Global Shallow Groundwater Patterns From Soil Moisture Satellite Retrievals

Mehmet Evren Soylu and Rafael L. Bras

***Abstract*—Groundwater is the most significant freshwater source and plays a critical role in the earth's water and energy balance. The lack of groundwater observations with a high spatiotemporal resolution at a global scale hinders our ability to study and model the environment when shallow groundwater has a direct impact on surface soil moisture. This study aims to estimate the spatial and temporal distributions of shallow groundwater-influenced areas at a global scale. We trained an ensemble machine learning algorithm, using outputs from a variably saturated soil moisture flux model, to identify the shallow groundwater occurrence. Model simulations spanned various climate zones and soil types across the globe. The overall accuracy of the algorithm in reproducing the soil moisture flux model results was 95.5%. We applied the algorithm to space-borne soil moisture observations retrieved by NASA's SMAP satellite and present a global-scale shallow groundwater map derived from the SMAP observations. The derived global distribution of shallow groundwater identifies wetlands, large riparian corridors, and seasonally inundated lowlands. The results showed that 19% of terrestrial land cover had been influenced by shallow groundwater at some point in time during the period of interest (2015–2018). Temporally, shallow groundwater follows an annual cyclic pattern with 2% to 6% of the land surface being influenced globally. This study shows that SMAP observations could be used in estimating shallow groundwater in high spatiotemporal resolution at a global scale, potentially providing invaluable inputs for modeling and environmental monitoring studies.**

***Index Terms*—Hydrology, image processing, machine learning, shallow groundwater, soil moisture active passive (SMAP), soil moisture.**

I. Introduction

GROUNDWATER (GW) is a significant source of domestic, agricultural, and industrial consumption. It impacts the global hydrologic and energy cycles. Shallow GW, in particular, directly affects evaporation and evapotranspiration with corresponding effects on the biosphere. It profoundly impacts water, energy, and carbon cycles by providing additional water to ecosystems and the land surface. Shallow GW influences global food security [1], ecosystem services [2], flood regulation [3], and climate [4].

Manuscript received September 9, 2021; revised October 26, 2021; accepted October 26, 2021. Date of publication November 4, 2021; date of current version December 20, 2021. This work was supported by the National Aeronautics and Space Administration under Grant 80NSSC20K1795, and by the K. Harrison Brown Family Chair Funding at Georgia Institute of Technology. *(Corresponding author: Mehmet Evren Soylu.)*

The authors are with the School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0002 USA (e-mail: msoylu3@gatech.edu; rlbras@gatech.edu).

Digital Object Identifier 10.1109/JSTARS.2021.3124892

Modeling GW and shallow GW globally is hindered by computational complexity and the lack of high-resolution information on lithological boundaries, soil properties, and knowledge of boundary and initial conditions. Some terrestrial water flux models have tried to simulate the GW coupling with the atmosphere at the scales from the watershed [5] to global [6], but they still lack the global datasets with a high spatial and temporal resolution to calibrate and verify results.

GW interacts with the land surface fluxes only when the water table is near the land surface. It influences the global land surface by sustaining surface water features and elevating soil moisture within the vadose zone (a layer between the land surface and the water table). GW interacts with vegetation by altering the soil moisture state within the unsaturated root zone. Elevated root zone soil moisture due to exfiltration from shallow GW may benefit the plants by contributing to root water uptake or may be harmful by causing hypoxia if the root zone is at or near saturation [7], [8].

GW is the least observed component of the global water cycle. Monitoring GW levels requires installing GW wells or piezometers. Although wells are the most direct and accurate way of measuring water table depths (WTDs), they only provide WTD measurements at a point scale. At larger scales, shallow GW may be estimated using various indicators such as topography, vegetation type, soil salinity, and flood frequency. However, these indicators only provide qualitative information about shallow GW with considerable uncertainty. In order to obtain quantitative information about GW beyond the point scale, GW observation networks with multiple wells are needed to monitor the distribution of GW levels due to the heterogeneous nature of the subsurface structure. Installing and maintaining GW well networks is labor-intensive and expensive, limiting the availability of GW data in many regions across the world [9].

Remote sensing has been increasingly used to observe components of the hydrologic cycle. GW was not a subject of satellite remote sensing until the early 2000s when the National Aeronautics and Space Administration's (NASA) Gravity Recovery and Climate Experiment (GRACE) mission was launched [10]. The GRACE satellite can be used to estimate changes in GW storage by monitoring the variations in subsurface mass via satellite orbit perturbations due to land gravitational anomalies [11]. Although GRACE observations are instrumental in monitoring GW storage changes at a global scale, its coarse spatial ($\sim$150 000 km^2 at midlatitudes) and temporal (monthly) resolution make it less useful for hydrological studies

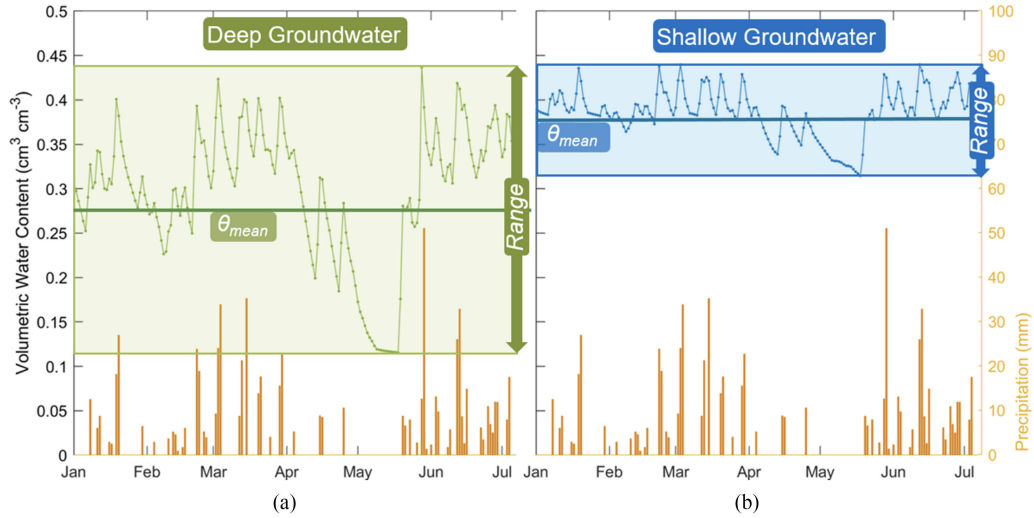


Fig. 1. Comparison of the impact of shallow and deep water table conditions on surface soil moisture. Model simulations of surface soil moisture (top 5 cm) variations under two WTD scenarios. (a) GW level is deep enough that it does not have any influence on surface soil moisture. (b) Water table is constant at 0.5 m from the surface increasing the overall average soil moisture and reducing its variation. Meteorological data to drive the Hydrus-1D model are obtained from a weather station located at Aiken, SC for the year 2001.

focusing on smaller scales. Higher-resolution but less direct remote sensing-based GW estimation methods are also available. Some of these methods include thermal remote sensing, capturing shallow GW via thermal anomalies on the surface [12], [13], [14]. Interferometric synthetic aperture radar can also be used to estimate GW storage changes depending on subtle surface elevation changes [15]. L-band microwave sensors to monitor soil moisture via measuring the natural thermal emission emanating from the soil surface have recently attracted attention as a promising approach to estimate shallow GW more accurately [16], [17].

Soil moisture at the vadose zone is a strong proxy for GW. GW influence may be captured from the surface soil moisture signals in areas where GW is shallow enough to couple with surface fluxes. Remote sensing of soil moisture, therefore, can be a viable option to detect shallow GW. NASA's Soil Moisture Active Passive (SMAP) mission monitors global soil moisture by measuring the natural thermal emission of the surface. The sensors use L-band microwave emissions, focusing on the top 5 cm of the soil column. The SMAP mission L-band microwave satellite images have been recently used to capture the signals of shallow GW at a global scale with high temporal (3 days) and spatial (∼9 km) resolution when the level 2 enhanced passive soil moisture (L2_SM_P_E) product is used [17]. L2_SM_P_E data are derived from lower resolution (36 km) SMAP passive microwave radiometer L-band brightness temperatures by applying the Backus-Gilbert optimal interpolation technique to the antenna temperature measurements in the original SMAP product. The main advantage of using the enhanced high-resolution product is to obtain a less pixelated representation of the soil moisture distributions, which can reveal more detailed spatial features that are not immediately obvious from the native resolution product. To identify the areas under the influence of shallow GW and their temporal fractions during a year, Soylu and Bras [17] used an empirical-based shallow GW detection algorithm that processes the moving average and the variation of the surface soil moisture from the SMAP retrievals. They showed that both the monthly moving average of soil moisture observations and their coefficient of variation are sensitive to shallow GW regardless of the prevailing climate. Fig. 1 shows how GW influences the mean surface soil moisture and its variation based on soil moisture flux simulations. Although [17] have tested their algorithm in locations across three different continents, using this novel approach to detect shallow GW from soil moisture signals across the globe would require a larger-scale data sampling and algorithm training to properly represent an extensive range of climate, soils, and ecosystems. A machine learning approach may be a viable option to process the large amount of data collected from SMAP soil moisture retrievals at a global scale and to analyze shallow GW patterns more accurately and more widely under various climate and soil conditions.

Adopting machine learning algorithms for exploring complex quantitative relationships between hydrological and meteorological components and their feedback dynamics is becoming widely used by Earth scientists [18]. Ensemble machine learning (EML) methods are among the most popular machine learning methods in Earth sciences. Such methods have been successfully used to solve various problems in hydrology, including streamflow prediction [19], [20], drought monitoring [21], generating GW potential maps [22], soil moisture estimation [23], and GW contamination risk assessment [24], among other areas of research in water resources, climate, and ecology. Zounemat-Kermani *et al.* [25] provided an extensive review of the EML methods employed in the hydrological sciences.

Our main objective is to estimate shallow GW-influenced terrestrial land surfaces by processing SMAP soil moisture retrievals at a global scale with high spatial and temporal resolution. To achieve this objective, we used results from a variably saturated soil moisture flow model to train a tree EML method.

The resulting model is then used to detect shallow GW signatures using SMAP soil moisture retrievals at the global scale as input. We focused on analyzing SMAP soil moisture retrievals for four years between 2015 and 2018 and compared our findings with multiple shallow GW indicators, including baseflow estimations inferred from the USGS' streamflow observations across the southeastern US watersheds, global wetland datasets, and global clay horizon distributions for potential perched aquifer formations.

We explain our approach and describe the tools we used, including the variably saturated soil moisture flow model, its atmospheric forcing data, the EML method, and its verification specifics in the following Section II. Section III discusses the method performance, the spatiotemporal distribution of shallow GW patterns at a global scale, and the EML model verification. We provide a comparison between the outputs from the EML model and baseflow contributions of some streamflow observations and global wetland and clay horizon distribution data. Finally, we discuss the implications of the findings, some restrictions of the proposed approach, and its potential use from a broader perspective.

II. Methods

We applied a supervised EML algorithm to process SMAP soil moisture retrievals to detect areas under the influence of shallow GW. Simulations mimicking SMAP observations under controlled lower boundary conditions representing various GW depth scenarios are used to train a supervised EML model. The simulations covered various soil textures, climate regimes, and GW depths scenarios. Below, we describe the model, simulations, settings, and details of the EML approach that is adopted.

A. Soil Moisture Model Simulations

1) Variably Saturated Water Flow Model: The Hydrus-1D model was used to simulate the surface soil moisture in a wide range of climate and soil conditions in response to shallow GW influence. Hydrus-1D numerically solves the Richards equation to simulate variably saturated flow within a soil profile [26]. The Richards equation in the vertical dimension is

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial x}\left[K(h)\left(\frac{\partial h}{\partial z}+1\right)\right] - S(h) \tag{1}$$

where θ is volumetric water content [L^3L^{-3}], $K(h)$ is unsaturated hydraulic conductivity [L T^{-1}], h is pressure head [L], z is vertical coordinate positive upward [L], and $S(h)$ is the sink term [$L^3L^{-3}T^{-1}$] accounting for plant root water uptake. In order to relate hydraulic conductivity, pressure head, and volumetric water content, we used the soil parameter functions of [27], which are

$$E_s(h) = \frac{\theta(h)-\theta_r}{\theta_s-\theta_r} = \begin{cases} (1+|\alpha h|^n)^{-m} & h<0 \\ 1 & h \geq 1 \end{cases}$$

$$K(h) = K_s\, E_s^l\left[1-\left(1-E_s^{1/m}\right)^m\right]^2 \tag{2}$$

where $E_s(h)$ is effective saturation [–], θ_r and θ_s are the residual and saturated volumetric water contents, respectively, [L^3 L^{-3}], K_s is the saturated hydraulic conductivity [LT^{-1}], α is the inverse of the air-entry value [L^{-1}], n is the pore size distribution index [–], l [–] is a pore-connectivity parameter, which was taken to be 0.5 as an average [28], and m is a function of n, which is equal to 1–1/n [–]. The sink term in the Hydrus-1D model represents actual transpiration, which is calculated based on a step-wise limiting factor proposed by [29] as

$$S(h) = \mu(h)\, T_p \tag{3}$$

where $\mu(h)$ is a prescribed stress factor, a function of the pressure head that changes between 0 and 1, T_p is the transpiration rate representing the potential volume of water removed by plant's root uptake from soil. Plant roots are distributed within the soil profile by following an asymptotic equation proposed by [30], which is

$$Y = 1-\beta^d \tag{4}$$

where Y is the cumulative root fraction, β is the extinction coefficient, and d is the soil depth (cm). β is a function of biomes. In our simulations, we adopted a β value of 0.96, which is considered as the value corresponding to the globally averaged root distribution for all ecosystems [31].

The upper boundary condition of the Hydrus-1D is specified as atmospheric boundary conditions controlled by evaporative demand and rainfall input without any surface ponding. Surface runoff is immediately removed if rainfall exceeds infiltration capacity or soil reaches saturation conditions.

2) Atmospheric Forcing, Soil Data, and Simulation Setup: Hydrus-1D model simulations were driven by the US Geological Survey (USGS) Famine Early Warning System Network daily global potential evapotranspiration (ET_p) dataset.[1] In this dataset, the Penman-Monteith equation is used to calculate ET_p data based on atmospheric forcing input data, including air temperature, atmospheric pressure, wind speed, relative humidity, and solar radiation generated by the Global Data Assimilation System analysis fields. ET_p is calculated for every 6 h and accumulated into daily temporal resolution. The data is provided on a 1° spatial resolution globally.

The Hydrus-1D model uses potential evaporation (E_p) and potential transpiration (T_p) separately to calculate actual evapotranspiration. E_p is used to calculate actual water loss from the soil surface as part of an upper boundary condition, and T_p is used to calculate the sink term as given in (3). In order to partition the ET_p data into E_p and T_p, we used enhanced vegetation index (EVI) estimations obtained from MODIS vegetation index products (MOD13C2). MODIS EVI data are calculated as

$$\mathrm{EVI} = 2.5\frac{\rho_{\mathrm{NIR}}-\rho_{\mathrm{Red}}}{\rho_{\mathrm{NIR}}+6.0\rho_{\mathrm{Red}}-7.5\rho_{\mathrm{Blue}}+1.0} \tag{5}$$

where ρ is the surface reflectance in the wavelength band, coefficients represent adjustment and scaling factors [32]. EVI provides information about the intensity of vegetation coverage

[1] [Online]. Available: https://earlywarning.usgs.gov/fews/product/81

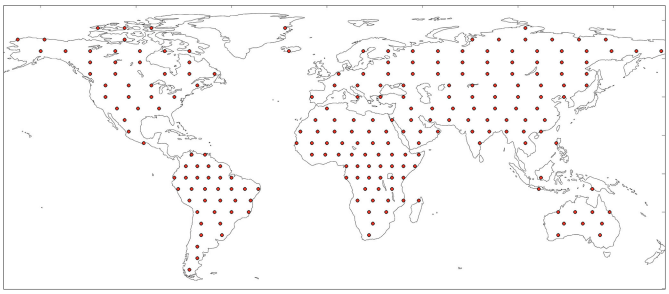


Fig. 2. Global distribution of sampled soil moisture model simulation locations (red dots).

on the land surface. Following [33], T_p is given by

$$T_p = ET_p\ (\mathrm{EVI}^*)^{\eta} \tag{6}$$

the exponent η, which is expected to be close to 1 [33], is determined based on the relationship between transpiration and the vegetation index [34]. EVI$*$ is an enhanced vegetation index stretched between no vegetation and fully covered unstressed vegetation and calculated by

$$\mathrm{EVI}^* = 1 - \frac{\mathrm{EVI}_{\max} - \mathrm{EVI}}{\mathrm{EVI}_{\max} - \mathrm{EVI}_{\min}} \tag{7}$$

where $\mathrm{EVI}_{\max}$ and $\mathrm{EVI}_{\min}$ were determined by using MODIS EVI retrieval statistics for each pixel.

The Global Precipitation Measurement (GPM) mission precipitation data (GPM IMERG Final Precipitation L3, [35]) was used to drive the model. The GPM mission was designed to unify precipitation measurements from a constellation of microwave sensors and deployed by NASA and the Japan Aerospace and Exploration Agency (JAXA) in 2014. Since then, the mission has been collecting global precipitation data with high spatial and temporal resolution [36].

Finally, soil hydraulic parameters for the van Genuchten soil retention function (2), including θ_r, θ_s, α, n, and K_{sat} were obtained from a dataset of soil hydraulic properties derived by [37]. This dataset has global coverage with a 0.25° spatial resolution and is available for seven soil layers extending from the land surface down to 2 meters depth in the soil profile. Soil hydraulic parameters were estimated based on the ROSETTA pedotransfer function of [38] applied to the SoilGrids1km dataset generated by [39].

3) Lower Boundary Conditions and Simulation Locations: The Hydrus-1D model was used to simulate surface soil moisture for six different GW level scenarios across the globe. We kept the water pressure head values constant at the lower boundary of the model domain to represent fixed WTDs for each scenario. Depth to water table was varied between 0.5 and 2.5 m, in increments of 50 cm, with an additional no GW scenario for which the lower boundary condition was assumed to be free drainage. In order to ensure that the model simulations represent as many distinct climate and soil conditions as possible, we distributed the sampling points uniformly across the globe. We simulated with the Hydrus-1D model in 226 locations (see Fig. 2). The simulations extended over five years between 2015 and 2019 for each location and each scenario. To obtain a realistic initialization of soil moisture distribution across the soil profile for specified constant WTDs, the simulation of the first year was used as a spin-up run which was not included in the training or the testing of the EML model. Each simulation was conducted using a 3 meters soil profile and 700 discretization nodes across the soil profile. Soil hydraulic parameters for van Genuchten's soil retention function for each location (as seven different soil layers) were assigned based on [37]'s dataset. To mimic SMAP observations, we used the top 5 cm average soil moisture model results in our analysis, similar to [17]'s approach.

B. EML Method

We adopted an EML algorithm to classify surface soil moisture signals based on GW influence. The adaptive boosting (AdaBoost) method is an iterative boosting EML algorithm introduced by [40]. The boosting algorithms combine a set of weak classifiers into a strong unique classifier through a weighted linear combination. The accuracy of classification increases as more weak classifiers are sequentially added. AdaBoost is considered one of the best statistical classifiers due to its accuracy, simplicity, and solid theoretical foundations [41].

The GentleBoost (gentle adaptive boosting) method, a variant of the AdaBoost technique, was used in this study. GentleBoost is used for binary classifications. It minimizes the exponential loss function of AdaBoost at each iteration by using weighted least-squares regression [42]. MATLAB R2018b was used for building the ensemble classifier. To develop the EML model, we used 1.98 million data points from the Hydrus-1D model simulations. Sixty percent of the data points were used to train and the rest of them to test the EML model. We used the top 5 cm average soil moisture simulation results in our analysis, similar to Soylu and Bras's [17] approach to replicate SMAP soil moisture observations. The predictors to train the EML algorithm were the daily values of coefficient of variation, minimum, maximum, and average of soil moisture. Each predictor was calculated based on a 30-day moving window for four years. Only surface soil moisture data statistics were used to train the EML model. Auxiliary predictors, such as soil properties, atmospheric variables, land use/land cover data, or topography data, were not used. However, we trained the algorithm to detect shallow GW occurrences for a wide variety of climate and soil conditions across the globe, as shown in Fig. 2.

After the initial classification was completed, we adopted the data correction scheme recommended by [17] to prevent potential misclassifications due to data uncertainties originating from satellite observations. The central assumption was that the seasonal GW influence on soil moisture can not frequently switch between GW-influenced and GW-free states within short time intervals because the temporal changes in GW storage are usually slow. The number of consecutive data points within each state was calculated. If the time span of any GW state is shorter than a certain threshold, the sequence of points is reclassified as being the complementary GW state within that period. In summary, to be classified as a shallow or deep GW period, the system must remain in that state for a minimum amount of

time. We used a 30-day threshold to ensure that the temporal variations of GW impact on land surfaces were realistically represented and applied this data correction scheme at a global scale for the four years of interest. It is important to note that we selected an average threshold value. Still, this threshold might vary depending on the rate of recharge (a function of rainfall regime, hydraulic conductivity, preferential flows, etc.) and discharge (a function of vegetation type, climate, etc.) of specific aquifers and their size and geometrical characteristics.

C. Method Tests

One of the direct impacts of shallow GW on surface hydrology is to sustain river flows by providing baseflow. We can assume that the high baseflow rates can be directly related to shallow GW at a watershed scale. This assumption is used to evaluate the findings of the EML model over the Southeastern US, where the EML model detected shallow GW impacts and streamflow observations collected by the USGS[2] are readily available. Although the EML model does not provide any direct streamflow estimations, a comparison of the baseflow estimations derived from the streamflow data with the EML identification of shallow GW provides a reality check on the model's ability to detect shallow GW impact in time and space.

We analyzed watersheds that were under shallow GW influence for at least five months and those that were not under shallow GW influence for, again, at least five months. In this way, we only accounted for the USGS stream gauges with their drainage areas partially influenced by shallow GW for some time during the evaluated period. One hundred and forty three USGS stream gauges across the Southeastern US were used to compare their baseflow and the EML shallow GW detection findings. Monthly baseflow averages for each watershed were calculated using the recursive digital filter technique [48].

III. Results and Discussion

A. Prediction Performance of the EML Model

We have tested multiple ML approaches with a wide range of complexities, including nearest neighbor, discriminant analyses, logistic regression, ensemble, and support vector machine approaches to conduct the binary classification of GW-influenced and GW-free states. We adopted the GentleBoost ensemble ML algorithm because it performed slightly better than the other methods. On the other hand, all tested methods performed satisfactorily with most of them exceeding 95% accuracy.

The overall accuracy in estimating shallow GW for the trained EML algorithm was 95.5%. However, the prediction accuracy was not homogeneously distributed across the globe. It varied based on the prevailing climate and WTD of the location of interest. The AI was used to represent the dryness of the climate at a given location. AI is a simple numerical indicator of aridity based on climatic water deficit. It is calculated as a ratio of precipitation (P) over potential evapotranspiration (ET_p), which is a drying power of the atmosphere. Higher (lower) AI values represent wetter (dryer) environments.

[2][Online]. Available: http://waterdata.usgs.gov

Fig. 3 shows the relation between aridity and the estimation accuracy across different shallow GW depths. Overall estimation accuracy reaches its highest value when the AI is around 0.8. As the AI deviates from this value, the overall EML estimation accuracy reduces slightly [see Fig. 3(a)]. We found that WTDs play a role in estimation accuracy. For example, the EML model can estimate shallow GW in arid environments with very high accuracy if the GW level is close to the surface [Fig. 3(b)]. The EML model can easily distinguish the statistical signal of the shallow GW on surface soil moisture on an otherwise dry surface. However, as AI increases (climate becomes wetter), the detection accuracy of the EML model decreases (though the accuracy is still over 0.8), even if GW levels are very near the surface. This is due to the increased climate-induced wetness in the surrounding area regardless of the shallow GW influence [see Fig. 3(b)].

If the WTD is deeper but still coupled with the surface (i.e., WTD $= 2.5$ m for this scenario), the EML estimation accuracy reduces by 20%, especially for the arid and wet environments. Fig. 3(c) shows that the EML model detection ability reduces with deeper water tables. The leading cause for this reduction in estimation accuracy stems from the fading influence of the shallow GW on surface soil moisture. The main reason causing the fading GW impact is that the rate of capillary upward flux from the water table to the land surface is inversely related to WTD. As the water table becomes deeper, the impact of GW on surface soil moisture reduces its significance. Model estimation accuracies were reduced even further for the wet and dry end of the AI range in the deeper WTD scenario [see Fig. 3(c)]. It is challenging to differentiate the source of elevated soil moisture from climate-induced to shallow GW-influenced in the wet end of the AI. It is also challenging to capture the shallow GW-influenced soil moisture signals in the dry end of the AI due to the limited moisture supply from GW and enhanced climatic water deficit, and increased demand from the vegetation.

B. Global Coverage of Shallow GW Influenced Land Surfaces

We applied the trained EML model to the terrestrial land surface using SMAP data retrievals from the date when the SMAP mission was launched in 2015 until 2019 (between 04/2015 and 03/2019) to estimate the spatial distribution and length of temporal influence of shallow GW at the global scale. The duration of the GW influence on soil moisture ranges from days to the entire year, depending on the region. The EML model output showed that 19% of terrestrial land cover had been influenced by shallow GW at some point in time during the four years of the period of interest [see Fig. 4(a)]. While 23% of the detected areas were under the influence of shallow GW less than one month, only 2.3% of them were under continuous influence of shallow GW representing regions such as floodplains, wetlands, and large riparian areas [see Fig. 4(b)]. We examined the temporal patterns of shallow GW connection to the surface soil moisture at the global scale and for various climate zones, including subarctic (latitudes between 50° and 70°), temperate (35° and 50°), subtropical (23.43° and 35°), tropics (−23.43° and 23.43°), and south of tropics

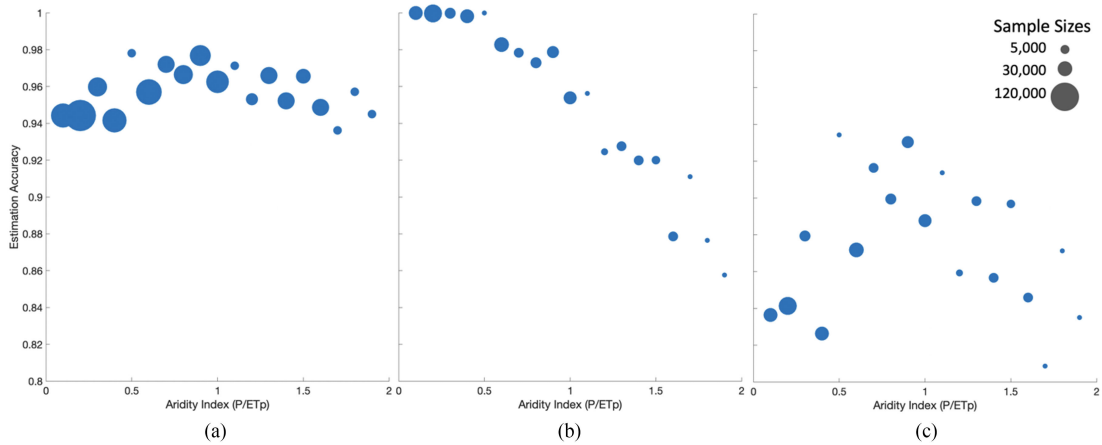


Fig. 3. Relationships between shallow GW detection accuracy and aridity (described as P/ETp). Plots represent average EML model prediction accuracy for each aridity index bin, in intervals of 0.1. The size of each dot represents the number of samples. The graph in (a) shows overall shallow GW detection accuracy against aridity index (AI); (b) represents GW detection accuracy when GW is relatively shallower (WTD is at 1 m from the surface), and (c) represents GW detection accuracy when GW is relatively deeper (WTD is at 2.5 m from the surface).

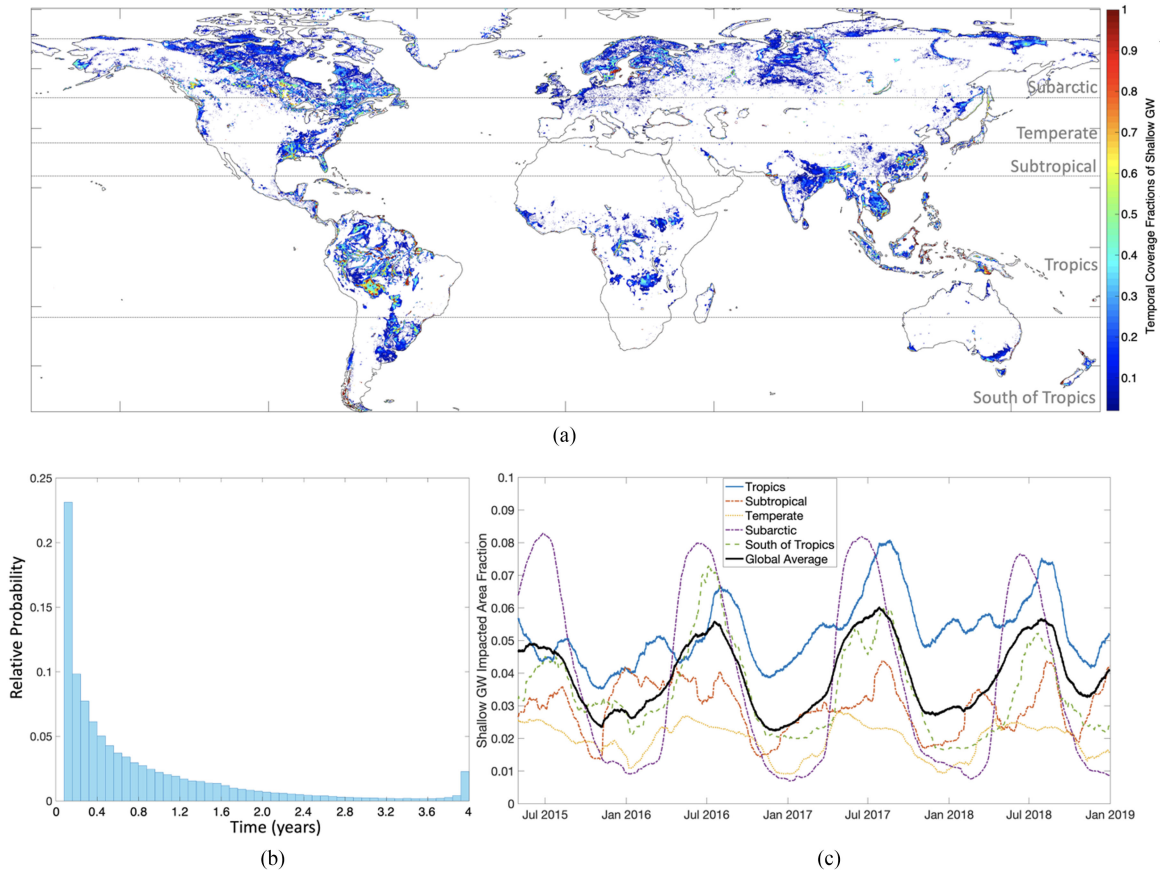


Fig. 4. (a) Spatial and temporal coverage of shallow GW over the globe based on SMAP data. Temporal fractions of shallow GW are shown in different colors, which are assigned depending on the persistence of the shallow GW for the specific SMAP cells over 4 years of time between 04/2015 and 03/2019 (bin sizes are approximately equal to a month). (b) Histogram of the relative probability of shallow GW influenced areas' temporal persistence detected by the EML model. (c) Average temporal fractional changes of shallow GW areas through time for various climate zones.

(−60° and −23.43°). We found that shallow GW influenced the surface soil moisture in an annual cyclic pattern, as illustrated in Fig. 4(c). The shallow GW areas reached their peak coverage (6%) during the northern hemisphere summers and regressed to 2.2% during the northern winters. Amplitudes and the cyclicity of the shallow GW impact varied depending on the climate zones.

SMAP soil moisture data are scarce at high latitudes and high elevations due to permafrost conditions and snow cover. However, the EML model still captures lowlands and riparian

corridors in the subarctic zone with infrequent SMAP soil moisture observations. For example, the EML model captured the lowlands of Canadian and Scandinavian boreal forests as areas under shallow GW influence. Similarly, shallow GW areas were identified in the riparian areas of large rivers such as the Lena, the Yenisei, and Ob Rivers in Northern Russia and downstream riparian areas of the Amur River in Eastern Russia as expected [see Fig. 4(a)]. Temporally, shallow GW influence can reach up to 8.3% of the total area of the subarctic zone during the summer months and regress down to 0.7% during the winter months periodically [see Fig. 4(c)].

We found that shallow GW primarily influenced low elevation plains and flat coastal areas in temperate and subtropical zones [illustrated as Temperate, Subtropics, and the South of Tropics in Fig. 4(a)]. On the southern hemisphere, shallow GW seasonally influences the Pampas Plains in Argentina and Uruguay, the southeast and southwest Australia coastline, most of Tasmania, North Island, and the south agricultural lowlands of New Zealand. In the temperate zone of the northern hemisphere, the northern margins of Western Europe, especially northern parts of France, Belgium, and The Netherlands, were found to be influenced seasonally. Vast plains of Southeast US, primarily downstream of the Lower Mississippi River basin, southern parts of the Yangtze River basin, China, and the vast riparian corridor of the Brahmaputra River and its tributaries at Bangladesh and Northeastern India were influenced by shallow GW.

In the tropics, shallow GW-influenced areas mostly correspond to forest-covered wetlands and riparian corridors, including Amazonia, the central part of the Congo River basin, the Indochinese peninsula, and the Indonesia islands' margins, especially the southern part of Papua Island. We found a more persistent impact of shallow GW along the river corridors of the Amazon River and its tributaries and floodplains close to the Amazon River basin's south and west upstream margins. Similarly, corridors of the Congo River and its tributaries, especially the Uele River located in the north of Congo River, are identified as areas influenced by shallow GW. We found a seasonal or temporary impact of shallow GW in riparian zones of Zambia, coastal lowlands of West Africa, especially around Guinea, and the Sudd, a vast wetland in South Sudan and northeastern India and Bangladesh. Coverage of shallow GW-influenced areas reached up to 8.1% of the total area of the tropics during the late summer and reduced to a minimum of 3.5% during the early winter months. Tropical areas receive frequent and large rainfall amounts that may cause wet bias in soil moisture and may over-estimate shallow GW areas in the tropics due to the nature of the EML model, which picks up areas with sustaining wet soil conditions. Nevertheless, shallow GW areas detected in this study aligned with the previous efforts of wetland delineations in the tropics [3], [49], [45].

South of the tropics, shallow GW seasonally influences the Pampas Plains in Argentina and its northern areas, including some parts of Uruguay and Paraguay. The coastlines of southeast and southwest Australia, including most of Tasmania, North Island, and the southern lowlands of New Zealand, were also classified as shallow GW-influenced regions. Shallow GW impacted area coverages in the south of the tropics fluctuate between 1.7% and 7.3% of their total area annually [see Fig. 4(c)]. Interestingly, shallow GW-influenced area coverage reached its maximums during winters, unlike the temporal fluctuation pattern of the northern hemisphere.

Temporal change in coverage of shallow GW impacts on the land surface in all different climate zones follows a similar pattern, regardless of their geographical location. The basic pattern illustrates that the coverage of shallow GW-influenced areas reaches its peak value during the northern hemisphere summer months and then gradually decreases until the winter months. However, this pattern is different in the subtropical climate zone, where two peaks of shallow GW-influenced area coverage exist. We found that temporal variation of shallow GW influences follows a different pattern in the Southeast US, where the coverage of shallow GW areas reaches their peak value during late winter to spring months. This is consistent with the rainfall pattern of the Southeastern Coastal Plain of the US, where a significant fraction of the annual rainfall is received during the winter months. Another potential contributor may be the extensive distribution of clay horizons across the Southeastern US, which may operate as perched aquifers during the winter and spring months when rainfall is usually high, and evapotranspiration is relatively lower.

C. Tests for Consistency With Other Datasets

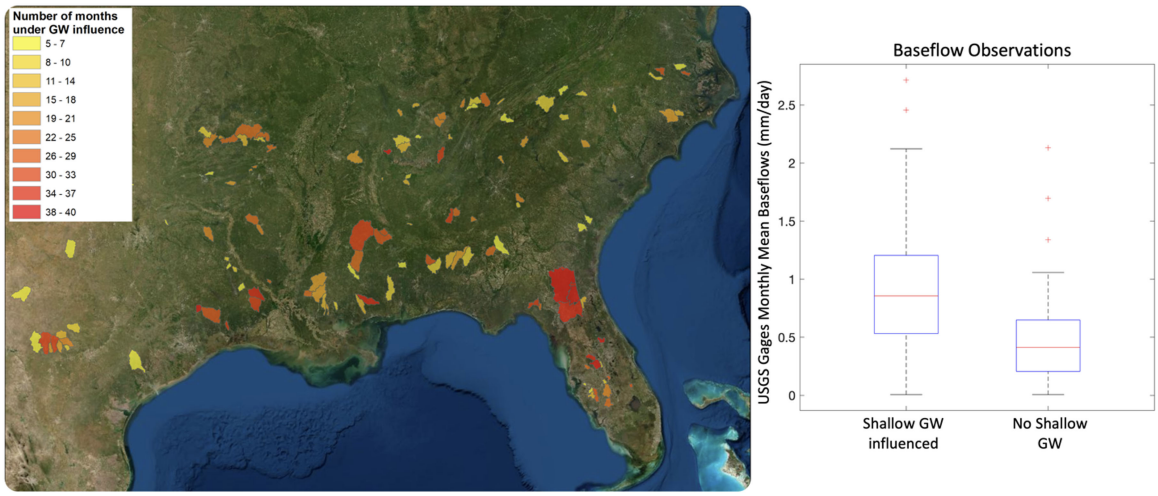
The output of the EML approach is the spatiotemporal distribution of the shallow GW-influenced areas during the 45 months between 2015 and 2018 at a global scale. We evaluated our findings against three datasets, namely:

1) baseflow estimations collected by the USGS stream gauges across the southeast US, where shallow GW impact was detected, especially during winter months;
2) global wetland distributions, where GW is the dominant source of elevated soil moisture states or saturated conditions; and
3) distributions of clay-enriched horizons potentially forming perched aquifers resulting in shallow GW.

1) Baseflow Comparisons: The streamflow data were obtained from the USGS water data website.[3] Stream gages were selected to have drainage areas corresponding to SMAP cells where shallow GW influence was detected. We analyzed USGS stream gages only located at the Southeast of the US with drainage areas partially or entirely influenced by shallow GW spatially, and at least for 5 months and at most 40 months during the analyzed 45 months of period temporally. Ultimately, 143 USGS stream gages in the region were used [see Fig. 5(a)]. We compared the average monthly baseflows during the periods of no shallow GW influence and periods with at least partially shallow GW influence detected within the catchment area of given stream gages. The baseflows were estimated by using the recursive digital filter technique [48].

If the EML model detected any GW influence within the boundaries of a given watershed on a given month, we classified it as "shallow GW month"; otherwise, as "no shallow GW

[3][Online]. Available: https://waterdata.usgs.gov/



(a) (b)

Fig. 5. Comparisons of monthly mean baseflow (mm/day) estimations from the Southeast US watersheds with USGS Gages, which were under at least 5 months (and at most 40 months out of 45 months of interest) of shallow GW influence (or no GW influence) as detected by the SMAP-based EML model. (a) Each catchment area of the USGS stream gage locations is shown in a map with colors representing the duration of shallow GW influence on a given area captured by the EML model. (b) The monthly mean baseflow estimation distributions for both "shallow GW months" and "No GW months" are represented in a box plot.

month." We compared the EML model-based GW classifications against monthly baseflow estimations. The boxplot illustrated in Fig. 5(b) represents the data variability characteristics for both cases. The figure shows that the estimated baseflows during periods with shallow GW influence within the drainage area of a given watershed were more than twice (0.93 mm/day) as high as the baseflows generated from the watersheds with no shallow GW influence (0.45 mm/day) on average for a given month.

The result shows that river baseflows tend to contribute more to the total streamflow discharge when GW is shallow enough to influence the surface soil moisture directly, as expected. Baseflow varies seasonally as a function of changes in GW storage. As GW recharge increases, the water table nears the surface increases the baseflow contribution to streamflow. However, this mechanism takes some time, and this time depends on the hydrological system. For example, the deeper the vadose zone, the longer time needed for rainfall to impact the baseflow.

The streamflow from December to April in watersheds throughout the Southeastern Coastal Plain is usually much greater than the remainder of a given year [43], [44]. Sheridan [43] found that 54% of the rainfall received during the first four months of a given year end up in streamflow, while only 12% becomes streamflow throughout the other eight months of a year. The coincidence of the peak rainfall season and the peak streamflow discharge illustrates that streamflow responds to the incoming rainfall quickly in the Southeastern Coastal Plain. This statement was further tested statistically. We found that high cross-correlation between the monthly rainfall averages and the EML detected shallow GW distributions throughout the Southeastern US (not shown here). The quick response can be explained by the shallow aquifer systems feeding the rivers across the region. The wide distribution of shallow clay-enriched horizons throughout the region causes water to perch, contributing to the fast transport of the incoming rainfall to the streams.

2) Global Wetland Distribution and Clay-Enriched Horizons: A significant challenge in validating our shallow GW detection model findings is the lack of an observation-based dataset at the global scale. We used two different datasets to get a sense of the performance of our approach at the global scale. First, we used global wetland data, assuming that wetlands are GW-driven saturated areas and captured with our model. Second, we used subsurface clay horizon distributions as a proxy for the tendency to form perched pockets of water due to the low permeability of shallow high clay fraction soil layers that inhibit the water percolation to deeper soil layers. Perched aquifers formed as a result of accumulating water above the clay horizons may directly impact surface soil moisture by elevating and sustaining the surface soil moisture.

First, we compared our findings with a global composite wetland map developed by [45], as shown in Fig. 6(a). They delineated the wetland areas where the water table is shallower than 20 cm by using a global scale GW model [45] based on the assumption that these areas are persistently at or near saturation due to shallow GW or regular flooding. The figure shows that the overall extent of the wetlands [see Fig. 6(a)] and the shallow GW areas [see Fig. 6(b)] are generally consistent at a global scale. However, estimated shallow GW areas by our SMAP-based EML model are more extensive than the wetlands. This may be explained as we trained the algorithm to detect regions with ranges of WTDs from 50 cm to 2.5 m, which does not necessarily cause persistent near saturation at the surface, unlike wetlands. The differences in area coverage between the maps can also be attributed to the steady-state assumption in the global GW model by [46]. While the global wetland map was developed based on steady-state GW model outputs, our map also captures transient GW conditions based on SMAP observations. Fig. 6(b) depicts all the areas influenced by shallow GW; even if this influence prevailed for a short time period. The discrepancy between the spatial resolutions might also contribute to the differences in maps. While our shallow GW map has the same grid size as the

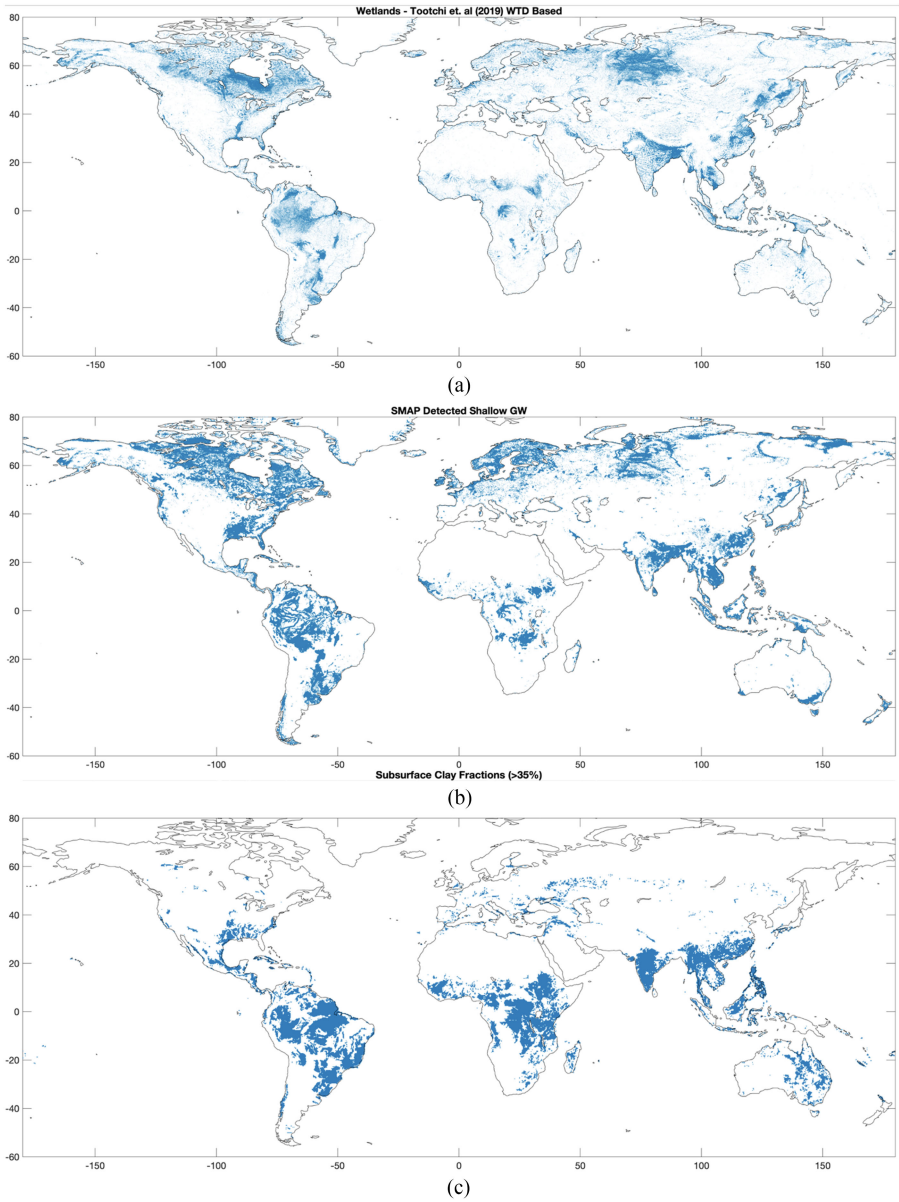


Fig. 6. Global distributions of (a) shallow GW detected by SMAP-based EML model, (b) GW driven-wetlands estimated by [48] based on a global GW model of [49], (c) subsurface clay fractions larger than 0.35 based on a dataset by [45].

SMAP enhanced resolution product (∼9 km), the wetland map has a finer spatial resolution (∼500 m), capturing the wetlands in smaller scales.

Second, we used the global subsurface distribution of clay-sized minerals to compare our findings. Forming clay-sized particles via physical or chemical processes in the subsurface creates clay-enriched horizons, usually found at 0.1 to 3 m depths from the surface across at least 25% of the Earth's surface [50]. Clay horizons may play an essential role in sustaining high soil moisture levels at the land surface [51]. They may form perched water bodies with limited storage capacity. Fig. 6(c) represents soils with high subsurface clay fraction (>%35) content at a global scale. The figure shows that the clay horizons tend to be formed across tropical and subtropical climate zones due to climate-related favorable pedogenetic processes. The distribution of clay horizons can further help to explain some of the differences between shallow GW areas and the wetlands coverage. For example, relatively extensive shallow GW coverage on the Southeastern US may not be explained solely by the existence of wetlands, concentrated mainly in the Mississippi alluvial plain. Moreover, spatial dissimilarities between wetlands coverage and shallow GW coverage due to relatively extensive coverage of shallow GW-influenced areas in Africa, southeastern Asia, and South America may be explained by the altered soil drainage characteristics caused by clay-enriched horizons. We compared the subsurface clay fractions between the EML-detected shallow GW areas and the areas free from any detected GW influence (see Table I). The table shows that the average subsurface clay fractions are 12% higher in the South of Tropics and about 7% higher in Tropical and Subtropical

TABLE I
SUBSURFACE CLAY FRACTIONS

	South of Tropics	Tropics	Subtropics	Temperate	Subarctic
Average Subsurface Clay fractions of Shallow GW-influenced areas	35.8% (12.1%)	36.4% (11.5%)	32.5% (8.3%)	20.13% (8.3%)	18.9% (7.1%)
Average Subsurface Clay fractions of areas with no GW influence	23.9% (11.5%)	29.3% (12.2%)	25.7% (7.5%)	25.7% (6.6%)	19.3% (7.2%)

Subsurface clay fraction percentages of areas influenced by shallow GW and areas with no GW influence. Values represent average percentages for each climate zone and their standard deviations given in parenthesis.

regions for the shallow GW areas. This finding implies that some shallow GW-influenced areas may be linked to the clay-enriched horizons, especially at the Tropics and Subtropics.

D. Broader Implications and Method Limitations

High-resolution spatial and temporal distribution of shallow GW data based on observations can be utilized in global scale flux models either by providing the model with lower boundary conditions or calibrating the models as discussed in detail in [17]. Besides, monitoring shallow GW may also provide critical information to the global change research from water resources to ecology. For example, GW depletion, due to excessive irrigation and climate change, has become a vital issue threatening global food production [1], [52], [53]. Unsustainable water table depletion due to excessive GW abstraction may induce irreversible loss of GW storage, causing partial loss or the entire disappearance of wetlands and shrinking in riparian zones. The method presented here may be a valuable tool to monitor the temporal changes in global shallow GW coverages, possibly linked to excessive GW use. Another area of research where the results might be helpful is in studying the thawing of permafrost areas. Shallow GW systems form in arctic and subarctic regions due to climate-induced lowering of permafrost table [54]. Progressive lowering of the permafrost table results in a shallow GW flow system [55]. As climate warming continues, deeper GW systems will eventually establish. However, monitoring the coverage of shallow GW-influenced areas during the transition period may provide important information about the climate change impacts on sensitive permafrost systems. It is important to note that the SMAP data is scarce at high latitudes due to prolonged snow cover periods. Although this scarcity of data may restrict the EML model from estimating shallow GW-influenced periods accurately, observation-based shallow GW data may still be helpful to monitor the changes in the permafrost areas.

The developed model may be improved by introducing more data points, more extended observations as SMAP data collection continues, and adding more auxiliary predictors to the training. We simulated 226 randomly selected data points across the globe and used the model outputs to train the EML model for the four years' worth of data points. Adding more points by spatially increasing the number of simulation points or temporally extending the data period might help improve the overall accuracy of the EML model. In addition, different sampling strategies, such as using more data from the areas representing steeper climate gradient and higher soil heterogeneity, might also improve the accuracy. Rather than training the EML model for the global scale, training it for smaller regions with more homogeneous geophysical and hydroclimatic characteristics might also enhance the accuracy of the EML model for the specific regions of interest. Moreover, we used only the predictors statistically derived from the Hydrus-1D model simulations mimicking the SMAP observations, including daily values of coefficient of variation, minimum, maximum, and averages of soil moisture to train the EML algorithm. Training the algorithm with different predictors such as soil properties, atmospheric variables, land use/land cover data, or topography data might also improve the shallow GW detection accuracy.

Another essential point to note is that the shallow GW varies in space and time depending on a given region's topography, lithology, and climate. The water table can vary significantly at a finer spatial scale than the SMAP resolution causing subgrid variability, a common but critical issue when working with gridded datasets such as satellite images. Therefore, shallow GW areas may partly fall into a given SMAP cell, resulting in the machine learning algorithm capturing weaker signals from the SMAP retrievals. As better downscaling methods for SMAP retrievals or higher resolution soil moisture observations become available, this limitation may further be relaxed, and more realistic shallow GW distributions can be achieved. The one-dimensional unsaturated flow model used in this work served only to detect the impact of shallow GW on the surface soil moisture signal. The behavior of the surface soil moisture, under different assumptions of the GW depth, is what drives the statistical approach. For the purposes of training the EML algorithm, it does not matter how the GW came to be where it is, whether by lateral fluxes or not. What matters is how the GW at that level impacts the surface. Our interest is to capture whether shallow GW exists at a certain location regardless of the prevailing subsurface conditions that lead to the surface conditions that the satellite sees at the satellite resolution. On the other hand, once a shallow water table for a given SMAP pixel is detected using the method proposed in this study, subgrid variability of the GW depths can be estimated with 3-D models during a post-processing effort. As long as a 3-D GW flow model is fed with accurate model inputs such as topography, lithostratigraphy, atmospheric forcing, boundary conditions, and surface water and GW level and discharge measurements, high-resolution WTD distributions are possible to achieve.

It is also worth noting that we used gridded soil hydraulic parameters [37]. However, they cannot realistically represent pedogenic variability, especially in areas where soil textural characteristics vary heterogeneously on a vertical or horizontal

scale. Therefore, there is an associated difficulty representing the variation of the soil hydraulic parameters in studies like this one that uses gridded datasets.

IV. Conclusion

Shallow GW influences the land surface in several different ways. For example, it increases latent heat fluxes, provides additional water to the plants, and alters the components of the hydrologic cycle by increasing the surface evaporation. Monitoring shallow GW is limited by the scarcity of GW observations. Global GW products are largely based on large-scale GW modeling efforts. However, the scarcity of data also hampers the calibration and verification of such models.

For the first time, we estimated the high-resolution spatiotemporal distribution of shallow GW based on SMAP soil moisture observations at a global scale. GW levels affect the land surface only if the water table is close enough to the surface to be coupled with surface processes (shallow GW). Our primary assumption was that shallow GW levels could be captured by processing the soil moisture signals. More specifically, shallow GW causes lower variations and a higher average surface soil moisture [17], and these statistical features can be attributed to the shallow GW's existence. We used an EML algorithm to process the soil moisture signals to make a binary classification if a given time series of soil moisture observations are either shallow GW-influenced or GW-free. To classify the soil moisture data series, we trained an EML model by using statistical features of the SMAP data such as coefficient of variation, minimum, maximum, and averages of soil moisture values. We found that the performance of the EML model satisfactorily estimated the shallow GW state after the model training with approximately 1.19 million data points obtained from the Hydrus-1D simulations. The simulations represented various soil textures, climate zones, and GW states across the globe. We tested the performance of the EML model with about 0.79 million data points and found that the model tends to perform better in drier climates and shallow GW conditions with reaching the 95.5% overall estimation accuracy.

We used the trained EML model with SMAP data retrievals to estimate the spatial and temporal distribution of shallow GW influence at the global scale. We found that 19% of terrestrial land cover had been influenced by shallow GW at some point in time during the four years of the period of interest (2015–2018). While 23% of the detected areas were under the influence of shallow GW less than one month, only 2.3% of them were under continuous influence of shallow GW representing regions such as floodplains, wetlands, and large riparian areas.

Our model revealed that shallow GW followed an annual cyclic pattern during the examined four years period. The shallow GW areas reached their peak coverage (6%) during the northern hemisphere summers and regressed to 2.2% during the northern winters. Amplitudes and the cyclicity of the shallow GW impact varied depending on the climate zones. The highest amplitude was observed in the subarctic, and the highest overall mean shallow GW coverage was found in the tropics.

We compared the findings of the EML model against three different datasets, including baseflow estimations across the southeast US, global wetland distributions, and the global distributions of clay-enriched horizons. The results showed that the derived global distribution of shallow GW identifies wetlands, large riparian corridors, and seasonally inundated/influenced lowlands. Moreover, we found that shallow GW occurrences were detected in the watersheds during their baseflow dominated periods and coincided with the subsurface clay-enriched horizon distributions due to the impact of accumulated water above the clay layers with low permeability on the surface soil moisture.

Our findings show that NASA's SMAP soil moisture retrievals can be used to estimate shallow GW with high spatiotemporal resolution at a global scale, potentially providing invaluable inputs for modeling and environmental change monitoring studies. While the relatively short observation period of the SMAP mission, which has been collecting soil moisture data since 2015, preclude the study of impacts of climate change and land use/land cover changes on water resources, we showed that the SMAP-based shallow GW detection model like the one we presented in this study might provide invaluable insights on short-term GW dynamics. With more SMAP data in years to come, we can hope for a better understanding of global changes in shallow GW areas and their associated causes.

Acknowledgment

SMAP data are available via.[4] Precipitation data (GPM IMERG product) is available at.[5] MODIS enhanced vegetation index estimations data can be accessed at.[6] The authors would like to thank Randall Koster and the other two referees, and the associate editor for their constructive comments, which helped improve the clarity of the manuscript, and Oguz Sariyildiz (Commonwealth of Kentucky) for his support on USGS baseflow analyses.

References

[1] C. Dalin, Y. Wada, T. Kastner, and M. J. Puma, "Groundwater depletion embedded in international food trade," *Nature*, vol. 543, no. 7647, pp. 700–704, 2017.

[2] J. Qiu, S. C. Zipper, M. Motew, E. G. Booth, C. J. Kucharik, and S. P. Loheide, "Nonlinear groundwater influence on biophysical indicators of ecosystem services," *Nature Sustain.*, vol. 2, no. 6, pp. 475–483, Jun. 2019.

[3] G. Miguez-Macho and Y. Fan, "The role of groundwater in the Amazon water cycle: 1 Influence on seasonal streamflow, flooding and wetlands," *J. Geophys. Res. Atmos.*, vol. 117, no. 15, pp. 1–30, 2012.

[4] M. O. Cuthbert *et al.*, "Global patterns and dynamics of climate–groundwater interactions," *Nature Clim. Change*, vol. 9, no. 2, pp. 137–141, Feb. 2019.

[5] R. M. Maxwell, F. K. Chow, and S. J. Kollet, "The groundwater–land-surface–atmosphere connection: Soil moisture effects on the atmospheric boundary layer in fully-coupled simulations," *Adv. Water Resour.*, vol. 30, no. 12, pp. 2447–2466, Dec. 2007.

[6] I. E. M. de Graaf, E. H. Sutanudjaja, L. P. H. van Beek, and M. F. P. Bierkens, "A high-resolution global-scale groundwater model," *Hydrol. Earth Syst. Sci.*, vol. 19, no. 2, pp. 823–837, 2015.

[7] M. E. Soylu, C. J. Kucharik, and S. P. Loheide, "Influence of groundwater on plant water use and productivity: Development of an integrated ecosystem – variably saturated soil water flow model," *Agricultural Forest Meteorol.*, vol. 189–190, pp. 198–210, Jun. 2014.

[8] S. C. Zipper, M. E. Soylu, E. G. Booth, and S. P. Loheide, "Untangling the effect of shallow grounwater and soil texture as drivers of subfield-scale yield variability," *Water Resour. R*, vol. 51, pp. 1–21, 2015.

[4][Online]. Available: https://nsidc.org/data/smap/smap-data.html

[5][Online]. Available: https://pmm.nasa.gov/data-access/downloads/gpm

[6][Online]. Available: https://lpdaac.usgs.gov/products/mod13c2v006/

[9] M. Rodell, J. Chen, H. Kato, J. S. Famiglietti, J. Nigro, and C. R. Wilson, "Estimating groundwater storage changes in the Mississippi River Basin (USA) using GRACE," *Hydrogeol. J.*, vol. 15, pp. 159–166, 2007.

[10] M. W. Becker, "Potential for satellite remote sensing of ground water," *Ground Water*, vol. 44, no. 2, pp. 306–318, 2006.

[11] M. Rodell and J. S. Famiglietti, "The potential for satellite-based monitoring of groundwater storage changes using GRACE: The high plains aquifer, Central US," *J. Hydrol.*, vol. 263, pp. 245–256, 2002.

[12] A. G. Bobba, R. P. Bukata, and J. H. Jerome, "Digitally processed satellite data as a tool in detecting potential groundwater flow systems," *J. Hydrol.*, vol. 131, pp. 25–62, 1992.

[13] A. G. Bobba, R. P. Bukata, and J. E. Bruton, "Application of LANDSAT data to the location of hydrologically active source areas" in *Proc. 5th Annu. William T. Pecora Memorial Symp. Remote Sens.*, Jun. 1979, p. 407.

[14] C. R. Hain, W. T. Crow, M. C. Anderson, and M. Tugrul Yilmaz, "Diagnosing neglected soil moisture source-sink processes via a thermal infrared-based two-source energy balance model," *J. Hydrometeorol.*, vol. 16, no. 3, pp. 1070–1086, 2015.

[15] D. L. Galloway *et al.*, "Detection of aquifer system compaction and land subsidence using interferometric synthetic aperture radar, Antelope Valley, Mojave Desert, California," *Water Resour. Res.*, vol. 34, pp. 2573–2585, 1998.

[16] E. H. Sutanudjaja, S. M. De Jong, F. C. Van Geer, and M. F. P. Bierkens, "Using ERS spaceborne microwave soil moisture observations to predict groundwater head in space and time," *Remote Sens. Environ.*, vol. 138, pp. 172–188, 2013.

[17] M. E. Soylu and R. L. Bras, "Detecting shallow groundwater from spaceborne soil moisture observations," *Water Resour. Res.*, vol. 57, 2021, Art. no. e2020WR029102.

[18] K. J. Bergen, P. A. Johnson, M. V. de Hoop, and G. C. Beroza, "Machine learning for data-driven discovery in solid Earth geoscience," *Science*, vol. 363, no. 6433, Mar. 2019, Art. no. eaau0323.

[19] Y. Zhang, F. H. S. Chiew, M. Li, and D. Post, "Predicting runoff signatures using regression and hydrological modeling approaches," *Water Resour. Res.*, vol. 54, no. 10, pp. 7859–7878, 2018.

[20] H. I. Erdal and O. Karakurt, "Advancing monthly streamflow prediction accuracy of CART models using ensemble learning paradigms," *J. Hydrol.*, vol. 477, pp. 119–128, 2013.

[21] G. Konapala and A. Mishra, "Quantifying climate and catchment control on hydrological drought in the continental United States," *Water Resour. Res.*, vol. 56, no. 1, pp. 1–25, 2020.

[22] S. A. Naghibi, D. D. Moghaddam, B. Kalantar, B. Pradhan, and O. Kisi, "A comparative assessment of GIS-based data mining models and a novel ensemble model in groundwater well potential mapping," *J. Hydrol.*, vol. 548, pp. 471–483, 2017.

[23] C. Carranza, C. Nolet, M. Pezij, and M. van der Ploeg, "Root zone soil moisture estimation with Random Forest," *J. Hydrol.*, vol. 593, Feb. 2021, Art. no. 125840.

[24] K. P. Singh, S. Gupta, and D. Mohan, "Evaluating influences of seasonal variations and anthropogenic activities on alluvial groundwater hydrochemistry using ensemble learning approaches," *J. Hydrol.*, vol. 511, pp. 254–266, 2014.

[25] M. Zounemat-Kermani, O. Batelaan, M. Fadaee, and R. Hinkelmann, "Ensemble machine learning paradigms in hydrology: A review," *J. Hydrol.*, vol. 598, no. 33, Apr. 2021, Art. no. 126266.

[26] J. Šimůnek, A. M. Šejna, H. Saito, M. Sakai, and M. T. Van Genuchten, "The HYDRUS-1D software package for simulating the movement of water, heat, and multiple solutes in variably saturated media," *Version 4.17, HYDRUS Software Series 3, Department of Environmental Sciences, University of California Riverside*, Riverside, CA, USA. 2013.

[27] M. Th. van Genuchten, "A closed-form equation for predicting the hydraulic conductivity of unsaturated soils," *Soil Sci. Soc. Amer. J.*, vol. 44, pp. 892–898, 1980.

[28] Y. Mualem, "A new model for predicting the hydraulic conductivity of unsaturated porous media," *Water Resour. Res.*, vol. 12, no. 3, pp. 513–522, 1976.

[29] R. A. Feddes, P. J. Kowalik, and H. Zaradny, *Simulation of Field Water Use and Crop Yield*. Hoboken, NJ, USA: Wiley, 1978.

[30] M. R. Gale and D. F. Grigal, "Vertical root distributions of northern tree species in relation to successional status," *Can. J. Forest Res.*, vol. 17, no. 8, pp. 829–834, 1987.

[31] R. B. Jackson, J. Canadell, J. R. Ehleringer, H. A. Mooney, O. E. Sala, and E. D. Schulze, "A global analysis of root distributions for terrestrial biomes," *Oecologia*, vol. 108, no. 3, pp. 389–411, Nov. 1996.

[32] K. Didan, A. B. Munoz, R. Solano, and A. Huete, "MODIS vegetation index user's guide (MOD13 series)," Vegetation Index and Phenology Lab., Univ. AZ, Tucson, AZ, USA, 2015.

[33] B. J. Choudhury, "Synergism of multispectral satellite observations for estimating regional land surface evaporation," *Remote Sens. Environ.*, vol. 49, no. 3, pp. 264–274, 1994.

[34] E. P. Glenn, P. L. Nagler, and A. R. Huete, "Vegetation index methods for estimating evapotranspiration by remote sensing," *Surv. Geophys.*, vol. 31, no. 6, pp. 531–555, 2010.

[35] G. J. Huffman, E. F. Stocker, D. T. Bolvin, E. J. Nelkin, and T. Jackson, "GPM IMERG final precipitation L3 1 day 0.1 degree x 0.1 degree V06," Edited by, and A. Savtchenko, , NASA Goddard Earth Sci. Data Inf. Serv. Center (GES DISC), Greenbelt, MD, USA,2019.

[36] A. Y. Hou *et al.*, "The global precipitation measurement mission," *Bull. Amer. Meteorol. Soc.*, vol. 95, no. 5, pp. 701–722, 2014.

[37] C. Montzka, M. Herbst, L. Weihermüller, A. Verhoef, and H. Vereecken, "A global data set of soil hydraulic properties and sub-grid variability of soil water retention and hydraulic conductivity curves," *Earth Syst. Sci. Data Discuss.*, vol. 9, no. 2, pp. 529–543, 2017.

[38] M. G. Schaap, F. J. Leij, and M. T. Van Genuchten, "ROSETTA: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions," *J. Hydrol.*, vol. 251, pp. 163–176, 2001.

[39] T. Hengl *et al.*, "SoilGrids1km - Global soil information based on automated mapping," *PLoS One*, vol. 9, no. 8, 2014, Art. no. e105992.

[40] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proc. 13th Int. Conf. Mach. Learn.*, 1996, pp. 148–156.

[41] X. Wu *et al.*, "Top 10 algorithms in data mining," *Knowl. Inf. Syst. Volume*, vol. 14, no. 1, pp. 1–37, 2008.

[42] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: A statistical view of boosting," *Ann. Statist.*, vol. 28, no. 2, pp. 337–407, 2000.

[43] R. J. Nathan and T. A. McMahon, "Evaluation of automated techniques for base flow and recession analyses," *Water Resour. Res.*, vol. 26, no. 7, pp. 1465–1473, Jul. 1990.

[44] J. R. B. Bwangoy, M. C. Hansen, D. P. Roy, G. De Grandi, and C. O. Justice, "Wetland mapping in the Congo Basin using optical and radar remotely sensed data and derived topographical indices," *Remote Sens. Environ.,* vol. 114, no. 1, pp. 73–86, 2010.

[45] A. Ito and R. Wagai, "Global map of clay minerals in terrestrial soils," National Institute for Environmental Studies, Tsukuba, Japan: PANGAEA. 2016. [Online]. Available: https://doi.org/10.1594/PANGAEA.868929

[46] J. M. Sheridan, "Rainfall-streamflow relations for Coastal Plain watersheds," *Appl. Eng. Agric.*, vol. 13, no. 3, pp. 333–344, 1997.

[47] D. D. Bosch, J. G. Arnold, P. G. Allen, K. -J. Lim, and Y. S. Park, "Temporal variations in baseflow for the Little River experimental watershed in South Georgia, USA," *J. Hydrol. Reg. Stud.*, vol. 10, pp. 110–121, 2017.

[48] A. Tootchi, A. Jost, and A. Ducharne, "Multi-source global wetland maps combining surface water imagery and groundwater constraints," *Earth Syst. Sci. Data*, vol. 11, no. 1, pp. 189–220, 2019.

[49] Y. Fan, H. Li, and G. Miguez-Macho, "Global patterns of groundwater table depth," *Science*, vol. 339, no. 6122, pp. 940–943, 2013.

[50] N. C. Brady, R. R. Weil, and R. R. Weil, *The Nature and Properties of Soils*, vol. 13. Upper Saddle River, NJ, USA: Prentice Hall, 2008.

[51] R. J. Schaetzl, F. J. Krist, K. Stanley, and C. M. Hupy, "The natural soil drainage index: An ordinal estimate of long-term soil wetness," *Phys. Geography*, vol. 30, no. 5, pp. 383–409, 2009.

[52] I. E. M. de Graaf *et al.*, "A global-scale two-layer transient groundwater model: Development and application to groundwater depletion," *Adv. Water Resour.*, vol. 102, pp. 53–67, 2017.

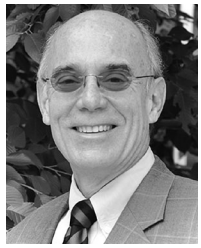
[53] J. S. Famiglietti, "The global groundwater crisis," *Nature Clim. Change*, vol. 4, no. 11, pp. 945–948, 2014.

[54] V. F. Bense, G. Ferguson, and H. Kooi, "Evolution of shallow groundwater flow systems in areas of degrading permafrost," *Geophys. Res. Lett.*, vol. 36, no. 22, pp. 2–7, 2009.

[55] F. A. Michel and R. O. Van Everdingen, "Changes in hydrogeologic regimes in permafrost regions due to climatic change," *Permafr. Periglac. Process.*, vol. 5, no. 3, pp. 191–195, 1994.

Mehmet Evren Soylu was born in Ankara, Turkey. He received the Engineering and the M.Sc. degrees in the hydrogeological engineering program from Hacettepe University, Ankara, Turkey, in 2001 and 2004, respectively, and the Ph.D. degree in hydrologic sciences from the School of Natural Resources, University of Nebraska-Lincoln with a specialization in hydrologic sciences in 2011.

He continued his studies as a post-doctoral researcher with the University of Wisconsin-Madison between 2011 and 2014. In addition, he worked at various environmental consulting companies as an Engineer and Project Manager. He is currently a Research Engineer with the School of Civil and Environmental Engineering, Georgia Institute of Technology. His research interests include hydrogeology, ecohydrology, remote sensing, climate change, and ecosystem modeling. More specifically, he focuses on both theoretical and applied aspects of environmental research, with a strong focus on groundwater – soil – plant interactions in the Earth's critical zone.



Rafael L. Bras was born in San Juan, Puerto Rico. He received the Engineering, M.Sc., and Sc.D. degrees in water resources and hydrology from the Department of Civil Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 1972, 1974, and 1975, respectively.

He was a Professor with MIT between 1976 and 2008. During that period, he served as a Director of Ralph M. Parsons Laboratory (1983–1991), Department Head (1992–2001) of Civil and Environmental Engineering at MIT, and Chair of the MIT Faculty (2003–2005). He was a Distinguished Professor and Dean of engineering with the University of California, Irvine, CA, USA (2008–2010). Currently, he is the Professor with the Georgia Institute of Technology, where he served as a Provost and Executive Vice President for Academic Affairs (2010–2020). He has authored 240 papers in prestigious refereed journals and has written and contributed to numerous books, conference proceedings, and technical reports. His research interests include hydrometeorology, fluvial geomorphology, and ecohydrology.

Prof. Bras is the K. Harrison Brown Family Chair. He is recipient of many awards, honors, and recognition, including several honorary degrees, Macelwane and Horton Medals of AGU, Clarke Prize, NASA Public Service Medal, elected Fellow of AMS, ASCE, AAAS, member of the U.S. National Academy of Engineering, National Academy of Arts and Sciences of Puerto Rico, National Academies of Science and Engineering of Mexico.