WATERAGRI

D7.5: Data Assimilation System for Physically Based Models

June 2023

WP7 Framework

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 858375.

Author(s)/Organisation(s)	Richard Hoffmann (FZJ), Harrie-Jan Hendricks-Franssen (FZJ), Philip Brunner (UNINE), Qi Tang (UNINE), Björn Klöve (OULU), Anna Autio (OULU), Wieslaw Fialkiewicz (UPWR), Arkadiusz Głogowski (UPWR), Attila Nagy (UNIDEB), Andrea Szabo (UNIDEB), Erik Nilsson (ULUND, „QuantClim“), Diego Guidotti (Agricolus)
Contributor(s)	
Work Package	WP7
Delivery Date (DoA)	30/06/2022
Actual Delivery Date	30/06/2022
Abstract:	D7.5 reports on the data assimilation systems for physically based models developed in WATERAGRI. The data assimilation systems are designed to provide the best possible estimates of hydrologic conditions in agricultural watersheds. For WATERAGRI, we assimilate soil moisture and groundwater levels in near real-time to quantify model uncertainty and create less uncertain initial conditions (states) for site-specific forecasting models. D7.5 shows the first simulations with and without data assimilation and how agricultural decision-making at the plot scale could look like when physically based models are used operationally.

Document Revision History			
Date	**Version**	**Author/Contributor/ Reviewer**	**Summary of main changes**
08/01/2022	V0	Richard Hoffmann (FZJ) – Editor	Template
23/05/2023	V1	Richard Hoffmann (FZJ) – Editor	Compiled version with input from all partners involved
05/06/2023	V2	Harrie-Jan Hendricks-Franssen (FZJ)	Review of V1
19/06/2023	V3	Richard Hoffmann (FZJ) – Editor	Revised draft after feedback from Reviewer
23/06/2023	V4	Richard Hoffmann (FZJ) – Editor	Final draft after revisions from all partners
26/06/2023	V5	Sebastian Puculek (ULUND) – Contributor	Quality Control
27/06/2023	V6	Rolf Larsson (ULUND) – Contributor	Final Check
28/06/2023	VF	Richard Hoffmann (FZJ) – Editor	Final version

Dissemination Level		
PU	Public	X
CI	Classified information as referred to in Commission Decision 2001/844/EC	
CO	Confidential, only for members of the consortium (including the EC)	

WATERAGRI Consortium			
Participant Number	Participant organisation name	Short name	Country
1	LUNDS UNIVERSITET	ULUND	SE
2	EDEN MICROFLUIDICS	EDEN	FR
3	FORSCHUNGSZENTRUM JULICH GMBH	FZJ	DE
4	TEKNOLOGIAN TUTKIMUSKESKUS VTT Oy	VTT	FI
5	DEBRECENI EGYETEM	UNIDEB	HU
6	ALCHEMIA-NOVA GMBH	ALCN	AT
7	AGROGEO AGARFEJLESZTO-FOLDTANI-FOVALLALKOZO KORLATOLT FELELOSSEGU TATRSASAG	AGROGEO	HU
8	UNIVERSITAET FUER BODENKULTUR WIEN	BOKU	AT
9	ALMA MATER STUDIORUM UNIVERSITA DI BOLOGNA	UNIBO	IT
10	THE UNIVERSITY OF SALFORD	USAL	UK
11	COCONSORZIO DI BONIFICA DI SECONDO GRADO PER IL CANALE EMILIANO ROMAGNOLO CANALE GIANDOTTI	CER	IT
12	CENTRUM DORADZTWA ROLNICZEGO W BRWINOWIE	CDR	PL
13	INOSENS DOO NOVI SAD	INOSENS	RS
14	UNIWERSYTET PRZYRODNICZY WE WROCLAWIU	UPWr	PL
15	BAY ZOLTAN ALKALMAZOTT KUTATASI KOZHASZNU NONPROFIT KFT	BZN	HU
16	VULTUS AB	VULTUS	SE
17	TECHNISCHE UNIVERSITEIT DELFT	TU DELFT	NL
18	UNIVERSITE DE NEUCHATEL	UNINE	CH
19	AB GARDSTANGA NYGARD	GN	SE
20	OULUN YLIOPISTO	OULU	FI
21	AGRICOLUS SRL	AGRICOLUS	IT
22	INSTITUT NATIONAL DE RECHERCHE POUR L'AGRICULTURE, L'ALIMENTATION ET L'ENVIRONNEMENT	INRAE	FR
23	MARTIN REGELSBERGER	TBR	AT

Funding Scheme: Research and Innovation Action (RIA) ● Theme: SFS-23-2019
Start date of project: 01 May 2020 ● Duration: 48 months

Table of contents

WATERAGRI

List of figures

List of tables

List of Abbreviations and Acronyms	

1 General Introduction

WATERAGRI is developing a data assimilation framework to continuously update numerical models using real-time data obtained from in-situ and remote sensors. We use the physically based models developed for WATERAGRI case study sites in Germany, Switzerland, Finland, Poland, and Hungary, which we presented in D6.2 and D7.2. This document, D7.5, updates the data assimilation system's current status and application at the WATERAGRI case study sites.

The data assimilation system for WATERAGRI allows the correction of simulated terrestrial system conditions to measured values in near real-time. The pilot sites are the sites in Germany and Switzerland, as FZJ and UNINE are mainly responsible for the development of the data assimilation framework. We quantify and reduce model uncertainty through an ensemble of model runs to use the best possible initial conditions for site-specific forecast models. Therefore, we use locally adjusted information on hydraulic and soil conditions and continuously update simulations of soil moisture and/or groundwater level with measurements stored in a cloud database that the models can access directly. In addition, we use as atmospheric forcings in-situ meteorological measurements or an ensemble of mid-term weather forecasts to provide optimal estimates of current and future conditions in agricultural watersheds. For example, we forecast soil moisture, groundwater levels and surface runoff up to 10 days in advance. Such information is becoming increasingly important to stakeholders like farmers as weather variability, and the frequency and intensity of weather extremes increase and have already negatively impacted crop yields worldwide in the past. In addition, stakeholders are calling for site-specific simulations, as we have learned in WATERAGRI's Stakeholder Engagement Workshop #3 (D1.6). In this context, the developments for the sites in Germany and Switzerland will be successively used at other WATERAGRI case study sites, and thus tested for other hydrological and crop conditions. This will allow to provide a generic data assimilation system that can be used at other sites than the WATERAGRI sites. Therefore, in D7.5, we explain the data assimilation developments and implementation for the sites in Germany and Switzerland, while we emphasize on the perspective of data assimilation for the sites in Finland, Poland and Hungary.

Note that some of the material presented and the description of the physically based models are relevant to other deliverables, e.g., D5.2, D5.3, D6.1 (factsheets), D6.2 and 7.2. Some descriptions of the codes, text, figures, and tables used are included in other deliverables. In particular, D6.2 - model-based assessment, provided the description of the physically based models and was submitted only two months before this report, so there could be unavoidable repetitions.

2 Methodological background

2.1 Physically based models

WATERAGRI develops site-specific land surface models and fully distributed 3D integrated surface-subsurface hydrological (ISSHM) models to improve irrigation planning and runoff management and to reduce the impact of drought stress and overly wet conditions in the face of increasing weather extremes now and in the future. We generate forecasts starting from initial conditions obtained after simulations with on-site measurements, classical calibration procedures and data assimilation. We develop the models using the code HydroGeoSphere (HGS, Aquanty 2020) or the Terrestrial System Modeling Platform (TSMP, Shrestha et al. 2014). The Terrestrial System Modeling Platform (TSMP, Shrestha et al. 2014) combines compartmental models for the atmosphere (COSMO) (Baldauf et al. 2011), land surface (CLM, Lawrence et al. 2019), and subsurface (ParFlow) (Ashby und Falgout 1996; Kollet und Maxwell 2006; Maxwell 2013). HGS and TSMP simulate the water balance and/or the flow field of a given hydrogeological environment (details in D7.2), surface water flow, water movement in the unsaturated zone (e.g., in the soil) and in the saturated zone (e.g., in an aquifer). TSMP has a modular setup and allows using individual model components like its land surface component CLM5 which simulates crop states and, for example, evapotranspiration, Leaf Area Index or soil moisture content. Thus, TSMP, where COSMO, CLM and ParFlow are coupled, can also simulate energy, carbon and nitrogen cycles and vegetation states.

Model inputs are spatially varying information on precipitation, land use, plant species and subsurface hydraulic properties. Site-specific conditions, water, climate, and land use scenarios constrain the physically based models developed in WATERAGRI. In addition, we drive our forecasts models using short-term weather forecasts, long-term climate change and different water and land use scenarios (D6.2 and D7.2). We use HGS, especially for simulating and forecasting shallow groundwater levels in agricultural catchments and TSMP for soil moisture simulations and forecasts at the plot-scale.

HGS and TSMP use an adaptive time step. The time required to build physically based models using one of the codes (or any other code or software) is typically significantly longer than the computational time required to calculate hydrologic conditions and crop yields. Catchment or even plot scale models are very data intensive compared to analytical solutions but can provide reliable information with high temporal and/or spatial resolution. Computational times depend on the spatial and temporal resolution, the number of ensemble members, and the machine used to solve the numerical problem. For example, more than 200 working hours for model generation are quite typical, whereas a developed plot-scale model or smaller catchment model (i.e., one model realisation) can be computed in about 30 minutes or faster.

2.1.1 Terrestrial System Modelling Platform (TSMP, CLMv5)

(The text below is mainly identical to the text of the Terrestrial Model Platform description in Deliverables D3.1, D6.2 and D7.2)

The Terrestrial System Modeling Platform (TSMP, Shrestha et al. 2014) is a modular Earth compartment model, coupling the convection-permitting atmospheric model COSMO (Baldauf et al. 2011), the land surface model Community Land Model version 3.5 (CLM, Oleson et al. 2008) and the integrated hydrological model ParFlow ((Ashby und Falgout 1996; Kollet und Maxwell 2006; Maxwell 2013). COSMO is not needed for the research questions of WATERAGRI. The land surface model CLM (Community Land Model) simulates the exchange of water, energy, carbon and nitrogen between the land surface and atmosphere, i.e., CLM calculates land-atmosphere fluxes, vegetation states, carbon and nitrogen pool dynamics, soil temperature and land surface temperature (Oleson et al. 2008). In CLM, plant physiological and crop parameters can be defined by specifying the percentage of a predefined plant functional type (pft) on the natural vegetation unit (% of area unit) and the percentage of a predefined crop functional type (cft) on the cropland unit for each grid cell. ParFlow calculates variably saturated groundwater flow by solving the Richards equation, and the calculation of surface water flow is based on the 1-D kinematic wave approximation of the shallow water flow equations. Equations are solved using a cell-centred finite difference scheme and implicit time integration (Kollet und Maxwell 2006; Maxwell 2013; Ashby und Falgout 1996). When compartment models are coupled, information about fluxes and state variables is exchanged at the conceptual boundaries of the respective compartment models (Valcke 2013; Gasper et al. 2014; Shrestha et al. 2014). For example, CLM calculates net infiltration, provided to ParFlow, which calculates pressure in the unsaturated zones and overland flow. So, ParFlow provides pressure and soil moisture contents to CLM, which uses this in various calculations, for example, for ET, i.e., the hydrology of CLM is replaced with the hydrology of ParFlow (Kurtz et al. 2016). The modular character of TSMP also allows the compartment models to operate independently of each other or to use only two neighbour compartment models (e.g., CLM-ParFlow or COSMO-CLM).

It should be noted that CLM3.5 in TSMP is being updated to CLM5, which has some fundamental improvements over version 3.5. The new features allow for more realistic, site-specific simulations to meet the goals of WATERAGRI, i.e., the needs of stakeholders consulted in workshop #3 (we refer to Lawrence et al. (2019) for details on CLM5). CLM5 split evapotranspiration into transpiration, canopy evaporation, and soil evaporation. New in CLM5 is that soil evaporation is now controlled by the diffusion rate of water vapour through a dry surface layer (Kennedy et al. 2019; Swenson et al. 2019; Lawrence et al. 2019). CLM5 considers soil layers of different thicknesses, and the vertical discretisation consists of 25 layers by default. The upper 20 layers are hydraulically and biogeochemically active by default, while the deepest 5 layers are only considered in the thermodynamical calculations. The deepest 5 layers represent an impermeable bedrock or a zero-flow soil boundary condition. CLM5 can now explicitly simulate saturated and unsaturated zones and the associated groundwater level (Lawrence et al. 2019). The vertical soil moisture transport is governed by infiltration, surface and sub-surface runoff, gradient diffusion, gravity, and canopy transpiration through root extraction. The accuracy and stability of the numerical soil water solution in CLM5 have been improved by introducing an adaptive time-stepping solution to Richard's equation (Lawrence et al. 2019). In addition, a prognostic LAI is utilised by the biophysical model that couples carbon, water, and energy cycles. So, CLM5 can calculate the vegetation states prognostically now and has increased the number of plant functional types and crop functional types, which allows a more fine-grained representation of the different vegetation types and the yearly cycles of vegetation states (Lawrence et al. 2019; Kennedy et al. 2019; Swenson et al. 2019; Boas et al. 2021). As a result, CLM5 standalone can provide complete site-specific information on decision-critical parameters, e.g., soil moisture, leaf area index and crop biomass, which, compared to CLM3.5, are expected to be more reliable, given the implemented model improvements. Compared to CLM3.5-ParFlow, a significant advantage is that less compute time is needed, and less time has to be spent on model setups.

2.1.2 HydroGeoSphere

(The text below is mainly identical to the text of the HydroGeoSphere model code description in Deliverables D3.1, D6.2 and D7.2)

HydroGeoSphere (HGS) (Brunner und Simmons 2012; Aquanty 2020; Schilling et al. 2019) is a physically based and fully distributed integrated hydrological model that has been successfully applied in many different hydrogeological contexts and at many different spatial and temporal scales. HGS can explicitly simulate the interactions between groundwater, surface water and vegetation under consideration of variably saturated subsurface flow and complex heterogeneous subsurface properties (e.g., Schilling et al. 2017; Schilling et al. 2014; Ala-aho et al. 2017; Tang et al. 2018. HGS can also consider fully explicit contaminant or nutrient transport and irrigation and tile drainage in agricultural contexts (e.g., Schepper et al. 2017; Bonton et al. 2012). HGS has been coupled to the Weather Research and Forecast (WRF) model for the integrated simulation of the atmosphere, surface, and subsurface interactions (Davison et al. 2015) and has been used for data assimilation experiments using the ensemble Kalman filter (EnKF) (Kurtz et al. 2017; Tang et al. 2018; Tang et al. 2017).

In HGS, the surface and subsurface are represented by two domains, the overland domain and the porous medium domain. Surface water flow, i.e., flow within the overland domain, is represented with the following diffusion-wave approximation of the two-dimensional Saint-Venant equation:

$$\frac{\partial \varphi_o h_o}{\partial t} = -\overline{\nabla} \cdot d_o \mathbf{q}_\mathrm{o} - d_o \Gamma_\mathrm{ex} \pm Q_o$$

where $\overline{\nabla}$ is the two-dimensional differential operator, d_o [m] is the depth of surface water (excluding rill storage height that represents microtopography), φ_o [-] is the surface flow equivalent porosity that accounts for microtopography, h_o [m] is the total head ($\equiv\ z + d_o$) for given water depth d_o and elevation z, $\mathbf{q}_\mathrm{o}$ [m/d] is the average surface water flow velocity, Γ_ex [m/d] is the volumetric fluid exchange rate between the surface and subsurface domains per unit surface area (positive when water flows from the surface to the subsurface), and Q_o [(m^3/d)/m^3] represents sources and sinks (volumetric flux per unit surface area). The surface flow equivalent porosity ranges between 0 and 1, depending on whether the surface water depth is below or above the microtopography.

The average surface water flow velocity $\mathbf{q}_\mathrm{o}$ is given by:

$$\mathbf{q}_\mathrm{o} = -\mathbf{K}_\mathrm{o} \cdot k_{ro} \nabla(h_o)$$

where k_{ro} is a dimensionless factor accounting for obstructed flow and microtopography, and $\mathbf{K}_\mathrm{o}$ [m/d] is the surface conductance solved using Manning's equation.

Irrigation is simulated in HGS by aligning the numerical grid during mesh generation with irrigation infrastructure, e.g., drip irrigation piping, and by subsequently specifying representative irrigation water fluxes [m^3/d] via a second-type (Neumann) boundary condition at discrete model nodes located within the surface domain.

Variably-saturated groundwater flow in HGS is simulated using Richards' equation:

$$\frac{\partial}{\partial t}(\theta_s S_w) = -\nabla \cdot \mathbf{q} + \Gamma_\mathrm{ex} \pm Q_o$$

where θ_s [-] is the saturated water content, S_w [-] is the water saturation, $\mathbf{q}$ [m/d] is the groundwater flux (i.e. Darcy flux), and Q_o [(m^3/d)/m^3] represents sinks and sources (volumetric flux per unit volume).

The groundwater flux $\mathbf{q}$ is given by:

$$\mathbf{q} = -k_r(S_w)\mathbf{K} \cdot \nabla(\psi_w + z)$$

where $k_r(S_w)$ [-] is the relative permeability of the medium, $\mathbf{K}$ [m/d] is the saturated hydraulic conductivity tensor of the porous medium, and ψ_w and z [m] are the pressure and the elevation head, respectively. Surface water flow and groundwater flow equations (i.e., the overland and the porous medium domains) are fully coupled with the dual-node approach (Rooij 2017) and are solved simultaneously without requiring iteration. The relationship between the relative permeability of the porous medium, the soil water content and pressure can be given in tabular form or can be parametrised using the Van Genuchten functions (van Genuchten 1980).

According to the approach of van Genuchten (1980), the saturation S_w is related to the matric suction ψ and the relative permeability k_r by:

$$S_w = \begin{cases} S_{wr} + (1 - S_{wr})\left[1 + |\alpha\psi|^{\beta}\right]^{-\upsilon}, & \psi < 0 \\ 1, & \psi \geq 0 \end{cases}$$

$$k_r(\psi) = S_e{}^{(l_p)}[1 - \left(1 - S_e{}^{1/\upsilon}\right)^{\upsilon}]^2$$

$$S_e = (S_w - S_{wr})/(1 - S_{wr})$$

where S_{wr} [-] is the residual saturation, α [L^{-1}] and υ [-] are the van Genuchten parameters, υ is given as $1 - {}^{1}/_{\beta}$ with $\beta > 1$, S_e [-] is the effective saturation and l_p [-] is the pore-connectivity parameter (which is 0.5 for the van Genuchten model).

Evapotranspiration is modeled as a combination of evaporation and transpiration, affecting both the surface and the subsurface. Transpiration T_p [LT^{-1}] is simulated based on the implementation of (Kristensen und Jensen 1975):

$$T_p = f_1(LAI)\, f_2(\theta)\, RDF\, \left[E_{pot} - E_{canopy}\right]$$

$$f_1(LAI) = \max\{0, \min[1, (\mathrm{C}_2 + \mathrm{C}_1 LAI)]\}$$

where *LAI* [-] is the leaf area index, θ [-] the soil moisture content, *RDF* [-] the root decay function, E_{pot} [LT^{-1}] the potential evapotranspiration, E_{canopy} [LT^{-1}] interception and canopy evaporation, and C_1 [-] and C_2 [-] are coefficients which express the relation of transpiration on LAI. C_1 allows accounting for transpiration limiting vegetation characteristics (e.g., height, development stage, age of vegetation, degradation) and C_2 for transpiration from vegetation for which LAI cannot be defined. The *RDF* describes the decrease of root density with depth. $f_2(\theta)$ takes on values between zero and one according to:

$$f_2(\theta) = \begin{cases} 0, & 0 \leq \theta \leq \theta_{wp} \\ f_3, & \theta_{wp} \leq \theta \leq \theta_{fc} \\ 1, & \theta_{fc} \leq \theta \leq \theta_{ox} \\ f_4, & \theta_{ox} \leq \theta \leq \theta_{an} \\ 0, & \theta_{an} \leq \theta \end{cases}$$

$$f_3 = 1 - \left[\frac{\theta_{fc} - \theta}{\theta_{fc} - \theta_{wp}}\right]^{\mathrm{C}_3}$$

$$f_4 = 1 - \left[\frac{\theta_{an} - \theta}{\theta_{an} - \theta_{ox}}\right]^{\mathrm{C}_3}$$

Below the wilting point θ_{wp} [-], transpiration is zero, maximum transpiration is reached between the field capacity θ_{fc} [-] and the oxic limit θ_{ox} [-], and if the soil moisture content is above the anoxic limit θ_{an} [-], root stress is so high that transpiration is again 0 (Feddes et al. 2004). C_3 [-] is a fitting parameter with a recommended value of 1, making the ramping functions f_3 and f_4 linear (Feddes et al. 1978; Panday und Huyakorn 2004).

Agricultural drainage infrastructure such as tile drains can be efficiently simulated in HGS using so-called 1-D pipe elements, which circumvents the computationally intensive calculations that would be necessary if drainage flow were considered explicitly as part of the variably saturated subsurface porous medium flow domain. Instead, drainage flow in tile drains is simulated using the efficient one-dimensional Hazen-Williams equation (Aquanty 2020). This, however, requires that the numerical grid is aligned with drainage infrastructure already during mesh generation, such that the specification of the 1-D drainage network is possible via discrete model node selections.

2.2 Observations

2.2.1 In situ measurements at the site

For large-scale hydro(geo)logical models, typical data sources are databases from authorities and geological, soil or weather services. These data sets often have a spatial resolution of kilometres to be used at the global or country scale. These data sources are often of limited use for simulations on the plot and catchment scale as done in WATERAGRI. Therefore, we also use in-situ measurements from sensors and weather stations installed on the WATERAGRI case study sites. We also consider feedback from individual stakeholders on, for example, crop rotation. Typical in-situ measurements are groundwater level, soil moisture, meteorological states (e.g., precipitation, air temperature, etc.), groundwater and soil temperature, and other hydrological flow and flux measurements.

Soil moisture is typically measured with sensors installed directly in the soil at different depths but can also be measured with non-invasive methods, such as Cosmic Ray Neutron Sensors (CRNS) (Fersch et al. 2020). CRNS determines volume-integrated soil moisture estimates. The horizontal footprint is 130 to 240 m depending on air humidity, and the vertical footprint is tens of centimetres (upper root zone) (Köhli et al. 2015). Multi-parameter probes are typically used to measure groundwater heads, temperature, salinity, and oxygen (Redox potential). Sensors and weather stations can be equipped with, for example, narrow-band IoT (Internet of Things) modems to transfer measurements directly to a database (e.g., Bogena et al. 2018). Measurements collected in a database are then available for calibration or data assimilation in near real-time and can be accessed via machine-specific protocols that are not widely available.

2.2.2 Weather data

Physically based models are driven by atmospheric forcings. Atmospheric forcings, in combination with data assimilation, are perturbed in situ measurements (past) or an ensemble of short to medium-range weather forecasts (future). Ensembles of weather forecasts are, for example, provided by the German weather service or twice per day (2 x 50 ensemble members) by the ECMWF for the next two weeks. Typically, these weather forecasts have higher reliability for the first 5 days but have some positive skill for a longer period. Many discussions with stakeholders, e.g., in the context of

WATERAGRI stakeholder engagement workshop #3, showed an apparent demand for seasonal weather forecasts or at least weather forecasts for more than 14 days.

In this context, the WATERAGI partner QuantClim (ULUND) acquired historical precipitation and temperature data for the field sites in Finland (Siikajoki Ruuki), Sweden (Gårdstånga Nygård), Poland (Wroclaw), and Germany (Julich) for 1981-2020 to evaluate the performance of the SEAS5 seasonal forecasts for forecast steps of 1-3 months. This, and a discussion on the benefits of seasonal weather forecasts for farmers, was mainly part of WATERAGRI WP2 and its related deliverables. However, the data developed by QuantClim is also potentially interesting to extend the forecast period of some of the physically based models of WATERAGRI, like, for example, the model developed in Finland. Therefore, we summarise some details here.

Seasonal forecasts by QuantCLim for WATERAGRI:

The SEAS5 forecast is ECMWF's fifth generation of seasonal forecast systems and is based on numerical models that solve hydrodynamic equations for processes in the atmosphere and ocean, consisting of 51 ensemble members (see Johnson et al. 2019 for full description). To derive a single value forecast from the 51 ensemble members for each forecast step, the Expert Advice Algorithm was used (Cheng und AghaKouchak 2015). This algorithm uses a pool of expert estimates, in this case, the forecast ensemble members, to arrive at a single estimate for each iteration. The mathematical property of the algorithm ensures that the single estimate is at least as good as the best expert estimate at each iteration.

The evaluation results found that while the root mean squared errors (RMSE) were low for temperature forecasts, the forecasts' capacity to accurately predict monthly variations were low for both precipitation and temperature (WP2). Due to the stability of the temperature variation over time, the forecasts can still generate low RMSE, making the forecasts useful for seasonal agricultural planning. On the other hand, precipitation forecasts displayed too large RMSE for the evaluated field sites to be useful for seasonal agricultural planning. For full evaluation results, see D2.2.

2.3 Data Assimilation

Uncertain initial and boundary conditions, as well as input parameters and their spatial distribution, can limit the reliability of numerical models. Within WATERAGRI, data assimilation techniques are used in addition to classical calibration methods to quantify and reduce model uncertainties. We use the ensemble Kalman filter, a sequential method where nonlinear dynamical systems are stochastically approximated using Monte Carlo methods (Burgers et al. 1998; Reichle 2008). Here, an ensemble of model runs, which approximate the model uncertainty, is run and used to estimate the model error covariance matrix, which is essential in the data assimilation procedure to weigh, on the one hand, the model predictions and, on the other hand, the measurements. Deviations between measurements and model predictions are used to update the model predictions, and the correcting influence of the measurements depends on the relative weights assigned to the model prediction and measurements, as calculated by the data assimilation algorithm.

2.3.1 Data Assimilation with Terrestrial System Modelling Platform: TSMP(CLM)-PDAF

TSMP, just like CLM5, was coupled with the generic data assimilation library "Parallel Data Assimilation Framework" (PDAF) (Kurtz et al. 2016; Nerger und Hiller 2013; Strebel et al. 2022). TSMP-PDAF or CLM5-PDAF allows correcting model simulations with measurements to update initial conditions and parameters. TSMP-PDAF coupled by Kurtz et al. (2016) uses a multiple program multiple data approach (MPMD), i.e., the Ocean Atmosphere Sea Ice Coupler – Model Coupling Toolkit (OASIS-MCT)). In contrast, Strebel et al. (2022) developed protocols to couple CLM5-PDAF by using a single executable built out of modified, pseudo-library versions avoiding the use of OASIS-MCT as well as I/O intensive model re-initialisations. Strebel et al. (2022) did not change the PDAF core functions from Nerger et al. (2005) but provided minor modification to the TSMP-PDAF driver, the PDAF user functions (Kurtz et al. 2016) and the CLM5 code (“libclm5, Lawrence et al. 2019) as well as significant modifications to the TSMP-wrapper. The TSMP-wrapper contains the model-specific routines for initialisation, time stepping and clean-up, and managing the PDAF state vector. All model data is kept in the main memory. Strebel et al. (2022) developed a subroutine to set the state vector and to call the data assimilation method to update the model variables contained in the state vector at time steps when new measurements are available, i.e., CLM5-PDAF uses stop alarms to allow for automatically correcting simulations with observations at specific time steps. CLM5-PDAF uses the ensemble mode of CLM5, where every ensemble member has name lists with identifier suffixes, i.e., every ensemble member consists of its own name lists assessing individual data sets, e.g., perturbed atmospheric forcings (i.e., multiple realisations). In parallel, parameters such as sand and clay content can be additionally disturbed and considered in data assimilation runs with and without parameter updating. For WATERAGRI, we use CLM5-PDAF because, as described earlier in this report and D6.2, we can provide faster results at the plot scale, which stakeholders clearly demand. We merge soil moisture simulations and observations, e.g., for Selhausen, to provide reliable soil moisture forecasts for the next 14 days.

2.3.2 Data Assimilation with HydroGeoSphere: HGS-PDAF

A new data assimilation framework is being developed for HGS. The data assimilation module is based on the parallel data assimilation framework (PDAF, Nerger und Hiller 2013), which allows state-of-the-art data assimilation algorithms. The newly developed HGS-PDAF system consists of four modules: the HGS executables, the PDAF source code, the model bindings, which are a set of user-specified subroutines that act as an interface to the PDAF source code, and the 'driver' to manage the data assimilation simulation runs. The driver can be written as bash or python scripts. PDAF allows the application of several types of observations that can be assimilated together or separately. This is achieved by implementing the observation modules for each observation type separately.

Similarly, the combination of variables in the state vector can be flexible. So far, the joint assimilation of several observations, such as piezometric head and soil moisture, is possible with the current version of the developed data assimilation system. The state vector can include the model simulated states (e.g., hydraulic heads and water saturation) and the model parameters (e.g., hydraulic conductivity).

3 Quasi-operational application of Data Assimilation System and discussion

3.1 Germany-Selhausen (T5.5): Soil moisture forecasting in near-real time using TSMP(CLM5)-PDAF

3.1.1 Site description

The description for the Selhausen observatory is mainly identical to the text of the official TERENO description (website) and the descriptions provided in Deliverables D7.2 and D5.2.

The German case study site Selhausen (T5.5) is part of the TERENO Rur Hydrological Observatory (Bogena et al. 2018) and represents a heterogeneous agricultural, rural area located in the lower Rhine valley (Figure 1). The mean annual temperature is 10 °C, the annual precipitation is 700 mm from 1961 to 2014, and the climate is temperate maritime. The most important crops are sugar beet, winter wheat, winter barley, maise and rapeseed. Quaternary fluvial deposits build the subsurface at the Selhausen site and are covered by loess. The primary soil types are luvisols and gleyed cambisols, which can contain large gravel. The topography is flat with maximum slopes of 4° in the area of a former channel of the Rur River (local catchment).

The Selhausen site is intensively instrumented, allowing access and use of a broad range of data for data analysis and modelling activities (Bogena et al. 2018). The Selhausen test site is equipped with an Eddy Covariance station, measuring sensible heat, latent heat and carbon fluxes and meteorological variables such as the three-dimensional wind component, air temperature, air humidity, net radiation (incoming and outgoing short- and long-wave radiation), photosynthetic photon flux density and precipitation (Bogena et al. 2018). An operating wireless sensor network consisting of five profiles (-0.01 m, -0.05 m, -0.1m, -0.2 m, -0.5 m and -1 m) provides soil moisture and temperature measurements in near-real-time (Bogena et al. 2018). Soil heat flux is measured with a ground heat flux plate (Bogena et al. 2018). The sensors host narrow-band IoT modems (Wireless sensor network), allowing data transmission to a database and inclusion in a model in near real-time. In addition, a Cosmic Ray Neutron Sensor is installed at Selhausen, providing soil moisture area measurement with a radius of about 200 m. The measurement depth depends on soil moisture content and is typically between 12 cm and 70 cm depth.

In addition, the phenological development of crops and farming activities are recorded on a weekly to monthly basis. Typical groundwater information, such as the groundwater level, groundwater electrical conductivity, and groundwater temperature, is continuously measured using a multi-parameter probe installed next to the Eddy Covariance station (Bogena et al. 2018). CO_2 emissions from the soil have been continuously measured since 2015 using four automated closed dynamic chambers (Bogena et al. 2018). Data of the described measurements are available from the TERENO Data Discovery Portal (https://ddp.tereno.net/ddp/). An overview of the location of the Rur catchment and the Selhausen site is provided in Figure 1.

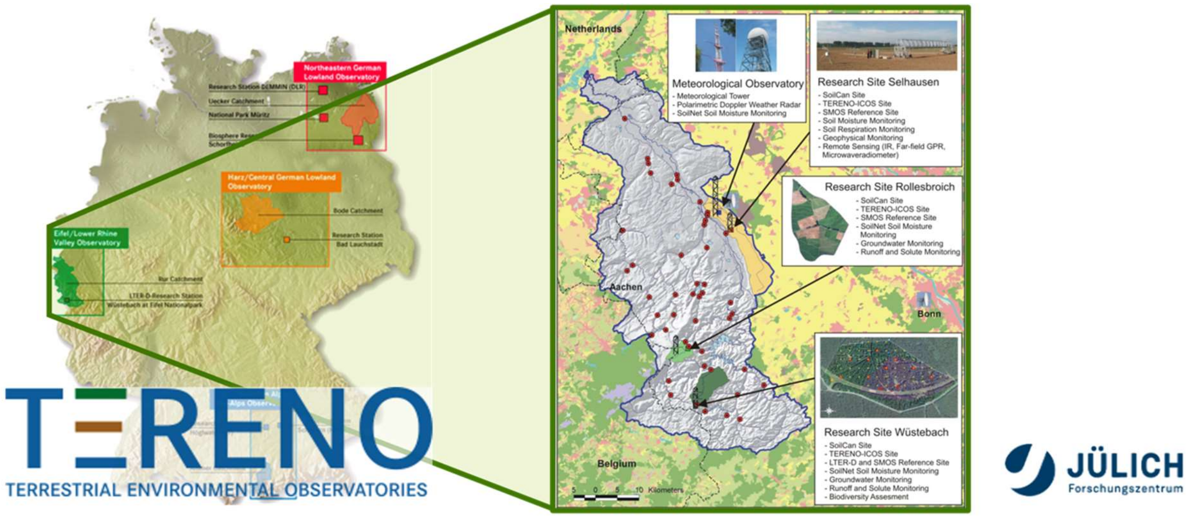


Figure 1: Overview of Terrestrial Environmental Observatories (TERENO) network in Germany (left) and close up of the Rur catchment observatories (right), of which the Selhausen case study site is a part of (image sources: TERENO & FZJ).

3.1.2 Model description

FZJ uses the open-source code "Community Land Model" version 5 (CLM5, Lawrence et al. 2019) for operational site-specific forecasts of relevant variables for agricultural practice (Figure 2). We couple observational data with physically based models in near-real time using the same plot-scale model for the Selhausen observatory presented in D6.2 (Figure 2). The model has a dimension of 100 m x 100 m x 40 m, considers 20 hydraulic active layers with vertically increasing thickness, and is constrained by site-specific conditions like soil texture data for different layers or land use and its change from season to season. We prognostically simulate water and energy transport in the unsaturated zone, crop growth and yield, snow depth and groundwater depth, vegetation states and changes in carbon and nitrogen pools after a 1000-year model spin-up. Such a long spin-up is needed to reach an equilibrium for the carbon and nitrogen pools. The criteria used here is that the total ecosystem carbon disequilibrium of the simulated land surface is less than 3 %. At least 97 % of the simulated surface of the Selhausen CLM5 model is in total ecosystem carbon equilibrium after 938 years. The spin-up run of 1000 years was about 6 hours on the machines of the Juelich Supercomputing Center (High-Performance Computing). Currently it is investigated whether the spin-up can be accelerated with help of machine learning methods to reduce the needed computing time.

In D6.2, we presented forecasts starting from an initial condition obtained from a simulation using as atmospheric forcings in-situ measurements from 2011 onwards instead of atmospheric model-based data. For that, we developed a toolbox with generic Python scripts (Jupyter Notebooks, Figure 3). Documentation for working on Juelich Supercomputers (or Linux machines) and processing model input and output data is available to allow replication at other sites (Figure 4). The documentation and the toolbox will be online and available with the scientific publication, which is in preparation. This makes it possible to use CLM5 to create site-specific models for other sites in a short period of time (e.g., 1 day if data is available for modelling or stored in cloud storage).

In the context of D7.5, we now develop generic scripts for simulation experiments with soil moisture assimilation extending the toolbox. Note that the delay in the hiring process at FZJ was 14 months (D5.1). Therefore, we present only preliminary results using CLM5-PDAF developed by Strebel et al.

2022 (see also Chapter 2.2 in this deliverable). We started with the perturbation of in-situ precipitation, global radiation, and air temperature measurements to generate 50 ensemble members for runs with and without (i.e., open loop) data assimilation. We use a multiplicative normal distribution for the perturbation of precipitation and global radiation and an additive normal distribution for the perturbation of air temperature (Table 1), while we use in-situ measurements for air pressure, relative humidity, and wind speed. Future conditions are simulated using a meteorological forecast from the German weather service.

Table 1: Statistical properties used to perturb the atmospheric forcings for Selhausen site.

Atmospheric Forcings	***Perturbation***	***Mean***	***Standard deviation***
Precipitation	Multiplicative normal distribution	1.0	0.5
Global Radiation	Multiplicative normal distribution	1.0	0.3
Air Temperature	Additive normal distribution	0.0	1.0

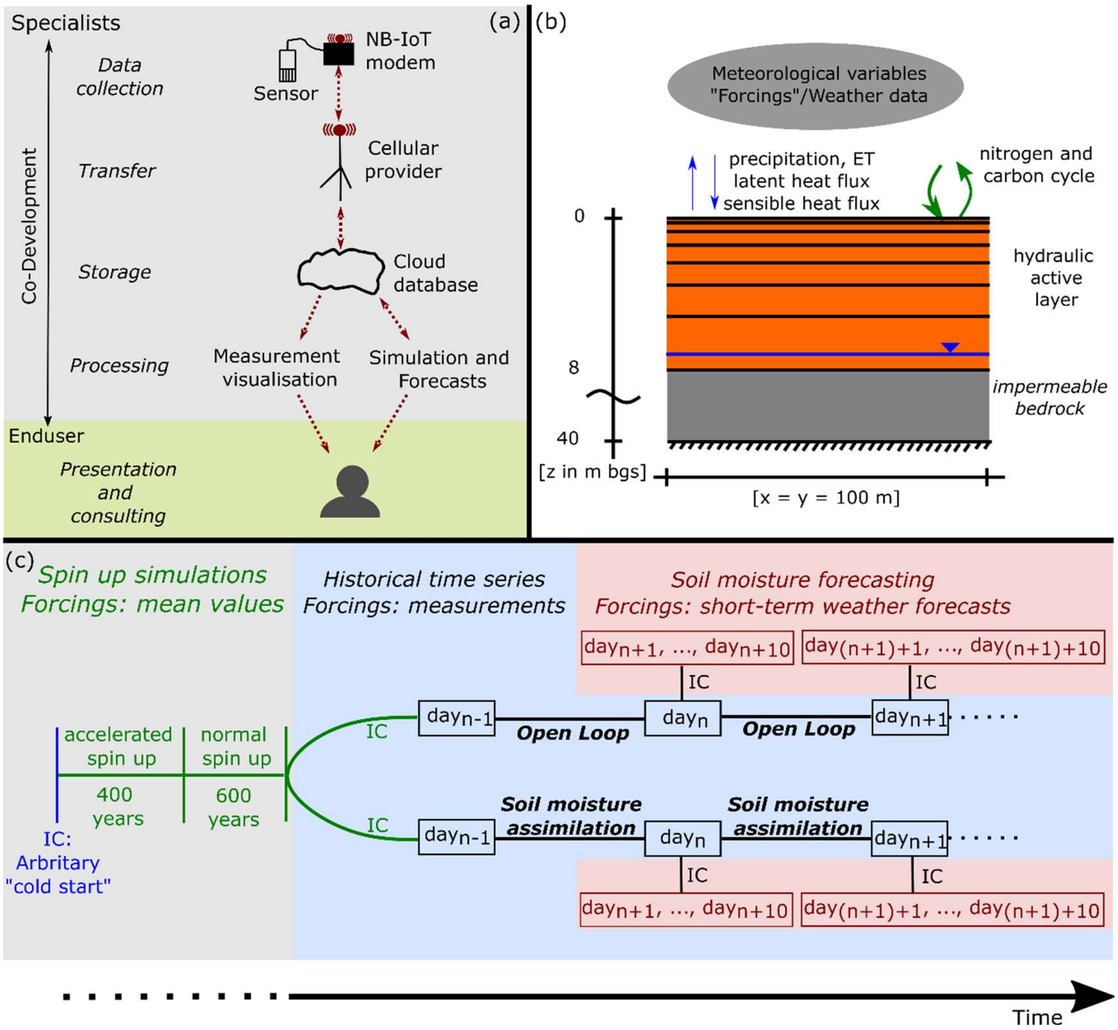


Figure 2: Automated data pipeline for operational site-specific soil moisture ensemble forecasts (Hoffmann et al., in prep. for GMD). (a) Scheme for data transmission from sensors to the end user; (b) Conceptual model for a plot-scale model of Selhausen; (c) Simulation routine for multiple ensemble members to provide daily forecasts of the hydrologic, crop, carbon, and nitrogen conditions and fluxes for the next 10 days.

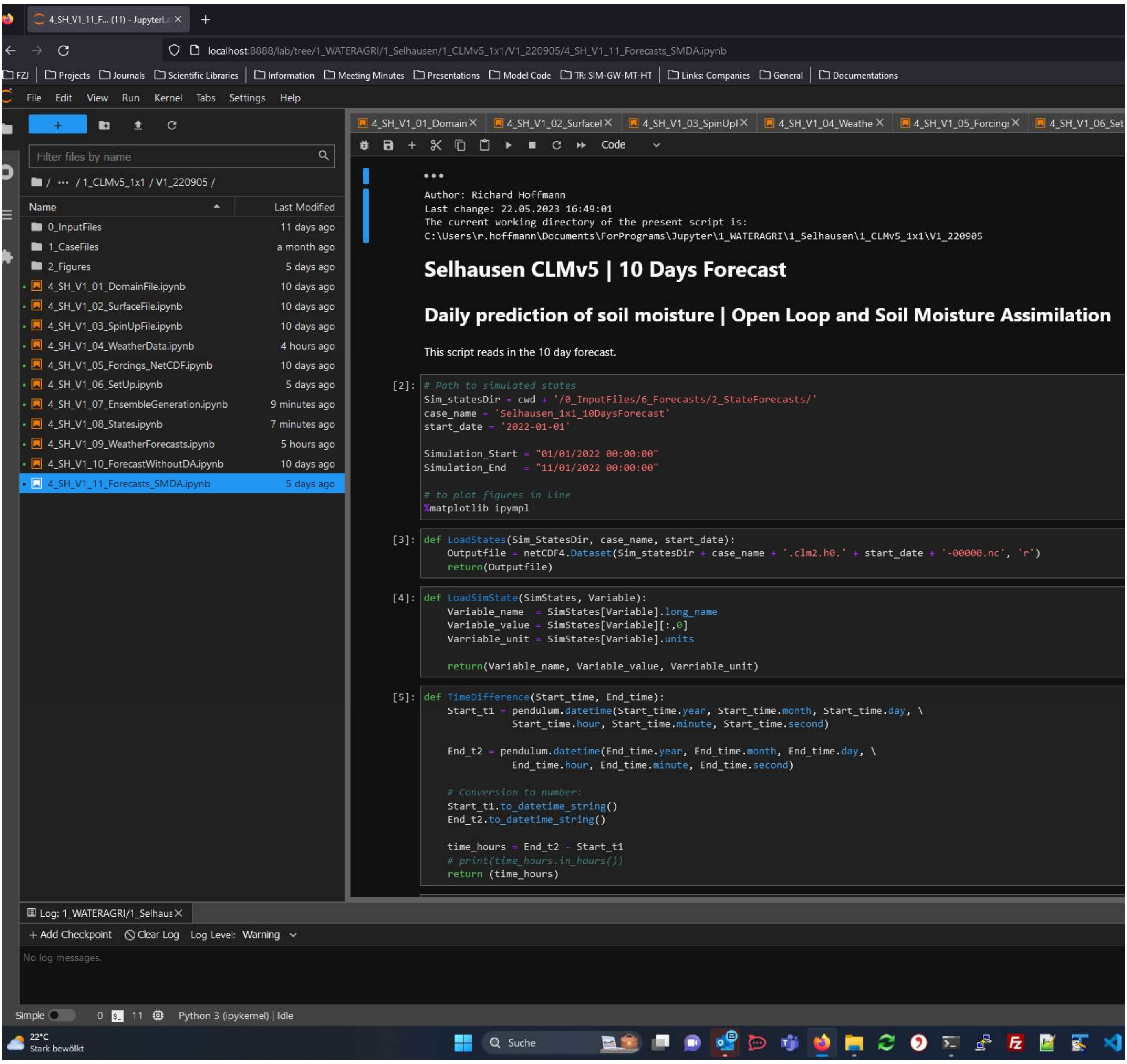
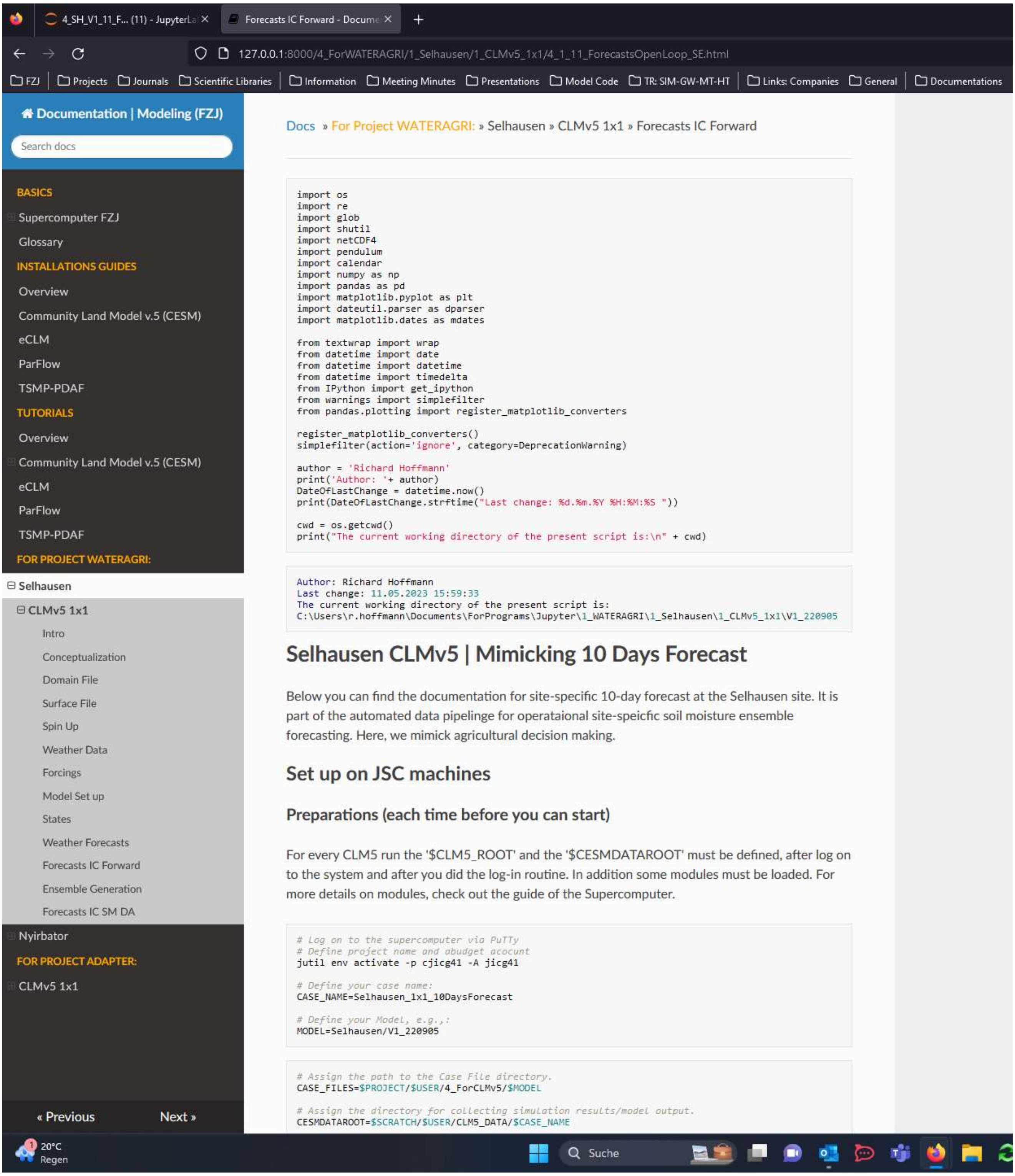


Figure 3: Screenshot of the toolbox under development for operational site-specific soil moisture ensemble forecasting using the "Community Land Model version 5" coupled with the parallel data assimilation framework (CLM-PDAF).



Docs » For Project WATERAGRI: » Selhausen » CLMv5 1x1 » Forecasts IC Forward

```
import os
import re
import glob
import shutil
import netCDF4
import pendulum
import calendar
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import dateutil.parser as dparser
import matplotlib.dates as mdates

from textwrap import wrap
from datetime import date
from datetime import datetime
from datetime import timedelta
from IPython import get_ipython
from warnings import simplefilter
from pandas.plotting import register_matplotlib_converters

register_matplotlib_converters()
simplefilter(action='ignore', category=DeprecationWarning)

author = 'Richard Hoffmann'
print('Author: '+ author)
DateOfLastChange = datetime.now()
print(DateOfLastChange.strftime("Last change: %d.%m.%Y %H:%M:%S "))

cwd = os.getcwd()
print("The current working directory of the present script is:\n" + cwd)
```

```
Author: Richard Hoffmann
Last change: 11.05.2023 15:59:33
The current working directory of the present script is:
C:\Users\r.hoffmann\Documents\ForPrograms\Jupyter\1_WATERAGRI\1_Selhausen\1_CLMv5_1x1\V1_220905
```

Selhausen CLMv5 | Mimicking 10 Days Forecast

Below you can find the documentation for site-specific 10-day forecast at the Selhausen site. It is part of the automated data pipelinge for operataional site-speicfic soil moisture ensemble forecasting. Here, we mimick agricultural decision making.

Set up on JSC machines

Preparations (each time before you can start)

For every CLM5 run the '$CLM5_ROOT' and the '$CESMDATAROOT' must be defined, after log on to the system and after you did the log-in routine. In addition some modules must be loaded. For more details on modules, check out the guide of the Supercomputer.

```
# Log on to the supercomputer via PuTTy
# Define project name and abudget acocunt
jutil env activate -p cjicg41 -A jicg41

# Define your case name:
CASE_NAME=Selhausen_1x1_10DaysForecast

# Define your Model, e.g.,:
MODEL=Selhausen/V1_220905
```

```
# Assign the path to the Case File directory.
CASE_FILES=$PROJECT/$USER/4_ForCLMv5/$MODEL

# Assign the directory for collecting simulation results/model output.
CESMDATAROOT=$SCRATCH/$USER/CLM5_DATA/$CASE_NAME
```

Figure 4: Screenshot of the documentation platform under development connected to the toolbox for operational site-specific soil moisture ensemble forecasts.

3.1.3 Soil moisture predictions

In D6.2, we already successfully demonstrated the forecast system starting from an initial condition obtained from a simulation run using as atmospheric forcings unperturbed in-situ meteorological measurements. Figure 5 shows preliminary time series from developments runs with and without soil moisture assimilation for the years 2016 to 2021, using soil moisture measurements from the cosmic neutron sensor installed at Selhausen. The root mean square errors are 0.049 cm^3/cm^3 and 0.035 cm^3/cm^3 for the open loop and soil moisture assimilation run, respectively. Simulations are closer to observations when including soil moisture assimilation (Figure 5a). Next, we will also do joint state and parameter updates (e.g., update sand and clay content) to further improve the simulations. Soil moisture assimilation improves simulations more significantly for summer months than for winter months and brings a less uncertain initial condition for the forecast model. The system could be used to provide every calendar day a forecast of soil moisture for the following days, using as atmospheric forcings, e.g., weather forecasts from the German weather service, as mimicked in Figure 5b.

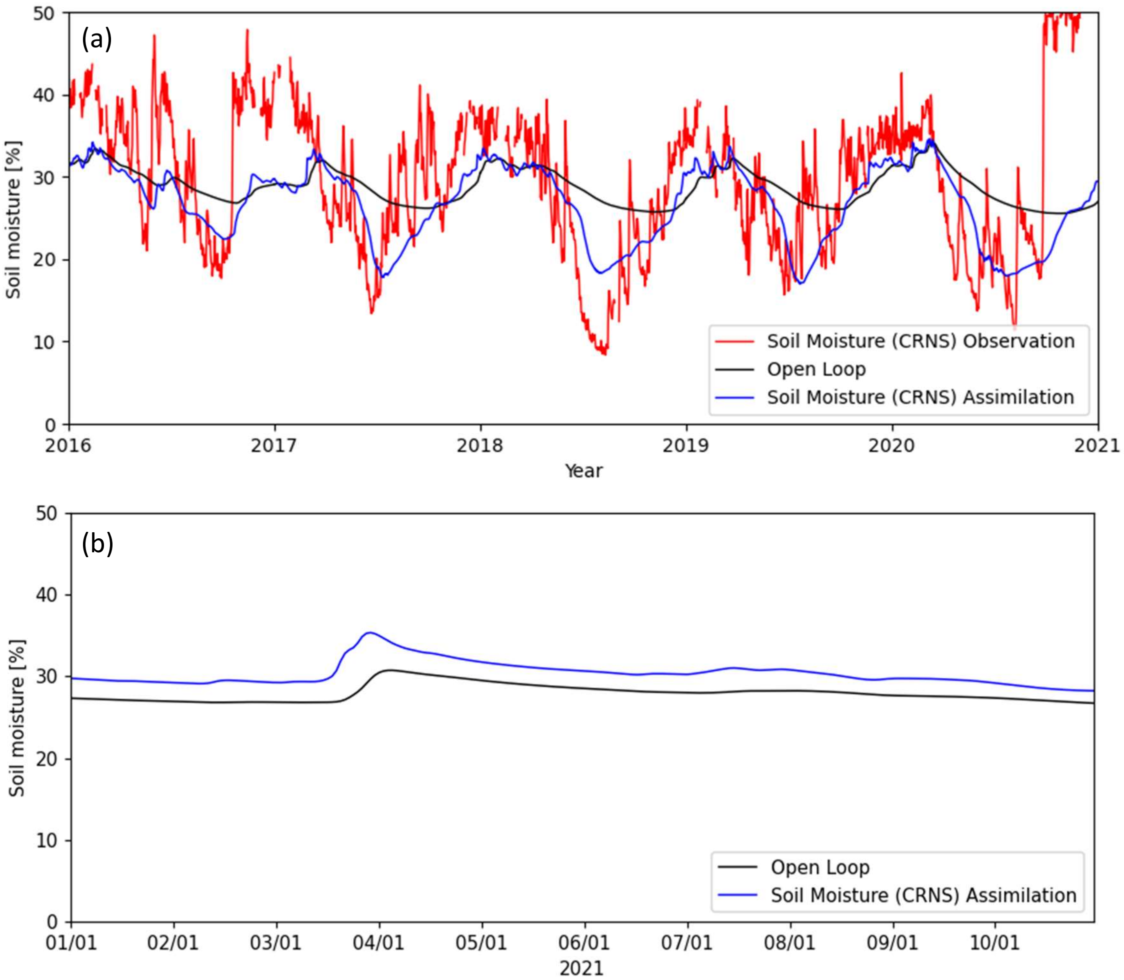


Figure 5: Preliminary time series from a development run mimicking the automated pipeline. (a) Preliminary comparison of the open loop and a development run with soil moisture assimilation, using soil moisture observations from the Cosmic Ray Neutron Sensor installed at Selhausen (RMSE for open loop: 0.049 cm^3/cm^3 and for soil moisture assimilation: 0.035 cm^3/cm^3). (b) Preliminary forecasts starting from the initial conditions obtained after open loop and soil moisture assimilation and using as atmospheric forcings weather forecasts from the German weather service.

Further refinements to the data assimilation system and model are currently being developed, especially using meteorological ensemble weather forecasts as atmospheric forcings and simulation experiments with the assimilation of further measurements besides soil moisture, such as leaf area index. Considering in situ measurements and local conditions provides less uncertain conditions for a predictive model and accurate local information, which cannot be derived from global maps having a too-coarse resolution. For example, operational site-specific soil moisture ensemble forecasts can contribute to quantitative, long-term water resources management in near real-time, e.g., to support management activities such as irrigation scheduling as an on-demand service. This may help a farmer plan irrigation and drainage and estimate crop development and yield trends. Potentially, site-specific models may also help with regional planning of water availability for irrigation and in quantifying a price indexation for agricultural water demand (i.e., “water pricing in agriculture”).

However, the toolbox updated with the new protocols will be released to allow trained stakeholders to set up the system on other sites. In parallel, FZJ started a bilateral exchange with a German insurance company to test the system in practice and to include potential feedback from a real-case scenario. It is in discussion to install sensors at a site of a farmer insured by this insurance company, and later, when data can be collected, the model developments of FZJ (toolbox) be tested in practice.

3.2 Switzerland-Seeland (T5.7): Groundwater level prediction in near-real time using HGS-PDAF

3.2.1 Site description

The text below is mainly identical to the text of the Seeland site description in Deliverables D5.1 and D6.2

The Seeland site (Figure 6) is Switzerland's largest vegetable farming area. It is characterised by very fertile peat soils, which are extensively used for farming. The Seeland has historically been subject to significant floods. Three major, regional-scale water correction projects carried out in the last 150 years reduced the flood risk and made the region arable. The corrections included the construction of several canals, an extensive drainage channel network to lower the groundwater table and the redirection of major waterways away from the heart of the Seeland, resulting in a much more predictable and manageable agricultural landscape ideal for vegetable farming. However, the lower water tables in the peat soils significantly increased the decomposition of organic matter, resulting in a loss of up to 1.6 meters of soil. The remaining fertile peat soils now represent a scarce resource in urgent need of protection. If current agricultural practices are continued, soil resources will be depleted in the next 10-15 years.

(a) (b)

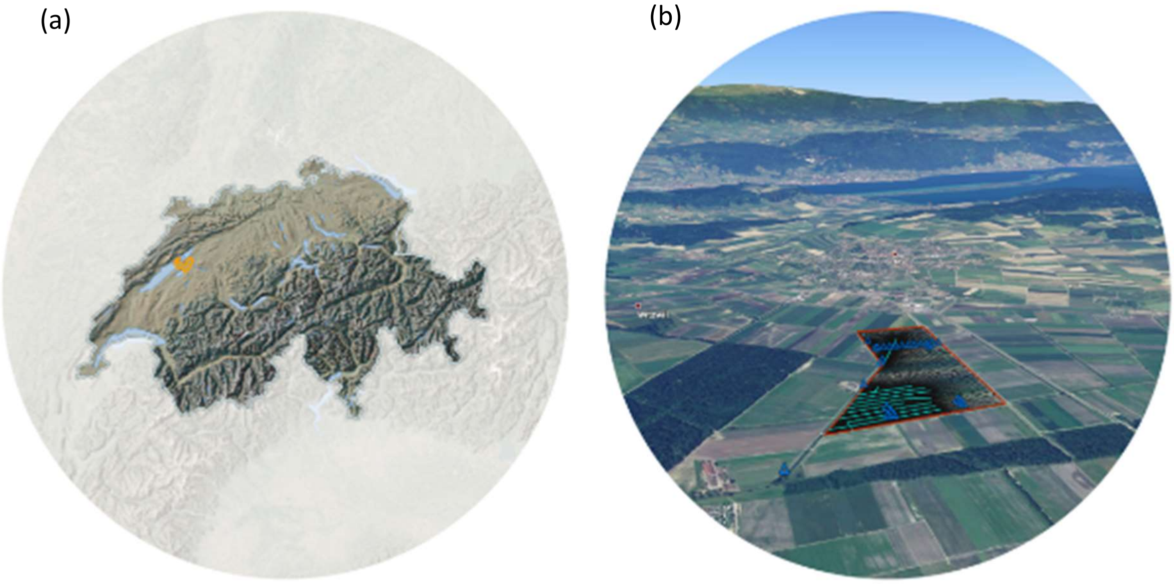


Figure 6: (a) Location of the agricultural area 'Seeland - Grosses Moos' within Switzerland, outlined in orange and (b) view of the study field (outlined in red). Image source: Google earth.

3.2.2 Model description

(The text below is mainly identical to the text of the model description in Deliverable D6.2)

A surface-subsurface model is being developed by HGS for the Seeland test site. The goal of the model is twofold: We use it as a test site to explore to what extent water management can be improved under consideration of agricultural efficiency and conservation of soil. Specifically, we can test to what extent the joint management of the operation of the drains and the regulation of the water level in the surrounding channels can be used to maximise efficiency. In addition, a parallel ensemble-based data assimilation system, HGS-PDAF, is being developed to facilitate real-time operational simulations of water quantity and quality.

Figure 7 shows the overview of the domain size and the topography of the model of the Seeland site. The model is discretised in a horizontal direction by triangles composed of nodes with inter-nodal spacing ranging from 5 to 15 m, resulting in 21 903 elements and 11 219 nodes per layer. In the vertical direction, the model is divided into 19 layers at depths of 0.15, 0.30, 0.45, 0.60, 0.75, 0.90, 1.05, 1.20, 1.50, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0 and 10.0 meters. The depth for the bottom layer is 15 meters. The model is constrained by three open channels simulated as head-dependent boundary conditions. Recently created 3-D subsurface maps of soil hydraulic properties are implemented in the model to significantly improve the representation of heterogeneity in the vadose zone.

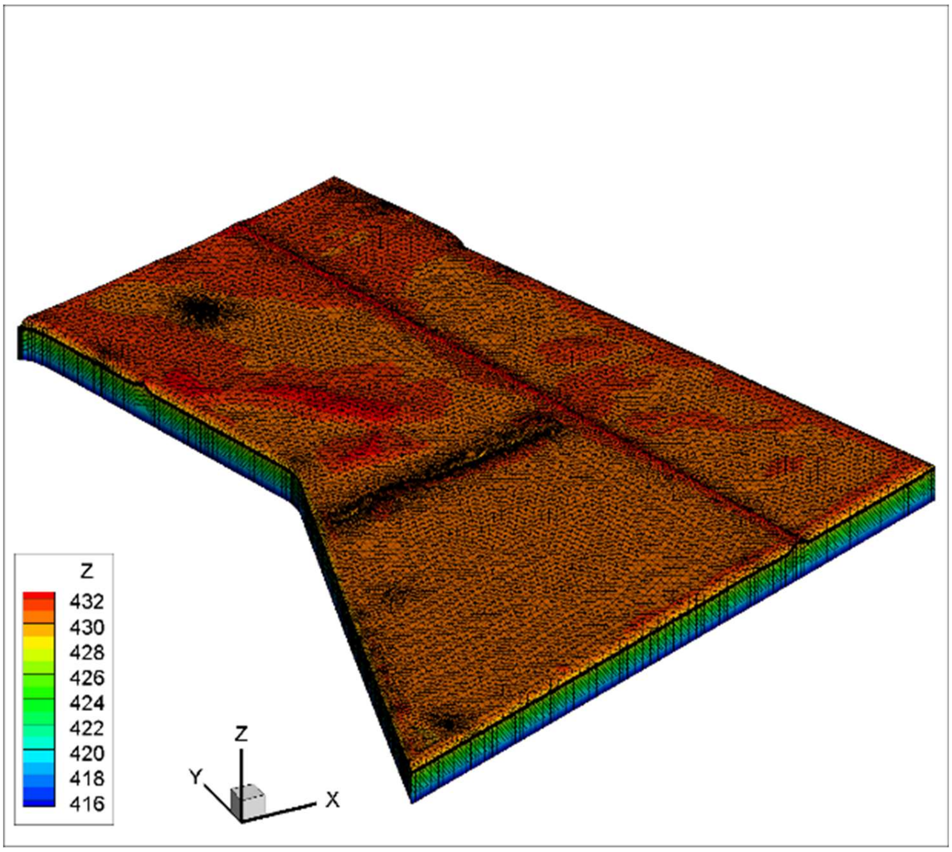


Figure 7: Numerical grid with high-resolution topography shading of the physically-based model of the Seeland case study site.

3.2.3 Prediction of Groundwater levels (Current state/First results)

So far, HGS-PDAF was tested with a synthetic model which simulates the surface-water and groundwater interactions. Such a model is ideally suited to develop and test the integration of DA to HGS. The conceptual model is designed as a 3-D rectangular synthetic alluvial plain model with a domain size of 500 m x 300 m x 30 m. In total, the model is discretised into 544,000 elements. A river is placed on top of the model layer. A detailed description of the model setup is given in (Delottier et al. 2022). Figure 8a shows the overview of the model. In reality, the heterogeneity of such systems is characterised by continuous, categorical structures that heavily influence observations that are assimilated. Therefore, in order to properly represent a realistic distribution of hydraulic properties, parameter fields like hydraulic conductivity (*K*) have been parameterised following these categorical features. See, e.g., Figure 8b and c, which provides two examples of the heterogeneous *K* fields used in the model. In total, 100 realisations are considered for such *K* fields for our DA experiments. Meanwhile, one more realisation of such *K* fields is generated and considered as the 'synthetic truth'. At eight observation points, synthetic hydraulic head measurements have been generated with this synthetic truth *K* field using the same boundary condition and parameter fields as for the ensemble runs. These observations are assimilated in a similar approach that is expected to be done in a real-world test case.

Two assimilation scenarios were carried out with different variables included in the state vector: Scenario 1: only model state, i.e. the simulated hydraulic head (denoted by 'DA_h') and Scenario 2: both the model state and parameters, i.e. the simulated hydraulic heads and K (denoted by 'DA_hk'). The state vector is updated by the covariance matrix using the ensemble Kalman filter (Evensen 2003). For each of these two scenarios, two types of initial *K* fields were considered. Besides the heterogeneous (denoted by 'hete_') fields, the homogeneous (denoted by 'homo_') fields are used for comparison purposes. Additionally, the free runs (i.e., without DA) were done for the two *K* fields, respectively. Results were evaluated by comparing the estimated heads with observations between the assimilation and free runs.

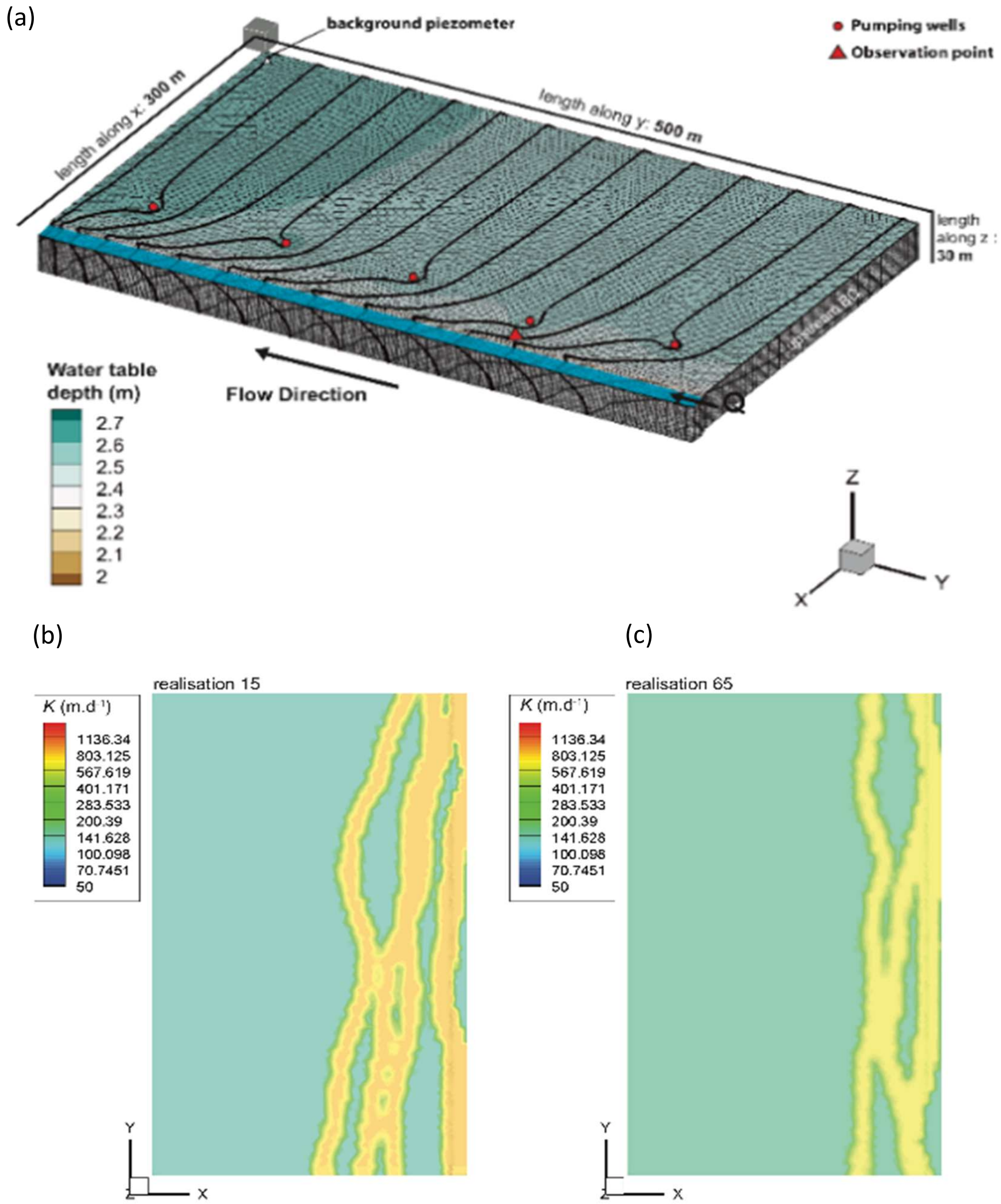


Figure 8: (a) Setup of the synthetic 3D model; (b, c) Two stochastic realisations of the initial K fields.

Figure 9 shows the average absolute difference between the simulated and observed heads calculated over the eight observation points. Errors (difference between the simulation and the observation) are higher for all simulation runs in the pumping period (first 15 days) than in the no-pumping period (after 15 days), especially for the free run with homogeneous parameter fields. With DA, the errors are reduced by 40 % for the homogeneous case in the pumping period. The improvement is even more significant for the heterogeneous case, with a reduction of up to 90 %. In the non-pumping period, no improvement is observed for the homogeneous case, while for the heterogeneous case, the improvement is 60 %. We have successfully demonstrated the DA system with synthetic models. However, the implementation to real-world systems still remains to be done.

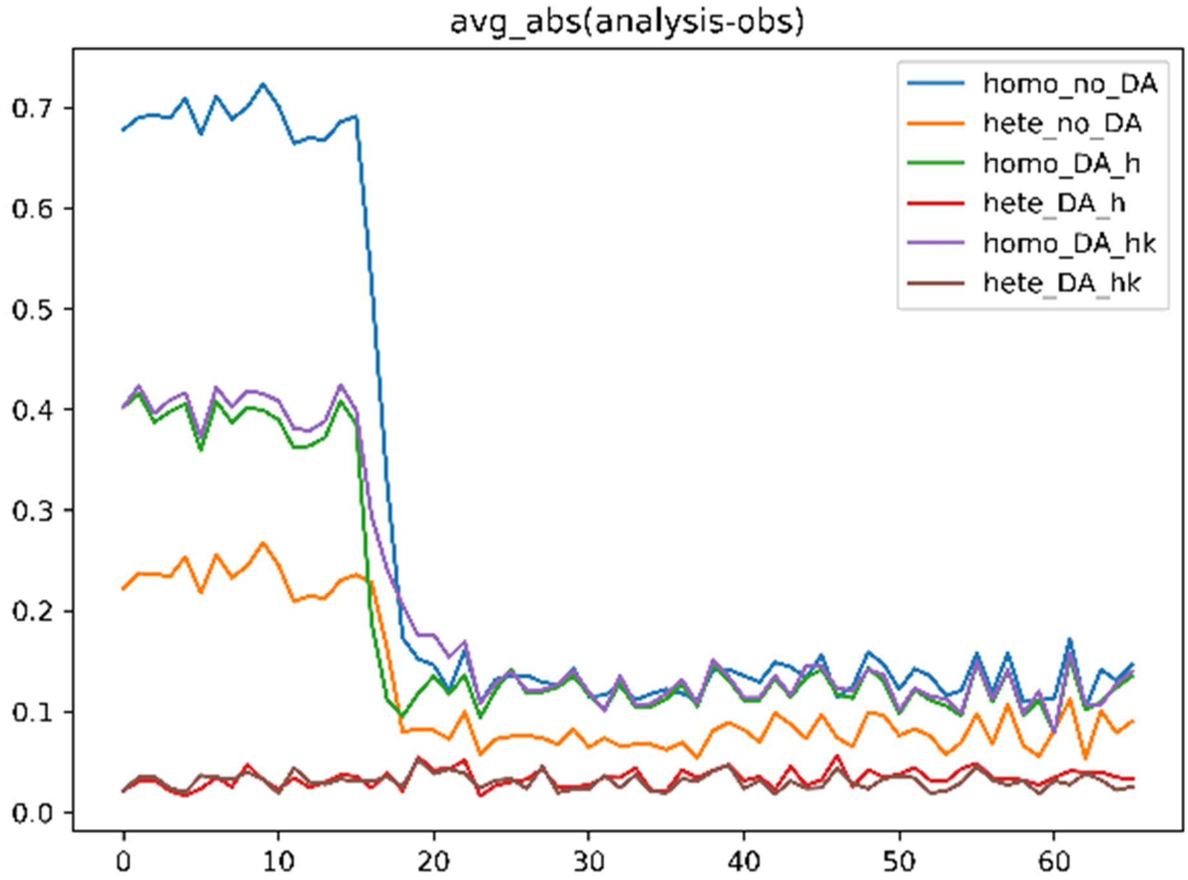


Figure 9: Average of the absolute difference of the hydraulic head between the simulation and the observation calculated over eight observation points. Unit for x-axis is time in days, and meters for y-axis. Note that the synthetic model was forced with transient boundary conditions to reproduce a controlled pumping test spanning 95 days. During the experiment, the maximum pumping has been turned off for 50 days after the first 15 days.

The developments at the Seeland site, i.e., coupling HGS with PDAF, enable more accurate simulations of surface and subsurface processes and, for example, better estimation of drainage volumes and groundwater recharge at the agricultural catchment scale. Accurate estimation of groundwater recharge is critical to irrigation planning in agricultural watersheds. Irrigation is the first measure of protection for sensitive crops worldwide, but irrigation is controversial because groundwater quality can be affected in irrigated areas. Accurate simulation of groundwater recharge, for example, can help size irrigation rates and regulate groundwater available for irrigation within an agricultural catchment. It may also be possible to make preliminary estimates of groundwater recharge in winter and its pattern in summer.

3.3 Poland-Lower Silesia (T5.6): Catchment scale efficiency of soil water retention solutions (HGS-PDAF)

3.3.1 Site description

The description for the Poland-Lower Silesia site is mainly identical to the text of the official TERENO description (website) and the descriptions provided in Deliverables D7.2 and D5.2

The Lower Silesia agricultural case study site is located in South-West Poland around Lubnów village, which is approximately 20 km North of Wrocław. In hydrological terms, the studied farm is located on the border of 2 different hydrological catchments. However, as 90 % of the area of the farm is located in the Ślęganina river catchment, only the Ślęganina river catchment was considered for the fully coupled and physically-based modelling experiments. The Ślęganina river is a tributary to the Odra river, which is Poland's second-largest river. The entire surface area of the catchment is 17.4 km^{2,} but for modelling purposes, the catchment was limited to the 14.6 km^2 to model the catchment outlet in the location of the installed limnigraph (roughly 500 m upstream from the joint of Ślęganina and the Odra rivers). According to the climate classification by Okołowicz (1977) the climate of the catchment is temperate warm transitional. The modelled Ślęganina river catchment and the currently implemented numerical grid are illustrated in Figure 10.

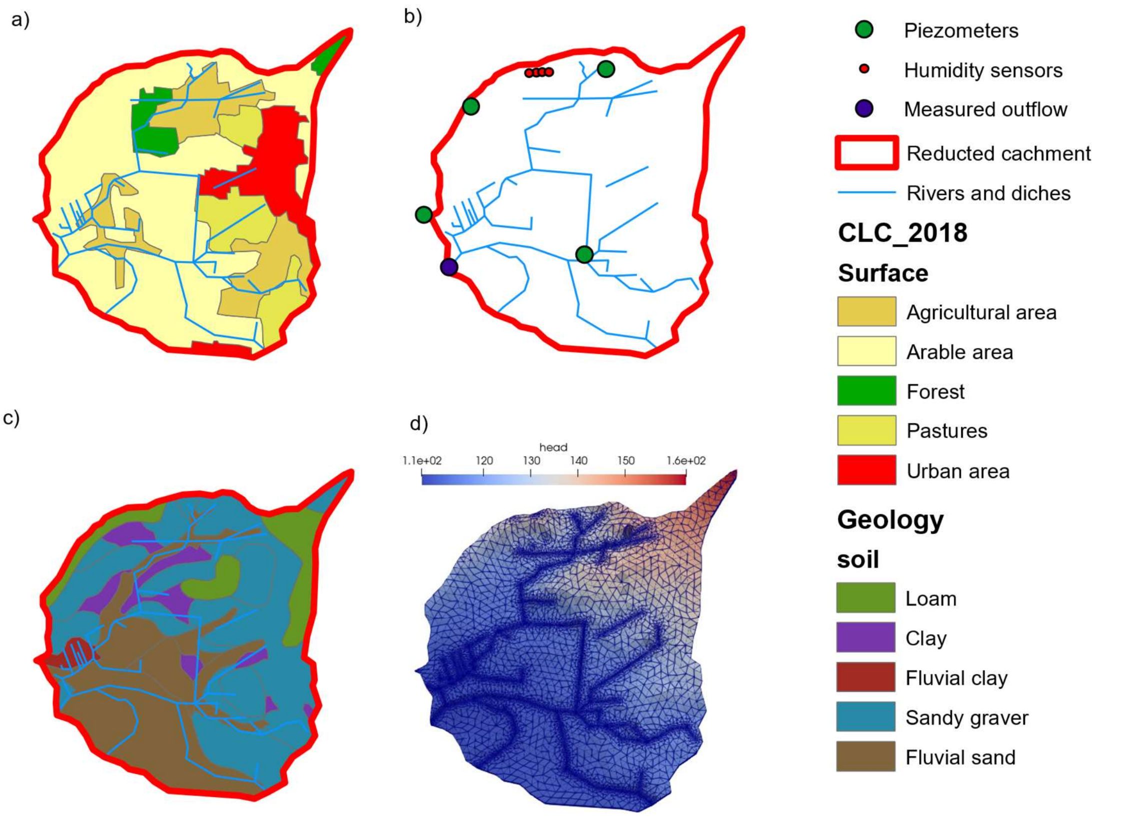


Figure 10: Maps of land cover (a), measurement network (b), soil types (c) and conceptual 3-D model of the catchment generated with HGS (d).

The mean annual precipitation at the study site, measured over 1991-2020, was 541 ± 95 mm. The mean air temperature over the same period was 9.7 ± 1 °C. With an average depth to groundwater, table of 1-2 m, a significant amount of groundwater resides in the shallow aquifer of the study site. According to the information obtained from 20 boreholes, the shallow aquifer is limited underneath by impermeable bedrock sitting at a depth of 2-5 m, depending on the location within the catchment. The catchment area is covered with topsoil consisting of loamy sand about 80 %, clay 5 % and silt 15 %. The soil is mainly composed of clay below a 3-3.5 m depth.

3.3.2 Model description

3d catchment model for an assessment of irrigation water demand under different climatic and agricultural land use scenarios (using HGS):

Catchment geometry is defined by 37027 elements (Figure 10d) generated in AlgoMesh, where the smallest triangle is 0.56 m^2 and the largest is 97.28 m^2. In the first version of the model the calibration was made with standard approach. Using geological maps 7 soil zones were defined of which 5 zones on the top (10 layers 0.1 m each) of porous media were identical to soil types (Figure 10c) while additional zones corresponded to shallow aquifer (3 layers approx. 0.5m each) and impermeable bedrock (approx. 3 layers 0.5m each). Calibration of the model was done for a 30 days period from 18.07 to 18.08.2021. Precipitation was defined as equally distributed based on on-site measurements. Evapotranspiration was calculated accordingly to FAO Penman–Monteith method and Leaf Area Index (LAI) was obtained from AGRICOLUS platform. To calibrate the model PEST++ software was used (Doherty 2015). Calibration was made against 3 piezometers (Figure 11) located in the upper, middle and lower part of the catchment. Root mean square error for PIEZO1, PIEZO3 and PIEZO4 was respectively 0.69, 0.28, 0.38. For all calibrated data correlation coefficient was equal 0.9989 and standard error of weight residuals varied from 0.8 to 1.3.

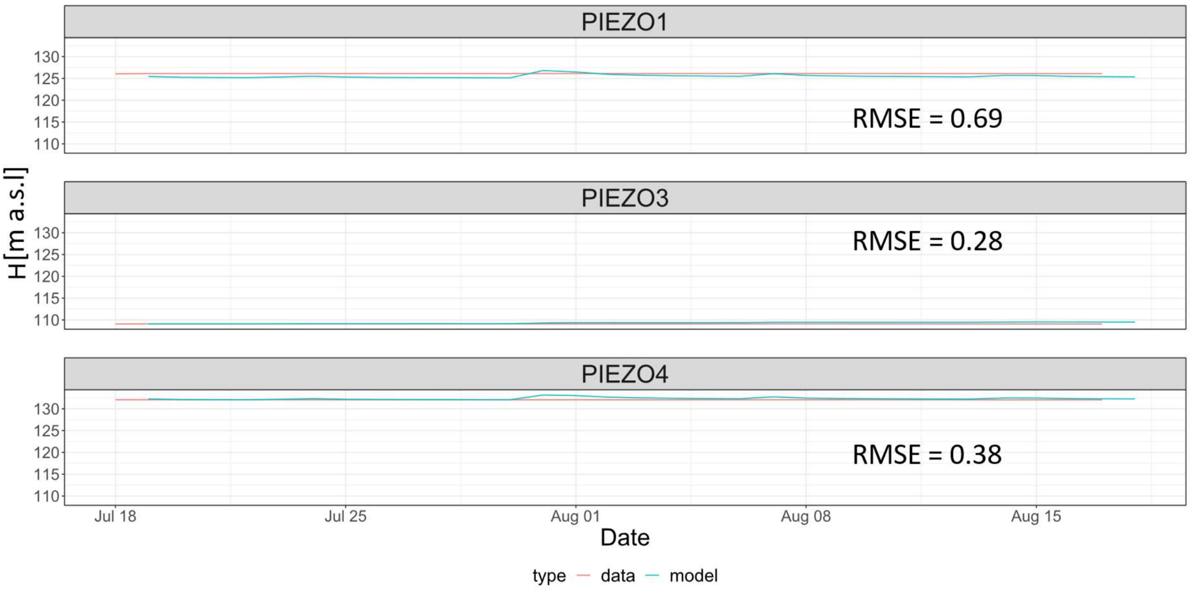


Figure 11: Preliminary comparison of the modeled and measured hydraulic head base on 3 piezometers (w located on a far north of the catchment and one on a far south Figure 10b)

For the second model number of layers was reduced to three to shorten the computation time. To obtain heterogeneous structure of soil type, hydraulic conductivity(k) and Van Genuchten parameters (alfa, beta) were calibrated with the SVD method (Moeck et al. 2015) based on pilot points (Figure 12). In addition, the hydraulic head will serve as initial conditions for selected areas of the catchment to run more precise simulation with particular regard to cultivating crops and assessing the volume of water needed for irrigation. SVD calibration should define different zones from presented at Figure 10c but it is expected to significantly lower uncertainty of the model. Setup of the second method of calibration is still under development and for this approach there are no preliminary results. Developed catchment model will serve, first of all, to identify the areas where irrigation is necessary and assess the volume of water needed to satisfy crop water requirements.

Both models run as physically based and fully distributed integrated surface-subsurface hydrological models (ISSHM).

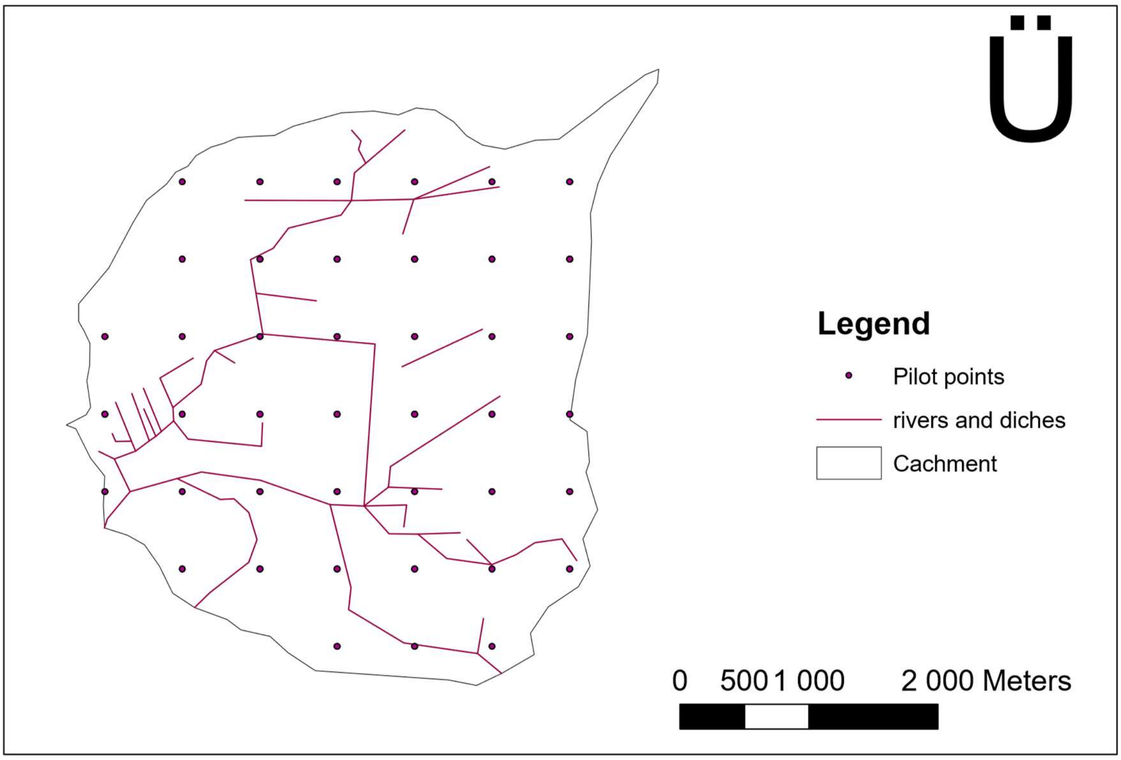


Figure 12: Pilot point distribution for SVD calibration with PEST ++.

3.3.3 The perspective of using data assimilation

UNINE is developing Data Assimilation (DA) HGS-PADF; while the DA system will be ready and the catchment model will be calibrated with PEST, UPWr plans to update groundwater level and soil water saturation through the DA system. Perhaps it is also possible to perform cases with parameter estimation, e.g., updating hydraulic conductivity.

3.4 Finland-Tyrnävä (T5.1): Efficiency of soil water retention in subsurface drains (HGS-PDAF)

3.4.1 Site description

The description for the Finland-Tyrnävä site is mainly identical to the text of the official TERENO description (website) and the descriptions provided in Deliverables D7.2 and D5.2)

The Tyrnävä study site (6.4 ha) is an agricultural field located in the municipality of Tyrnävä in the Northern Ostrobothnia region, Finland (Figure 13), used for industrial potato production. The Tyrnävä region is an important seed potato area for Europe and a significant national producer of potatoes for food and industrial uses. The Tyrnävä area is predominantly a low-laying plain with minor topographic relief dominated by agriculture on mineral soils and with a low occurrence of peatlands.

The field soil is classified as "löyhä karkea hieta" (Liedes et al. 2020), which is a term that refers to loose fine sands or coarse silts. The study site is located within the Muhos sedimentary rock formation of silicate siltstone and is characterised by thick unconsolidated Quarternary sediments (Johansson et al. 2005; Eskola und Lunkka 2022). Typical soil horizon consists of thick fine-graded sediments in topsoil (fine sand with silt and clay layers) deposited during and after the last deglaciation period and underlying more permeable sand deposits (Geological Survey of Finland 2017; Eskola und Lunkka 2022).

The climate of the Tyrnävä field is classified as boreal. The average annual temperature is 3.4 °C and precipitation of 582 mm for 2000-2020 (derived from daily 1km x 1 km gridded time series FMI_ClimGrid product; Aalto et al. 2016). The average annual reference evapotranspiration for 2015-2020 was around 480 mm (derived from daily 1 km x 1 km gridded reference evapotranspiration time series estimated by FAO Penman–Monteith method; Pirinen et al. 2022).

The field is surrounded by open ditches connected to the vast network of open ditch drainage covering the whole sub-catchment. The primary water management at the field is carried out by applying control tile drainage. The purpose of the controlled drainage is to control the amount of water in the field, thereby optimising soil water conditions for the potato crop. The system consists of two separate networks regulated by manual control wells (Figure 12, see also D5.2 Fig. 5) discharging to the open ditch at the west border of the field. Typically, at this particular field, the farmer keeps the control well gate open throughout the winter and closes it at the end of the snowmelt period before starting agricultural operations. The well is adjusted to a target height and only opened or lowered in case of forecasted heavy rainstorms. The control wells were predominantly adjusted to their maximum heights in recent years due to the relatively dry summers and corresponding low groundwater tables (GWT).

The study site hydrogeological monitoring consists of shallow groundwater observational pipes (AA1-AA3 and AV1-AV3 for years 2018-2019, GWp1-GWp4 for years 2022), ditch water level monitoring (two locations for the year 2022), water level observations in the tile drainage control wells (AA and AV for 2018-2019 and 2021), and soil water content (SWC) observations (AA1-AA3, AV1-AV3 for years 2019 and AA and AV for years 2021-2023).

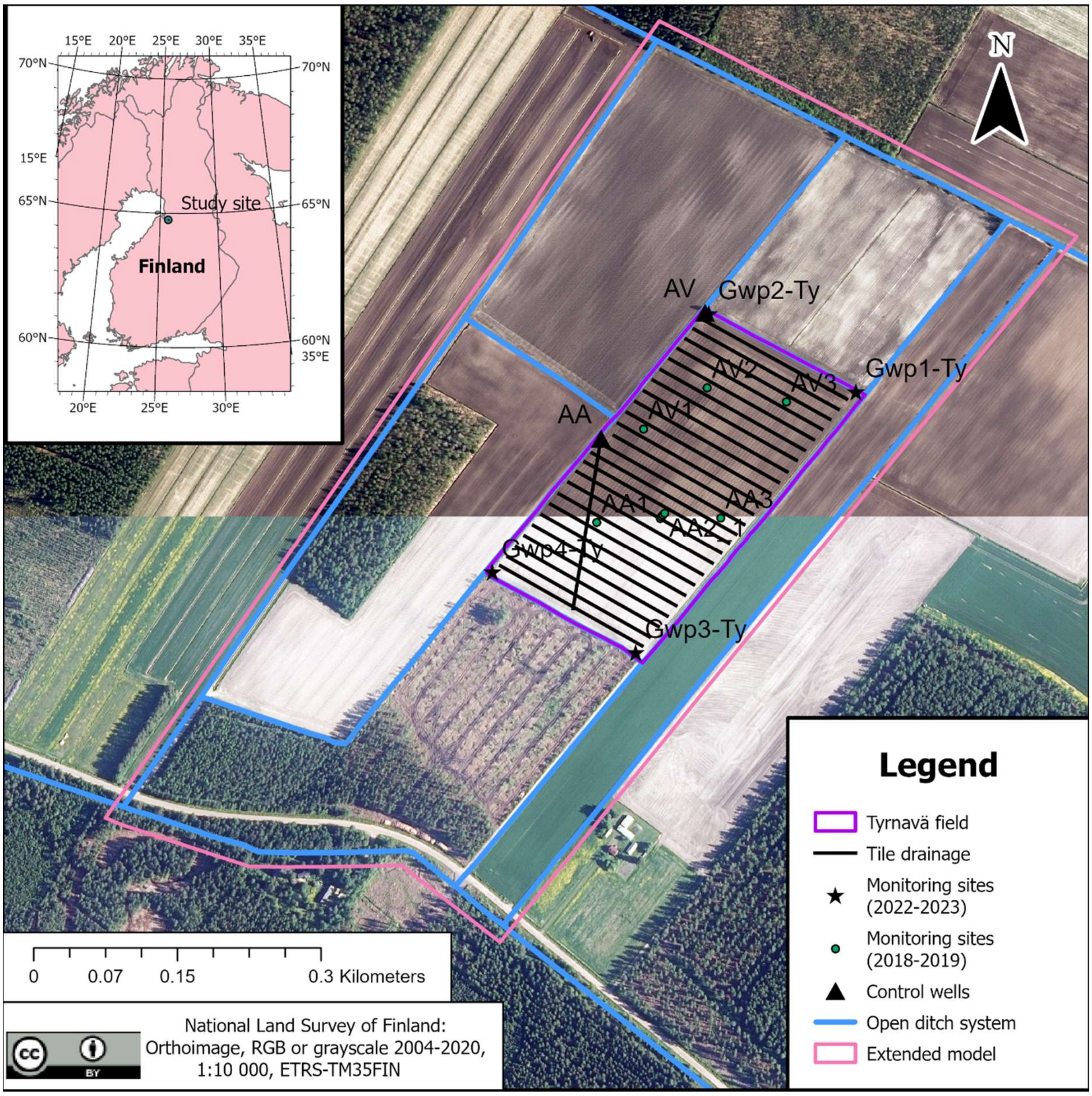


Figure 13: The study field and its surroundings. Two versions of the model were built for the WATERAGRI project: a field-scale model (violet outline) and an extended area (pink outline).

3.4.2 Model description

The base case model developed for WP6 Task 6.2 (D6.2, Numerical Modelling Assessment) was extended to decrease the effect of applied boundary conditions on the simulations and to allow including the GWp1-GWp4 GWT observations into the DA procedure. The two-dimensional triangular mesh for the extended model area (0.43 km^2) was built in AlgoMesh v 2.0.19.19384 and consists of 5214 elements and 2748 nodes. The mesh is more refined for the study field (average cell size 6.1 m^2) and coarser for the surrounding areas (average cell size 9.3 m^2). The mesh was designed so that the mesh nodes coincide with the location of monitoring sites, the open ditch network and the tile drainage network (Figure 14).

The extended model setup follows the field scale model. The subsurface was separated into five homogenous soil zones to represent a typical agricultural soil profile of the Tyrnävä region, including a relatively permeable furrow zone, less permeable subsoil, tightly compacted undersoil ("jankko"), deep subsoil, and sandy soil of the Muhos formation. The overland flow and ET domains consist of single homogenous zones representing potato crops, i.e., the forest patches and other crop types tapes are not included in the model for simplicity. No flow boundary conditions are assigned to the bottom of the model and lateral boundaries for the porous domain. For the overland domain, we used critical depth boundary conditions at the main ditch outlets (Figure 14), allowing water to leave the model domain freely. The applied parameter values and other detailed information on the model setup are presented in the model described in detail in D6.2, Numerical Modelling Assessment.

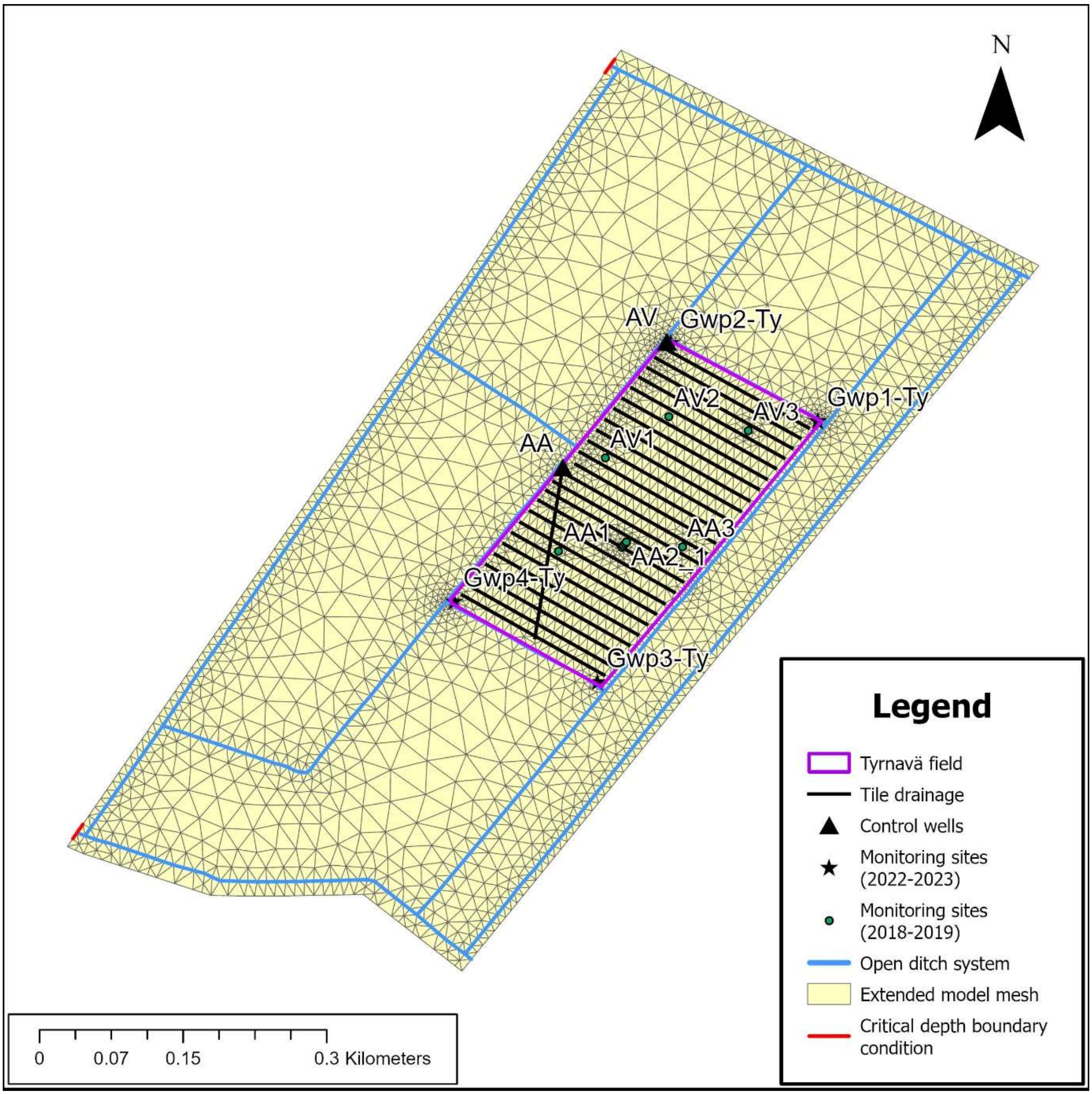


Figure 14: Two-dimensional triangular mesh of the extended model; the Tyrnävä study field is shown by violet outline.

3.4.3 The perspective of using data assimilation

The soil water content (SWC) is essential for potato crop production. Soil moisture is the predominant source of water for crop growth. Excessive or limited SWC is often one of the main factors affecting potato yields. Excessive SWC can favour the spread of diseases (Fiers et al. 2012) but also affects plants directly by causing plant drowning and, in extreme cases, die-offs. Limited water availability was found to correlate with field yield directly, and similarly, in extreme conditions, lead to plan die-offs (e.g., Martin et al. 1992). The SWC is partly controlled by groundwater table elevation in shallow groundwater areas. The groundwater provides water to the unsaturated zone through the capillary rise, and soil water retention properties define the SWC in the soil.

Whereas drainage control is a standard management solution in northern Europe and Finland, irrigation infrastructure is rare as irrigation is not needed annually. In order to not increase the costs of infrastructure, it was proposed to use control drainage systems as below-ground irrigation systems. In such an irrigation approach, water is introduced to the control well, and tile drainage piping allows water distribution to the whole field. Based on SWC, farmers can decide on possible drainage/irrigation operations. However, the best benefits can be expected when SWC is not only measured at a few locations but also extrapolated to the whole field area and forecasted for the short and the long term.

Fully-integrated hydrological modelling provides a realistic framework to predict SWC, but it is also associated with high uncertainties related to uncertain drivers (e.g., precipitation), parameters (e.g. hydraulic conductivity), model initial and boundary conditions and model structural errors. Data assimilation reduces the uncertainty of the estimated model states and parameters by merging observation with model predictions. Data assimilation has been proven to be an effective method to improve fully-integrated model predictions in synthetic case studies at various scales (Kurtz et al. 2016; Gebler et al. 2019; Hung et al. 2022; Zhang et al. 2018). In contrast, the DA's applications for 3D real-world experiments are rare and challenged by errors in the observed variables and model structural errors (Shi et al. 2015; Gebler et al. 2019).

The starting point for practical applications in agriculture is different than for highly instrumentalised research sites for which the previous DA experiments were conducted or mimicked in synthetic cases. The monitoring network investment costs will likely limit the amount of the monitoring infrastructure that can be applied, especially if the benefits of the investments are uncertain. For this reason, from a farmer's perspective, especially beneficial it would be to focus on what could be accomplished with the minimum infrastructural investment. In practice, this could mean the monitoring network consisting of 2-4 soil moisture IoT sensors installed at two locations and various depths and one observational groundwater pipe per field or using spatially distributed remote sensing data of limited resolution and temporal coverage.

In the quasi-operational application of the DA, we will attempt to address the following questions:

1. Can the characterisation of the SWC be improved at a field scale with DA of limited SWC observations in a homogenously set model?
2. Could joint assimilation of GWT (and model parameters, e.g., hydraulic conductivity) improve further SWC characterisation?

3. What benefits moderate and long-term weather forecasts can add to the DA? Can the forecasted SWC support water management operations at the site?

The DA experiments will be based on the HGS-PDAF framework developed by UNINE. The DA will be applied for the past summers of 2019 and 2021, for which most comprehensive observations of the SWC and GWT are currently available. The main focus will be assimilating SWC measurements from various depths and locations. GWT data assimilation will be attempted but challenged by missing a significant part of data due to prevailing dry conditions and corresponding low GWT. In addition, we plan to conduct real-time DA of the SWC if the HGS-PDAF framework will be available for other users in summer 2023 (crop collection in August/September). The IoT network at the Tyrnävä site will be extended to support real-time DA: 1) by replacing GWT and WT offline sensors with IoTs and 2) by installing additional sensors within the field area.

The state update in DA will be conducted daily for SWC and GWT data. The DA experiments will include at least 96 ensemble members, possibly more depending on the overall simulation times. For each member, the random perturbation will be assigned to forcing terms (historical precipitation and reference evapotranspiration), soil properties (hydraulic conductivity) and Leaf Area Index (derived from NDVI index). The experiment utilising forecast data will be based on the ensemble members of the System 5 forecasts (SEAS5) from the European Centre for Medium-Range Weather Forecasts (ECMWF) provided by "QuantClim" (ULUND) or alternatively using 15-day forecasts by Finnish Meteorological Institute.

3.5 Hungary-Nyírbátor (T5.10): Crop yield prediction and updating Irrigation schedules (CLM5-PDAF)

The site and model description for the Hungary-Nyírbátor site is mainly identical to the text of the official TERENO description (website) and the descriptions provided in Deliverables D7.2 and D5.2

3.5.1 Site description

The study area is situated in the Pannonian region with a continental climate, in the Northern Great Plain region in Szabolcs-Szatmár-Bereg county, next to Nyírbátor city (47°48'18.60"N, 22° 9'43.89"E) (Figure 15), which has a moderately cool and dry climate in Hungary. The case study site is in a nitrate-sensitive area (based on European guidelines) and owned by a private company: Bátortrade Ltd. The case study site comprises 16 ha of pasture with sprinkler irrigation (on fixed hydrants) for cattle grazing and 50 ha of irrigated arable land with a lateral moving irrigation system for feedstocks. The case study site is situated at an alluvial cone plain covered mainly with sand. The area's current rivers and partially deflated wetlands are indications of the region's historically rich water network, but its active water network is limited, so its horizontal fragmentation is modest. The micro-region is characterised by parabolic sand dunes and closed depressions, also found at the case study site. The altitude of the pasture site is between 151 and 156 m; for the site with the lateral moving sprinkler irrigation system, it is between 146 and 150 m. The average annual amount of sunshine hours is around 1875, with approximately 750-780 hours in the summer (June- August) and 165-170 hours in the winter (December-February). The area's wind averages 2.5 m/s, and predominant wind directions are north and southeast. In summary, the case study site with no irrigation is suitable for drought-tolerant species and varieties with lower water demand.

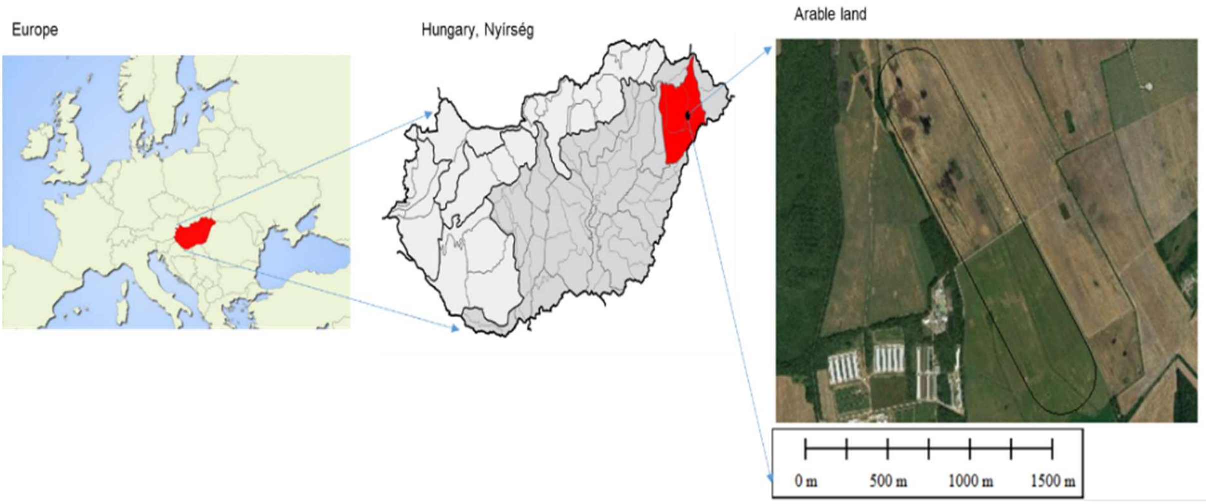


Figure 15: Location of the meteorological research station Nyírbátor in Eastern Hungary

3.5.2 Model description

For the Nyírbator site, a one-grid cell model for optimising irrigation scheduling is being developed using the Community Land Model version 5 (CLM5). CLM5 is a land surface model, and its main features were described earlier in this deliverable. Here a physically based CLM5 model for the

Nyírbátor site was set up, simulating hydraulic and plant conditions for 2020 and 2021. The model mesh consists of a single grid cell of 100x100x40 m and considers 20 hydraulically active layers with increasing thicknesses. The coordinates of the Nyírbátor weather station form the grid cell centre. The grid cell is bare soil during winter and rain-fed corn in summer. Later simulations with irrigated-fed corn are planned. The observation period is simulated starting from the 1000-year spun-up states. The spin-up run initialises vegetation, carbon, and nitrogen pools. On-site instrumentation provides in-situ measurements of precipitation, wind speed, air pressure, air temperature and relative humidity with a time step of 1 hour, and global radiation as a daily average (Figure 16 and Figure 17). All meteorological observations were aggregated in daily time steps to match the temporal resolution of global radiation. We used these observed meteorological conditions as atmospheric forcings and used them to simulate the years 2020 and 2021, starting from the conditions at the end of the spin-up phase. The calculation time step was 1 day, like the temporal resolution of global radiation. Other input parameters are aggregated from high-resolution input data sets to the model grid cell and time step, as typically performed by CLM5 users.

3.5.3 The perspective of using data assimilation

The first results allowed us to simulate the soil moisture and soil temperature observation data over two years in the test area. We now begin to down-sample daily information on global radiation to time-steps of 1 hour, using the Meteorological Simulator "METSIM" (https://metsim.readthedocs.io/en/develop/). This will significantly increase the temporal resolution of the CLM5 model for the Nyírbátor site and allow us to mimic how irrigation could be optimised in the Hungarian test area using a site-specific physically based model. For example, we can model the irrigation data of the area by optimising the soil data, thus predicting the irrigation period and the amount to be irrigated. We also plan to use data assimilation later to reduce model uncertainties, considering developments from FZJ.

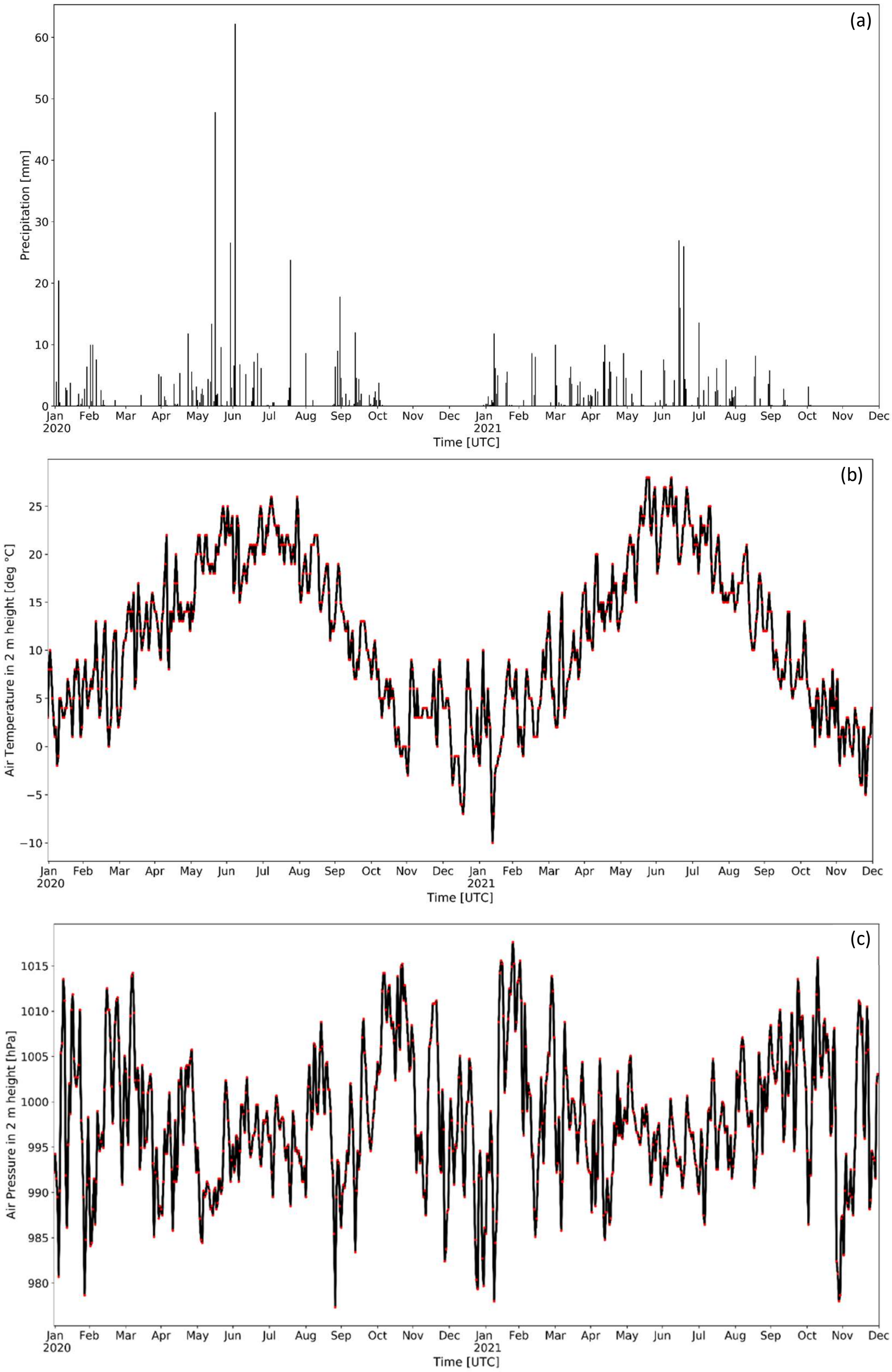


Figure 16: Input atmospheric forcings for the CLM5 model of the Nyírbátor site: daily average (a) precipitation; (b) air temperature (c) air pressure.

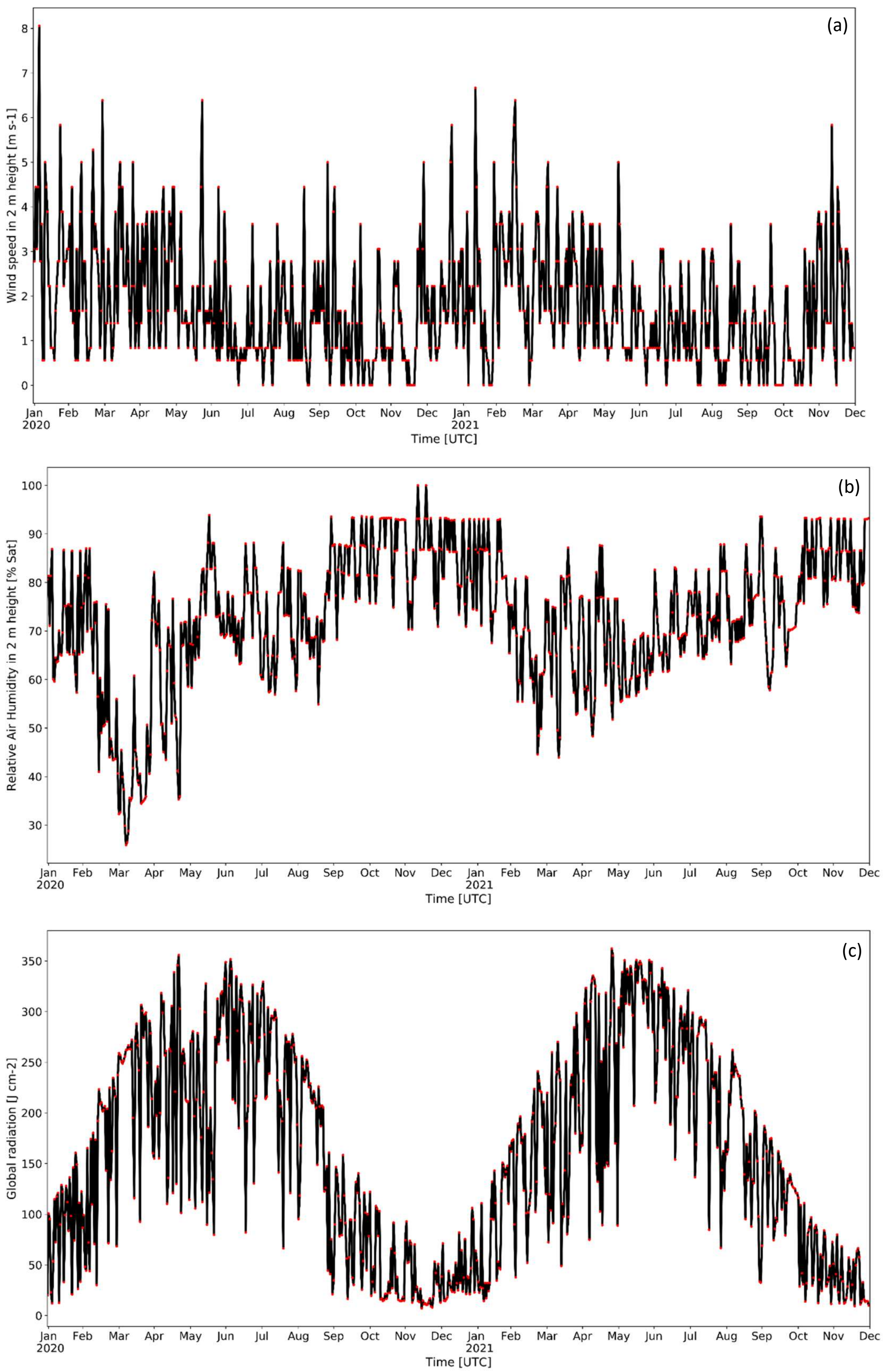


Figure 17: Input atmospheric forcings for the CLM5 model of the Nyírbátor site: daily average (a) wind speed; (b) relative humidity; (c) global radiation.

4 Conclusion and Outlook

WATERAGRI develops a data assimilation system for physically based models, including the assimilation of groundwater levels, soil moisture data and other measurements. Model simulations are continuously corrected to measured values in near-real time, quantifying and reducing uncertainty. We are mimicking operational site-specific forecasts of typical agricultural watershed conditions. Developments have progressed differently at the WATERAGRI locations with physically based models.

The successful coupling of HydroGeoSphere with the Parallel Data Assimilation Platform (HGS-PDAF) now enables the assimilation of groundwater levels for the WATERAGRI site "Seeland". The developments allow a local drainage system to be managed based on model results. Data assimilation significantly reduces the mean square error by about 60 % compared to simulations without data assimilation at the Seeland site, which can lead to more reliable forecasts that allow timely action to be taken. Better estimates of drainage volumes and groundwater recharge at the catchment scale can thus support, for example, near real-time drainage control management and irrigation scheduling.

FZJ mimics operational site-specific soil moisture forecasts for the next 10 days, using the Community Land Model version 5 coupled to PDAF (TSMP(CLM5)-PDAF, Strebel et al. 2022). Accurate site-specific soil moisture ensemble forecasts (e.g., assimilation of soil moisture data) at the plot scale and in near-real time are not yet widely available. Compared to large-scale models, local conditions can be better reflected and targeted plot-scale management options can be provided based on model results. Such site-specific simulations are potentially interesting for long-term quantitative support at the plot scale, as simulation results can be provided relatively quickly. Such forecasts can, for example, contribute to better irrigation management during drought periods and optimised irrigation scheduling during periods of limited water supply for irrigation.

The models in Finland, Poland and Hungary have been further refined, and there is a perspective for data assimilation using the data assimilation frameworks HGS-PDAF or CLM5-PDAF. The data assimilation systems can support their near real-time agricultural decision-making, like optimisation of irrigation rates to minimise wastewater or drainage control. While WATERAGRI will provide the technical resources, long-term support is beyond the capacity of the scientists. The data assimilation system can be set up for sites other than the WATERAGRI sites by trained professionals such as consultants. Consultants should also process the results, e.g., farmers, and provide quantitative support for agricultural decision-making. Scientists can support the work of consultants, i.e., by providing training and technical updates on the codes. The required instrumentation and computational resources may limit the commercial use for single farms, but it is also possible to setup models for larger areas, which are for example operated by consultants or specialized boards, and the improvement of soil moisture characterization around soil moisture sensors can also improve soil moisture characterization in a larger region around the sensor.

In this context, UNINE will organise a workshop on HGS-PDAF and CLM5-PDAF in the summer of 2023. This would allow the application of the data assimilation frameworks in Poland and Finland and maybe later by agricultural consultants. FZJ will release generic Python scripts (toolbox), which could, for example, be used in Hungary. In parallel, FZJ started an initiative with a German insurance company and a related farmer to disseminate the system (i.e., the toolbox) and will test it in an actual practice case.

5 References

Aalto, Juha; Pirinen, Pentti; Jylhä, Kirsti (2016): New gridded daily climatology of Finland: Permutation-based uncertainty estimates and temporal trends in climate. In: *J. Geophys. Res. Atmos.* 121 (8), S. 3807–3823. DOI: 10.1002/2015JD024651.

Ala-aho, Pertti; Soulsby, Chris; Wang, Hailong; Tetzlaff, Doerthe (2017): Integrated surface-subsurface model to investigate the role of groundwater in headwater catchment runoff generation: A minimalist approach to parameterisation. In: *Journal of Hydrology* 547, S. 664–677. DOI: 10.1016/j.jhydrol.2017.02.023.

Aquanty, I. (2020): HydroGeoSphere: A three-dimensional numerical model describing fully-integrated subsurface and surface flow and solute transport. Waterloo: Aquanty Inc.

Ashby, Steven F.; Falgout, Robert D. (1996): A Parallel Multigrid Preconditioned Conjugate Gradient Algorithm for Groundwater Flow Simulations. In: *Nuclear Science and Engineering* 124 (1), S. 145–159. DOI: 10.13182/NSE96-A24230.

Baldauf, Michael; Seifert, Axel; Förstner, Jochen; Majewski, Detlev; Raschendorfer, Matthias; Reinhardt, Thorsten (2011): Operational Convective-Scale Numerical Weather Prediction with the COSMO Model: Description and Sensitivities. In: *Monthly Weather Review* 139 (12), S. 3887–3905. DOI: 10.1175/MWR-D-10-05013.1.

Boas, Theresa; Bogena, Heye; Grünwald, Thomas; Heinesch, Bernard; Ryu, Dongryeol; Schmidt, Marius et al. (2021): Improving the representation of cropland sites in the Community Land Model (CLM) version 5.0. In: *Geosci. Model Dev.* 14 (1), S. 573–601. DOI: 10.5194/gmd-14-573-2021.

Bogena, H. R.; Montzka, C.; Huisman, J. A.; Graf, A.; Schmidt, M.; Stockinger, M. et al. (2018): The TERENO-Rur Hydrological Observatory: A Multiscale Multi-Compartment Research Platform for the Advancement of Hydrological Science. In: *Vadose Zone Journal* 17 (1), S. 180055. DOI: 10.2136/vzj2018.03.0055.

Bonton, Alexandre; Bouchard, Christian; Rouleau, Alain; Rodriguez, Manuel J.; Therrien, René (2012): Calibration and validation of an integrated nitrate transport model within a well capture zone. In: *Journal of Contaminant Hydrology* 128 (1-4), S. 1–18. DOI: 10.1016/j.jconhyd.2011.10.007.

Brunner, Philip; Simmons, Craig T. (2012): HydroGeoSphere: A Fully Integrated, Physically Based Hydrological Model. In: *Ground Water* 50 (2), S. 170–176. DOI: 10.1111/j.1745-6584.2011.00882.x.

Burgers, Gerrit; van Jan Leeuwen, Peter; Evensen, Geir (1998): Analysis Scheme in the Ensemble Kalman Filter. In: *Mon. Wea. Rev.* 126 (6), S. 1719–1724. DOI: 10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2.

Cheng, Linyin; AghaKouchak, Amir (2015): A methodology for deriving ensemble response from multimodel simulations. In: *Journal of Hydrology* 522, S. 49–57. DOI: 10.1016/j.jhydrol.2014.12.025.

Davison, Jason H.; Hwang, Hyoun-Tae; Sudicky, Edward A.; Lin, John C. (2015): Coupled atmospheric, land surface, and subsurface modeling: Exploring water and energy feedbacks in three-dimensions. In: *Advances in Water Resources* 86, S. 73–85. DOI: 10.1016/j.advwatres.2015.09.002.

Delottier, Hugo; Peel, Morgan; Musy, Stéphanie; Schilling, Oliver S.; Purtschert, Roland; Brunner, Philip (2022): Explicit simulation of environmental gas tracers with integrated surface and subsurface hydrological models. In: *Front. Water* 4, Artikel 980030. DOI: 10.3389/frwa.2022.980030.

Doherty, John (2015): Calibration and uncertainty analysis for complex environmental models. Brisbane, Australia: Watermark Numerical Computing.

Eskola, Tiina; Lunkka, Juha P. (2022): Sediment sequence at Muhos, western Finland – a window to the Pleistocene history of the Scandinavian Ice Sheet. In: *Boreas* 51 (2), S. 332–349. DOI: 10.1111/bor.12560.

Evensen, Geir (2003): The Ensemble Kalman Filter: theoretical formulation and practical implementation. In: *Ocean Dynamics* 53 (4), S. 343–367. DOI: 10.1007/s10236-003-0036-9.

Feddes, R. A.; Kowalik, P. J.; Zaradny, Henryk (1978): Simulation of field water use and crop yield. New York: Wiley (Simulation monographs).

Feddes, Reinder Auke; Rooij, Gerrit Huibert de; van Dam, Jan Cornelis (Hg.) (2004): Unsaturated-zone modeling. Progress, challenges and applications. Dordrecht: Kluwer academic Publishers (Wageningen UR frontis series, volume 6).

Fersch, Benjamin; Francke, Till; Heistermann, Maik; Schrön, Martin; Döpper, Veronika; Jakobi, Jannis et al. (2020): A dense network of cosmic-ray neutron sensors for soil moisture observation in a highly instrumented pre-Alpine headwater catchment in Germany. In: *Earth Syst. Sci. Data* 12 (3), S. 2289–2309. DOI: 10.5194/essd-12-2289-2020.

Fiers, Marie; Edel-Hermann, Véronique; Chatot, Catherine; Le Hingrat, Yves; Alabouvette, Claude; Steinberg, Christian (2012): Potato soil-borne diseases. A review. In: *Agron. Sustain. Dev.* 32 (1), S. 93–132. DOI: 10.1007/s13593-011-0035-z.

Gasper, F.; Goergen, K.; Shrestha, P.; Sulis, M.; Rihani, J.; Geimer, M.; Kollet, S. (2014): Implementation and scaling of the fully coupled Terrestrial Systems Modeling Platform (TerrSysMP v1.0) in a massively parallel supercomputing environment – a case study on JUQUEEN (IBM Blue Gene/Q). In: *Geosci. Model Dev.* 7 (5), S. 2531–2543. DOI: 10.5194/gmd-7-2531-2014.

Gebler, S.; Kurtz, W.; Pauwels, V. R. N.; Kollet, S. J.; Vereecken, H.; Hendricks Franssen, H.-J. (2019): Assimilation of High-Resolution Soil Moisture Data Into an Integrated Terrestrial Model for a Small-Scale Head-Water Catchment. In: *Water Res* 55 (12), S. 10358–10385. DOI: 10.1029/2018WR024658.

Geological Survey of Finland (2017): Stratum data for superficial deposits [database].

Hung, Ching Pui; Schalge, Bernd; Baroni, Gabriele; Vereecken, Harry; Hendricks Franssen, Harrie-Jan (2022): Assimilation of Groundwater Level and Soil Moisture Data in an Integrated Land Surface-Subsurface Model for Southwestern Germany. In: *Water Res* 58 (6). DOI: 10.1029/2021WR031549.

Johansson, P.; Räisänen, J.; Väisänen, U. (2005): Superficial deposits of Tyrnävä area, Superficial deposits 1:20 000 interpretation. Hg. v. Geological Survey of Finland, Rovaniemi (sheet 342106 [In Finnish]).

Johnson, Stephanie J.; Stockdale, Timothy N.; Ferranti, Laura; Balmaseda, Magdalena A.; Molteni, Franco; Magnusson, Linus et al. (2019): SEAS5: the new ECMWF seasonal forecast system. In: *Geosci. Model Dev.* 12 (3), S. 1087–1117. DOI: 10.5194/gmd-12-1087-2019.

Kennedy, Daniel; Swenson, Sean; Oleson, Keith W.; Lawrence, David M.; Fisher, Rosie; Da Lola Costa, Antonio Carlos; Gentine, Pierre (2019): Implementing Plant Hydraulics in the Community Land Model, Version 5. In: *J. Adv. Model. Earth Syst.* 11 (2), S. 485–513. DOI: 10.1029/2018MS001500.

Köhli, M.; Schrön, M.; Zreda, M.; Schmidt, U.; Dietrich, P.; Zacharias, S. (2015): Footprint characteristics revised for field-scale soil moisture monitoring with cosmic-ray neutrons. In: *Water Resour. Res.* 51 (7), S. 5772–5790. DOI: 10.1002/2015WR017169.

Kollet, Stefan J.; Maxwell, Reed M. (2006): Integrated surface–groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model. In: *Advances in Water Resources* 29 (7), S. 945–958. DOI: 10.1016/j.advwatres.2005.08.006.

Kristensen, K. J.; Jensen, S. E. (1975): A model for estimating actual evapotranspiration from potential evapotranspiration. In: *Hydrology Research* 6 (3), S. 170–188. DOI: 10.2166/nh.1975.0012.

Kurtz, Wolfgang; He, Guowei; Kollet, Stefan J.; Maxwell, Reed M.; Vereecken, Harry; Hendricks Franssen, Harrie-Jan (2016): TerrSysMP–PDAF (version 1.0): a modular high-performance data assimilation framework for an integrated land surface–subsurface model. In: *Geosci. Model Dev.* 9 (4), S. 1341–1360. DOI: 10.5194/gmd-9-1341-2016.

Kurtz, Wolfgang; Lapin, Andrei; Schilling, Oliver S.; Tang, Qi; Schiller, Eryk; Braun, Torsten et al. (2017): Integrating hydrological modelling, data assimilation and cloud computing for real-time management of water resources. In: *Environmental Modelling & Software* 93, S. 418–435. DOI: 10.1016/j.envsoft.2017.03.011.

Lawrence, David M.; Fisher, Rosie A.; Koven, Charles D.; Oleson, Keith W.; Swenson, Sean C.; Bonan, Gordon et al. (2019): The Community Land Model Version 5: Description of New Features, Benchmarking, and Impact of Forcing Uncertainty. In: *J. Adv. Model. Earth Syst.* 11 (12), S. 4245–4287. DOI: 10.1029/2018MS001583.

Liedes, T.; Pylvänäinen, J.; Torvela, J.; Aurio, V.; Tuula, L.; Sakari, N.; Esa, T. (2020): Final report of the TIMAKO project on the data driven soil moisture control.

Martin, R. J.; Jamieson, P. D.; Wilson, D. R.; Francis, G. S. (1992): Effects of soil moisture deficits on yield and quality of 'Russet Burbank' potatoes. In: *New Zealand Journal of Crop and Horticultural Science* 20 (1), S. 1–9. DOI: 10.1080/01140671.1992.10422319.

Maxwell, Reed M. (2013): A terrain-following grid transform and preconditioner for parallel, large-scale, integrated hydrologic modeling. In: *Advances in Water Resources* 53, S. 109–117. DOI: 10.1016/j.advwatres.2012.10.001.

Moeck, Christian; Hunkeler, Daniel; Brunner, Philip (2015): Tutorials as a flexible alternative to GUIs: An example for advanced model calibration using Pilot Points. In: *Environmental Modelling & Software* 66, S. 78–86. DOI: 10.1016/j.envsoft.2014.12.018.

Nerger, Lars; Hiller, Wolfgang (2013): Software for ensemble-based data assimilation systems—Implementation strategies and scalability. In: *Computers & Geosciences* 55, S. 110–118. DOI: 10.1016/j.cageo.2012.03.026.

Nerger, Lars; Hiller, Wolfgang; SCHRÖTER, JENS (2005): A comparison of error subspace Kalman filters. In: *Tellus A* 57 (5), S. 715–735. DOI: 10.1111/j.1600-0870.2005.00141.x.

Okołowicz, W. (1977): The climate of Poland, Czechoslovakia and Hungary. In: C. C. Wallén (Hg.): Climates of central and southern Europe. Amsterdam, Oxford, New York: Elsevier Scientific; Distributors for the U.S. and Canada, Elsevier-North Holland (World survey of climatology, v. 6), S. 75–124.

Oleson, K. W.; Niu, G.-Y.; Yang, Z.-L.; Lawrence, D. M.; Thornton, P. E.; Lawrence, P. J. et al. (2008): Improvements to the Community Land Model and their impact on the hydrological cycle. In: *J. Geophys. Res.* 113 (G1), n/a-n/a. DOI: 10.1029/2007JG000563.

Panday, Sorab; Huyakorn, Peter S. (2004): A fully coupled physically-based spatially-distributed model for evaluating surface/subsurface flow. In: *Advances in Water Resources* 27 (4), S. 361–382. DOI: 10.1016/j.advwatres.2004.02.016.

Pirinen, Pentti; Lehtonen, Ilari; Heikkinen, Risto K.; Aapala, Kaisu; Aalto, Juha (2022): FMI's Climate Bulletin Research Letters 2/2022. In: *Ilmastokatsaus* 4 (2), S. 35–37. DOI: 10.35614/ISSN-2341-6408-IK-2022-11-RL.

Reichle, Rolf H. (2008): Data assimilation methods in the Earth sciences. In: *Advances in Water Resources* 31 (11), S. 1411–1418. DOI: 10.1016/j.advwatres.2008.01.001.

Rooij, Rob de (2017): A consistent implementation of the dual node approach for coupling surface-subsurface flow and its comparison to the common node approach.

Schepper, G. de; Therrien, R.; Refsgaard, J. C.; He, X.; Kjaergaard, C.; Iversen, B. V. (2017): Simulating seasonal variations of tile drainage discharge in an agricultural catchment. In: *Water Resour. Res.* 53 (5), S. 3896–3920. DOI: 10.1002/2016WR020209.

Schilling, O. S.; Doherty, J.; Kinzelbach, W.; Wang, H.; Yang, P. N.; Brunner, P. (2014): Using tree ring data as a proxy for transpiration to reduce predictive uncertainty of a model simulating groundwater–surface water–vegetation interactions. In: *Journal of Hydrology* 519, S. 2258–2271. DOI: 10.1016/j.jhydrol.2014.08.063.

Schilling, Oliver S.; Irvine, Dylan J.; Hendricks Franssen, Harrie-Jan; Brunner, Philip (2017): Estimating the Spatial Extent of Unsaturated Zones in Heterogeneous River-Aquifer Systems. In: *Water Resour. Res.* 53 (12), S. 10583–10602. DOI: 10.1002/2017WR020409.

Schilling, Oliver S.; Park, Young-Jin; Therrien, René; Nagare, Ranjeet M. (2019): Integrated Surface and Subsurface Hydrological Modeling with Snowmelt and Pore Water Freeze-Thaw. In: *Groundwater* 57 (1), S. 63–74. DOI: 10.1111/gwat.12841.

Shi, Yuning; Davis, Kenneth J.; Zhang, Fuqing; Duffy, Christopher J.; Yu, Xuan (2015): Parameter estimation of a physically-based land surface hydrologic model using an ensemble Kalman filter: A multivariate real-data experiment. In: *Advances in Water Resources* 83, S. 421–427. DOI: 10.1016/j.advwatres.2015.06.009.

Shrestha, P.; Sulis, M.; Masbou, M.; Kollet, S.; Simmer, C. (2014): A Scale-Consistent Terrestrial Systems Modeling Platform Based on COSMO, CLM, and ParFlow. In: *Mon. Wea. Rev.* 142 (9), S. 3466–3483. DOI: 10.1175/MWR-D-14-00029.1.

Strebel, Lukas; Bogena, Heye R.; Vereecken, Harry; Hendricks Franssen, Harrie-Jan (2022): Coupling the Community Land Model version 5.0 to the parallel data assimilation framework PDAF: description and applications. In: *Geosci. Model Dev.* 15 (2), S. 395–411. DOI: 10.5194/gmd-15-395-2022.

Swenson, Sean C.; Clark, Martyn; Fan, Ying; Lawrence, David M.; Perket, Justin (2019): Representing Intrahillslope Lateral Subsurface Flow in the Community Land Model. In: *J. Adv. Model. Earth Syst.* 11 (12), S. 4044–4065. DOI: 10.1029/2019MS001833.

Tang, Q.; Kurtz, W.; Schilling, O. S.; Brunner, P.; Vereecken, H.; Hendricks Franssen, H.-J. (2017): The influence of riverbed heterogeneity patterns on river-aquifer exchange fluxes under different connection regimes. In: *Journal of Hydrology* 554, S. 383–396. DOI: 10.1016/j.jhydrol.2017.09.031.

Tang, Qi; Schilling, Oliver S.; Kurtz, Wolfgang; Brunner, Philip; Vereecken, Harry; Hendricks Franssen, Harrie-Jan (2018): Simulating Flood-Induced Riverbed Transience Using Unmanned Aerial Vehicles, Physically Based Hydrological Modeling, and the Ensemble Kalman Filter. In: *Water Resour. Res.* 54 (11), S. 9342–9363. DOI: 10.1029/2018WR023067.

Valcke, S. (2013): The OASIS3 coupler: a European climate modelling community software. In: *Geosci. Model Dev.* 6 (2), S. 373–388. DOI: 10.5194/gmd-6-373-2013.

van Genuchten, M. Th. (1980): A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils. In: *Soil Science Society of America Journal* 44 (5), S. 892–898. DOI: 10.2136/sssaj1980.03615995004400050002x.

Zhang, Hongjuan; Kurtz, Wolfgang; Kollet, Stefan; Vereecken, Harry; Franssen, Harrie-Jan Hendricks (2018): Comparison of different assimilation methodologies of groundwater levels to improve predictions of root zone soil moisture with an integrated terrestrial system model. In: *Advances in Water Resources* 111, S. 224–238. DOI: 10.1016/j.advwatres.2017.11.003.