Evaluating predictions of the soil moisture model with data assimilation by the triple collocation method

Olena Kozhushko, Mykhailo Boiko, Petro Martyniuk,
Olha Stepanchenko
EOS Data Analytics,
Department of Computer Sciences and Applied Mathematics
National University of Water and Environmental Engineering
Rivne, Ukraine
ol.d.kozhushko@nuwm.edu.ua

Mykola Kovbasa
EOS Data Analytics,
Division of Physico-Technological Problems in
Semiconductor IR Engineering
V.Ye. Lashkaryov Institute of Semiconductor Physics
of the NAS of Ukraine
Kyiv , Ukraine

Mykola Uvarov
EOS Data Analytics,
Department of Computational Physics
G. V. Kurdyumov Institute for Metal Physics
Kyiv , Ukraine

Abstract— **This paper describes a nonlinear soil moisture transport problem, solved with addition of satellite soil moisture observations. The satellite data are assimilated into the model using Newtonian nudging method. Evaluation is done by the use of triple collocation method, which involves three data sources: model results, ground station measurements and satellite observations. The results demonstrates that the presented model is capable of producing results with accuracy close to the ground station measurements.**

Keywords— data assimilation; Newtonian nudging; mathematical modelling; soil moisture; triple collocation

I. 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 accurate data provided by either measurements or model simulations.

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 expensive and provide only isolated data at one observation point.

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. 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. However, they unavoidably contain errors due to generalizations of process physics and choice of model parameters.

Since each of the described methods has its own strengths and weaknesses, the best results are achieved by combining different data sources. For example, insertion of satellite data into the moisture model is widely used in practice. This approach, called data assimilation, enables the model to adapt to observed data, and to achieve higher precision level than provided by each data source alone. There are multiple data assimilation algorithms available, some of which are described in [1]. 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 [2]. 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 [3, 4].

II. MATHEMATICAL MODEL

A. 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:

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

$$\left(-k(h)\frac{\partial h}{\partial x} + k(h)\right)\bigg|_{x=0} = Q(t) - E_s(t), t > 0, \qquad (2)$$

$$\frac{\partial h}{\partial x}\bigg|_{x=0} = 0, t > 0, \qquad (3)$$

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

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 [6].

Since the Richards equation requires a translation rule between moisture and pressure head values, we chose a widespread Mualem–van Genuchten model [7]. The model is represented by the following equations:

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

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

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 [8].

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

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

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

The one-dimensional problem described above is then discretized using an implicit time scheme and solved on a uniform grid with. 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 [9].

B. Newtonian nudging assimilation

Newtonian nudging is a smoother algorithm that modifies the simulation result on each time based on the past and future observations. The method is not as popular as the filtering methods, e.g. ensemble Kalman filters, but it is suited for boundary-value problems since it is incorporated directly into the model governing equation. Unlike the filtering methods that treat the solution as a random variable, Newtonian 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:

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

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 the 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 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 [10].

Original formulation of Newtonian nudging uses constant nudging factor. However, we use an adaptive relation to calculate the factor considering current soil retention and permeability:

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

III. SATELLITE MOISTURE RETRIEVAL

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.

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 [11] 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 [12], Dobson model [13] for C-band to convert dielectric permittivity to soil moisture content.

IV. RESULTS AND DISCUSSION

A. Evaluation metrics

We apply the triple collocation method to combine and assess the errors of ground station measurements, satellite observations and model simulations. The method requires three data sets θ_1, θ_2 and θ_3, each containing the same number of estimations of some variable. Each of them differs from the hypothetical truth θ by a residual r, 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 quality of each dataset can then be estimated as a variance of the residuals, denoted σ_1, σ_2, σ_3 respectively. After eliminating the hypothetical truth from the equations (10) and taking average over the resulting equation, we get the following expressions:

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

where $\langle \cdot \rangle$ denotes covariance operator. Note that the method statistical justification implies the datasets are unbiased, so any bias should be removed first [4].

The model results are also evaluated against the ground station data using the traditional metrics, such as average absolute deviation (AAD), root mean square deviation (RMSD), bias, Pearson correlation (R) and index of agreement (IoA).

B. Numerical experiments

To achieve statistically relevant accuracy evaluation, we conducted a large-scale numerical experiment using the open ground station data provided by the International Soil Moisture Network (ISMN) [14].

The weather data were downloaded from the NOAA database as for the nearest meteorological station. Soil parameters were assumed based on the soil type individually for each station, based on the data provided by SoilGrids. Initial conditions were set according to the satellite moisture data.

The experiment was conducted for 2018 over 659 ground stations in the USA. Out of them, 178 stations were excluded because of the issues with other data for the location. Therefore, the summary presented below includes experiment results for the remaining 481 stations. Note 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.

C. Results

We start from the comparison of model simulation results and ground station observations. The averaged metrics over all stations are presented in Table 1. The analysis shows that absolute deviation and RMSD are rather high, but they are mostly caused by bias. The bias 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 average unbiased RMSE is 0.08758, which indicates that the results can be improved greatly if the source of bias is eliminated. The average correlation is rather weak, however, the index of agreement implies a tolerably good convergence of the datasets.

TABLE I. EVALUATION OF THE MODEL RESULTS AGAINST GROUND STATION MEASUREMENTS

Metrics	AAD	RMSD	Bias	R	IoA
Value	0.10948	0.13039	0.06554	0.25661	0.46532

Further analysis of the latter two characteristics is shown on Fig. 1. The first frequency chart 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.
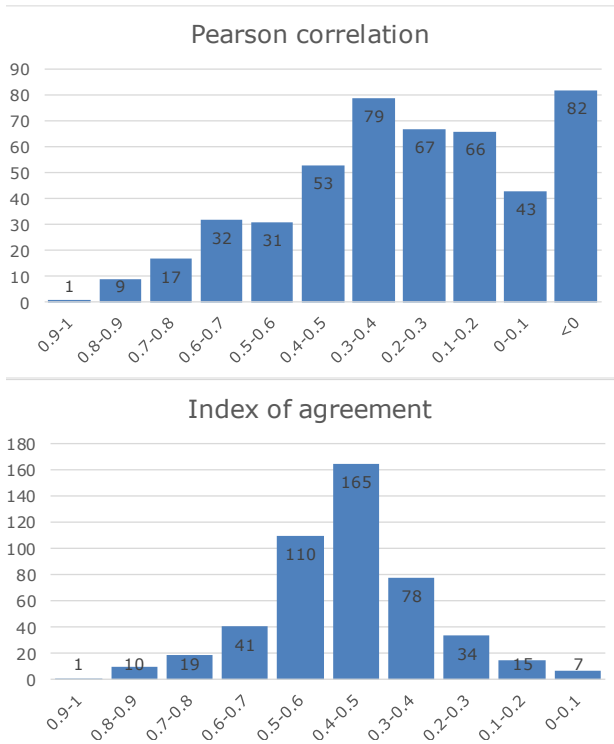
Figure 1. Frequency histogram of the correlation (top) and index of agreement (bottom) between the model and the ground station data
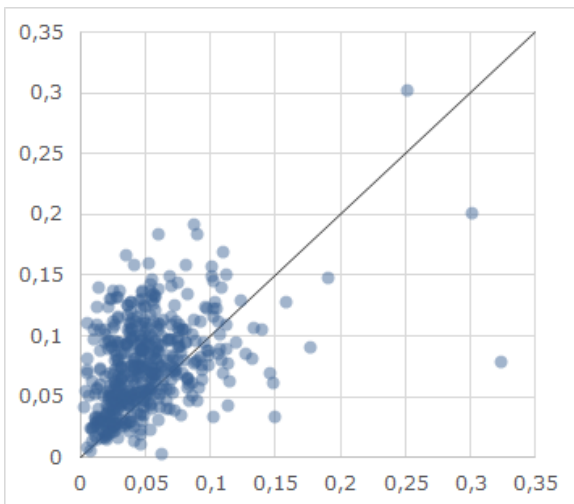

Figure 2. Scatterplot of the model variance (vertical axis) against the ground station measurements (horizontal axis), estimated by the triple collocation method

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 the model is the least accurate of the estimations, yet its error is comparable with that of in-situ measurements.

Fig. 2 demonstrates the relation between the variances of ground station and model data. The dots represent the variance pairs, and the thin black line represents $y = x$. 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 line, and more than 25% of the model variances are better than that of the measurements. This can be considered a favorable result, since it proves that in a sufficient number of cases model estimations are as accurate as the ground sensors.

REFERENCES

[1] P.R. Houser, G.J.M. de Lannoy and J.P. Walker, "Hydrologic data assimilation," in Approaches to Managing Disaster – Assessing Hazards, Emergencies and Disaster Impacts, J. Tiefenbacher, Ed. IntechOpen: Rijeka, 2012, pp. 41–64.

[2] B. Martens, D.G. Miralles, H. Lievens, R. van der Schalie, R.A.M. de Jeu, D. Fernández-Prieto et al., "GLEAM v3: satellite-based land evaporation and root-zone soil moisture," Geoscientific Model Development , vol. 10, pp. 1903–1925, 2017.

[3] K. Scipal, W. Dorigo and R. de Jeu, "Triple collocation — A new tool to determine the error structure of global soil moisture products," International Geoscience and Remote Sensing Symposium (IGARSS), pp. 4426-4429, 2010.

[4] M. Kedzior and J. Zawadzki, "Comparative study of soil moisture estimations from SMOS satellite mission, GLDAs database, and cosmic-ray neutrons measurements at COSMOS station in eastern Poland," Geoderma, vol. 283, pp. 21–31, 2016.

[5] C.-Y. Xu and V. P. Singh, "Cross comparison of empirical equations for calculating potential evapotranspiration with data from Switzerland," Water Resources Management, vol. 16, pp. 197–219, 2002.

[6] R.A. Feddes, P.T. Kowalik and H. Zaradny, "Simulation of field water use and crop yield," Pudoc: Wageningen, 1978.

[7] M.G. Schaap and M.Th. van Genuchten, "A modified Mualem–van Genuchten formulation for improved description of the hydraulic conductivity near saturation," Vadose Zone J., vol. 5, pp. 27–34, 2005.

[8] Y. Zhang and M.G. Schaap, "Weighted recalibration of the Rosetta pedotransfer model with improved estimates of hydraulic parameter distributions and summary statistics (Rosetta3)," Journal of Hydrology, vol. 547, pp. 39–53, 2017.

[9] A.A. Samarskiy, "The theory of difference schemes," New York: Marcel Dekker, 2001, pp. 145-223.

[10] C. Paniconi, M. Marrocu, M. Putti and M. Verbunt, "Newtonian nudging for a Richards equation-based distributed hydrological model," Advances in Water Resources, vol. 26, pp. 161–178, 2003.

[11] T.R.H. Holmes, R.A.M. de Jeu, M. Owe and A.J. Dolman, "Land surface temperature from Ka band (37 GHz) passive microwave observations," J. Geophys. Res., vol. 114, no. D04113, 2009.

[12] S. Chan, R. Bindlish, P. O'Neill, E. Njoku, T. Jackson, A. Colliander et al., "Assessment of the SMAP passive soil moisture product," IEEE Transactions on Geoscience and Remote Sens., vol. 54, pp. 1–14, 2016.

[13] V.L. Mironov, Y. Kerr, J.-P. Wigneron, L. Kosolapova and F. Demontoux, "Temperature- and texture-dependent dielectric model for moist soils at 1.4 GHz," IEEE Geoscience and Remote Sensing Letters, vol. 10, pp. 419–423, 2013.

[14] W.A. Dorigo, W. Wagner, R. Hohensinn, S. Hahn, C. Paulik, A. Xaver et al., "The International Soil Moisture Network: a data hosting facility for global in situ soil moisture measurements," Hydrol. Earth Syst. Sci., vol. 15, pp. 1675–1698, 2011