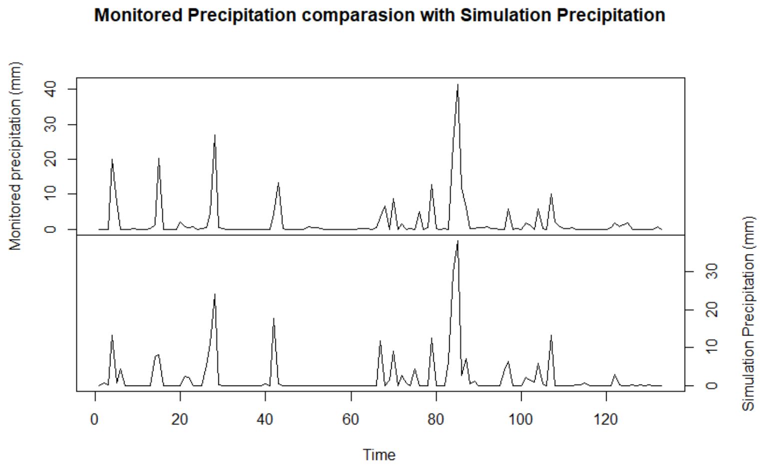

Figure 8 - Comparison of the simulated precipitation data with the field monitored ones

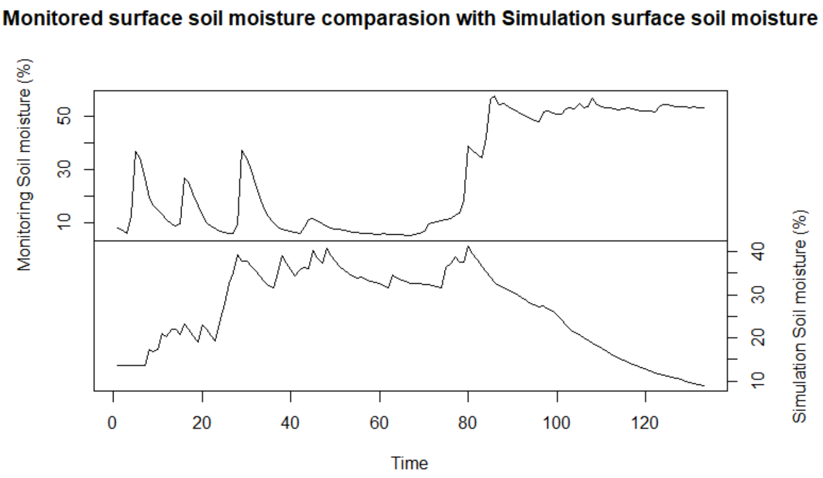

Figure 9 - Comparison of the simulated soil moisture content (surface layer) data with the field monitored ones

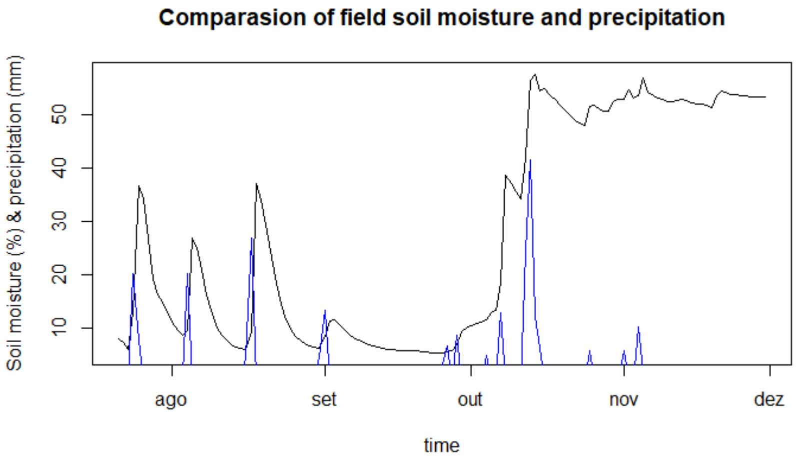

Figure 10 - Field (measured) soil moisture content for the surface layer (black) and precipitation (blue)

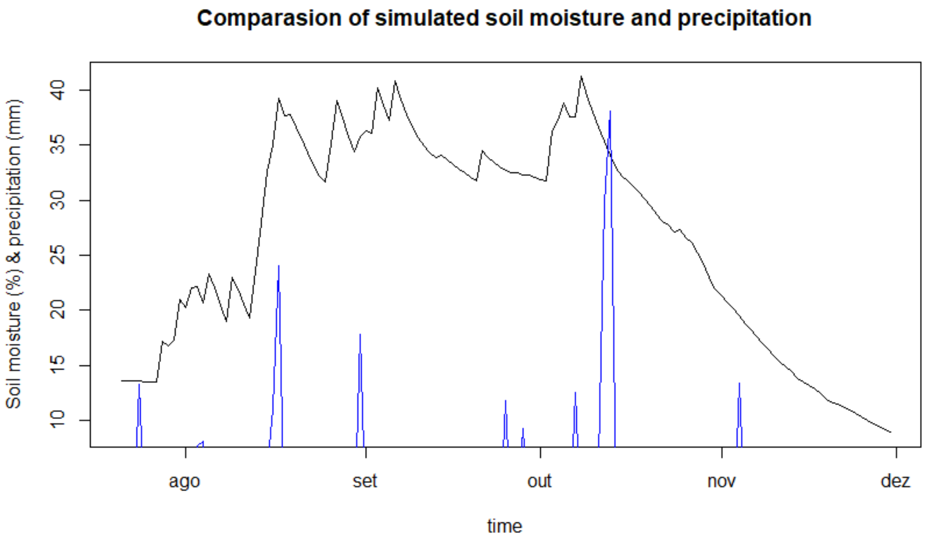

Figure 11 - Simulated soil moisture content for the surface layer (black) and precipitation (blue)

Figures 10 and 11 present a comparison of precipitation data and topsoil soil moisture content for measured and simulated data respectively. The rise in measured soil moisture for the middle of October happens at the same time as a big precipitation event occurs and after that the soil stays in a "saturated" state until the end of the time series. Before October both Figure 10 and 11 soil moisture peaks follow the precipitation peaks as well, but Figure 10 shows an overall lower variance, which could mean that other inputs (such as soil parameters) are making this difference between the two graphs occur, since precipitation is practically similar between them. This soil parameters assumption will be checked more in depth in the final study case of this thesis (as mentioned in prior paragraphs) and it will be seen that changing these soil parameters will make the simulated results fit better with the ones present in field.

While the above differences in modeled and measured soil moisture dynamics are significant, it is important to note that limitations of input data can clearly be the source of such errors. The differences are clearly driven by soil parameters, most of which used in this study have been derived from a 250 by 250m raster dataset that has been designed at a European level. This can clearly be the source of such errors and highlights the importance of scale and resolution. However, in this scenario (also limited by the applied model), the effects of topography and surface conditions (such as ruggedness) have not been considered. But according to Vereecken et al. this approach can have benefits, too, since the very detailed local parameters are usually not representative enough for the whole modeled environment. This issue with soil parameters dictating the error for the simulation is addressed in the last study case, as it is one of the main points observed in that study case.

3.3 Results third case and discussion

When comparing both of the results for AquaCrop and Hydrus that can be seen in figure 29 and 30, as mentioned AquaCrop shows a higher daily variation for the soil moisture results. Even when taking in consideration all the errors of the input data that were found after this short analysis was done, the importance of it still remains, as it can be seen a difference for both models' soil moisture results. The idea of utilizing another model to verify how well their outputs are when comparing to each other is to have a broader view of different methodologies. When taking in consideration different references the Hydrus model does present as an alternative for simulating soil moisture and it could be utilized together with AquaCrop for developing a methodology that would bring the best for bother models. But for this study that avenue was not chosen, as the development and usage of the SpatialAquaCrop package has been the aim of this PhD study, learning and focusing on another model would not be beneficial for this study in specific, but this small study led to some good ideas for some different methodologies, which were presented in the soil conference in 2022.

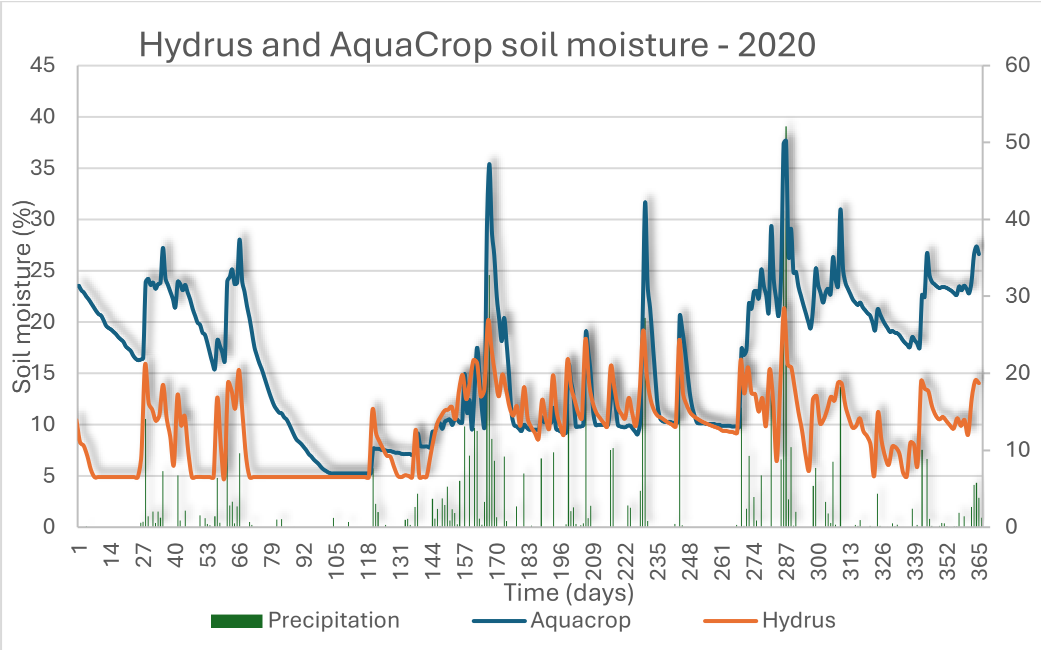

Figure 12 - Hydrus and AquaCrop Soil moisture - 2020

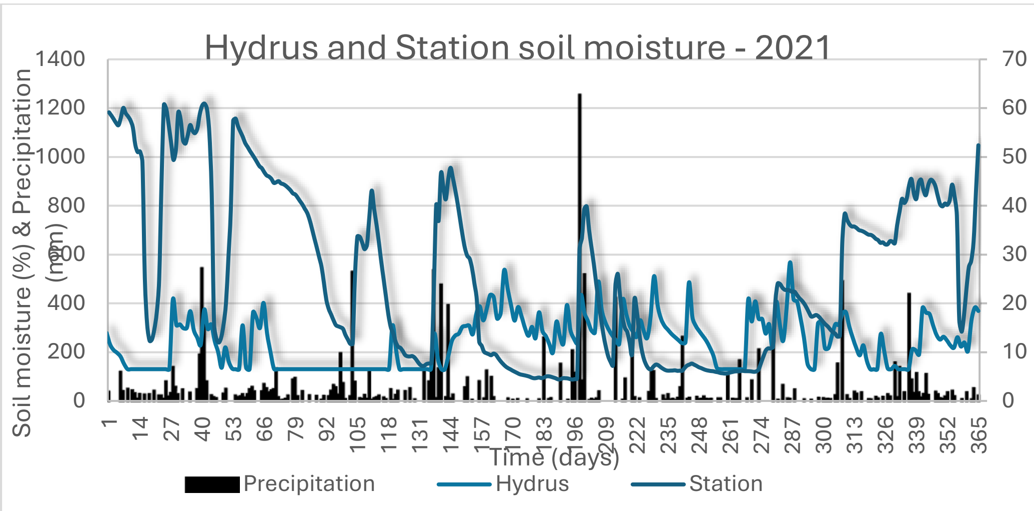

Figure 13 - Hydrus and Research station Soil moisture – 2021

3.4 Results fourth case and discussion

When evaluating the soil moisture for maize in 2020 at Martonvásár , it is possible to see some trends between both of the timeseries and significant and good correlation coefficient between them both. For the beginning of the timeseries they showed around a difference between both of the timeseries of 7%, but that value decreases as two “significant” precipitation events that occurred, one on May 24th and the other one in June 12th. It is interesting to see that the values tend to get closer after the precipitation events and another important point to take into consideration is how the AquaCrop model reacts faster to the rain in relation to what happens in reality in the field. This can be explained due to how AquaCrop calculates soil moisture and how it doesn’t take in consideration some soil characteristics that would “smoothen” this rise in soil moisture due to the precipitation. As the soil parameters for this comparison were taken from the work of Sándor we can exclude that this different from both timeseries comes from some difference in the soil parameters at the area.

One good conclusion that can be taken from this comparison is that the model works in a good manner when compared to data that was collected in the field, when comparing to the spatial data that was used for prior study cases, adding more to the many different model validations that were done before (. This difference will be analyzed better in the next, and last, study case of this PhD thesis, as utilizing raster datasets is one of the main goals of the methodology developed in this thesis.

When analyzing figures 32, 33 and 34 and later their corresponding correlation values it can be seen that there is a good correlation between modeled biomass and NDVI until the senescence period of winter wheat for this region in Hungary. It has been seen that this correlation can be seen with other crops as well in other different regions, but they might be region dependent (Abi Saab et al., 2021; Tenreiro et al., 2021). Despite this it is possible to say that AquaCrop can give good results for the modeled biomass as they correlate quite well with NDVI that was seen in the studied area.

As for green canopy crop cover the R^2 values were lower than biomass while considering a linear regression model but showed better results when considering a polynomial model of second order. When taking into consideration CC values over 80% an exponential regression model became a better fit and that can be seen better for the 2021 analysis, because there were more points that could be used for the comparison in relation to 2020. It is interesting to note that this exponential behavior starts to happen when CC is around 80%, which is close to when the model considers that crop reached its maximum rooting depth (parameters were established from winter wheat crop reference (Szasz, 1988).

Even though these correlations between biomass and CC with NDVI could be found for winter wheat in our chosen region in Hungary, this could change for other regions of the country, also depending on winter wheat varieties. Different studies support that NDVI and CC have a correlation for different crops, even for winter wheat as well (Tenreiro et al., 2021, Lykhovyd et al., 2022). This supports even more the data that shows this correlation for these two parameters. The same can be said for biomass and NVDI (Goswami et al., 2015; Farias et al., 2023).

An advantage of the applied methodology is the potential generation of spatially distributed daily output data, which essentially allows the generation of "data cubes" for the specific study area. This opens the possibility of comparison and validation with earth observation data, as well as for related agronomical applications, such as irrigation scheduling.

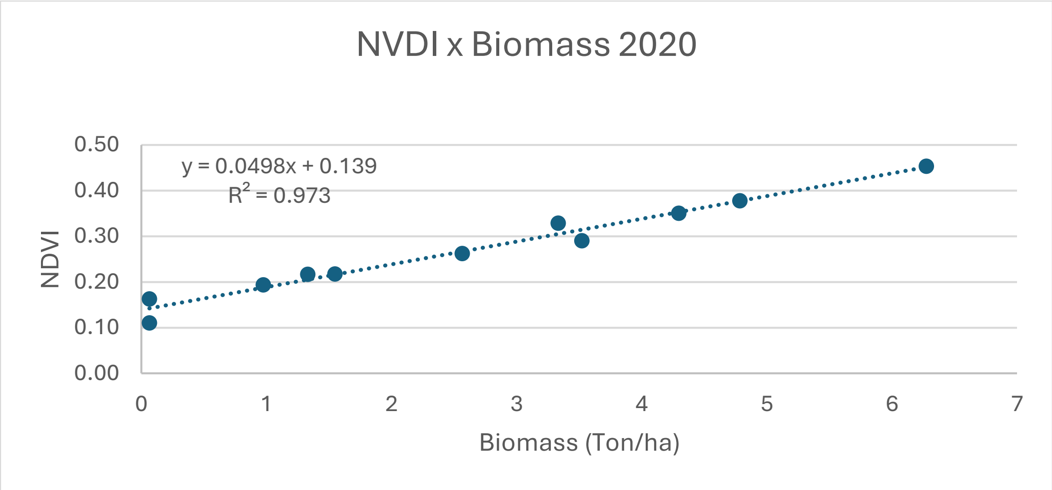

Figure 14 - Comparison between NDVI index and modeled biomass for winter wheat in 2020

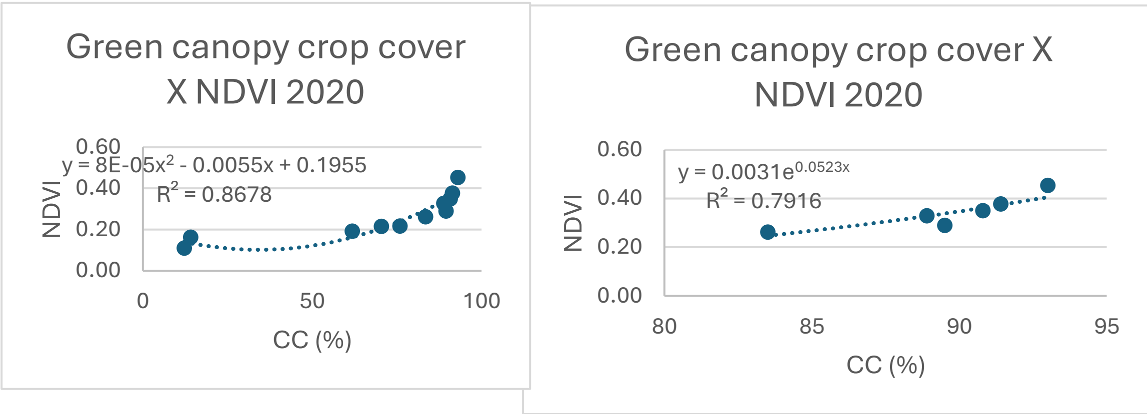

Figure 15 - Comparison between NDVI index and modeled green canopy crop cover for winter wheat in 2020 while: (15.a.) Considering the whole time series; (15.b.) Considering just when green canopy crop cover is above 80%.

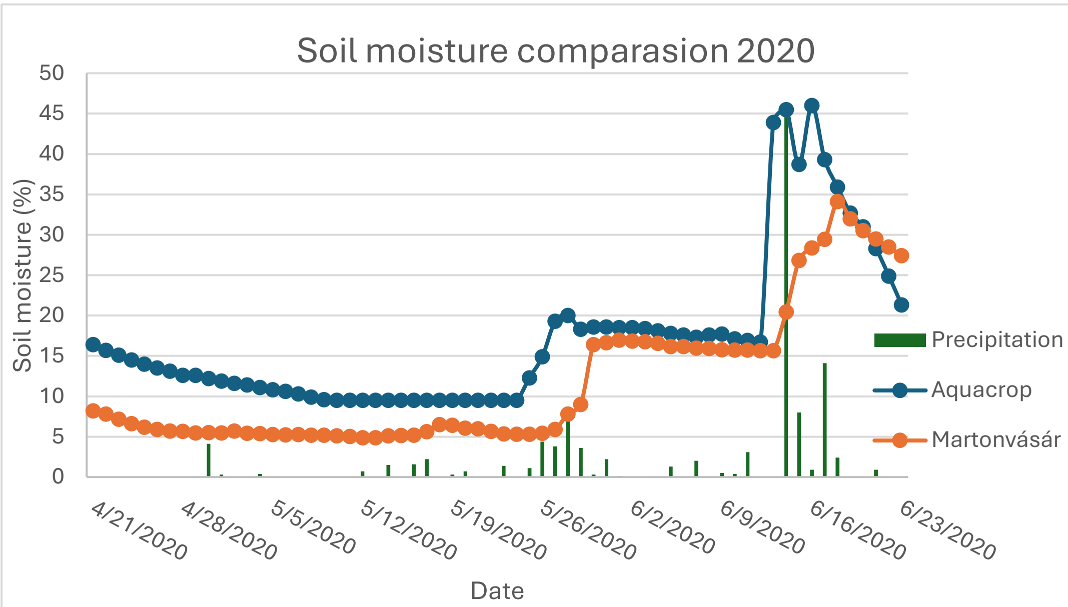

Figure 16 - Comparison between modeled soil moisture and measure soil moisture in Martonvásar for 2020.

3.5 Results fifth case and discussion

When first analyzing the results for the different meteorological stations, it was possible to see that there was a big discrepancy in how soil moisture behaved for the different years. Another thing that was noticed was that there was quite a difference in the behaviour of soil moisture for the different analyzed points, with correlations between the modeled soil moisture and the surveyed one varying from values of 0.4 to 0.8. The results with low correlation values showed similar results to past study cases, in which the soil moisture had a higher daily variation and response to different precipitation events or sometimes opposite behaving from the surveyed soil moisture from the sites.

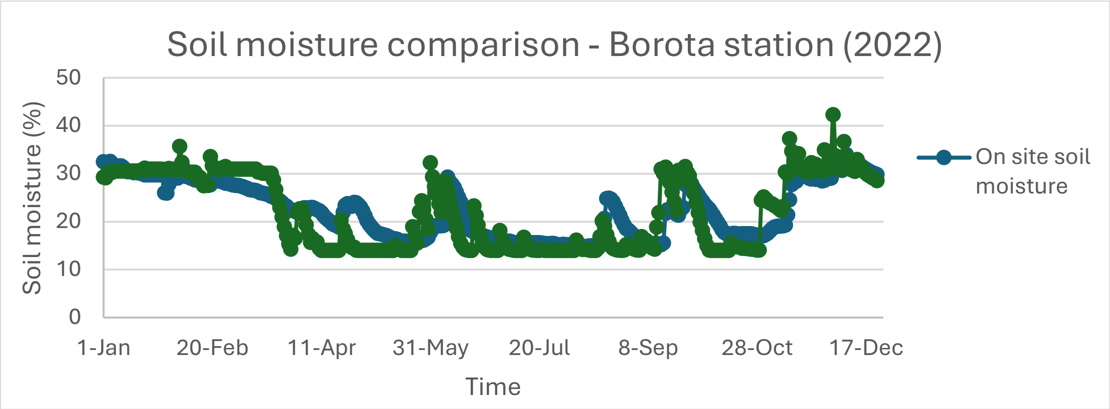

Figure 17 – Soil moisture comparison 2022 Borota

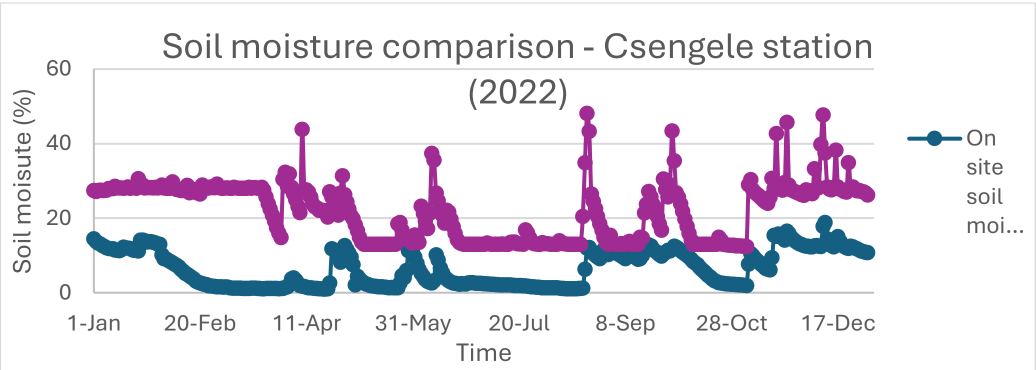

Figure 18 – Soil moisture comparison 2022 Csengele

Even though some of the station showed promising results with correlation values higher than 7.0, more than half of them showed correlation values lower than 7.0, not being satisfactory in theses cases. To improve the simulated soil moisture results, some changes in the soil parameters, and sometimes adding other effects such as mulching or the presence of groundwater in a specific depth, were made, so that the correlation between the modeled results and the onsite results would increase, having the the soil parameters being what mainly was affecting the difference in soil moisture results. These changes made the correlation between the two soil moistures for some of the points to improve up to around 0.15. Figures 40 and 41 shows the timeseries for the station in Csengele for 2022 after these changes were made.

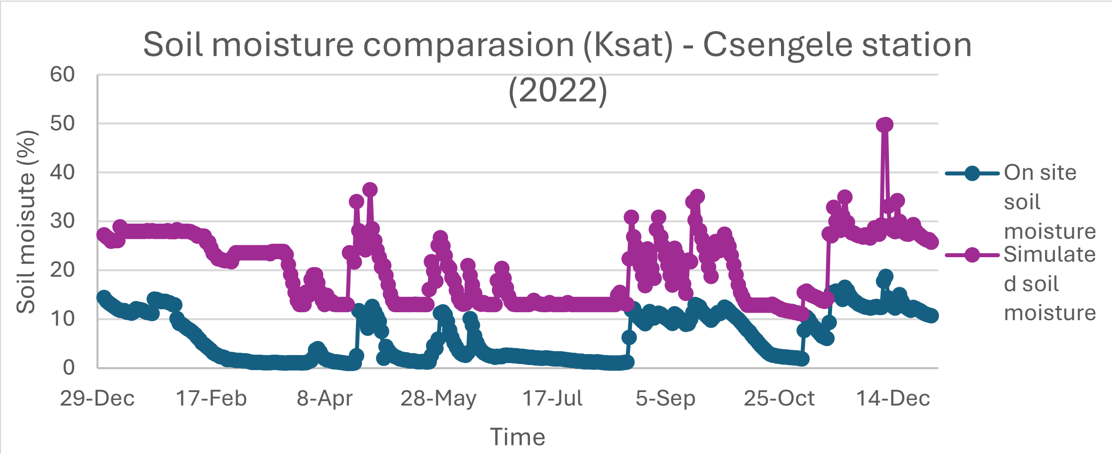

Figure 19 – 2nd Soil moisture comparison 2022 Csengele

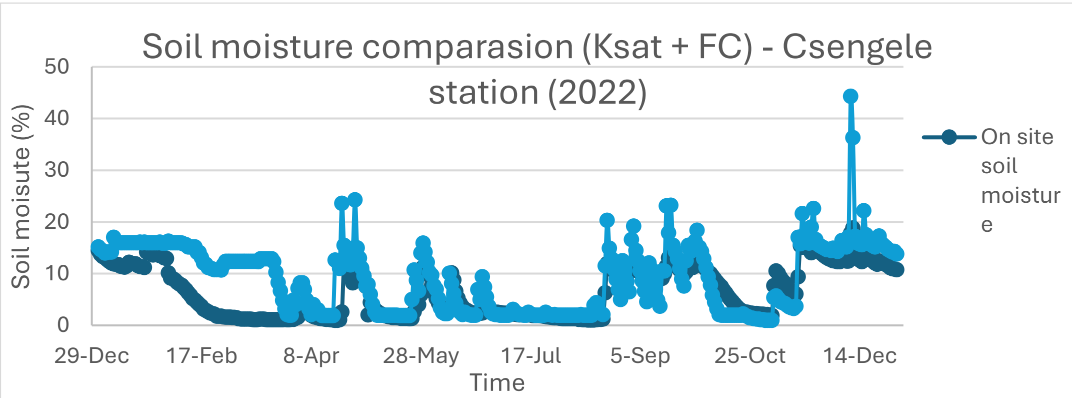

Figure 20 – 2nd Soil moisture comparison 2022 Csengele

Changing the Ksat value made the daily variation from the simulated soil moisture be closer to the one found on site , but still there was a big different in between their values. For that difference to be lower a change in the Field Capacity (FC), generally close to halving the FC value, and considering the Wilting Point as the lowest value for soil moisture that can be seen on site. These changes can be seen in Fig 20 for the station in Csengeled and it can be visually seen that the difference in value between the simulated results and on site results are quite lower for most of the year, presenting only a bigger difference for colder periods. Mulching and a set depth for groundwater were checked as well in a way to improve the correlation results for this analysis, but even though changing/considering some of them for some stations would improve their correlation value, this was not true for all the stations, so it was chosen to leave them out of the changes for a preferred change that could be used for all station and would have a similar effect to them (soil parameters).

Knowing these discrepancies the next step for this analysis was using another dataset for the soil parameters and check how well the results would correlate with the field data from the meteorological stations. The new soil parameter data has been taken from Hungarian research network (HUN-REN, 2024) and that dataset provided new soil parameters for the region, specifically for each of the points. To analyze as well how well the standard values that AquaCrop provides for Ksat for each of the different soil textures, two new simulations have been run for each of the sites. One while taking in consideration the USDA soil texture for each of the points and utilizing the standard values that AquaCrop provides for each texture and another utilizing the specific Ksat for each of the points. Table 3 shows the difference in soil parameters from each of the different datasets for the surface layer.

Table 1 – soil parameters simulations

	Agrotopo					hungary soil					
Stations	**Texture**	**Ksat - Aquacrop**	**FC**	**SAT**	**WP**	**Texture**	**Ksat USDA -Aquacrop**	**FC**	**SAT**	**WP**	**Ksat**
apaj	loam	500	34	50	15	Loam	500	30.56642	42.21951	12.83411	6.6125398
borota	loam	500	33	48	14	Clay loam	125	25.54282	44.38335	13.35102	10.522793
csavoly	loam	500	35	48	14	sandy clay loam	150	30.0395	46.6776	14.5914	30.295965
csengele	sand	3000	31	49	13	clay	35	16.30338	41.53843	5.381576	49.259346
csolyopalos	sand	3000	30	48	12	clay	35	11.96997	42.89684	6.309484	187.44196
csongrad	sand	3000	31	49	13	Loam	500	28.89514	43.79094	13.62467	63.440269
fajsz	silt clay	100	35	47	17	sandy clay	35	36.99158	46.56603	19.59976	0.4444771
fulophaza	sand	3000	30	49	13	Loam	500	25.65304	41.81246	12.07318	19.975574
harta	loamy sand	2200	34	49	15	sandy clay loam	150	32.79635	45.60336	14.28027	5.2184701
hernad	sand	3000	30	50	12	clay	35	8.939511	41.03044	3.578464	90.556374
homokmegy	loam	500	33	49	15	sandy clay loam	150	31.00703	47.71582	18.20502	3.190614
izsak	sand	3000	28	48	12	clay	35	10.99911	41.38602	7.872583	258.32324
kalocsa	silt clay	100	35	48	16	sandy clay loam	150	31.96735	47.51631	19.54557	0.7486087
kecel	sand	3000	30	47	12	clay	35	11.33064	41.85612	5.75299	22.994823
kiskunfelegyhaza	loam	500	32	49	13	Loam	500	30.33614	44.91705	18.09769	10.143685
kiskunhalas	loamy sand	2200	33	49	14	Clay loam	125	15.60596	39.7766	8.181291	14.607512
kisszallas	loamy sand	2200	30	49	12	Clay loam	125	15.86756	40.8507	9.310092	25.518831
Kunszemtkilos	loam	500	31	50	14	Loam	500	26.02725	44.43876	12.55921	39.377037
kunpeszer	loam	500	33	50	14	Loam	500	27.59867	45.91656	14.15188	8.8669472
lajosmize	loamy sand	2200	30	50	12	Loam	500	17.64167	40.98136	6.852107	64.807022
Melykut	loamy sand	2200	32	48	14	Clay loam	125	10.81333	41.03601	7.304429	105.51717
Nagykiros	loam	500	32	50	14	Loam	500	16.97335	41.35119	7.642901	376.91357
Palmonostora	loam	500	34	49	15	Clay loam	125	24.29067	41.88156	13.96841	559.08112
Ruzsa	loamy sand	2200	30	48	12	clay	35	9.918707	39.406	5.196022	23.181309
Sandorfalva	sand	3000	34	49	15	Clay loam	125	13.13267	41.72457	8.218563	195.68369
Solt	loam	500	34	50	15	Clay loam	125	23.67883	45.11568	10.7722	27.07959
Sukosd	sand	3000	34	50	15	sandy clay loam	150	33.71981	46.98284	17.67896	1.0207294
Tazlar	sand	3000	30	48	12	clay	35	10.80107	43.06703	5.559392	160.22644
Varosfold	loam	500	33	49	14	Loam	500	26.36108	46.04242	14.91554	37.459206

From looking at table 1 it is possible to see that there is a big difference for some stations regarding the soil texture and soil parameters. This indicates that there will be different results when running the SpatialAquaCrop package. Another point to check is that the soil texture variation is lower in the Agrotopo soil data when compared to the new one from HUN-REN, 2024, this might indicate that the second dataset is more closely related to the reality on the field due to this texture variation.

Table 2 shows all the difference in correlation for every different simulation for the points of this study case. As mentioned, before it can be seen an increase in correlation for most of the stations when utilizing the soil parameters for the Hungary research network and when changing just the Ksat for the standard values provided in AquaCrop as well.

Table 2 – Correlation values 2022 for all points

	Correlation (2022)		
Stations	**Agrotopo**	**USDA**	**HUN-REN, 2024**
Apaj	0.73	0.72	0.75
Borota	0.87	0.85	0.89
Csávoly	0.72	0.84	0.71
Csengele	0.47	0.77	0.47
Csólyopálos	0.60	0.76	0.66
Csongrád	0.47	0.58	0.57
Fajsz	0.67	0.81	0.55
Fülöpháza	0.61	0.76	0.78
Harta	0.78	0.85	0.85
Hernád	0.52	0.46	0.34
Homokmégy	0.42	0.76	0.51
Izsák	0.41	0.66	0.46
Kalocsa	0.6	0.67	0.71
Kecel	0.28	0.65	0.59
Kskunfélegyháza	0.41	0.71	0.73
Kiskunhalas	0.42	0.68	0.68
Kisszállás	0.48	0.53	0.53
Kunszemtkilós	0.66	0.74	0.73
Kunpeszér	0.56	0.69	0.73
Lajosmize	0.53	0.71	0.71
Mélykút	0.42	0.60	0.48
Nagykőrös	0.20	0.63	0.61
Pálmonostora	0.39	0.25	0.23
Ruzsa	0.31	0.64	0.56

Sándorfalva	0.25	0.65	0.52
Solt	0.63	0.75	0.75
Sükösd	0.46	0.78	0.87
Tázlár	0.34	0.59	0.44
Városföld	0.45	0.67	0.68

For this last study case of this PhD one important point is the availability of different soil datasets , for comparison, as some discrepancies due to, most likely, soil parameters in the past study cases have been seen. Even though both of these datasets have been obtained from interpolated data, it is possible to see from the results of the comparison of soil moisture, that the dataset from the Hungarian research network has shown better results than the one from Corine. As this comparison was done for the south region of Hungary it is not possible to say that for sure the Hungarian research network is more accurate than the other one in general, but for these points and region these parameters show better results when simulating soil moisture.

Besides the soil parameters, another point in which can be compared with prior study cases is the crop used in the simulation. As in the second study case a grassland simulation was attempted, in this last one grass was simulated as well utilizing AquaCrop. Even though AquaCrop is not first meant to simulate grass it has been possible as seen in different studies . The new grass crop that was created utilizing these studies in mind is not meant to completely follow reality but to try to create a grass crop which would always be present on the soil surface and mimic the water consumption/root depth and plant growth from the real grassy crops. As seen in Figure 36 the first year of the 3 year simulation showed a “build up” period for the simulated soil moisture, that is thought to be mainly a cause of the crop parameters, as the “new crop” has been designed to be and stay mature as fast as possible, but there might have been some issues with the different parameters used in the crop file, which may have caused this “build up” in the first year. For the second and third year it is possible to see that the soil moisture started behaving more accordingly to what is expected and because of that only the third year has been used for comparison.

As the main goal of this study case was the comparison of soil moisture from the simulated data and the on site data from the meteorological stations and that this soil moisture data from these stations may represent a base soil moisture value for the region, as the area whether the soil moisture has been assessed near the stations are most likely covered with grass. This base soil moisture is important as it shows a value of soil moisture that may be expected in the region and some different thresholds for when low infiltration or high runoff is to be expected.

What has been seen on the different comparisons is that the different points show different rates of correlation between the simulated datasets and the on site soil moisture. And that changing the soil moisture parameters change the correlation value. This aspect let to believe that some points in the raster datasets dot not really represent the parameters seen on the field. This shows that there has a change in which the soil in some of these stations may have been tempered, possible being affected by compaction or added organic matter in the soil, these affect on the value for Ksat and field capacity. Of course it is impossible to know for sure if these effects are actually happening in reality, but as two different soil parameters sets were used in this study (and the one from the Hungarian research network seems to be an improvement to the Corine one) it can be said that it is expected that the best results from then would properly simulate the soil parameters in the region. So, considering that the best correlation results represent soil parameters that match with reality, and that has happened for most of the points (table 4), so the points which do not show this good correlation can be considered to have something that changes the soil characteristics.

When looking at the results from the 3 different runs it is possible to see that when utilizing the soil parameters from the Hungarian research network the correlation results were generally better and showed a small or big correlation, higher than 0.7. And utilizing the Ksat from AquaCrop in correlation to the USDA texture the number of stations which show a good correlation increase. These variations in soil parameters are important for when analyzing the results from the AquaCrop model, as they show how different parameters, soil moisture in this case, are affected by the increase or decrease of a specific parameter. A parameter that made a big different in the correlation value has been the Ksat, as it mainly changed how the daily variation of soil moisture behaved.

This study case showed a more point based approach when in comparison to the other studies, as the other ones have already set the spatial capabilities with the methodology of utilizing the SpatiaAquaCrop package. So in this last study a point based approach was chosen, with a focus on the soil parameters modeling and how they affect the overall results of the AquaCrop model, and more specifically soil moisture. Besides showing how the soil parameters affected the soil moisture results, it showed as well the importance of the quality of input data, as there is a significant difference from both soil parameter datasets that were used in this study.

4. Conclusion and recommendations

Nowadays, given the increasing amount of available spatial and remotely sensed data, combined with the need of agricultural water management, an increased number of applications would require raster-based utilization of AquaCrop. The purpose of this PhD was to develop such a methodology. In data scarce regions where accurate yield and soil moisture measurements are not available we can utilize remote sensing based and also model based estimations.

The SpatialAquaCrop package was made in a way that it would give an easy and understandable option to utilize the AquaCrop model not just for a specific point but rather to an area (utilizing raster based inputs). In its current state the package can output all the outputs that the stand alone version of AquaCrop is capable to and besides this the package has a function to output the green water footprint. As shown in the different study cases the package has show the versatility and capability to properly run the AquaCrop-plugin and prepare all the input data that are necessary for the simulation. The advantage of being able to read and output data in a raster format or in different tables for the daily results is a great advantage, as it opens more advantages for the utilization of the AquaCrop model, instead of the point base software.

When seeing the results from the five different study cases from an overall viewpoint, one thing that stands out is the importance of input data to the outputs the AquaCrop model can give. Study cases 1 and 2 showed different correlations with different outputs, which had similar trends to what has been seen in some prior studies, and one important was that it was possible to simulate grass with not big errors. Study case 4 showed that it is possible to correlate and utilize satellite base data for input and correlation to the model results with good results. That is important as in areas with data scarcity it is possible to utilize satellite based data to make up for missing data. Study case 5 mainly addressed one point that was becoming apparent in the prior cases, which was the quality of the input data, as that could be seen that the HUN-REN soil dataset gave better results when compared to the CORINE dataset. As the Corine dataset has been used to all other studies, this shows that maybe utilizing this new dataset to prior studies might improve past results, especially when utilizing Ksat from the standard table that AquaCrop provides. One important thing to highlight is that with the new soil dataset that OVF has released recently it could open new possibilities for the analysis of the region of study case number 5, as now there is a detailed soil description and data for each of the different meteorological stations, which now can provide data for validating the modeled results, but due to the time this dataset has been released, validation would be done in a future research.

When taking in consideration the different results and the practical applicability of this methodology, one possibility which could be approached is utilizing it together with different climate prediction models, so in this case this methodology could be used to predict, for example, floods, necessity of irrigation for a particular period, general soil water content and other environmental characteristics which would be pertinent for the applied data.

Most of this research was done inside R environment where the package was created, which enhances the possibility of an easier way for analysing and utilizing the different outputs that SpatialAquaCrop provides. Also, if the user has some programing knowledge in R, it is possible to add additional outputs to the package if necessary.

5. New scientific results

1. I have successfully developed an R-based methodology and package (SpatialAquaCrop) for the raster-based implementation of the AquaCrop model, demonstrating that the model can be utilized to generate spatially explicit predictions where data is available.
2. I have demonstrated that the methodology is suitable for generating not only seasonal estimates, but also daily grid outputs of selected parameters, making it suitable for further research into drought management and irrigation applications.
3. I have demonstrated that the estimated biomass and green canopy cover values of AquaCrop (utilized with the SpatialAquaCrop package) correlated well with Sentinel-1 based NDVI values in case of winter wheat, proving that the developed methodology could be particularly useful for applications combined with remote sensing.
4. I have demonstrated that with the available data for Martonvásár 2020, under winter wheat, AquaCrop has systematically produced higher estimates of volumetric soil moisture content than the measured values.
5. I have developed a methodology to apply AquaCrop for selected points of the Hungarian National Drought Monitoring Network, and demonstrated that there are notable differences in the accuracy of the estimates, that can be most likely be associated with inaccurate soil data parameters, derived from spatial datasets. This indicates that spatial extension of the soil moisture data will likely carry similar limitations.
6. I have demonstrated that the utilization of the new HunSoilHydroGrids dataset has significantly improved the estimates of soil moisture, making this dataset more feasible for generating spatial estimates of soil moisture.

6. Publications

Deganutti de Barros, Vinicius & Waltner, István & Rakotoarivony, Ny & Halupka, Gábor & Sándor, Renáta & Kaldybayeva, Dana & Gelybó, Györgyi. (2022). SpatialAquaCrop, an R Package for Raster-Based Implementation of the AquaCrop Model. Plants. 11. 2907. 10.3390/plants11212907.

Hatvani, István & **Deganutti de Barros, Vinicius** & Tanos, Péter & Kovács, József & Székely, Ilona & Clement, Adrienne. (2020). Spatiotemporal changes and drivers of trophic status over three decades in the largest shallow lake in Central Europe, Lake Balaton. Ecological Engineering. 151. 105861. 10.1016/j.ecoleng.2020.105861.

Saeidi, Sahar & Grósz, János & Sebők, András & **Deganutti de Barros, Vinicius** & Waltner, István. (2020). A területhasználat változása a Rákos-patak vízgyűjtőjén 1990-től. Journal of Landscape Ecology. 17. 287-296. 10.56617/tl.3524.

Deganutti de Barros, Vinicius & Saeidi, Sahar & Amekelew, Shewaye & Grósz, János & Sebők, András & Waltner, István. (2020). WATER QUALITY MONITORING FOR RÁKOS STREAM IN THE WINTER OF 2019/2020.

Vinicius Deganutti De Barros; Malek Abidli ; Gábor Halupka;Dana Kaldybayeva ; János Grósz ; István Waltner .Spatial modeling of soil moisture dynamics for the Rákos catchment, Hungary
In: 22nd World Congress of Soil Science : poster book of abstracts Conference: Glasgow, United Kingdom / Scotland 2022.07.31. - 2022.08.05. (International Union of Soil Science, British Society of Soil Science) Glasgow: IUSS, p. 7. 1 p. (2022)