UDC [004.94+519.64]:556.3
DOI https://doi.org/10.26661/2413-6549-2021-2-03

EVALUATING PREDICTIONS OF THE SOIL MOISTURE MODEL WITH DATA ASSIMILATION BY THE TRIPLE COLLOCATION METHOD

Kozhushko O. D.
Postgraduate Student
National University of Water and Environmental Engineering
Soborna str., 11, Rivne, Ukraine
orcid.org/0000-0001-9934-1013
ol.d.kozhushko@nuwm.edu.ua

Boiko M. V.
Postgraduate Student
National University of Water and Environmental Engineering
Soborna str., 11, Rivne, Ukraine
orcid.org/0000-0002-0962-8625
m.v.boiko@nuwm.edu.ua

Kovbasa N. Yu.
Junior Research Fellow
V. Ye. Lashkaryov Institute of Semiconductor Physics
Nauky ave., 41, Kyiv, Ukraine
orcid.org/0000-0001-7988-0175
kovbasa_myu@nas.gov.ua

Martyniuk P. M.
Doctor of Technical Sciences, Professor,
Head of the Institute of Automatics, Cybernetics and Computer Engineering
National University of Water and Environmental Engineering
Soborna str., 11, Rivne, Ukraine
orcid.org/0000-0002-2750-2508
p.m.martyniuk@nuwm.edu.ua

Stepanchenko O. M.
Candidate of Technical Sciences, Associate Professor,
Doctorate Student
National University of Water and Environmental Engineering
Soborna str., 11, Rivne, Ukraine
orcid.org/0000-0001-6033-7404
o.m.stepanchenko@nuwm.edu.ua

Uvarov N. V.
Candidate of Physical and Mathematical Sciences
Research Fellow
G. V. Kurdyumov Institute for Metal Physics
Academician Vernadsky boul., 36, Kyiv, Ukraine
orcid.org/0000-0001-7988-0175
uvarovn@gmail.com

Key words: *Newtonian nudging, nonlinear mathematical modelling, Richards equation, Earth remote sensing, statistical model validation.*

For a long time, ground measurements are considered the most accurate method of regular monitoring of soil moisture. However, ground stations are expensive, require local calibration and thus often not practical to use. Other, more affordable means of soil moisture monitoring can be developed with the recent advancement of Earth remote sensing technologies.
In this paper, we describe a nonlinear problem of soil moisture transfer problem with addition of satellite soil moisture measurements. The mathematical model is based on

the Richards equation for soil moisture transport, and solved with the finite difference method on implicit iterative scheme. Satellite moisture retrievals are acquired by combining active and passive sensors data with the decomposition algorithm.
The satellite data are incorporated into the model with the data assimilation algorithm called Newtonian nudging. This method adds a special 'nudging' term into the model governing equation. This assures that the model is corrected by satellite measurements without affecting the process physics. Moreover, we look into the nudging factor problem, and propose a simple empirical relation based on the soil properties for more universally stable work of the method.
For validation purpose, we conduct a massive numerical experiments over all registered ground stations in the USA. Evaluation is done by the use of triple collocation method, which allows assessing the errors of three independent data sources. The data sources used for evaluation are the model results, ground station measurements and ERA5 satellite observations. The results demonstrate that the presented model is capable of producing results with close accuracy to the ground station measurements.

ОЦІНКА ПРОГНОЗУ МОДЕЛІ ВОЛОГОСТІ ҐРУНТУ ІЗ АСИМІЛЯЦІЄЮ ДАНИХ МЕТОДОМ ПОТРІЙНОЇ КОЛЛОКАЦІЇ

Кожушко О. Д.
аспірант
Національний університет водного господарства та природокористування
вул. Соборна, 11, Рівне, Україна
orcid.org/0000-0001-9934-1013
ol.d.kozhushko@nuwm.edu.ua

Бойко М. В.
аспірант
Національний університет водного господарства та природокористування
вул. Соборна, 11, Рівне, Україна
orcid.org/0000-0002-0962-8625
m.v.boiko@nuwm.edu.ua

Ковбаса М. Ю.
молодший науковий співробітник
Інститут фізики напівпровідників імені В. Є. Лашкарьова
просп. Науки, 41, Київ, Україна
orcid.org/0000-0001-7988-0175
kovbasa_myu@nas.gov.ua

Мартинюк П. М.
доктор технічних наук, професор,
директор інституту автоматики, кібернетики та обчислювальної техніки
Національний університет водного господарства та природокористування
вул. Соборна, 11, Рівне, Україна
orcid.org/0000-0002-2750-2508
p.m.martyniuk@nuwm.edu.ua

Степанченко О. М.
кандидат технічних наук, доцент, докторант
Національний університет водного господарства та природокористування
вул. Соборна, 11, Рівне, Україна
orcid.org/0000-0001-6033-7404
o.m.stepanchenko@nuwm.edu.ua

Уваров М. В.
кандидат фізико-математичних наук,
науковий співробітник
Інститут металофізики імені Г. В. Курдюмова,
бул. Академіка Вернадського, 36, Київ, Україна
orcid.org/0000-0001-7988-0175
uvarovn@gmail.com

Ключові слова: *ньютонівське підштовхування, нелінійне математичне моделювання, рівняння Річардса, дистанційне зондування Землі, статистична валідація.*

Упродовж довго часу наземні вимірювання вважалися найточнішим способом регулярного моніторингу вологості ґрунту. Проте наземні вимірювальні станції є дороговартісними та потребують калібрування на місці встановлення, що часто робить їх використання недоцільним на практиці. З нещодавнім розвитком технологій дистанційного зондування Землі можуть бути розроблені інші, більш дешеві методи.

У цій роботі представлено нелінійну задачу вологоперенесення із додаванням супутникових вимірювань вологості ґрунту. Математична модель базується на рівнянні Річардса для вологоперенесення та розв'язується методом скінченних різниць з використанням неявної ітераційної схеми Самарського. Супутникові оцінки вологості ґрунту отримані із поєднання даних активних та пасивних сенсорів за допомогою алгоритмів декомпозиції.

Супутникові дані включалися до моделі згідно з алгоритмом асиміляції даних, що називається ньютонівським підштовхуванням. Цей метод передбачає додавання особливого члену «підштовхування» до модельного рівняння. Таким чином, модель коригується згідно супутникових вимірювань та забезпечується дотримання фізики процесу. Крім того, було здійснено огляд проблеми із вибором фактору підштовхування та запропонована емпірична формула з використанням параметрів ґрунту, що дозволяє підвищити універсальність та стійкість методу.

З метою валідації моделі було здійснено масштабний числовий експеримент з використанням усіх зареєстрованих наземних станцій вимірювання вологості в США. Оцінку було здійснено методом потрійної коллокації, що дає можливість оцінити похибки у трьох незалежних наборах даних. Для оцінки було використано такі джерела даних: модельні результати, дані вимірювань наземних станцій та супутникові спостереження із бази даних ERA5. Отримані результати демонструють, що представлена модель здатна показувати результати із точністю, що наближається до точності наземних вимірювань.

Introduction. Soil moisture data can be valuable in numerous practical applications, from agriculture to climate forecasts. Up-to-date information on moisture can improve the precision of predictions, optimize water resource management and advice on irrigation planning. These applications demand frequent and precise data which may be provided by measurements or model simulations. However, soil moisture is a notoriously difficult parameter to measure [1], and the accuracy of collected data, especially if they are derived from a model simulation, is not easily evaluated. This is particularly true for practical cases, where data is limited due to economical reasons. For that reason, our study aims not only to provide soil moisture predictions, but also to evaluate the model results, as well as other data that may be available in a practical case, against the available historical datasets.

In-situ observations are usually considered the most reliable soil moisture measurements. To provide immediate observation data at different soil depth, ground stations use multiple measurement methods, such as oven-drying, neutron probe, capacitance method etc. However, they are not used systematically due to their cost, installation difficulties and other practical reasons. It has also been pointed out that in-situ data, measured at a single observation point, may fail to describe the state of the whole system, so the installation point should be considered carefully [2].

Another way to measure soil moisture is by satellite imagery. Recently, the method became popular in real-world applications due to its increasing quality and availability, and it is considered relatively cheap. Satellite microwave sensors are able to measure soil moisture in different spatial scales. The accuracy of these observations is satisfactory for global scale, but derivation of precise local data requires complex image processing algorithms. Moreover, microwave sensors measure only surface soil moisture (0-5 cm layer), and provide data on a few days interval, which cannot accurately represent the state of the system.

Land models are another alternative to assess soil moisture. Model simulations provide continuous data of all system states, and are cheap to perform. On the other hand, they unavoidably contain errors due to generalizations of physics since every participating process and effect cannot be accounted for. Moreover, the more complex and full the model is, the more parameters it requires, which, on their own turn, demand additional data. Therefore, modern soil moisture models are often combined with other data sources and measurements to estimate the parameters accurately. For example, studies [3; 4] propose formulas for calculating model coefficients with the use of satellite moisture data.

In a case where multiple methods are available, and each of them has its own strengths and weaknesses, combination of differently acquired data might provide substantial improvements. For example, addition of satellite data into a simulation model is lately becoming a widely used tool, called data assimilation. It adapts model to the observed data, and allows to achieve higher precision level than provided by each data source alone [5]. There are multiple data assimilation algorithms available, the most widely used of which are described in [6]. In our study, we implement a method called Newtonian relaxation or nudging.

Newtonian nudging is a rather simple and effective data assimilation method, first used for oceanography problems. It is also widely used in hydrology problems, and has been implemented into a number of hydrological and environmental models, such as GLEAM [7] and CATHY [8]. The method consists in adding a nudging term, multiplied by the difference between model prediction and observed value, into the governing equation. The term works as a physical force that relaxes the result towards observations.

In this study, we describe our model and data assimilation approach, and also perform a statistical verification test. Traditionally, the results are validated against in-situ measurements, which are taken as a benchmark for comparison. However, among the variety of statistical methods designed to evaluate prediction accuracy, we chose the triple collocation method as it does not imply knowing the absolute truth. The method involves comparison of three independent data sources, where each is assumed to contain errors of some sort. Therefore, the method is effective for real-world validation tests, and is often employed in evaluation of soil moisture models [9; 10].

Mathematical problem setting. Our model includes moisture transfer problem based on the Richards equation. The problem domain is one-dimensional of thickness l, with downward x axis and $x=0$ at the soil surface. The boundary value problem setting is as follows:

$$\frac{\partial\theta(x,t)}{\partial t} = \frac{\partial}{\partial x}\left(k(h)\frac{\partial h}{\partial x}\right) - \frac{\partial k(h)}{\partial x} - \mathrm{S}(h,x,t), \quad (1)$$

$$\left.\left(-k(h)\frac{\partial h}{\partial x} + k(h)\right)\right|_{x=0} = \mathrm{Q}(t) - \mathrm{E}_s(t), \mathrm{t} > 0, \quad (2)$$

$$\left.\frac{\partial h}{\partial x}\right|_{x=0} = 0, \mathrm{t} > 0, \quad (3)$$

$$h(x,0) = h_0(x),\ x \in [0;l]. \quad (4)$$

Here θ is absolute soil moisture, h – pressure head, k – soil hydraulic conductivity, $S(h, x, t)$ – root water uptake, $Q(t)$ – precipitation rate, $E_s(t)$ – soil evaporation rate, $h_0(x)$ is initial condition for pressure head. Potential evaporation E_s is derived from meteorological parameters [5], root water uptake S is calculated according to potential evapotranspiration and water availability based on the Feddes model [12].

Since the Richards equation requires a translation rule between moisture and pressure head values, we chose a widely used Mualem–van Genuchten model [13]. The model is represented by the following equations:

$$\theta(h) = \theta_{\min} + \frac{\theta_{\max} - \theta_{\min}}{\left(1 + (-\alpha h)^n\right)^m},\ m = 1 - \frac{1}{n}, \quad (5)$$

$$k(h) = k_s S^l \left(1 - \left(1 - S^{\frac{1}{m}}\right)^m\right)^2,\ \ S = \frac{\theta - \theta_{\min}}{\theta_{\max} - \theta_{\min}}, \quad (6)$$

where θ_{min}, θ_{max} are residual and saturation water content, k_s – saturated soil hydraulic conductivity, S – saturation degree, α, n, l – empirical model parameters. These parameters define water retention curve of the soil. The reliable values of Mualem–van Genuchten model parameters for basic soil types can be found in the various researches on the topic, as well as estimated by the program systems like Rosetta [14].

Substituting (5) into (1), the governing model equation is rewritten in pressure heads and becomes

$$\beta(h)\frac{\partial h}{\partial t} = \frac{\partial}{\partial x}\left(k(h)\frac{\partial h}{\partial x}\right) - v(h)\frac{\partial h}{\partial x} - \mathrm{S}(h,x,t), \quad (7)$$

where values of coefficients $\beta(h) = \partial\theta/\partial h$, $v(h) = \partial k(h)/\partial h$ can be calculated analytically according to equations (5), (6).

The one-dimensional problem described above is then discretized using an implicit time scheme and solved on a uniform grid. Specifically, the homogeneous finite difference scheme is used for numerical calculations, since it is rather cheap computationally and allows variable coefficients in equations. To deal with nonlinearity of Mualem–van Genuchten relations, we use an implicit iterative scheme described by Samarskiy. The scheme requires doing additional solving iterations on each time step, recalculating parameters until the solution converges. This scheme is not optimal in the sense of computational time, but is simple to implement and has good convergence [15].

Newtonian nudging assimilation. Newtonian nudging is a smoother algorithm that modifies the simulation result on each time based on the past and future observations. Its original application was for determining best initial conditions in oceanography and meteorology problems, but later was adapted for updating current model state in hydrology problems, including soil moisture [16].

Despite being not as popular as the filtering methods, e.g. ensemble Kalman filters [17], Newtonian nudging is one of the prominent 4-dimensional data assimilation (4DDA) methods. It is incorporated directly into the model governing equation, which makes it better suited for use in boundary-value problems [18]. Unlike the filtering methods that treat the solution as a random variable, nudging introduces a physical force into the equation, yielding smooth and physically sensible results. The nudging term, inserted into the Richard equation (1), has the following form:

$$\frac{\partial\theta}{\partial t}=\frac{\partial}{\partial x}\left(k(h)\frac{\partial h}{\partial x}-k(h)\right)-S(h,x,t)+G\cdot W(x,t)\varepsilon(x)(\theta_{obs}-\theta), \quad (8)$$

where θ_{obs} is observed surface soil moisture, G – nudging factor, $W(x, t)$ – weight function, and $\varepsilon(x)$ is the degree of trust to the observations, limited by the interval of [0; 1].

The core of the term is the difference between observed and simulated soil moisture. In a more general case, the term can include a few observations, past and future, at the same time. The nudging factor represents the magnitude of the nudging force. It is recommended by the authors that this force should correspond to the slowest process in the model. The weight function corrects the force of nudging based on the distance from observation point and time span between the simulation and observation time [19].

Original formulation of Newtonian nudging uses constant nudging factor. However, while testing the method, we found out the constant factor causes irregular behavior, as the nudging seems stronger on lower moisture values and weaker near saturation values. This can be attributed to the fact that coefficients of the model equation (7) are nonlinear and can vary in the order of magnitude.

Some studies suggest choosing nudging coefficient from optimality conditions [20] or by solving an adjoint problem [21]. While these methods are well-justified and proved to be accurate, they introduce a significant computational overhead from solving the additional problem. Here we suggest a completely different approach to tackle the nudging factor problem. The idea is to calculate the suitable factor value using the current model equation coefficients. In our numerical experiments, we employed the following adaptive relation to calculate the factor considering the current soil retention and permeability characteristics:

$$G=100\left(\beta(h)\cdot|h|+0.5\frac{k(h)}{k_s}\right). \quad (9)$$

Here, as above, h is pressure head, k and k_s are saturated and actual soil hydraulic conductivity, and β is defined as $\beta(h)=\partial\theta/\partial h$ according to (5).

Though this ‘adaptive nudging’ relies on unverified relations and cannot offer the optimal accuracy, it can be calibrated further to provide stable and accurate results, and it does not add any computation difficulty to the problem since the values of coefficients $\beta(h)$ and $k(h)$ are already calculated during the solving process.

Satellite moisture retrieving. Data assimilation procedure requires low-noise and frequent satellite data. Active instruments on Sentinel-1, RADARSAT, RISAT-1 etc. can provide high-resolution soil moisture data with appropriate algorithms; however, as a rule, they have a sparse repeated interval around 10 days worldwide. On the other hand, passive instruments on SMAP, SMOS, AMSR-E and AMSR2 can provide data with repeated intervals of a couple of days worldwide, although without disaggregation algorithms these instruments provide low resolution of about tens of kilometers [22].

A disaggregation method is applied to obtain high-resolution soil moisture data from passive sensors AMSR-E, AMSR2 and SMAP, land surface temperature from AMSR2 and AMSR-E data. To calculate dielectric permittivity of the soil, we applied Single Channel Algorithm – Vertical [23] for SMAP disaggregated data and Land Parameter Retrieval Model for AMSR-E and AMSR2 data. The method allows us to get high-resolution dielectric permittivity maps with 250×250 m resolution, which is close to the field scale. We applied the Mironov model for L-band [24], Dobson model [25] for C-band to convert dielectric permittivity to soil moisture content.

Model evaluation methods. We apply the triple collocation method to combine and assess the errors of ground station measurements, satellite observations and model simulations. The method considers at least three data sets, represented here by random variables θ_1, θ_2 and θ_3, each containing the same number of estimations of some variable. It is also required that the datasets are unbiased, which cannot be guaranteed in practice, so the bias is removed artificially. Each of the variables θ_1, θ_2 and θ_3 differs from the hypothetical truth θ by a residual r_i, $i=\overline{1,n}$, as shown by the equations

$$\begin{aligned}\theta_1&=\theta+r_1,\\ \theta_2&=\theta+r_2,\\ \theta_3&=\theta+r_3.\end{aligned} \quad (10)$$

The method does not allow to find the truth value θ, only to evaluate the quality of each dataset through the estimation of the random errors r_i. After eliminating the hypothetical truth from the equations (10) and taking average over the resulting equation, we can estimate the variances of residuals, denoted σ_1, σ_2, σ_3 respectively, with the following formula:

$$\begin{aligned}\sigma_1^2&=\langle(\theta_1-\theta_2)(\theta_1-\theta_3)\rangle,\\ \sigma_2^2&=\langle(\theta_1-\theta_2)(\theta_2-\theta_3)\rangle,\\ \sigma_3^2&=\langle(\theta_1-\theta_3)(\theta_2-\theta_3)\rangle,\end{aligned} \quad (11)$$

where $\langle\bullet\rangle$ denotes covariance operator. Hence, using the random error variance, we can estimate the residuals as

$$\begin{aligned}\langle r_1^2\rangle&=\langle\theta_1'^2\rangle-\langle\theta_1'\theta_2'\rangle\langle\theta_1'\theta_3'\rangle/\langle\theta_2'\theta_3'\rangle,\\ \langle r_2^2\rangle&=\langle\theta_2'^2\rangle-\langle\theta_2'\theta_1'\rangle\langle\theta_2'\theta_3'\rangle/\langle\theta_1'\theta_3'\rangle,\\ \langle r_3^2\rangle&=\langle\theta_3'^2\rangle-\langle\theta_3'\theta_1'\rangle\langle\theta_3'\theta_2'\rangle/\langle\theta_1'\theta_2'\rangle,\end{aligned} \quad (12)$$

where prime sign means anomaly, for example $\theta_1'=\theta_1-\langle\theta_1\rangle$ [10].

Though in this work we use the triple collocation analysis only to verify and compare the accuracy of datasets, its application in fact is much more wider. The supposed variance of given datasets can be employed for further calibration. For example, authors of [26] propose using the variances provided by triple collocation analysis as weights to merge active and passive sensor retrieved values.

The model results are also evaluated against the measurements using the following traditional metrics: average absolute deviation (AAD), root mean square deviation (RMSD), bias, Pearson correlation (R) and the index of agreement (IoA), defined as follows:

$$IoA = 1 - \frac{\sum_t \left(\theta^{DA}(t) - \theta^{obs}(t)\right)^2}{\sum_t \left(\left|\theta^{DA}(t) - \overline{\theta^{obs}}\right| + \left|\theta^{obs}(t) - \overline{\theta^{obs}}\right|\right)^2}. \quad (13)$$

In the equation above $\theta^{DA}(t)$ are model simulated moisture values on time points t, $\theta^{obs}(t)$ are soil moisture measurements, and $\overline{\theta^{obs}}$ is their mean value over the simulation period.

All of the metrics mentioned above are calculated using the Pytesmo library, developed specifically for evaluation of soil moisture observations [27].

Numerical experiments. To achieve statistically relevant accuracy evaluation, we conducted a large-scale numerical experiment using the open ground sensor data provided by the International Soil Moisture Network (ISMN) [28]. The database includes information from hundreds of in-situ soil moisture sensors over the world, that is intended to be used as validation data for various modelling problems.

We chose ERA5 Climate Reanalysis data as a third dataset for triple collocation analysis. It contains spatial data of various meteorological and climate variables at 0.25° spatial resolution, including satellite-derived soil moisture on four soil layers [29]. Here, we use only the satellite topsoil moisture data for comparison.

The weather data were acquired from the NOAA database as for the nearest meteorological station. The NOAA provides one of the most full open meteorological datasets; however, the format of data is not easy to read. We downloaded the weather data from Lametsy API service [30], which provides the same NOAA data with daily aggregation and in a more readable format. Soil parameters were assumed based on the soil type individually for each station, based on the data provided by SoilGrids [31]. Initial conditions were set according to the satellite moisture data used for data assimilation.

The experiment was conducted for 2018 over all 659 ground stations in the USA, registered in the ISMN database. Out of them, 14 stations were excluded due to the issues with data availability, and another 83 stations – because of inconsistent or scarce measurement data, severe model errors etc. Therefore, the summary presented below includes experiment results for the remaining 562 stations.

Note also that comparison has been done only for the soil surface moisture since some ground stations provided no belowground data, and satellites sense moisture at the top soil layer only.

Results and discussion. First, we perform a traditional comparison of model simulation results against ground station measurements. The averaged metrics over all stations are presented in Table 1. Analysis shows that absolute deviation and RMSD values are rather high, and must be caused chiefly by the lack of correlation. Moreover, the deviation can be attributed to incorrect initial conditions or soil parameters. These are two of the key model parameters, but the chosen values were rough due to the great number of stations in the experiment. Nevertheless, the bias between the datasets is very low, meaning their average values are the same, and only the deviations from the mean might not be represented correctly by the model.

The average correlation demonstrated by the datasets is 26%, which is a rather weak correlation. However, the index of agreement is about 50%, which implies a tolerably good convergence.

Further analysis of the latter two characteristics is shown on Fig. 1. The first frequency histogram indicates that many simulations demonstrated negative correlation with ground station data. This may be primary due to imprecise weather data, e.g. when the meteorological station is very remote, and its data differ from actual situation on the site. Another reason may be the groundwater, which can cause a significant influence and is not yet accounted for in the model. Most of the positive correlation values are near the 0.3-0.4 interval, which is medium correlation. The index of agreement demonstrates a normal-like frequency distribution, clustered around the 50% value. It indicates that the datasets demonstrate a stable agreement with each other, even if correlation is weak.

Table 1

Evaluation of the model results against ground station measurements

Metrics	AAD	RMSD	Bias	R	IoA
Values	0.100001	0.11830	0.01782	0.26080	0.48001

Resuming the discussion of nudging factor issue, we can now compare the verification results for both nudging methods. It should be pointed out that bias seems to be the chief indicator of the problem with constant factor. Original calculations with constant factor resulted in the bias value of 0.06554, whereas adapted formulation reduced it to 0.01782, which is approximately 3.5 times less. RMSD is lowered accordingly, whereas other metrics are only slightly improved.

We also provide here likewise comparison of model results with ERA5 surface satellite data. As an evaluation benchmark, ERA5 data have an advantage of being consistent and harmonized with each other, whereas ISMN in-situ measurements are provided by various types of sensors, having different measurement methods and limitations. At the same time, ERA5 data represent satellite measurements, but are not used in data assimilation process, therefore they can serve as another benchmark for model predictions. The average key metrics of model evaluation against the independent satellite data are presented in Table 2.

Surprisingly, we acquired less convergence between the model and satellite data than between the model and in-situ measurements. Average deviation and RMSD are nearly 50% higher in this case, but this must be due to the significant bias between the data. In fact, while revising the ERA5 soil moisture dataset, we found that the values presented there are significantly higher than in the other datasets, especially our satellite-retrieved moisture. However, ERA5 database is aimed for climatic and meteorological analysis, so the difference might be caused by the different speciality of the datasets. While analyzing the data further with triple collocation method, we obviously remove the bias, so this should no longer present a problem.

Table 2

Evaluation of the model results against ERA5 satellite measurements

Metrics	AAD	RMSD	Bias	R	IoA
Values	0.14119	0.16052	0.12023	0.35014	0.46031

As for the correlation, the model and ERA5 satellite data agree by 35%, which is significantly better than in the previous case. Fig. 2 presents the frequency histograms of Pearson correlation and index of agreement results. Index of agreement chart is very similar to that on the Fig. 1, and its average value (46%) is also nearly the same. As for the correlation, its values are distributed rather evenly

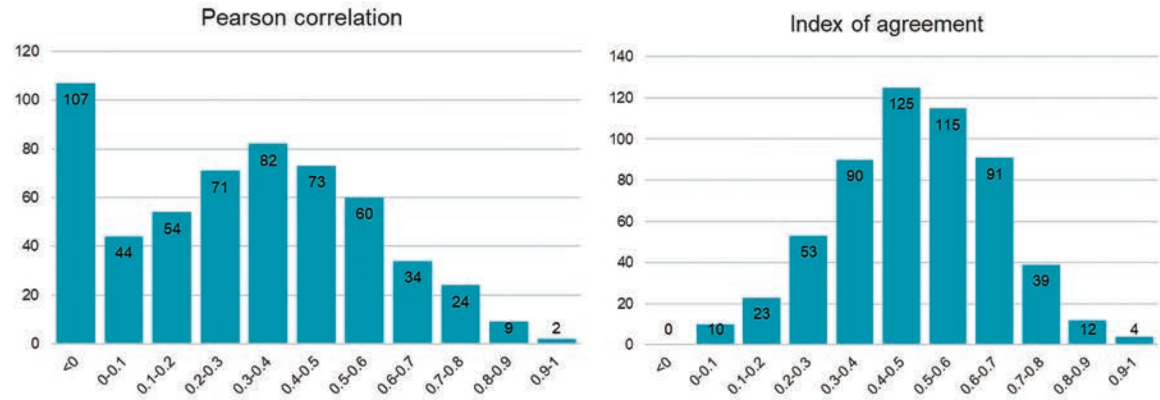

Fig. 1. Frequency histogram of the correlation (left) and index of agreement (right) between the model and the ground station data

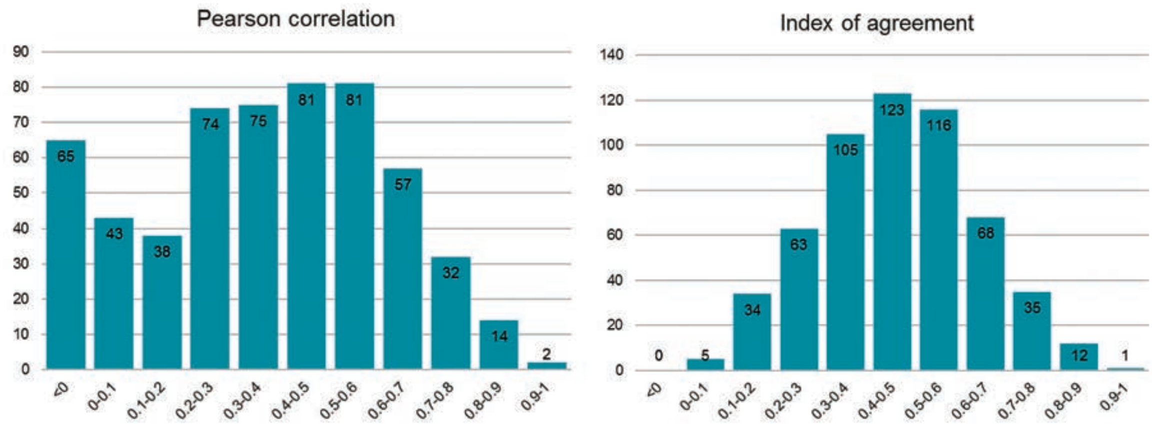

Fig. 2. Frequency histogram of the correlation (left) and index of agreement (right) between the model and ERA5 satellite data

on the histogram. Nevertheless, low and negative correlation values are less frequent than in the previous case.

The triple collocation analysis yielded an average 0.05258, 0.04290 and 0.07473 variances for ground station measurements, satellite observations and model simulations, respectively. These results suggest that ERA5 satellite data are the most accurate of the three, and ground sensor observations only slightly behind. The model appears to be the least accurate of the estimations, yet its error is comparable with that of in-situ measurements.

The most peculiar result of this triple collocation analysis might be the variance of the satellite observations. The lowest variance value implies that satellite data must be the most accurate of all considered data sources. However, it seems very likely that our satellite soil moisture data and those provided by the ERA5, though calculated independently and by different algorithms, might be derived from the same satellite images. If that is the case, then the three datasets used for analysis are not truly independent, as is required by the triple collocation methods. Recent research shows, that thought actual independence of data sources cannot be guaranteed, it does not influence the results significantly [32]. To avoid making unfounded inferences, we leave out the results of triple collocation analysis for satellite retrievals and further discuss only the interdependence between model results and in-situ measurements.

Fig. 3 demonstrates the relation between the variances of ground station and model data. The dots represent the variance pairs, and the thin black line represents the equality $y = x$. Most of the pairs are above the line, meaning model estimations are mostly less credible than in-situ measurements. However, the points are mostly clustered near the equality line, and 29% of the model variances are better than that of the measurements. We consider it a favorable result, since it proves that in a sufficient number of cases model estimations are as accurate as the ground sensors.

Conclusion. The chief purpose of our article and the presented experiment was to compare the model performance with the ground sensor measurements, and to determine whether the satellite model can be a reliable alternative for the ground station. Thought direct comparison of the model predictions against the benchmarks suggested rather low accuracy of the former, triple collocation analysis showed that overall accuracy of the model was not far from the ground sensors. Nearly 30% of model results were estimated to have even better accuracy than in-situ measurements.

As suggested by our results, the mathematical model still faces a number of problems, such as determining model parameters, taking into account all essential physical processes, finding reliable weather and assimilation data etc. However, it still has advantages over the traditional measurements methods as it does not require installing additional sensors, and can predict moisture even in belowground layers. The results are less accurate then that acquired by the direct measurements, but still the model accuracy is comparable to measured results.

In addition, the model has a vast potential for improvement. Boundary conditions may be considering groundwater level, soil parameters on different soil depths and factors such as soil hysteresis and temperature driven water flow may be taken into account. Assimilation methods and satellite accuracy are likewise improving, which is hopefully indicating the high potential of current research.

Acknowledgements. This research is supported and funded by EOS Data Analytics.

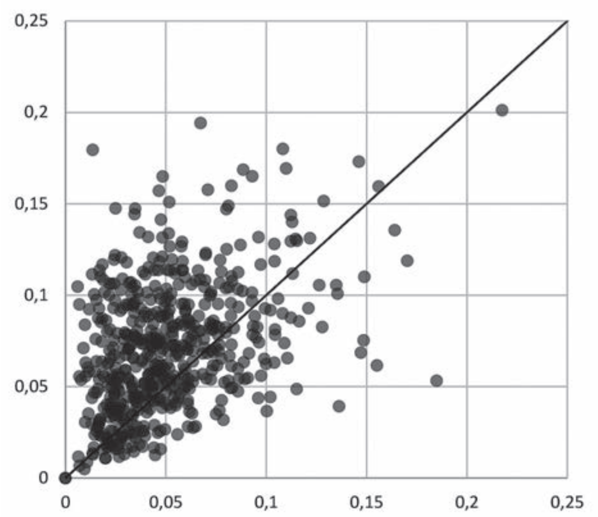

Fig. 3. Scatterplot of the model variance (vertical axis) against the ground station measurements (horizontal axis), estimated by the triple collocation method

BIBLIOGRAPHY

1. Lekshmi S., Singh D., Baghini M. S. A critical review of soil moisture measurement. *Measurement*. 2014. Vol. 54. P. 92–105. URL: https://doi.org/10.1016/j.measurement.2014.04.007.
2. Shen X., Liang J., Zeleke K. T. et al. Optimizing the Positioning of Soil Moisture Monitoring Sensors in Winter Wheat Fields. *Water*. 2018. Vol. 10. No. 12/1707. URL: https://doi.org/10.3390/w10121707.
3. Tian Y., Xiong L., Bin X., Zhuang R. A prior estimation of the spatial distribution parameter of soil moisture storage capacity using satellite-based root-zone soil moisture data. *Remote Sensing*. 2019. Vol. 11, No. 2580. URL: https://doi.org/10.3390/rs11212580.

4. Chen W., Shen H., Huang C., Li X. Improving soil moisture estimation with a dual ensemble Kalman smoother by jointly assimilating AMSR-E brightness temperature and MODIS LST. *Remote Sensing*. 2017. Vol. 9, No. 273. URL: https://doi.org/10.3390/rs9030273.
5. Nearing G., Yatheendradas S., Crow W. et al. The efficiency of data assimilation. *Water Resources Research*. 2018. Vol. 54. No. 9, P. 6374–6392. URL: https://doi.org/10.1029/2017WR020991.
6. Houser P. R., de Lannoy G. J. M., Walker J. P. Hydrologic data assimilation. *Approaches to Managing Disaster – Assessing Hazards, Emergencies and Disaster Impacts* / ed. by J. Tiefenbacher. IntechOpen: Rijeka, 2012. P. 41–64. URL: https://doi.org/10.5772/31246.
7. Martens B., Miralles D. G., Lievens H. et al. GLEAM v3: satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development*. 2017. Vol. 10. P. 1903–1925. URL: https://doi.org/10.5194/gmd-10-1903-2017.
8. Paniconi C., Marrocu M., Putti M., Verbunt M. Newtonian nudging for a Richards equation-based distributed hydrological model. *Advances in Water Resources*. 2003. Vol. 26. P. 161–178. URL: https://doi.org/10.1016/S03091708(02)00099-4.
9. Scipal K., Dorigo W., de Jeu R. Triple collocation — a new tool to determine the error structure of global soil moisture products. *Geoscience and Remote Sensing Symposium (IGARSS)*: conference proceedings, Honolulu, 25-30 July 2010. IEEE, 2010. P. 4426–4429. URL: https://doi.org/10.1109/IGARSS.2010.5652128.
10. Kedzior M., Zawadzki J. Comparative study of soil moisture estimations from SMOS satellite mission, GLDAs database, and cosmic-ray neutrons measurements at COSMOS station in eastern Poland. *Geoderma*. 2016. Vol. 283. P. 21–31. URL: https://doi.org/10.1016/j.geoderma.2016.07.023.
11. Hang G., Ng J. L., Huang Y., Yong S. Performance of potential evapotranspiration models in Peninsular Malaysia. *Journal of Water and Climate Change*. 2021. URL: https://doi.org/10.2166/wcc.2021.018. (Preprint. IWA Publishing).
12. Melo M., de Jong van Lier Q. Revisiting the Feddes reduction function for modeling root water uptake and crop transpiration. *Journal of Hydrology*. 2021. Vol. 603. No. 126952. URL: https://doi.org/10.1016/j.jhydrol.2021.126952.
13. Terleev V., Mirschel W., Badenko V., Guseva I. An improved Mualem–Van Genuchten method and its verification using data on Beit Netofa clay. *Eurasian Soil Science*. 2017. Vol. 50, P. 445–455. URL: https://doi.org/10.1134/S1064229317040135.
14. Zhang Y., Schaap M. G. Weighted recalibration of the Rosetta pedotransfer model with improved estimates of hydraulic parameter distributions and summary statistics (Rosetta3). *Journal of Hydrology*. 2017. Vol. 547, P. 39–53. URL: https://doi.org/10.1016/j.jhydrol.2017.01.004.
15. Samarskiy A. A. *The theory of difference schemes*. New York: Marcel Dekker, 2001. 786 p.
16. Houser P. R., Shuttleworth W. J., Famiglietti J. S. et al. Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. *Water Resources Research*. 1998. Vol. 34. No. 12. P. 3405–3420. URL: https://doi.org/10.1029/1998WR900001.
17. Conti G., Aydoğdu A., Gualdi S. et al. On the physical nudging equations. *Climate Dynamics*. 2021. URL: https://doi.org/10.1007/s00382-021-05972-w. (Preprint. Springer Nature).
18. Camporese M., Paniconi C., Putti M., Salandin P. Comparison of data assimilation techniques for a coupled model of surface and subsurface flow. *Vadose Zone Journal*. 2009. Vol. 8. No. 4. P. 837–845. URL: https://doi.org/10.2136/vzj2009.0018.
19. Mirosław-Świątek D. Application of Newtonian nudging data assimilation method in hydrodynamic model of flood flow in the lower Biebrza basin. *Studia Geotechnica et Mechanica*. 2012. Vol. 34, No. 2. P. 91–105. URL: https://doi.org/10.2478/sgm021208.
20. Vidard A., le Dimet F. X., Piacentini A. Optimal determination of nudging coefficients. *Tellus A*. 2003. Vol. 55. No. 1, P. 1–15. URL: https://doi.org/10.3402/tellusa.v55i1.14576.
21. Stauffer D. R., Bao J. W. Optimal determination of nudging coefficients using the adjoint equations. *Tellus A: Dynamic Meteorology and Oceanography*. 1993. Vol. 45. No. 5. P. 358–369. URL: https://doi.org/10.3402/tellusa.v45i5.14899.
22. Reichle R. H., Koster R. D., Liu P. et al. Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR). *Journal of Geophysical Research*. 2007. Vol. 112. No. D09108. URL: https://doi.org/10.1029/2006JD008033.
23. Holmes T. R. H., de Jeu R. A. M., Owe M., Dolman A. J. Land surface temperature from Ka band (37 GHz) passive microwave observations. *Journal of Geophysical Research*. 2009. Vol. 114, No. D04113. URL: https://doi.org/10.1029/2008JD010257.

24. Chan S., Bindlish R., O'Neill P. et al. Assessment of the SMAP passive soil moisture product. *IEEE Transactions on Geoscience and Remote Sensing*. 2016. Vol. 54. P. 1–14. URL: https://doi.org/10.1109/TGRS.2016.2561938.
25. Mironov V. L., Kerr Y., Wigneron J.-P. et al. Temperature- and texture-dependent dielectric model for moist soils at 1.4 GHz. *IEEE Geoscience and Remote Sensing Letters*. 2013. Vol. 10. P. 419–423. URL: https://doi.org/10.1109/LGRS.2012.2207878.
26. Gruber A., Dorigo W. A., Crow W., Wagner W. Triple collocation-based merging of satellite soil moisture retrievals. *IEEE Transactions on Geoscience and Remote Sensing*. 2017. Vol. 55. P. 6780–6792. URL: https://doi.org/10.1109/TGRS.2017.2734070.
27. Paulik C., Preimesberger W., Hahn S. et al. TUW-GEO/pytesmo: v0.12.0 (Python library). *Zenodo*. URL: https://doi.org/10.5281/zenodo.596422.
28. Dorigo W., Wagner W., Hohensinn R. et al. The International Soil Moisture Network: a data hosting facility for global in situ soil moisture measurements. *Hydrology and Earth System Sciences*. 2011. Vol. 15. P. 1675–1698. URL: https://doi.org/10.5194/hess-15-1675-2011.
29. Hersbach H., Bell B., Berrisford P. et al. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*. 2020. Vol. 146, P. 1999–2049. URL: https://doi.org/10.1002/qj.3803.
30. Thea C. M. LaMetSy, Aggregated meteorogical database. URL: https://lametsy.pp.ua.
31. Poggio L., de Sousa L., Batjes N. et al. SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. *Soil*. 2021. Vol. 7, P. 217–240. URL: https://doi.org/10.5194/soil-7-217-2021.
32. Yilmaz M., Crow W. Evaluation of assumptions in soil moisture triple collocation analysis. *Journal of Hydrometeorology*. 2014. Vol. 15, P. 1293–1302. URL: https://doi.org/10.1175/JHM-D-13-0158.1.

REFERENCES

1. Lekshmi, S., Singh, D., & Baghini, M. S. (2014). A critical review of soil moisture measurement. *Measurement, 54*, 92–105. https://doi.org/10.1016/j.measurement.2014.04.007
2. Shen, X., Liang, J., Zeleke, K. T., Liang, Y., Wang, G., Duan, A., Mi, Z., Ning, H., Gao, Y., & Zhang, J. (2018). Optimizing the Positioning of Soil Moisture Monitoring Sensors in Winter Wheat Fields. *Water, 10*(12), 1707. https://doi.org/10.3390/w10121707
3. Tian, Y., Xiong, L., Bin, X., & Zhuang, R. (2019). A prior estimation of the spatial distribution parameter of soil moisture storage capacity using satellite-based root-zone soil moisture data. *Remote Sensing*, *11*, 2580. https://doi.org/10.3390/rs11212580
4. Chen, W., Shen, H., Huang, C., & Li, X. (2017). Improving soil moisture estimation with a dual ensemble Kalman smoother by jointly assimilating AMSR-E brightness temperature and MODIS LST. *Remote Sensing, 9*, 273. https://doi.org/10.3390/rs9030273
5. Nearing, G., Yatheendradas, S., Crow, W., Zhan, X., Liu, J., & Chen, F. (2018). The efficiency of data assimilation. *Water Resources Research, 54*(9), 6374–6392. https://doi.org/10.1029/2017WR020991
6. Houser, P. R., de Lannoy, G. J. M., & Walker, J. P. (2012). Hydrologic data assimilation. In J. Tiefenbacher (Ed.), *Approaches to Managing Disaster – Assessing Hazards, Emergencies and Disaster Impacts* (pp. 41–64). IntechOpen. https://doi.org/10.5772/31246
7. Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., Beck, H. E., Dorigo, W. A., & Verhoest, N. E. C. (2017). GLEAM v3: satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development, 10*, 1903–1925. https://doi.org/10.5194/gmd-10-1903-2017
8. Paniconi, C., Marrocu, M., Putti, M., & Verbunt, M. (2003). Newtonian nudging for a Richards equation-based distributed hydrological model. *Advances in Water Resources, 26*, 161–178. https://doi.org/10.1016/S03091708(02)00099-4
9. Scipal, K., Dorigo, W., & de Jeu, R. (2010). Triple collocation — a new tool to determine the error structure of global soil moisture products. *Geoscience and Remote Sensing Symposium (IGARSS), Honolulu, 2010*, 4426–4429. https://doi.org/10.1109/IGARSS.2010.5652128
10. Kedzior, M., & Zawadzki, J. (2016). Comparative study of soil moisture estimations from SMOS satellite mission, GLDAs database, and cosmic-ray neutrons measurements at COSMOS station in eastern Poland. *Geoderma, 283,* 21–31. https://doi.org/10.1016/j.geoderma.2016.07.023
11. Hang, G., Ng, J. L., Huang, Y., & Yong, S. (2021). Performance of potential evapotranspiration models in Peninsular Malaysia. *Journal of Water and Climate Change*, preprint. https://doi.org/10.2166/wcc.2021.018
12. Melo, M., & de Jong van Lier, Q. (2021). Revisiting the Feddes reduction function for modeling root water uptake and crop transpiration. *Journal of Hydrology, 603*, 126952. https://doi.org/10.1016/j.jhydrol.2021.126952

13. Terleev, V., Mirschel, W., Badenko, V., & Guseva, I. (2017). An improved Mualem–Van Genuchten method and its verification using data on Beit Netofa clay. *Eurasian Soil Science, 50*, 445–455. https://doi.org/10.1134/S1064229317040135
14. Zhang, Y., & Schaap, M. G. (2017). Weighted recalibration of the Rosetta pedotransfer model with improved estimates of hydraulic parameter distributions and summary statistics (Rosetta3). *Journal of Hydrology, 547*, 39–53. https://doi.org/10.1016/j.jhydrol.2017.01.004
15. Samarskiy, A. A. (2001). *The theory of difference schemes*. Marcel Dekker.
16. Houser, P. R., Shuttleworth, W. J., Famiglietti, J. S., Gupta, H. V., Syed, K. H., & Goodrich, D. C. (1998). Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. *Water Resources Research, 34*(12), 3405–3420. https://doi.org/10.1029/1998WR900001
17. Conti, G., Aydoğdu, A., Gualdi, S., Navarra, A., & Tribbia, J. (2021). On the physical nudging equations. *Climate Dynamics*, preprint. https://doi.org/10.1007/s00382-021-05972-w
18. Camporese, M., Paniconi, C., Putti, M., & Salandin, P. (2009). Comparison of data assimilation techniques for a coupled model of surface and subsurface flow. *Vadose Zone Journal, 8*(4), 837–845. https://doi.org/10.2136/vzj2009.0018
19. Mirosław-Świątek, D. (2012). Application of Newtonian nudging data assimilation method in hydrodynamic model of flood flow in the lower Biebrza basin. *Studia Geotechnica et Mechanica, 34*(2), 91–105. https://doi.org/10.2478/sgm021208
20. Vidard, A., le Dimet, F. X., & Piacentini, A. (2003). Optimal determination of nudging coefficients. *Tellus A, 55*(1), 1–15. https://doi.org/10.3402/tellusa.v55i1.14576
21. Stauffer, D. R., & Bao, J. W. (1993). Optimal determination of nudging coefficients using the adjoint equations. *Tellus A: Dynamic Meteorology and Oceanography, 45*(5), 358–369. https://doi.org/10.3402/tellusa.v45i5.14899
22. Reichle, R. H., Koster, R. D., Liu, P., Mahanama, S. P. P., Njoku, E. G., & Owe, M. (2007). Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR). *Journal of Geophysical Research, 112*, Article D09108. https://doi.org/10.1029/2006JD008033
23. Holmes, T. R. H., de Jeu, R. A. M., Owe, M., & Dolman, A. J. (2009). Land surface temperature from Ka band (37 GHz) passive microwave observations. *Journal of Geophysical Research, 114*, Article D04113. https://doi.org/10.1029/2008JD010257
24. Chan, S., Bindlish, R., O'Neill, P., Njoku, E., Jackson, T., Colliander, A., Chen, F., Burgin, M., Dunbar, S., Piepmeier, J., Yueh, S., Entekhabi, D., Cosh, M., Caldwell, T., Walker, J., Wu, X., Berg, A., Rowlandson, T., Pacheco, A., & Kerr, Y. (2016). Assessment of the SMAP passive soil moisture product. *IEEE Transactions on Geoscience and Remote Sensing, 54*, 1–14. https://doi.org/10.1109/TGRS.2016.2561938
25. Mironov, V. L., Kerr, Y., Wigneron, J.-P., Kosolapova, L., & Demontoux, F. (2013). Temperature- and texture-dependent dielectric model for moist soils at 1.4 GHz. *IEEE Geoscience and Remote Sensing Letters, 10*, 419–423. https://doi.org/10.1109/LGRS.2012.2207878
26. Gruber, A., Dorigo, W. A., Crow, W., & Wagner, W. (2017). Triple collocation-based merging of satellite soil moisture retrievals. *IEEE Transactions on Geoscience and Remote Sensing, 55*, 6780–6792. https://doi.org/10.1109/TGRS.2017.2734070
27. Paulik, C., Preimesberger, W., Hahn, S., Baum, D., Plocon, A., Mistelbauer, Th., & Schmitzer M. (2021). *TUW-GEO/pytesmo: v0.12.0* (Python library). Zenodo. https://doi.org/10.5281/zenodo.596422
28. Dorigo, W., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., Gruber, A., Drusch, M., Mecklenburg, S., Oevelen, P., Robock, A., & Jackson, Tj. (2011). The International Soil Moisture Network: a data hosting facility for global in situ soil moisture measurements. *Hydrology and Earth System Sciences, 15*, 1675–1698. https://doi.org/10.5194/hess-15-1675-2011
29. Hersbach, H., Bell, B., Berrisford, P., Hirahara, Sh., Horányi A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X, Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J. N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society, 146*, 1999–2049. https://doi.org/10.1002/qj.3803
30. Thea C. M. (2019). *LaMetSy, Aggregated meteorogical database* [Data set]. https://lametsy.pp.ua.
31. Poggio, L., de Sousa, L., Batjes, N., Heuvelink, G., Kempen, B., Ribeiro, E., & Rossiter, D. (2021). SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. *Soil, 7*, 217–240. https://doi.org/10.5194/soil-7-217-2021
32. Yilmaz, M., & Crow, W. (2014). Evaluation of assumptions in soil moisture triple collocation analysis. *Journal of Hydrometeorology, 15*, 1293–1302. https://doi.org/10.1175/JHM-D-13-0158.1