Global Soil Salinity Estimation at 10 m Using Multi-Source Remote Sensing

Nan Wang, Songchao Chen, Jingyi Huang, Frédéric Frappart, Ruhollah Taghizadeh, Xianglin Zhang, Jean-Pierre Wigneron, Jie Xue, Yi Xiao, Jie Peng, et al.

Journal of Remote Sensing
A SCIENCE PARTNER JOURNAL

RESEARCH ARTICLE

Global Soil Salinity Estimation at 10 m Using Multi-Source Remote Sensing

Nan Wang[1,2], Songchao Chen[3], Jingyi Huang[4], Frederic Frappart[2], Ruhollah Taghizadeh[5], Xianglin Zhang[1], Jean-Pierre Wigneron[2], Jie Xue[1], Yi Xiao[1], Jie Peng[6], and Zhou Shi[1,7*]

[1]Institute of Agricultural Remote Sensing and Information Technology Application, College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310058, China. [2]Interactions Sol Plante Atmosphère, Bordeaux Sciences Agro, National Research Institute for Agriculture, Food and Environment, 33140, Villenave d'Ornon, France. [3]Zhejiang University-HangzhouGlobal Scientific and Technological Innovation Center, Hangzhou 311200, China. [4]Department of Soil Science, University of Wisconsin-Madison, 1525 Observatory Drive, Madison, WI 53706, USA. [5]Department of Geosciences, Soil Science and Geomorphology, University of Tübingen, 72070 Tübingen, Germany. [6]College of Plant Science, Tarim University, Alar 843300, China. [7]Key Laboratory of Spectroscopy Sensing, Ministry of Agriculture, Hangzhou 310058, China.

*Address correspondence to: shizhou@zju.edu.cn

Citation: Wang N, Chen S, Huang J, Frappart F, Taghizadeh R, Zhang X, Wigneron JP, Xue J, Xiao Y, Peng J, et al. Global Soil Salinity Estimation at 10 m Using Multi-Source Remote Sensing. *J. Remote Sens.* 2024;4:Article 0130. https://doi.org/10.34133/remotesensing.0130

Submitted 1 August 2023
Accepted 26 February 2024
Published 28 March 2024

Salinization is a threat to global agricultural and soil resource allocation. Current investigations of global soil salinity are limited to coarse spatial resolution of the available datasets (>250 m) and semiqualitative classification rules (five ranks). Based on these two limitations, we proposed a framework to quantitatively estimate global soil salt content in five climate regions at 10 m by integrating Sentinel-1/2 remotely sensed images, climate, parent material, terrain data, and machine learning. In hyper-arid and arid region, models established using Sentinel-2 and other geospatial data showed the highest accuracy with R^2 of 0.85 and 0.62, respectively. In semi-arid, dry sub-humid, and humid regions, models performed best using Sentinel-1, Sentinel-2, and other geospatial data with R^2 of 0.87, 0.80, and 0.87, respectively. The accuracy of the global models is considerable with field validation in Iran and Xinjiang, and compared with digitized salinity maps in California, Brazil, Turkey, South Africa, and Shandong. The proportion of extremely saline soils in Europe is 10.21%, followed by South America (5.91%), Oceania (5.80%), North America (4.05%), Asia (1.19%), and Africa (1.11%). Climatic conditions, groundwater, and salinity index are key covariates in global soil salinity estimation. Use of radar data improves estimation accuracy in wet regions. The map of global soil salinity at 10 m provides a detailed, high-precision basis for soil property investigation and resource management.

Introduction

The increasing global soil salinization that affects soil health and constrains agricultural production has become a major global land degradation problem. Over 1 billion hectares of soil suffer from salinization [1]. Saline soils are predominantly found in China, India, Pakistan, Iran, Australia, and the United States [2]. The consequences of soil salinity extend to environmental and social issues, particularly in arid and semi-arid regions [3]. Globally, about 20% of irrigated soils are salinized, with this figure rising to over 30% in Egypt, Iran, and Argentina [3]. Nevertheless, the availability of soil samples with salinization measurements remains limited, and comprehensive data products on soil salinity content are scarce. High-quality assessments of global soil salinization are lacking. Hence, a rapid and accurate detection method of soil salinity is needed to support sustainable agricultural practices and ecological protection.

Currently, a limited number of datasets qualitatively describe saline soils. The World Reference Base for Soil Resources (WRB) and Harmonized World Soil Database (HWSD) classify solonchaks at a spatial resolution of 250 m and 1 km, respectively. But salt also tends to accumulate in other types of soils such as Calcisols, Solonetz, and Arenosols. Although WRB and HWSD are important references for soil type classification, they are limited in expressing the continuity of soil salinity content [1]. The World Soil Information Service (WoSIS) offers a valuable resource with over 100,000 georeferenced soil salinity data points, but the sparse samples are limited to describe global soil salinization [4]. In areas with insufficient data, the European Union (EU) developed an area-frame randomized soil sampling (ARRSS) method, which functions to randomly generate samples with geographical information [5–7]. Although the soil dataset published by WoSIS is not evenly distributed, it has been widely used to effectively estimate soil attributes. Recent studies

are focusing on estimating soil salinity from regional to global scales by remotely sensed images.

The accumulation of salt in soil is affected by soil texture, moisture, and meteorological conditions [8]. Many studies have integrated specific covariates and machine learning (ML) regression algorithms to quantitatively estimate soil salinity at regional scales [4]. Image reflectance and salinity index derived from remote sensing data can be used to estimate the soil salinity, as soil with a higher salt content has a larger reflectance in the visible–near-infrared region of bands [9]. The composition of soil particle fractions and topographical and geological conditions affect the accumulation of water in soil [2]. As a high salt content in soils negatively affects the growth of crops, various vegetation indices have been used to characterize soil salinity in previous studies [10]. In arid and semi-arid regions, the high evapotranspiration and low precipitation cause rapid accumulation of salt in sandy soils [11]. In drier soils, salinity estimation frequently relies on soil texture information and salinity indexes calculated from remote sensing data. Conversely, in humid regions such as salt marshes in coastal areas, vegetation indexes, soil moisture content, and climatic conditions are decisive estimators of salinity [8]. Tree-based algorithms (for example, classification and regression trees), random forest (RF), cubist, and neural networks (for example, artificial and convolutional neural networks and temporal convolutional networks are most popular models used in the quantitative description of soil salinity) [9,12]. Ivushkin et al. [1] mapped global soil salinity changes at 250 m, by applying RF classification with soil property maps, field observations, and thermal infrared images between 1986 and 2016, with R^2 of 0.67 to 0.70. This ranked map of global soil salinity level is an important reference for soil property surveys, but it lacks a quantitative and high-resolution description of soil salinity content and cannot provide field-level actionable information for soil management. Hoa et al. [13] estimated soil salinity in Vietnam at 10 m, using Sentinel-1 images and ML approaches, and obtained an R^2 of 0.66. Nabiollahi et al. [11] calculated soil salinity in Iran at 10 m using Landsat 8 and Sentinel-2 data, with relatively good accuracy (R^2 ~ 0.60 to 0.80, when compared against in situ data). Quantitative inversion techniques have shown benefits in creating detailed soil salinity maps, and high-resolution remote sensing images with global coverage now provide valuable data on a global scale.

There is a lack of quantitative estimation of soil salinity at a high resolution and large scale. Based on this research gap, we proposed an innovative methodology to produce global soil salinity map at a spatial resolution of 10 m. This approach relies on four main steps: (a) collect and calculate global covariates derived from multi-source remote sensing and other geospatial data, (b) construct soil salinity training and testing datasets over sub-regions according to global climatic conditions, (c) quantitatively map global soil salinity using multi-source covariates and RF regression at each sub-region, and (d) evaluate the precision and uncertainty of the models using the bootstrap method. Our results provide fine resolution data for soil salinization management.

Materials

Global climate class data

The Global Aridity Index and Potential Evapo-Transpiration (AI-ET) Climate Database was used to assess global climate conditions at 1 km (https://cgiarcsi.community). The annual average of aridity index (AI; precipitation/evapotranspiration) over the 1970–2000 period was used to generalize climate classification according to United Nations Environment Programme (UNEP) [14]: hyper-arid (<0.03), arid (0.03 to 0.2), semi-arid (0.2 to 0.5), dry sub-humid (0.5 to 0.65), and humid regions (>0.65) (Fig. 1). It has been observed that salt is more likely to accumulate in climate zones with an AI lower than 0.5 [15]. Salinization can also affect soil health in naturally humid areas, such as coastal regions and river deltas, where soil may be inundated with seawater. Consequently, we classified global climate zones based on AI values to evaluate and estimate soil salinity content, considering climate regions with distinct dryness levels.

Ground soil salinity datasets

Global modeling datasets

WoSIS compiled about 96,000 geo-referenced soil profile samples around the world [4]. WoSIS record 18,148 electrical conductivity (EC_e) data at the top layer of soil profiles measured from a water-saturated soil paste extract [16]. Soil salinity samples were ranked into non-saline (<2 dS m^{-1}, 11,396 points), slightly saline (2 to 4 dS m^{-1}, 2,047 points), moderately saline (4 to 8 dS m^{-1}, 1,858 points), highly saline (8 to 16 dS m^{-1}, 1,524 points), and extremely saline (>16 dS m^{-1}, 1,323 points) [17]. The ground-truth dataset exhibits an uneven distribution due to the challenges of sampling locations, and in some cases, local data holders may be unwilling to make the data publicly available. To balance the number of points in different salinization levels with the five climate zones, we assume that there is no salinization in acidic soil in the arid region (defined by WRB) [18]; thus 3,000 samples were randomly generated in these regions and defined as "non-salinity" datasets to complement the original WoSIS datasets (Fig. 1). The sample selection procedure employs a randomized and unbiased points approach at the pixel scale. Notably, soil sampling sites are densely distributed in North America but sparsely distributed in Asia. At the global scale, most sampling points are concentrated in arid and semi-arid regions. Despite these limitations, the WoSIS database remains the most comprehensive and up-to-date source of global soil salinity investigation, thus serving as the foundation for this research [1].

Case studies and datasets

Previous research has predominantly focused on regional soil salinity estimation. Soil salinization typically affects soils in arid and semi-arid areas, leading previous studies to primarily concentrate on soil salinity in drier regions such as Iran, Iraq, Turkey, and Xinjiang in central Asia, as well as coastal areas like Shandong Province in China [19–23]. To address the regions where soil salinization is frequently observed or relatively severe, we have seven typical areas as case studies to evaluate the performance of our modeling approach. These cases include Xinjiang (China, Asia), the Yellow River Delta region (Shandong Province, China, Asia), Iran (Asia), Turkey (Asia), South Africa (Africa), Afonso Bezerra and Alto do Rodrigues (Brazil, South America), and California (United States, North America) (Fig. 2). The case of Iran primarily represents soil salinization in arid regions, offering valuable insights into salinity issues in such environments. Conversely, Shandong and Turkey, due to their proximity to the sea, serve as important cases for soil salinization studies in more humid regions, where unique challenges

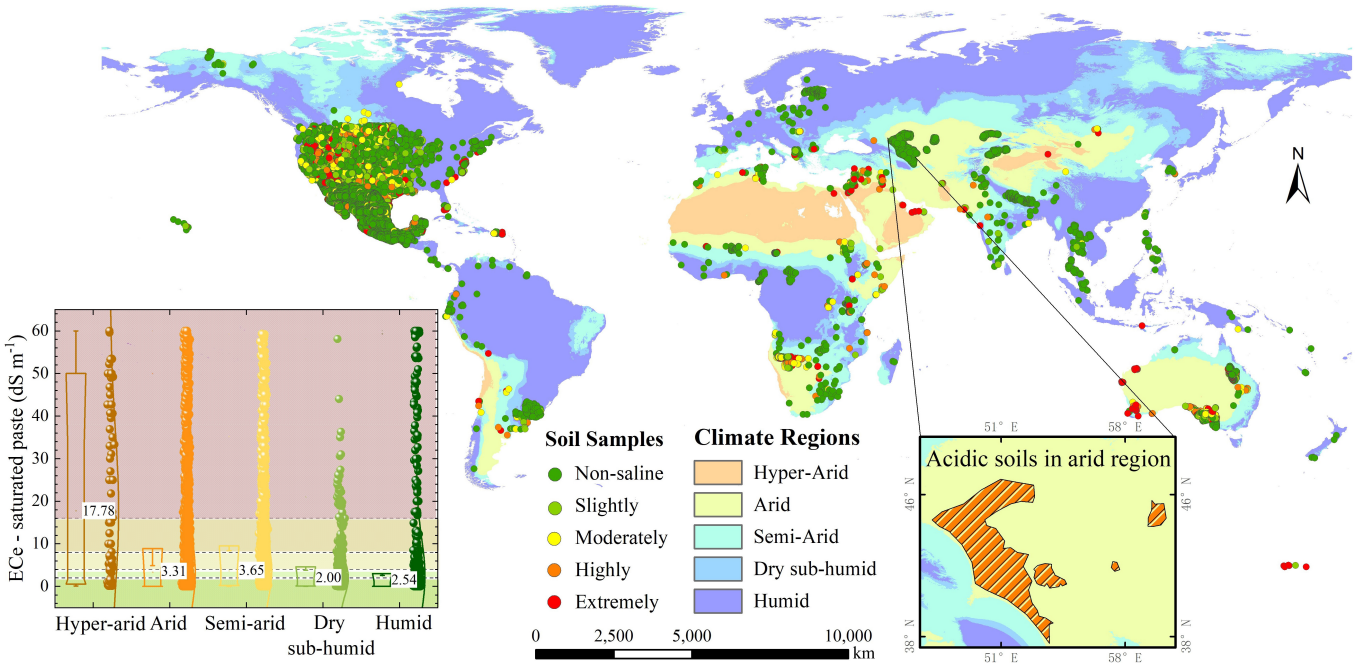
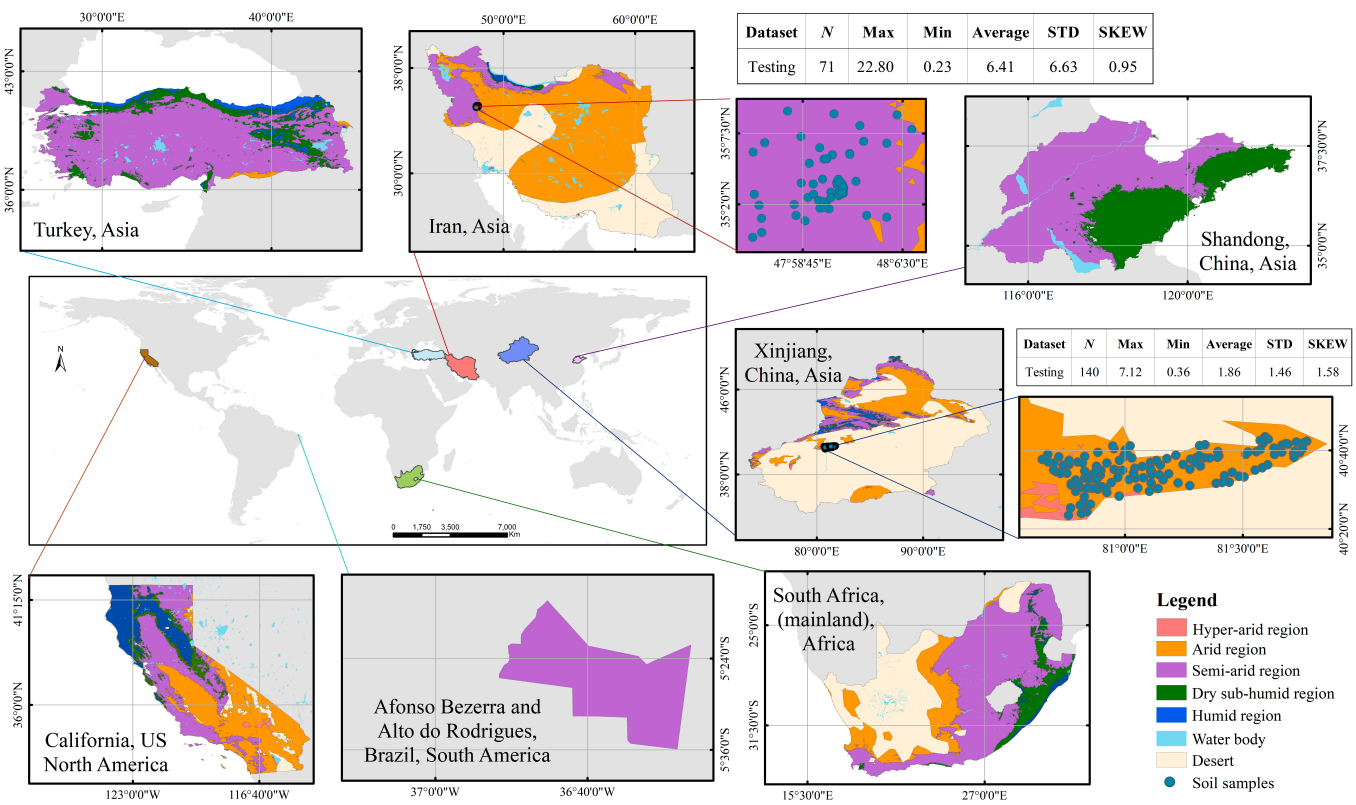


Fig. 1. Spatial distribution of the global climate class and the ground soil salinity observations. Notes: the subfigure shows an example of acidic soil areas defined by the World Reference Base for Soil Resources; the scatterplot shows the salt content of soil samples in five climate regions.



Dataset	N	Max	Min	Average	STD	SKEW
Testing	71	22.80	0.23	6.41	6.63	0.95

Dataset	N	Max	Min	Average	STD	SKEW
Testing	140	7.12	0.36	1.86	1.46	1.58

Fig. 2. Selected seven typical cases around the world (Turkey, Iran, Shandong, Xinjiang, California, Afonso Bezerra, and Alto do Rodrigues), with their climate regions, and validation samples in Iran and Xinjiang.

are encountered. Xinjiang, South Africa, and California represent different climatic regions, making them ideal candidates as representative cases for comprehensive inversion of soil salinity in diverse climatic zones across Asia, Africa, and North America. The case study in Brazil (Afonso Bezerra and Alto do Rodrigues) is especially relevant for understanding soil salinization in semi-arid regions in South America, providing crucial data and insights for this specific climatic context.

To verify the global model at regional scale and prove a robust estimation at spatial resolution of 10 m, we collected 71 topsoil samples (0 to 0.3 m) in Iran and 140 samples (0 to 0.3 m) in Xinjiang in 2019 as independent validation dataset. Field experiments and collection of regional soil samples of Iran were available in the central and western Isfahan Province. In Xinjiang, topsoil samples were collected through field surveys in arid and hyper-arid regions. Soil EC_e was measured in saturated paste. For cases in California [19], Afonso Bezerra and Alto do Rodrigues in Brazil [20], Turkey [21], South Africa [22], and Shandong province [23], the digital soil salinity maps published in previous studies were used as validation datasets. According to the published digital soil salinity maps, 50 verification points were randomly generated in each category of salinized areas (Fig. 3).

Soil salinity modeling covariates

Remote sensing data

Driven by the earth's water cycle, salt accumulates in soil driven by climate, soil conditions, vegetation, and topography. To identify soil salinity content, topography, and vegetation, soil properties based on remote sensing ground observation data and geographical auxiliary data are used for estimation. Optical Sentinel-2 and microwave Sentinel-1 images were employed to calculate various indices for observing soil salinity. Forty-four remote sensing covariates collected from previous studies are shown in Table 1. Sentinel-2 L2A Bottom of Atmosphere (BOA) reflectance data were computed by running sen2cor for the atmospheric correction, and were acquired using Google Earth Engine (GEE). To remove the clouds on the image, S2 images from 2019 August 1 to 2019 September 30 were selected to remove clouds with S2 cloud probability data. To enhance the robustness to outliers and extreme values and preserve the temporal patterns, the pixel-wise median value of the images was calculated at 10 m [24]. Sentinel-1 data obtained in dual-band cross-polarization and interferometric wide-swath mode were used. The median value of S1 images from 2019 August 1 to 2019 September 30 was calculated to obtain global-covered microwave remote sensing data at 10 m. A total of 44 covariates calculated using S2 and S1 data were used to estimate soil salinity (Table 1).

Other geospatial data

Additionally, numerous potential exhaustive Scorpan covariates relevant to soil formation are available at a global scale. The most common ones include climate, parent material, and terrain attributes [4,25]. Precipitation affects the soil evapotranspiration and changes the migration and accumulation of salt. Annual mean precipitation (mm) was obtained from the WorldClim 2.1 dataset at 1 km. The composition of soil particles

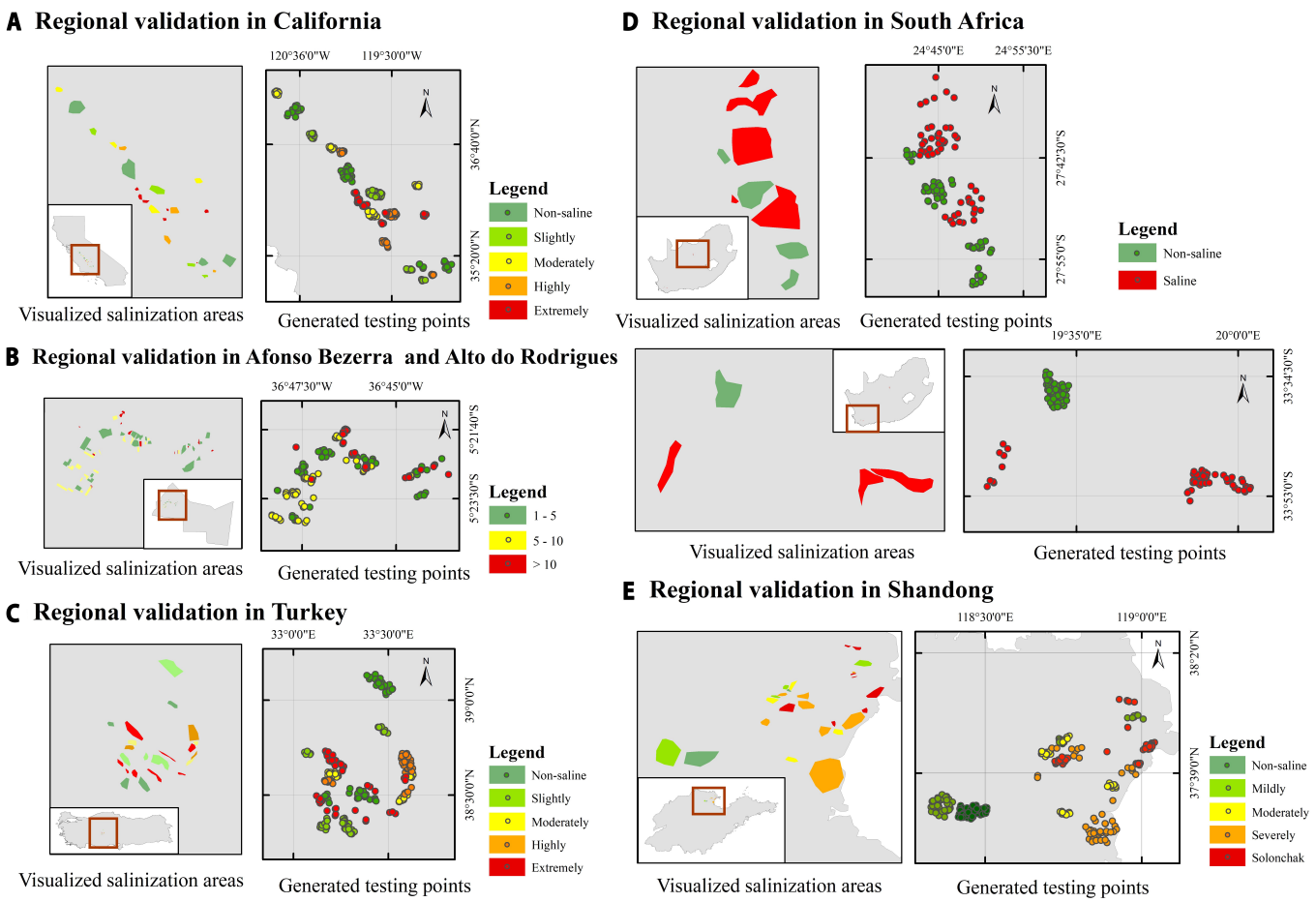


Fig. 3. Randomly generated salinization areas and validation samples from published soil salinity maps in (A) California, (B) Afonso Bezerra and Alto do Rodrigues, (C) Turkey, (D) South Africa, and (E) Shandong.

affects the water and fertility retention, and indirectly determines the presence of salt [11]. SoilGrids v2 provides one of the most recent and coherent global datasets of soil composition [25]. The soil composition patterns, including the proportion of silt, sand, and clay, were obtained from the SoilGrids dataset at a 250-m spatial resolution. Groundwater table depth (WTD) data reveal the control of the climate environment along with the distribution of global wetlands at a 1-km spatial resolution [26]. A multi-error-removed improved terrain (MERIT) digital elevation model (DEM), which provides elevation and slope data at about 90 m, was used as terrain-based auxiliary data [27]. All remote sensing data, environmental auxiliary data, and terrain auxiliary data (Table 2) were resampled to 10 m using bilinear resampling method to construct a covariate dataset for global soil salinity estimation.

Our study aimed at investigating the salinity in soils, so glaciers, permafrost, deserts, and seawater submerged areas are not considered. The climate type of global desert, tundra, and frost was classified by the Koppen–Geiger system at 0.1° × 0.1° grid [27]. Desert and tundra climates are considered extreme climate types, which are unsuitable for plant growth or land use, so they were masked in this study. The global extent of tidal flats from 2014 to 2016 was masked using the Murray Global Intertidal Change Dataset at 30 m using the GEE platform [28]. The flooded areas were masked using the Global Flood Database from Terra and Aqua MODIS sensors [29].

Table 1. Forty-four covariates with feature types, description, and formula

Feature types	Description	Formula	Reference
Original backscatter feature	Normalized backscatter coefficient	$\gamma^0_{vv}, \gamma^0_{vh}$	-
Radar index	Ratio of backscatter coefficient	$\text{Ratio} = \frac{\gamma^0_{vv}}{\gamma^0_{vh}}$	[56]
	Total scattering power	$\text{SPAN} = \gamma^{0\,2}_{vv} + \gamma^{0\,2}_{vh})^{0.5}$	[58]
	Normal difference index	$\text{NDI} = \gamma^0_{vv} - \gamma^0_{vh}$	[56]
	Radar vegetation index	$\text{RVI} = \gamma^0_{vv} / (\gamma^0_{vv} + \gamma^0_{vh})$	[59]
	Square difference index	$\text{SDI} = (\gamma^{0\,2}_{vv} - \gamma^{0\,2}_{vh}) / (\gamma^{0\,2}_{vv} + \gamma^{0\,2}_{vh})$	[56]
Salinity indices	Salinity index	$\text{SI} = (\text{G} \times \text{R})^{0.5}$	[60]
	Salinity index 1	$\text{SI1} = (\text{G} + \text{R})^{0.5}$	[61]
	Salinity index 2	$\text{SI2} = (\text{NIR}^2 + \text{G}^2 + \text{R}^2)^{0.5}$	[61]
	Salinity index 3	$\text{SI3} = (\text{G}^2 + \text{R}^2)^{0.5}$	[61]
	Salinity index 4	$\text{SI4} = \text{SWIR}_1 / \text{NIR}$	[62]
	Salinity index I	$\text{S1} = \text{B} / \text{R}$	[61]
	Salinity index II	$\text{S2} = (\text{B} - \text{R}) / (\text{B} + \text{R})$	[61]
	Salinity index III	$\text{S3} = \text{G} \times \text{R} / \text{B}$	[61]
	Salinity index V	$\text{S5} = \text{B} \times \text{R} / \text{G}$	[61]
	Salinity index VI	$\text{S6} = \text{R} \times \text{NIR} / \text{G}$	[61]
	Salinity index VII	$\text{S7} = (\text{SWIR}_1 - \text{SWIR}_2) / (\text{SWIR}_1 + \text{SWIR}_2)$	[63]
	Salinity index VII	$\text{S8} = (\text{G} + \text{R}) / 2$	[62]
	Salinity index IX	$\text{S9} = (\text{G} + \text{R} + \text{NIR}) / 2$	[62]
	Salinity index	$\text{SI-T} = \text{R} / \text{NIR} \times 100$	[60]
	Soil salinity and sodicity indices1	$\text{SSSI-1} = \text{R} - \text{NIR}$	[63]
	Soil salinity and sodicity indices2	$\text{SSSI-2} = (\text{R} \times \text{NIR} - \text{NIR} \times \text{NIR}) / \text{R}$	[63]
	Normalized difference salinity index	$\text{NDSI} = (\text{NIR} - \text{SWIR}_1) / (\text{NIR} + \text{SWIR}_1)$	[64]
	Canopy response salinity index	$\text{CRSI} = [(\text{NIR} \times \text{R} - \text{G} \times \text{B}) / (\text{NIR} \times \text{R} + \text{G} \times \text{B})]^{0.5d}$	[65]
	Salinization remote sensing index	$\text{SRSI} = [(\text{NDVI} - 1)^2 + \text{SI}^2]^{0.5}$	[66]
	Salinity ratio index	$\text{SAIO} = (\text{G} - \text{NIR}) / (\text{B} + \text{NIR})$	[3]
	Enhanced residues soil salinity index	$\text{ERSSI} = \text{G}^2 / (\text{B} \times \text{SWIR}_1)$	[57]
Soil indices	Clay index	$\text{CLEX} = \text{SWIR}_1 / \text{SWIR}_2$	[47]
	Gypsum index	$\text{GYEX} = (\text{SWIR}_1 - \text{NIR}) / (\text{SWIR}_2 + \text{NIR})$	[47]
	Carbonate index	$\text{CAEX} = \text{G} / \text{B}$	[47]
	Brightness index	$\text{BI} = (\text{G}^2 + \text{B}^2)^{0.5}$	[60]

(Continued)

Table 1. (Continued)

Feature types	Description	Formula	Reference
Vegetation indices	Ratio vegetation index	$RVI = NIR / R$	[67]
	Enhanced normalized differential vegetation index	$ENDVI = (NIR + SWIR_1 - R) / (NIR + SWIR_2 + R)$	[44]
	Infrared percentage vegetation index	$IPVI = NIR / (NIR + R)$	[68]
	Generalized difference vegetation index	$GDVI = (NIR^2 - R^2) / (NIR^2 + R^2)$	[69]
	Non-linear vegetation index	$NLI = (NIR^2 - R) / (NIR^2 + R)$	[69]
	Green atmospherically resistant vegetation index	$GARI = \{NIR - [G + \gamma \times (B - R)]\} / \{NIR + [G + \gamma \times (B - R)]\}$	[70]
	Normalized differential vegetation index	$NDVI = (NIR - R) / (NIR + R)$	[44]
	Differential vegetation index	$DVI = NIR - R$	[9]
	Enhanced vegetation index	$EVI = (1 + L) \times (NIR - R) / (NIR + C1 \times R - C2 \times B + L)$	[71]
	Soil-adjusted vegetation index	$SAVI = [(NIR - R) / (NIR + R + L)] \times (1 + L)$	[70]
	Optimized soil-adjusted vegetation index	$OSAVI = (NIR - R) / (NIR + R + 0.16)$	[72]
	Global vegetation moisture index	$GVMI = [(NIR + 0.1) - (SWIR_1 + 0.02)] / [(NIR + 0.1) + (SWIR_1 + 0.02)]$	[73]

Table 2. Geospatial covariates with spatial resolution, unit, and reference

Feature type	Covariate	Spatial resolution	Unit	Reference
Climate data	Annual mean precipitation (prec)	1 km	mm	WorldClim 2.1
	Evapo-transpiration (ET)	1 km	$mm\,day^{-1}$	The Global Aridity Index and Potential Evapo-Transpiration Climate Database
	Aridity index and potential evapo-transpiration index (AI-ET)	1 km	/	The Global Aridity Index and Potential Evapo-Transpiration Climate Database
Parent material	Silt	250 m	%	SoilGrids
	Sand	250 m	%	SoilGrids
	Clay	250 m	%	SoilGrids
Groundwater	Groundwater table depth (WTD)	1 km	m	[26]
Terrain data	Elevation	90 m	m	MERIT DEM
	Slope	90 m	°	MERIT DEM

Methods

RF algorithm

By integrating S1, S2, and other geographic data, we used the global ground soil salinity data to train the bootstrap-RF model. All data were divided into five sub-datasets according to the AI, and five estimation models were independently constructed in hyper-arid, arid, semi-arid, dry sub-humid, and humid regions. Then, we estimated soil salinity in five climate zones and verified the accuracy and uncertainty to obtain a global soil salinization map at 10 m. In addition, we conducted case analysis and validation on six regional-scale area. The methodology flowchart is shown in Fig. 4.

The RF is a classification and regression tree-based (CART) algorithm. It reprocesses the predictions from each generated tree to ensemble a stronger learner [4,30]. Two parameters (the number of trees, ntree, and the number of variables in each split, mtry) decided the largest possible size of the regression without being pruned [30]. RF randomly allocates features to trees in sub-datasets by bootstrap aggregating [31]. RF automatically processes various independent variables (including multicollinearity test) to produce a robust prediction [4]. Each tree in the forest prunes its branches according to the importance of the covariates to obtain predicted values. All trees are then averaged to provide the final result [11]. As the different indices are not sensitive to the same soil surface properties, we

evaluated the RF regressions of three datasets of modeling variables for the five climate regions. The three modeling datasets were as follows: (A) 7 indices calculated from S1 images and 8 covariate indicators derived from other geospatial data; (B) 37 indices calculated from S2 images and 8 covariate indicators derived from other geospatial data; (C) 37 indices calculated from S2 images, 7 indices calculated from S1 images, and 8 covariate indicators derived from other geospatial data. When modeling in each climate region, mtry was looped from 1 to 15, 45, and 52 for models (A), (B), and (C), respectively. ntree was optimized from 500 to 1,000.

For each RF, the variable importance (VI) shows the contribution of a predictive covariate [30]. For each tree, the variable is randomly permuted to test the importance of X^i before training, which is calculated as Eq. 1:

$$VI(X^i) = \frac{1}{ntree}\sum_t \left(errO\widetilde{O}B_t^i - errOOB_t\right) \quad (1)$$

where $errO\widetilde{O}B_t^i$ is the perturbed set with the permutation of X^i.

Accuracy and uncertainty control

The nonparametric bootstrapping technique was applied 20 times to mitigate the bias associated with the specific training set selection and assess the uncertainty linked to the estimation.

A 95% probability within the confidence interval (CI) limits was used to indicate that the true value of soil EC value has 95% possibility within the interval between upper and lower CI limits [12]. The uncertainty (Eq. 3) was used to assess the deviation of 20 bootstrap results about the possible changing ranges and reduce the estimation errors of soil salinity

$$CIs = \overline{Y} \pm a \times \frac{SD}{\sqrt{n}} \quad (2)$$

$$Uncertainty = \frac{CI_{upper} - CI_{lower}}{\overline{Y}} \quad (3)$$

where $\overline{Y}$ is the average value of the 20 estimations. The constant of a was 1.725, and SD was the standard deviation of estimations. CI_{upper} and CI_{lower} were the upper (95%) and lower (5%) bounds of CIs.

R^2 (Eq. 4) and root mean square error (RMSE) (Eq. 5) were used to evaluate the fitting accuracy of the models. Lin's concordance correlation coefficient (LC) (Eq. 6) was used to evaluate the accuracy between measured and estimated soil salinity values

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \quad (4)$$

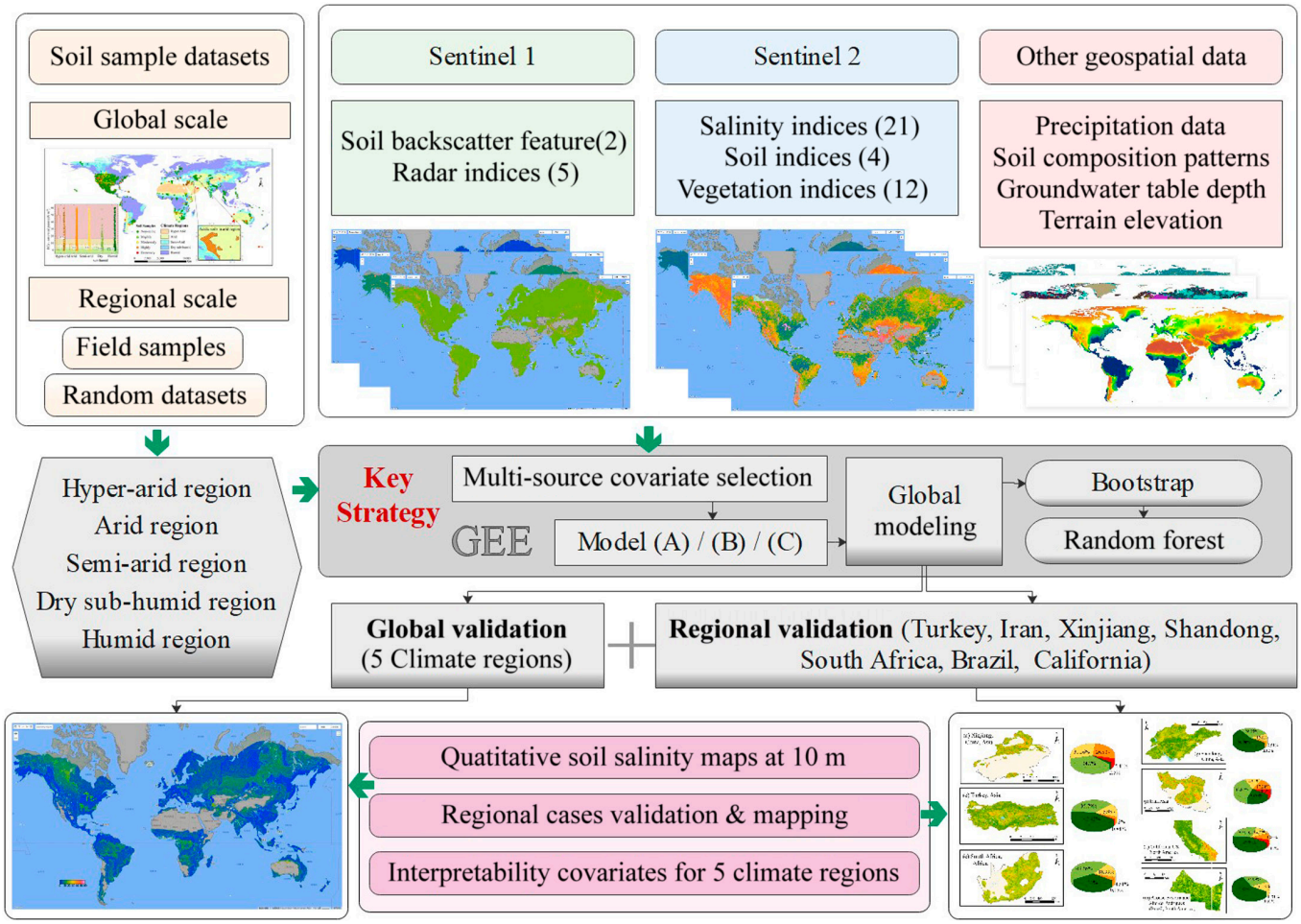


Fig. 4. Methodology flowchart developed for global soil salinity estimation.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(\left(y_i - \hat{y}_i\right)^2\right.}{n}}. \quad (5)$$

$$LC = \frac{2 \times r \times s_{Y_{model}} \times s_{Y_{testing}}}{s^2_{Y_{model}} + s^2_{Y_{testing}} + \left(\overline{Y}_{\text{model}} - \overline{Y}_{\text{testing}}\right)^2}. \quad (6)$$

where n is the sample number. y_i and $\hat{y}_i$ are the observed and predicted soil EC value, and $\overline{y}$ is the mean observed value. r is Pearson coefficient of the observation and prediction. s_{Ymodel} and $s_{Ytesting}$ are standard deviation of each datasets. The classification evaluation indicators are user's accuracy, producer's accuracy, overall accuracy, and Kappa coefficient in study cases [32].

Results

Soil salinity dataset construction

A total of 21,148 points were classified as five datasets by five climate classes to construct ground soil salinity datasets, including 90 points in hyper-arid region, 8,077 points in the arid region, 8,419 points in the semi-arid region, 1,702 points in the dry sub-humid region, and 2,860 points in the humid region. Training and testing datasets were randomly divided according to the ratio of 3:1 on the soil salinity content sequence in each climate region (Table 3). The statistical characteristics of sub-datasets were similarly distributed in different climate regions, which ensures the efficiency of model training. The overall mean and median values of soil salinity datasets increase with increasing levels of drought. However, although saline soils mainly harm arid regions, extremely salinized soil samples (>16 dS m^{-1}) also existed in humid and dry sub-humid regions.

Global modeling evaluation and uncertainty assessment

A total of 52 indices from multi-source data were used to model soil EC values, including 7 indices calculated from S1 images, 37 indices calculated from S2 images, and 8 covariate indicators derived from other geospatial data. The mean values of the 20 estimations were obtained as the final results using bootstrap methods. Table 4 lists the accuracy results in the five climate regions using the three modeling datasets. In hyper-arid and arid regions, modeling with indices derived from S2 and other geospatial data performed best among the three models with an R^2 of 0.85 and 0.62, and LC of 0.83 (strong agreement) and 0.70 (moderate agreement), respectively. As the humid level increases, the use of S1 images improved the estimation accuracy in semi-arid, dry sub-humid, and humid regions with an R^2 of 0.87, 0.80, and 0.87, and LC of 0.87, 0.80, and 0.87 (strong agreement), respectively. The scatterplots of the RF models in the five climate regions using the three modeling datasets are shown in Fig. 5. Although the estimation accuracy is high, there is a deviation from the 1:1 line. Especially for non-saline and extremely saline samples, an overestimation for low values and an underestimation for high values can be observed. Global soil salinity estimation models performed well for soils classified with low salinity, but the models did not do a good job for predictions of non-saline and extremely saline soils.

The uncertainty results (calculated by Eq. 3) are shown in Fig. 6. The mean uncertainty values for each validation data were 0.16, 0.23, 0.12, 0.13, and 0.20 for the five climate regions, respectively. Estimating soil EC in the semi-arid region represented the most stable modeling, with the mean uncertainty value of 0.12. Figure 6 also shows the uncertainty statistics estimated for soil samples with different salinization in the modeling of the five climate regions. For salinity estimation in hyper-arid region, the model showed the lowest average uncertainty in the extremely salinized soil dataset. For the arid, semi-arid, and dry sub-humid regions, the models showed low average uncertainty values in slightly salinized, moderate salinized, and highly salinized samples. For humid regions, however, the highly salinized soil dataset exhibited the lowest estimation uncertainty. The estimation uncertainties for the dry sub-humid and humid regions were generally higher than those for the other three climate regions. Also, the overall uncertainty values were relatively higher in estimating non-saline soils.

Explanatory covariates

Model (B) was used to estimate soil salinization in hyper-arid and arid regions, and model (C) was applied to semi-arid, dry sub-humid, and humid regions. For each climate region, the

Table 3. Descriptive statistics of the sub-datasets in the five climate regions. All in dS m^{-1} except *N*.

Climate class	Dataset	*N*	Max	Min	Average	STD	SKEW
Hyper-arid	Training	67	59.60	0.10	17.53	18.34	0.88
	Testing	23	60.00	0.10	18.52	19.31	0.89
Arid	Training	6,058	60.00	0.01	3.31	8.49	4.13
	Testing	2,019	60.00	0.01	3.32	8.50	4.14
Semi-arid	Training	6,314	59.30	0.01	3.65	6.44	4.33
	Testing	2,105	59.20	0.01	3.69	6.42	4.11
Dry sub-humid	Training	1,276	58.10	0.01	2.03	4.09	6.00
	Testing	426	30.60	0.01	1.93	3.56	4.57
Humid	Training	2,145	60.00	0.01	2.55	8.63	5.53
	Testing	715	60.00	0.01	2.51	8.50	5.60

Table 4. Accuracy assessment of the RF regression models in the five climate regions using the three modeling strategies: model (A): S1 + other geospatial data; model (B): S2 + other geospatial data; model (C): S1 + S2 + other geospatial data (RMSE in dS m^{-1}).

	Model (A)			Model (B)			Model (C)			Parameters	
	R^2	RMSE	LC	R^2	RMSE	LC	R^2	RMSE	LC	mtry	ntree
Hyper-arid	0.63	12.68	0.67	**0.85**	**9.18**	**0.83**	0.81	10.74	0.74	34	500
Arid	0.21	7.74	0.33	**0.62**	**5.61**	**0.70**	0.55	5.90	0.67	23	500
Semi-arid	0.70	3.75	0.75	0.80	2.87	0.86	**0.87**	**2.76**	**0.87**	28	800
Dry sub- humid	0.49	2.56	0.64	0.70	1.81	0.81	**0.80**	**1.99**	**0.80**	39	800
Humid	0.65	5.89	0.73	0.84	4.88	0.84	**0.87**	**3.65**	**0.87**	32	500

Note: Bold data indicate the highest accuracy of the model.

most important variable is shown in Fig. 7, and the top 20 relatively important ranking of covariates is shown in Fig. 8A to E. AI-ET, evapo-transpiration (ET), WTD, and slope played the most important role in quantitatively estimating soil salinity in five climate regions. Based on the range of climate divisions, slope dominates the largest areas (semi-arid and humid regions). Especially for hyper-arid and arid regions, climatic conditions were decisive to the formation and estimation of salinity, as a significantly larger proportion of importance showed (34% for AI-ET in hyper-arid region and 15% for ET in arid regions). AI-ET, WTD, S7, slope, and CLEX were included as indicative independent variables in soil salinity estimations for all five climate zones. In more arid regions (hyper-arid, arid, semi-arid, and dry sub-humid regions), WTD was the top five important independent variables for salinity estimation. Figure 8F shows the statistical proportion of importance in five categories (topography, climate conditions, vegetation indices, salinity indices, soil properties) calculated from 20 covariates derived from estimating five climate regions. In hyper-arid and arid regions, the cumulative contribution of climate conditions was high, especially for AI-ET and ET, because only S2 data were used for modeling, so climate auxiliary data played an important role. However, for semi-arid, dry sub-humid, and humid regions, soil properties were more indicative of estimating, because of the introduction of the S1 data to the characterization of soil backscattering coefficients. In summary, environmental climate conditions, soil moisture content, and groundwater were important covariates for estimating global soil salinity.

Global salinity maps at 10 m

We trained 20 bootstrap-RF models in the five climate regions with covariates and parameters shown in Table 3 and averaged to one prediction. The global soil salinity map at 10 m is presented in Fig. 9. After masking desert, snow, ice, intertidal, and flooded areas, the proportion of soils at the five salinization levels was calculated using GEE. The results of the seven continents are shown in Fig. 10.

The more arid and semi-arid regions are located in Oceania, North America, and Asia (except Sahara and Sahel in North Africa, not considered in this study), and higher proportions (>20%) of highly and extremely saline soils were found over these areas. Oceania has the lowest proportion of non-saline soils. The higher salinization in Africa is mainly located in the central and southern regions, such as Congo and Kenya. The saline soils of Asia are concentrated in Central Asia (China, Mongolia) and the Middle East (Iran, Iraq, Saudi Arabia). The saline soils of South America are distributed in the center part (Brazil) and the southern part (Argentina). In North America, higher soil salinity areas are observed in the United States.

Regional-scale validation

The estimated soil salinity content and the uncertainty values of the seven cases calculated from 20 times RF regressions, as well as the proportion of each salinization level, are shown in Fig. 11. For cases in California, Afonso Bezerra and Alto do Rodrigues, Turkey, South Africa, Shandong, Iran, and Xinjiang, the validation results based on the published soil salinity digital maps and field investigations are shown in Table 5.

Xinjiang is located in northwest China, with the Taklimakan Desert occupying most of the area. Figure 11A shows that the soil salinity content in Xinjiang is 0.015 to 33.44 dS m^{-1}, with uncertainty values of 0.0066 to 2.18 and an average of 0.97 (Fig. 11B). The soil salinity content in Turkey is 0.27 to 22.45 dS m^{-1} (Fig. 11C). In the random validation based on published digital maps of soil salinization on a regional scale in Turkey (Table 5C), underestimation of high values and overestimation of low values were also observed. The validation accuracy was lower in slightly salinized areas, while the highest results were validated in the non-saline and extremely saline areas. Figure 11E shows that the soil salinity content in South Africa is 0.15 to 26.61 dS m^{-1} and the uncertainty values are 0.022 to 1.25, with an average of 0.17. The zonal models for estimation performed a considerable ability in the qualitative classification of salinized and non-salinized soils (Table 5D). Figure 11G shows that the soil salinity content in Shandong is 0.089 to 22.78 dS m^{-1}. The models showed the lowest accuracy values in the classification validation of moderately salinized and solonchak soils. Especially in solonchak soils, most of the samples were estimated with lower values (mostly classified as severely saline soils) (Table 5E). Figure 11I shows the soil salinity content in Iran of 0.032 to 33.92 dS m^{-1}, with an average uncertainty of 0.12. Similar to the results of the 1:1 line in Fig. 12, underestimation and overestimation are significant at regional scale, especially in the non-saline soil region. Non-saline soils in Iran samples are overestimated, while extremely saline samples are underestimated. Figure 11K shows that the soil salinity content in California is 0.025 to 46.31 dS m^{-1} The average uncertainty value is 0.16. The California case area

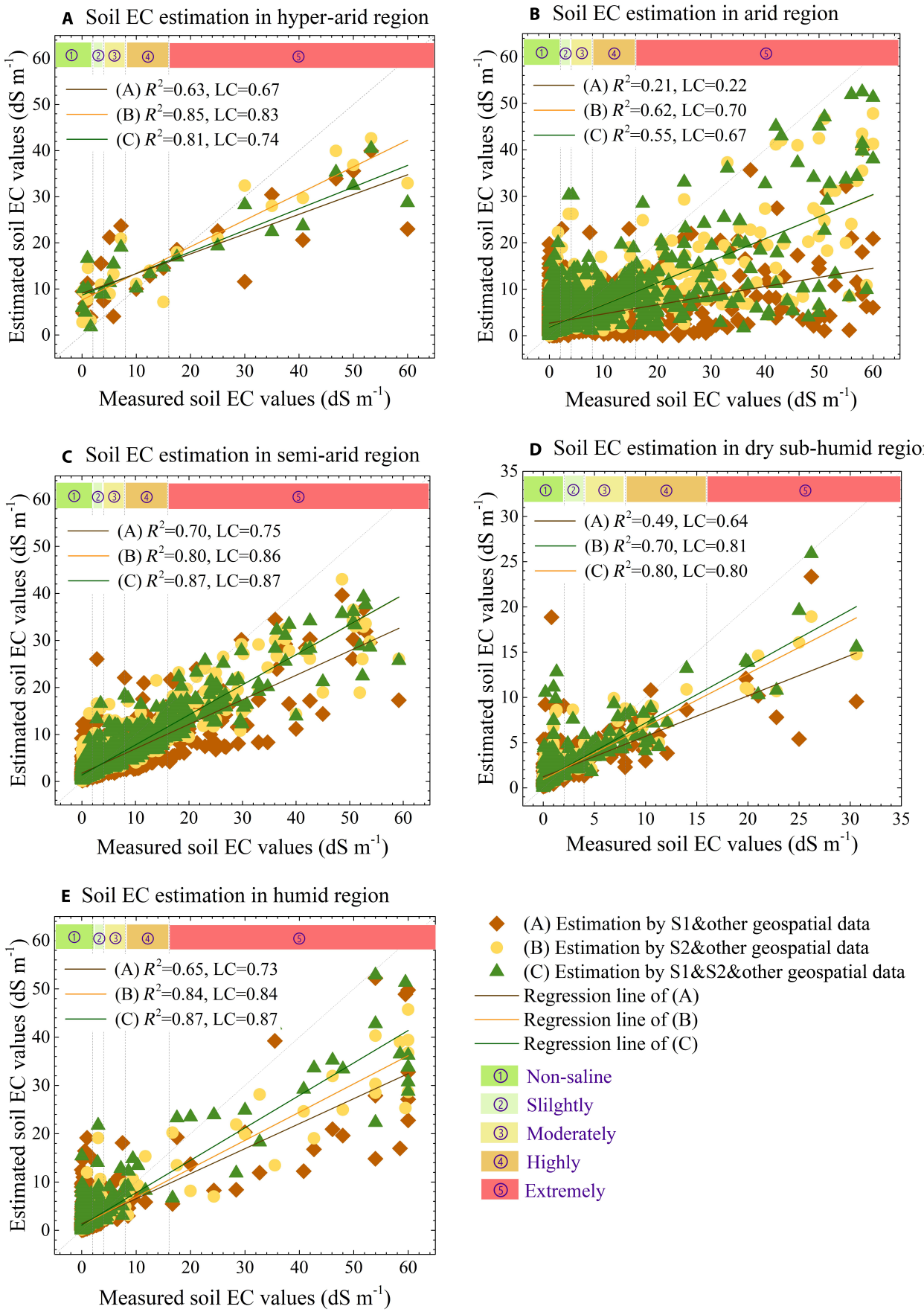


Fig. 5. Scatterplots of the RF models using model (A), model (B), and model (C) in (A) hyper-arid, (B) arid, (C) semi-arid, (D) dry sub-humid, and (E) humid region.

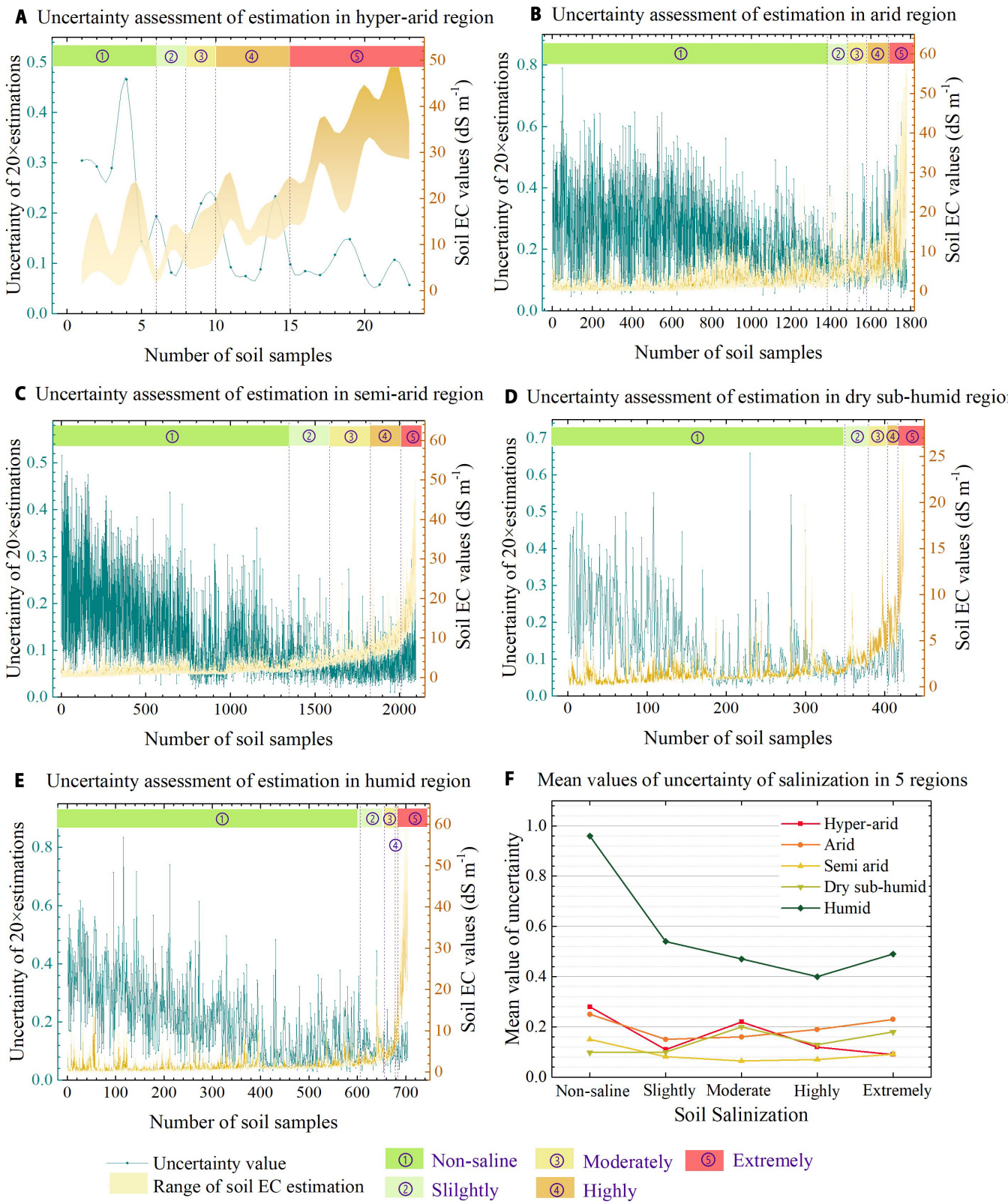


Fig. 6. The uncertainty values and estimated soil EC values for each validation sample with the 20 RF regressions in (A) hyper-arid, (B) arid, (C) semi-arid, (D) dry sub-humid, and (E) humid region and (F) mean values of uncertainty of salinization.

showed low rates of correct classification in the slightly and extremely saline datasets. An obvious low-value overestimation and high-value underestimation problem can be observed in the confusion matrix (Table 5A). The soil salinity content in Afonso Bezerra and Alto do Rodrigues is 0.22 to 20.25 dS m^{-1}, with a mean uncertainty of 0.11 (Fig. 11M). Field surveys, ground soil samples, and true salt content measurements obtained in Iran and Xinjiang were used to validate models at a field scale. The

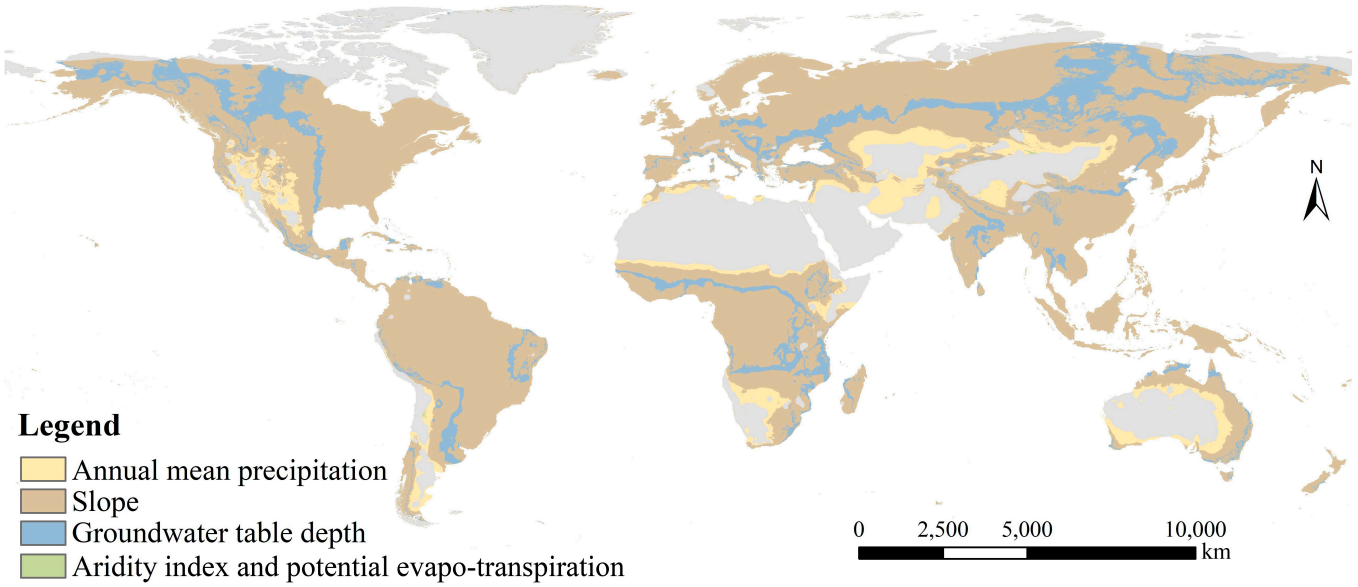


Fig. 7. The most important variables for estimating soil salinization.

validation R^2 values (0.60 and 0.53, respectively) and *LC* values (0.67 and 0.67, respectively) indicate that trained models are credible at 10 m. According to the random validation approach based on published soil salinity maps, the overall accuracy values ranged from 0.36 to 0.61. The method was verified to be credible at the landscape scale in South Africa and Afonso Bezerra and Alto do Rodrigues in Brazil.

Discussion

Interpretable covariate analysis

Salt in the soil migrates and accumulates with water. The degree of drought and humidity in the climate determines the distribution of water and affects the type and rate of salinization [33]. In arid and semi-arid regions, salinity control is generally a consequence of leaching driven by strong evaporation. However, soil salinization in humid and coastal areas is dominated by groundwater and seawater intrusion [34]. Estimating global soil salinization according to climate types revealed the mechanism of decisive covariates in primary and secondary soil salinization processes, which helped to improve the modeling [35]. The rate of evapotranspiration is a very relevant predictor of soil salinity modeling as it controls the processes of salt accumulation and leaching, which determines the zoning modeling of soil salinization [36]. Figure 8A to E shows that, in arid regions, climatic conditions and evapotranspiration contribute more to soil salinity estimation than other factors, with relative importance greater than 10% for each of them. The extremely dry climate makes the soil evapotranspiration rate much higher than the precipitation rate, resulting in a strong upward migration of soil moisture in soil layers. Moving along with water, salt accumulates to the surface and results in less leaching in soils. Soil salinity in arid climate was primarily controlled by drought and heat [37]. In contrast, topography, groundwater, and soil texture increasingly contribute to salinity as the climatic conditions become wetter. Historically, climate and topography are two controlling covariates to global salinization [38]. According to the soil salinity analysis in sub-humid regions by Nosetto et al. [39], topography and groundwater depth are critical to soil salinity distribution and content in humid regions. The enhanced groundwater may promote the intrusion of soluble salt groundwater from deeper levels of the aquifer, which enhanced salinization [40]. Soluble salts are easier to accumulate in areas with a low slope from the upslope areas with the move of water [11]. Figure 8 indicates the contribution of soil backscatter coefficients (VV, VH) ranked in the top 20 as key covariates in semi-arid, dry sub-humid, and humid regions. It can also be seen from Table 4 that in these regions, model (C) with S1 data performed best, exhibiting the highest accuracy. It shows that detecting soil moisture was conclusive for the improvement of soil salinity estimation in wetter areas as radar data were sensitive to soil moisture [41]. These results provide a new framework for feature selection in soil salinity modeling. In the tree-based algorithm, each decision tree trains a salt content estimation model based on the relative feature importance of covariates and assigns weights to indicators based on salinization characteristics under climate conditions [42]. According to covariates that are valid for soil salinity estimates (AI-ET, WTD, S7, slope, and CLEX) in the five climate regions, climate conditions, groundwater depth, topography, and soil surface reflectance characteristics are key parameters to salinity estimation at the global scale.

Global distribution of soil salinization

The soil salinity map indicates that Central Asia, Middle East, South America, North Africa, and the coastal areas in Australia are facing severe soil salinization, which is similar to the conclusion from the World Soil Day Conference (https://www.fao.org/world-soil-day/en/). Oceania has a higher proportion of saline soil because the desert of Australia occupies more than 40% of Australia's land area. In terms of the salt-affected lands at a national scale, China, Australia, Kazakhstan, Canada, and Iran are the more seriously affected. It is similar to the ranking by Hassani et al. [43]. At the global scale, soils in dry regions (hyper-arid, arid, semi-arid) are more salinized than in wet climates (dry sub-humid, humid).

Different from the types of salinization in most arid areas, the types of salinization in Shandong Province, China are

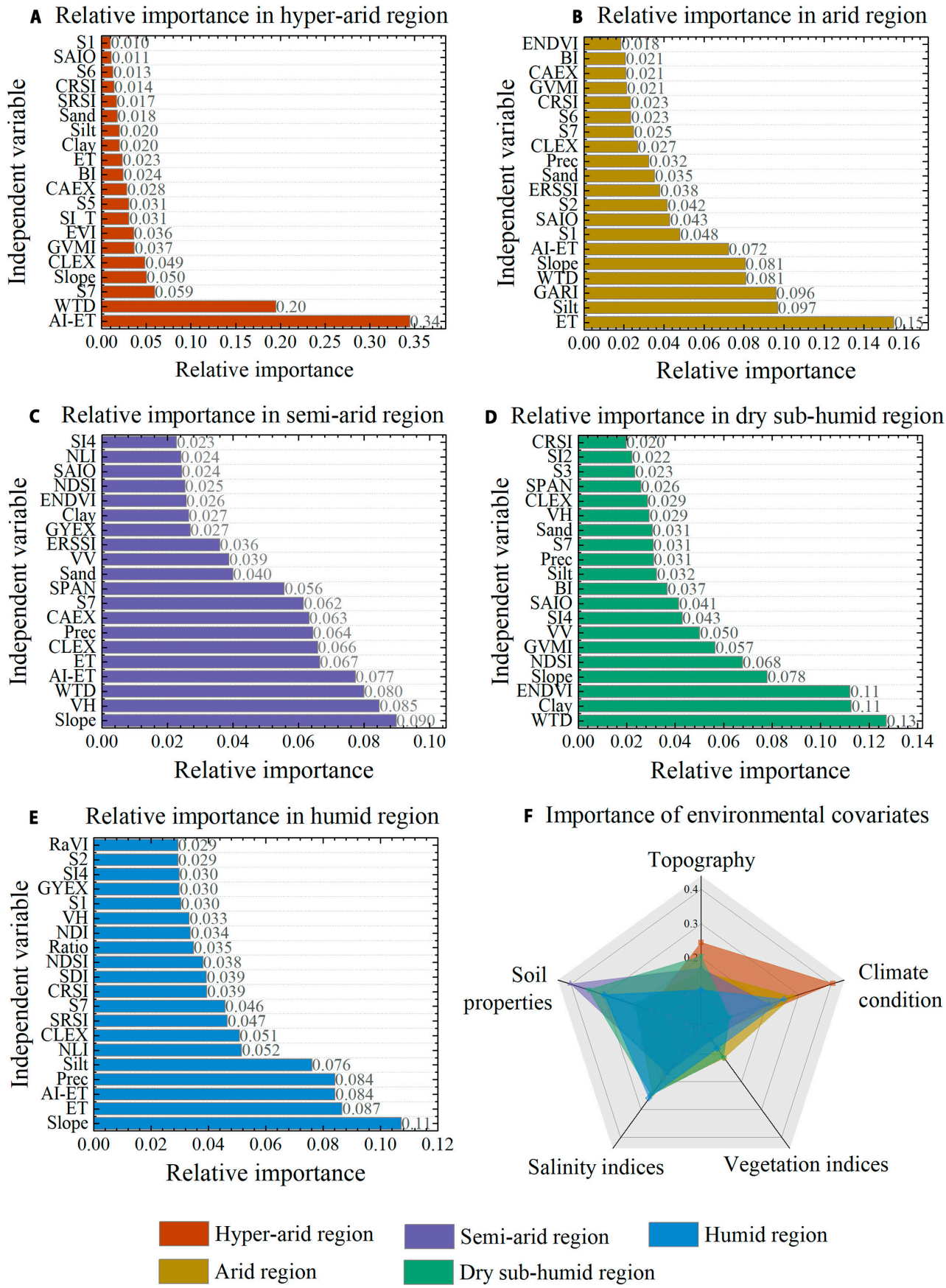


Fig. 8. The relative importance ranking of covariates (top 20) obtained by RF algorithm in (A) hyper-arid, (B) arid, (C) semi-arid, (D) dry sub-humid, and (E) humid region and (F) statistical proportion of importance in five categories (topography, climate conditions, vegetation indices, salinity indices, and soil properties).

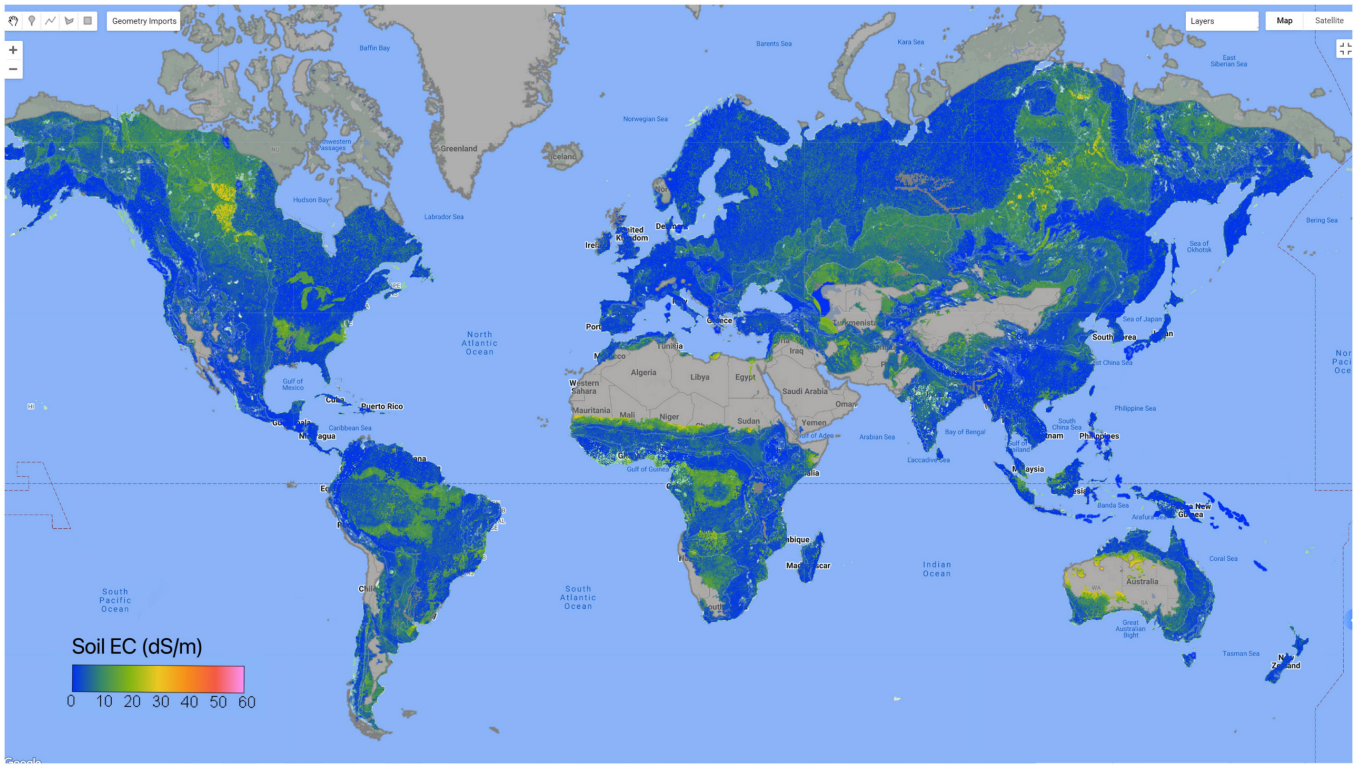


Fig. 9. Map of quantitative global soil salinity estimation at 10 m using Google Earth Engine.

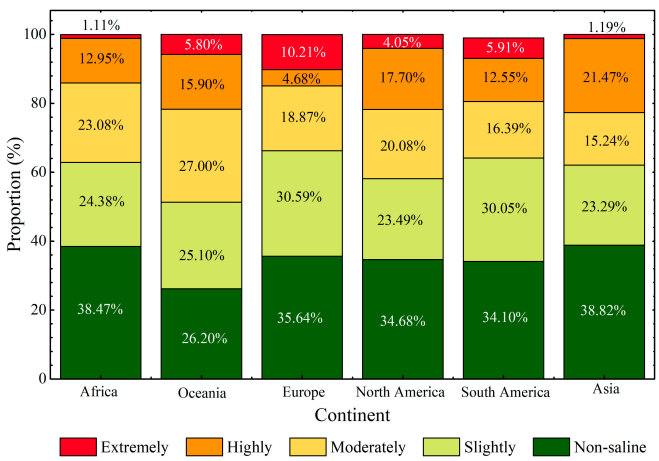


Fig. 10. The proportion of different salinized soils over the six continents.

mainly coastal saline soils. Consistent to the results from Chen et al. [23], but more refined, the saline soils were mostly identified along the banks of the Yellow River. In addition, high and extremely high saline soils were observed in offshore areas (northeast of the case study) [23]. The climatic types of non-desert areas are mainly arid and semi-arid in Xinjiang, China. Figure 11A shows the same conclusion as Wang et al. [44]. Saline soil was mostly found around Ebinur Lake, but with a higher salt content. The soil salinity map in Turkey (Fig. 11C) indicated that soils near Lake Tuz have higher salt content than other areas, which is the same as the result from Gorji et al. [21]. Kiliç and Kiliç [45] concluded that the groundwater table was the most relative indicator of soil salinization and alkalization in Turkey. Our result also shows that in semi-arid region (which is also the climate region in which most of Turkey is located), groundwater depth is the third influential covariate (0.08) in modeling. Lopes et al. [7] found that in the coastal area of Turkey, which was divided into dry sub-humid regions, the vegetation index derived from remote sensing was an important parameter in identifying salt marsh. The results in this study also proved that vegetation index contributes a lot to the estimation of soil salinity in dry sub-humid areas, as Fig. 8D shows. Mashimbye et al. [46] estimated soil salinity in South Africa based on soil samples, which concluded that high salinization occurred in central and southwestern South Africa. Vermeulen and Van Niekerk [22] further estimated soil salinization in central and southwestern South Africa and found the same distribution as the one presented in our result (Fig. 11E). Figure 11I shows that soil salinization is particularly severe in central Iran. Soils near Urmia Lake, in Western Iran, are at risk of moderate and high salinization [37]. Central Iran is a concave-shaped bare land with silty clay flat in arid climate, and under the combined influence of strong evapotranspiration and lower slopes, salt tends to accumulate in the central region [36,47]. Nabiollahi et al. [11] and Fathizad et al. [48] also reported a high salinization in the lower elevations of the region, mainly distributed in central Iran. As Fig. 11K indicates, the soils suffered from salinization in the southwest and southeast part. Lee and Hendricks [19] surveyed saline soils in southwestern California with the same distribution compared to our result.

Validity and implementation of multi-source data

The combinations of optical and radar remote sensing data help soil property cognition. Sensitive detection of moisture in soils

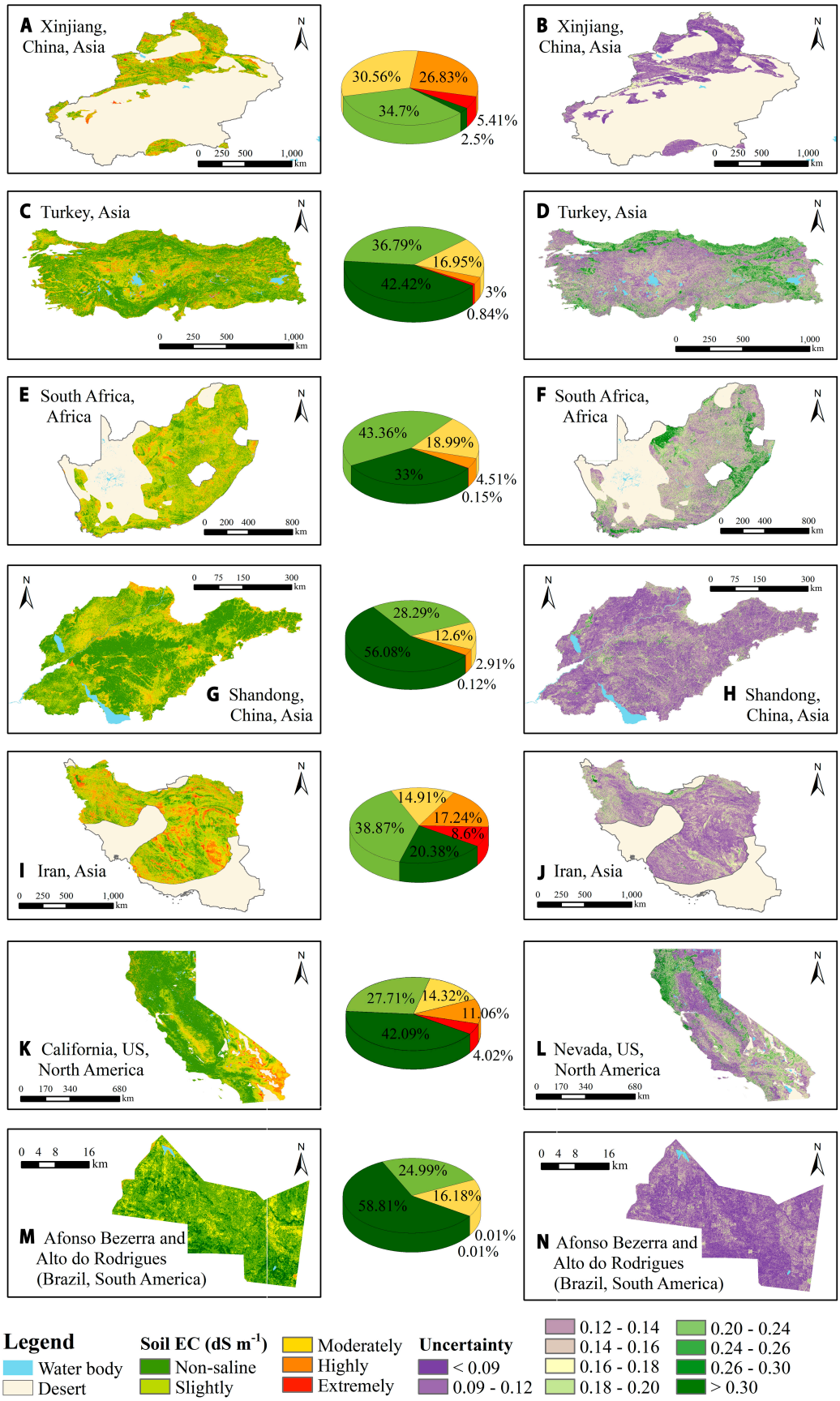


Fig. 11. The estimated soil salinity content and the proportion of each salinization levels in (A) Xinjiang, (C) Turkey, (E) South Africa, (G) Shandong, (I) Iran, (K) California, and (M) Afnoso Bezerra and Alto do Rodrigues with the uncertainty values calculated from 20 times RF regressions in (B) Xinjiang, (D) Turkey, (F) South Africa, (H) Shandong, (J) Iran, (L) California, and (N) Afnoso Bezerra and Alto do Rodrigues.

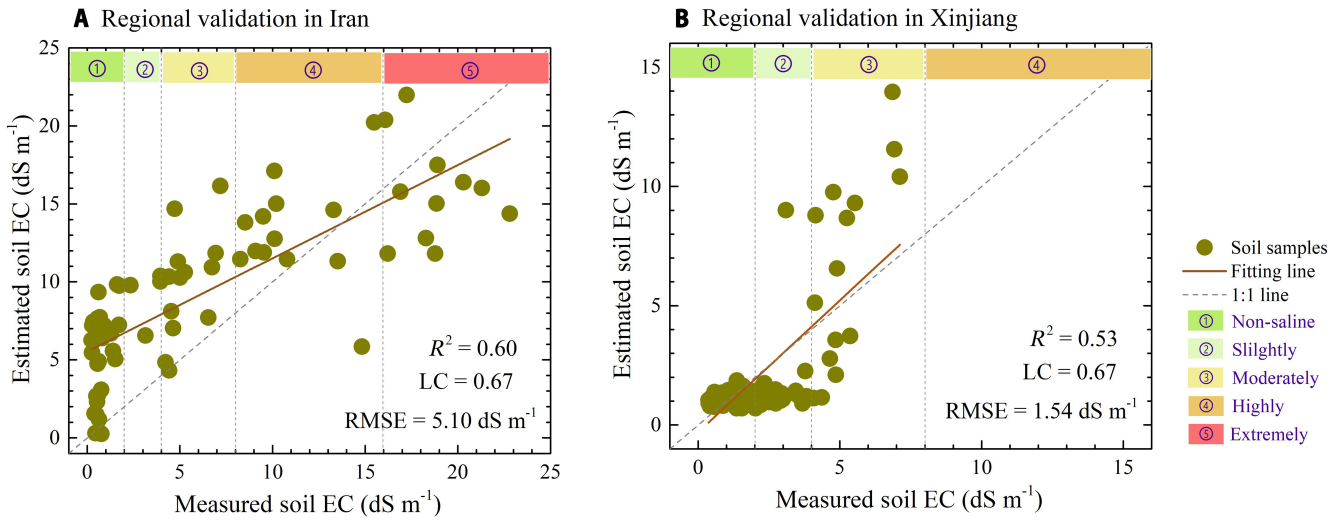


Fig. 12. Scatterplots of the RF model validated in the regional-scale dataset in (A) Iran and (B) Xinjiang.

by radar data provides highly contributing covariates to soil salinization estimates under arid and humid climate types [13]. The amplitude of backscatter derived from synthetic aperture radar (SAR) was decided by the dielectric properties of soil composition (for example, moisture and salinity), roughness, and the polarization sensitivity to sediments (e.g., texture information and salt crust) [49]. From theoretical modeling and experiments, Lasne et al. [50] concluded that the wetter soil indicated a stronger effect of salinity on imaginary part of the dielectric constant, while this effect decreases to the real part and the frequency. As Table 4 indicates, radar data show their effectiveness in soil salinity detection in humid regions. In humid climates (e.g., coastal saline soils), the soil moisture content is higher, which amplifies the effective detection of salinity through backscattering coefficient. Therefore, the fusion of multispectral and radar data helps to model the salinity accurately. In arid areas, radar data are less sensitive to salt crust and soil texture than to soil moisture detection, and only using optical remote sensing data can efficiently estimate soil salt content [50]. Field surveys, ground soil samples, and true salt content measurements obtained in Iran and Xinjiang were used to validate models at a field scale. The validation R^2 values (0.60 and 0.53, respectively) and *LC* values (0.67 and 0.67, respectively) indicate that trained models are credible at 10 m. According to the random validation approach based on published soil salinity maps, the overall accuracy values ranged from 0.36 to 0.61. The method was verified to be credible at the landscape scale in South Africa and Afonso Bezerra and Alto do Rodrigues in Brazil. In the remaining cases, the poor accuracy may be caused by the uncertainty introduced by digitizing maps and random selection. The spatially autocorrelated error is caused by unweighted mean of the prediction cross-validation errors [51]. The spatial autocorrelation of ground sampling points and random verification points reduces the verification accuracy. The models show low-value overestimation and high-value underestimation in Figs. 5 and 12. Soil salinization tends to be concentrated in arid or coastal areas, while soils in areas with sufficient rainfall and sustainable soil management are healthy. In the global salinization map, soil salinity may be overestimated in regions with abundant data points (such as Mexico), while it tends to be underestimated in regions with scarce data (such as Central Asia). As the global scale considered, estimated values for soil salinity can be considered to be balanced on a global scale [1].

Uncertainty and further improvement

The uncertainty of the estimation is a result of the uncertainty of soil data, covariates, sensors, and the models. First, the measured soil samples used in the training process are unevenly distributed in space, with a higher density in North America and an extreme scarcity in Asia and South America. Since soil salinization mostly occurs in arid areas, there are few non-saline soil samples in humid areas, resulting in an imbalance in the training datasets and introducing uncertainty to the estimation [43]. Although our study added samples that were classified as "non-saline soil" according to WRB, they were mainly distributed in arid regions and not sufficient for training at a global scale. Uncertainty values at low EC values of Fig. 6 and uncertainty maps (Fig. 11) in the five cases show higher values at lower soil salinity content. Soil salinity values may be overestimated in areas where soil samples are concentrated, such as North America, but in regions with sparse training data (e.g., Central Asia), the proposed models are able to calculate soil salinity content. This is considered to be a complementary and balancing effect of global estimations [1]. The unequal distribution of global training samples did influence the results, but not significantly. Second, inconsistent scales of covariates from remote sensing images increase uncertainty. Most ML models are robust to multicollinearity between covariates, and RF can automatically handle nonlinear relationships and interactions between predictor variables to obtain a high prediction accuracy [4,52]. But the various indices calculated from S1 and S2 data were at 10 m, while the other covariates from other geospatial data were at 1,000 m. This leads to incomplete training on the difference in soil salinity content, making the estimated results closer to each other [53]. Bilinear interpolation also produces jagged error values for pixels at the edges of remote sensing images. In addition, flooded and coastal agriculture caused soils in coastal areas to be submerged by salt water. Although they were masked in our research, uncertainty is introduced to the estimation of soil salt in its surrounding

Table 5. Accuracy assessment of the RF regression models validated in (A) California, (B) Afonso Bezerra and Alto do Rodrigues, (C) Turkey, (D) South Africa, (E) Shandong, and (F) Iran and Xinjiang

(A) California	Non-saline	Slightly saline	Moderately saline	Highly saline	Extremely saline	User's accuracy
Non-saline	17	13	16	4	0	0.34
Slightly saline	12	13	10	15	0	0.26
Moderately saline	1	2	24	23	0	0.48
Highly saline	0	1	14	34	1	0.67
Extremely saline	0	2	8	37	3	0.060
Producer's accuracy	0.57	0.42	0.33	0.30	0.75	
Overall accuracy				0.36		
Kappa				0.21		

(B) Afonso Bezerra and Alto do Rodrigues	1–5 dS m^{-1}	5–10 dS m^{-1}	>10 dS m^{-1}	User's accuracy
1–5 dS m^{-1}	44	6	0	0.88
5–10 dS m^{-1}	16	28	6	0.56
>10 dS m^{-1}	12	18	20	0.40
Producer's accuracy	0.61	0.54	0.77	
Overall accuracy			0.61	
Kappa			0.42	

(C) Turkey	Non-saline	Slightly saline	Moderately saline	Highly saline	Extremely saline	User's accuracy
Non-saline	15	21	13	1	0	0.30
Slightly saline	0	13	33	4	0	0.26
Moderately saline	0	4	20	26	0	0.40
Highly saline	0	0	5	44	1	0.88
Extremely saline	0	0	2	34	13	0.26
Producer's accuracy	1.00	0.34	0.27	0.40	0.93	
Overall accuracy				0.42		
Kappa				0.28		

(D) South Africa	Non-saline	Saline	User's accuracy
Non-saline	27	73	0.27
Saline	8	92	0.92
Producer's accuracy	0.77	0.56	
Overall accuracy		0.60	
Kappa		0.17	

(E) Shandong	Non-saline	Mildly Saline	Moderately Saline	Severely Saline	Solonchak	User's accuracy
Non-saline	23	18	7	2	0	0.46
Mildly saline	1	25	23	1	0	0.50
Moderately saline	0	12	21	15	2	0.42
Severely saline	0	2	7	35	6	0.70
Solonchak	0	0	8	35	7	0.14
Producer's accuracy	0.96	0.44	0.32	0.40	0.47	
Overall accuracy				0.44		
Kappa				0.31		

(F) Iran and Xinjiang	R^2 (dS m^{-1})	RMSE (dS m^{-1})	LC
Iran	0.60	5.10	0.67
Xinjiang	0.53	1.54	0.67

areas [54]. Global soil salinization estimated based on surface information from continuous observations of remote sensing images may intensify spatial heterogeneity and incorrectly estimate healthy soil pixels as salinized [24]. Our results aim to provide a global soil salinization map, and the uncertainty of multi-source remote sensing images can be balanced in the large-scale observation distribution [1]. At the field or landscape scale, in arid areas and coastal areas with typical severe salinization, the estimation accuracy based on remote sensing images will improve [34]. Exploring the contribution of covariates at finer scales and long time series will be helpful in future studies on the driving mechanisms of soil salinization. Third, the scatterplots in Figs. 5 and 12 show low-value overestimation and high-value underestimation, which commonly happened in ML methods [55]. RFs are limited by the distribution of training samples on the covariate shift problem and affect the performance in testing datasets. Each tree in the forest puts a heavy weight on them, thus leading to low-value overestimation and high-value underestimation [56]. Future research should focus on estimating soil salinity by applying a divide-and-conquer strategy in different types of soil salinization. Besides, soil salinization has strong spatial heterogeneity and temporal heterogeneity [57], and the time series prediction of the proposed framework for global estimation of soil salinity needs to prove its robustness with validation datasets in different periods. Future work should also emphasize on collecting field-level, geographically and feature-space well-distributed samples, and constructing time-series datasets at a global extent, especially adding non-saline soil samples.

Conclusions

Our study proposed a new framework for global soil salinity estimation and fine-resolution mapping by coupling multi-source data and bootstrapping-RF on the GEE. For modeling global soil salinity, both optical remote sensing data, radar data, and environmental covariates were used, with the feature selection based on different soil-environmental climate conditions. In hyper-arid and arid regions, modeling soil salinity using S2 and auxiliary data showed an R^2 of 0.85 and 0.62, respectively. In semi-arid, dry sub-humid, and humid regions, modeling using S1, S2, and auxiliary data showed the highest accuracy with an R^2 of 0.87, 0.80, and 0.87, respectively. Our models obtained considerable accuracy with an R^2 of 0.60 and 0.67 at independent validation in Iran and Xinjiang. The random validation results from published soil salinity maps in California, Afonso Bezerra and Alto do Rodrigues in Brazil, Turkey, South Africa, and Shandong showed the overall classification accuracy values of 0.36 to 0.61. The uncertainty values of the models decreased with the increasing salinization in the five climate regions. The global soil salinity map indicated that Central Asia, Middle East, South America, North Africa, and the Pacific areas are facing severe soil salinization. At the global scale, soils in dry regions (hyper-arid, arid, semi-arid regions) are easier to salinized than those in wet climates (dry sub-humid, humid regions). Climatic conditions (e.g., AI-ET and ET) played a primary role when estimating soil salinity in hyper-arid and arid regions, while in hyper-arid, arid, semi-arid, and dry sub-humid regions, WTD showed a stronger contribution. In conclusion, climatic conditions, groundwater, and salinity index play key roles in global soil salinity estimation, and in more humid regions, the use of radar data improves the model performance.

Acknowledgments

Funding: This study was supported by the National Key Research and Development Program (grant numbers 2018YFE0107000 and 2023YFD1900102), the National Science Foundation of China (grant numbers 42261016 and 41061031), the Bingtuan Science and Technology Program (grant number 2020CB032), the Tarim University President's Fund (grant number TDZKCX202205), the China Scholarship Council (CSC), the Academic Rising Star Program for Doctoral Students of Zhejiang University, and the Outstanding Ph.D. Dissertation Funding of Zhejiang University.
Author contributions: N.W. and Z.S. designed the research, carried out salinity estimation, and wrote the manuscript. X.Z. and Y.X. analyzed the data. S.C., F.F., and J.H. provided ideas to method development. R.T. and J.P. provided field soil data to model validation. J.X. and J.-P.W. evaluated the results. All authors reviewed and edited the manuscript.
Competing interests: The authors declare that they have no competing interests.

Data Availability

The code and the global soil salinity map can be accessed at https://code.earthengine.google.com/47b1cbccd96ddb90e5851bfcf79bc759. The Global AI-ET Climate Database can be accessed at https://cgiarcsi.community. WoSIS can be accessed at http://dx.doi.org/10.17027/isric-wdcsoils.20190901. The WorldClim 2.1 dataset can be accessed at http://worldclim.org. The SoilGrids v2 dataset can be accessed at https://soilgrids.org/. The WTD data can be accessed at http://worldclim.org. The SoilGrids v2 dataset can be accessed at https://glowasis.deltares.nl/thredds/catalog/opendap/opendap/Equilibrium_Water_Table/catalog.html. The WTD data can be accessed at http://worldclim.org. The Murray Global Intertidal Change Dataset can be accessed at https://code.earthengine.google.com/?scriptPath=Examples%3ADatasets%2FUQ%2FUQ_murray_Intertidal_v1_1_global_intertidal. The Global Flood Database can be accessed at https://code.earthengine.google.com/?scriptPath=Examples%3ADatasets%2FGLOBAL_FLOOD_DB%2FGLOBAL_FLOOD_DB_MODIS_EVENTS_V1.

References

1. Ivushkin K, Bartholomeus H, Bregt AK, Pulatov A, Kempen B, de Sousa L. Global mapping of soil salinity change. *Remote Sens Environ*. 2019;231:Article 111260.
2. Wang F, Yang S, Wei Y, Shi Q, Ding J. Characterizing soil salinity at multiple depth using electromagnetic induction and remote sensing data with random forests: A case study in Tarim River Basin of southern Xinjiang China. *Sci Total Environ*. 2021;754:Article 142030.
3. Metternicht GI, Zinck JA. Remote sensing of soil salinity: Potentials and constraints. *Remote Sens Environ*. 2003;85(1):1–20.
4. Wadoux AMJ-C, Minasny B, McBratney AB. Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth Sci Rev*. 2020;210:103359.
5. Berger T, Schreinemachers P. Creating agents and landscapes for multiagent systems from random samples. *Ecol Soc*. 2006;11(2):19.
6. Stolbovoy V, Montanarella L, Filippi N, Jones A, Gallego J, Grassi G. 2007. Soil sampling protocol to certify the changes oforganic carbon stock in mineral soil of the European Union.

Version 2. EUR21576 EN/2 (Office for Official Publications of the European Communities: Luxembourg).
7. Lopes CL, Mendes R, Caçador I, Dias JM. Assessing salt marsh extent and condition changes with 35 years of Landsat imagery: Tagus estuary case study. *Remote Sens Environ*. 2020;247, 111939.
8. Scudiero E, Skaggs TH, Corwin DL. Regional-scale soil salinity assessment using Landsat ETM + canopy reflectance. *Remote Sens Environ*. 2015;169:335–343.
9. Peng J, Biswas A, Jiang Q, Zhao R, Hu J, Hu B, Shi Z. Estimating soil salinity from remote sensing and terrain data in southern Xinjiang Province, China. *Geoderma*. 2019;337:1309–1319.
10. Ma L, Ma F, Li J, Gu Q, Yang S, Wu D, Feng J, Ding J. Characterizing and modeling regional-scale variations in soil salinity in the arid oasis of Tarim Basin, China. *Geoderma*. 2017;305:1–11.
11. Nabiollahi K, Taghizadeh-Mehrjardi R, Shahabi A, Heung B, Amirian-Chakan A, Davari M, Scholten T. Assessing agricultural salt-affected land using digital soil mapping and hybridized random forests. *Geoderma*. 2021;385:114858.
12. Wang N, Peng J, Xue J, Zhang X, Huang J, Biswas A, He Y, Shi Z. A framework for determining the total salt content of soil profiles using time-series Sentinel-2 images and a random forest-temporal convolution network. *Geoderma*. 2022;409:Article 115656.
13. Hoa PV, Giang NV, Binh NA, Hai LVH, Pham T-D, Hasanlou M, Tien Bui D. Soil salinity mapping using SAR Sentinel-1 data and advanced machine learning algorithms: A case study at Ben Tre Province of the Mekong River Delta (Vietnam). *Remote Sens (Basel)*. 2019;11(2):128–149.
14. Cherlet M, Hutchinson C, Reynolds J, Hill J, Sommer S, Von Maltitz G. *World atlas of desertification*. Luxembourg: Publication Office of the European Union; 2018.
15. Osman KT, Osman KT. Saline and sodic soils. In: Osman KT, editor. *Management of soil problems*. Cham: Springer International Publishing; 2018. p. 255–298.
16. Batjes U, Ribeiro E, Oostrum AV. Standardised soil profile data to support global mapping and modelling (WoSIS snapshot 2019). *Earth Syst Sci Data*. 2020;12:299–320.
17. Abrol I, Yadav J, Massoud F. *Salt-affected soils and their management. FAO Soils Bulletin 39*. Rome: Food and Agriculture Organization of the United Nations; 1988.
18. Stockmann U, Huang J, Minasny B, Triantafilis J. Utilizing a DUALEM-421 and inversion modelling to map baseline soil salinity along toposequences in the Hunter Valley Wine district. *Soil Use Manage*. 2017;33(3):413–424.
19. Lee J, Hendricks NP. Crop choice decisions in response to soil salinity on irrigated lands in California. Paper presented at: The 2022 Agricultural & Applied Economics Association Annual Meeting; 2002 July 31–August 2; Anheim, CA.
20. Barreto AC, Neto MF, de Oliveira RP, Moreira LCJ, de Medeiros JF, da Silva Sa FV. Comparative analysis of spectral indexes for soil salinity mapping in irrigated areas in a semi-arid region, Brazil. *J Arid Environ*. 2023;209:Article 104888.
21. Gorji T, Sertel E, Tanik A. Monitoring soil salinity via remote sensing technology under data scarce conditions: A case study from Turkey. *Ecol Indic*. 2017;74:384–391.
22. Vermeulen D, Van Niekerk A. Machine learning performance for predicting soil salinity using different combinations of geomorphometric covariates. *Geoderma*. 2017;299:1–12.
23. Chen H, Ma Y, Zhu A, Wang Z, Zhao G, Wei Y. Soil salinity inversion based on differentiated fusion of satellite image and ground spectra. *Int J Appl Earth Obs Geoinf*. 2021;101:102360.
24. Liu C, Zhang Q, Tao S, Qi J, Ding M, Guan Q, Wu B, Zhang M, Nabil M, Tian F, et al. A new framework to map fine resolution cropping intensity across the globe: Algorithm, validation, and implication. *Remote Sens Environ*. 2020;251:112095.
25. de Sousa L, Poggio L, Batjes NH, Heuvelink GBM, Kempen B, Ribeiro E, Rossiter D. SoilGrids 2.0: Producing quality-assessed soil information for the globe. 2020.
26. Fan Y, Li H, Miguez-Macho G. Global patterns of groundwater table depth. *Science*. 2013;339(6122):940–943.
27. Peel MC, Finlayson BL, McMahon TA. Updated world map of the Koppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.* 2007;11:1633–1644.
28. Murray NJ, Phinn SR, DeWitt M, Ferrari R, Johnston R, Lyons MB, Clinton N, Thau D, Fuller RA. The global distribution and trajectory of tidal flats. *Nature*. 2019;565:222–225.
29. Tellman B, Sullivan JA, Kuhn C, Kettner AJ, Doyle CS, Brakenridge GR, Erickson TA, Slayback DA. Satellite imaging reveals increased proportion of population exposed to floods. *Nature*. 2021;596(7870):80–86.
30. Song X-D, Wu H-Y, Ju B, Liu F, Yang F, Li D-C, Zhao Y-G, Yang J-L, Zhang G-L. Pedoclimatic zone-based three-dimensional soil organic carbon mapping in China. *Geoderma*. 2020;363:114145.
31. Breiman L. Random forest. *Mach Learn*. 2001:45.
32. Congalton RG. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens Environ*. 1991;37(1):35–46.
33. Yang R-M, Guo W-W. Using Sentinel-1 imagery for soil salinity prediction under the condition of coastal restoration. *IEEE J Sel Top Appl Earth Obs Remote Sens*. 2019;12(5):1482–1488.
34. Corwin DL. Climate change impacts on soil salinity in agricultural areas. *Eur J Soil Sci*. 2020;72(2):842–862.
35. Farzamian M, Paz MC, Paz AM, Castanheira NL, Gonçalves MC, Monteiro Santos FA, Triantafilis J. Mapping soil salinity using electromagnetic conductivity imaging—A comparison of regional and location-specific calibrations. *Land Degrad Dev*. 2019;30(12):1393–1406.
36. Roozitalab MH, Toomanian N, Dehkordi VRG, Khormali F. Major soils, properties, and classification. In: Roozitalab MH, Siadat H, Farshad A, editors. *The soils of Iran*. Cham: Springer; 2018. p. 93–147.
37. Taghizadeh-Mehrjardi R, Hamzehpour N, Hassanzadeh M, Heung B, Goydaragh MG, Schmidt K, Scholten T. Enhancing the accuracy of machine learning models using the super learner technique in digital soil mapping. *Geoderma*. 2021;399:Article 115108.
38. Schofield RV, Kirkby MJ. Application of salinization indicators and initial development of potential global soil salinization scenario under climatic change. *Global Biogeochem Cycles*. 2003;17(3).
39. Nosetto MD, Acosta AM, Jayawickreme DH, Ballesteros SI, Jackson RB, Jobbágy EG. Land-use and topography shape soil and groundwater salinity in Central Argentina. *Agric Water Manag*. 2013;129:120–129.
40. Jobbágy EG, Jackson RB. Groundwater and soil chemical changes under phreatophytic tree plantations. *J Geophys Res*. 2007;112(G2).
41. Zhao S, Ding J, Ge X, Huang S, Han L. Soil salinity estimation: Effects of microwave dielectric spectroscopy and important frequencies. *Land Degrad Dev*. 2023;34(6):1725–1739.
42. Jang E, Kim YJ, Im J, Park Y-G, Sung T. Global Sea surface salinity via the synergistic use of SMAP satellite and HYCOM

data based on machine learning. *Remote Sens Environ*. 2022;273:112980.
43. Hassani A, Azapagic A, Shokri N. Predicting long-term dynamics of soil salinity and sodicity on a global scale. *Proc Natl Acad Sci USA*. 2020;117(52):33017–33027.
44. Wang J, Ding J, Yu D, Teng D, He B, Chen X, Ge X, Zhang Z, Wang Y, Yang X, et al. Machine learning-based detection of soil salinity in an arid desert region, Northwest China: A comparison between Landsat-8 OLI and Sentinel-2 MSI. *Sci Total Environ*. 2020;707:Article 136092.
45. Kiliç K, Kiliç S. Spatial variability of salinity and alkalinity of a field having salination risk in semi-arid climate in northern Turkey. *Environ Monit Assess*. 2007;127(1–3):55–65.
46. Mashimbye ZE, Cho MA, Nell JP, De Clercq WP, Van Niekerk A, Turner DP. Model-based integrated methods for quantitative estimation of soil salinity from hyperspectral remote sensing data: A case study of selected South African soils. *Pedosphere*. 2012;22(5):640–649.
47. Taghizadeh-Mehrjardi R, Minasny B, Sarmadian F, Malone BP. Digital mapping of soil salinity in Ardakan region, central Iran. *Geoderma*. 2014;213:15–28.
48. Fathizad H, Ali Hakimzadeh Ardakani M, Sodaiezadeh H, Kerry R, Taghizadeh-Mehrjardi R. Investigation of the spatial and temporal variation of soil salinity using random forests in the central desert of Iran. *Geoderma*. 2020;365:Article 114233.
49. Singhroy V, Molch K. Geological case studies related to RADARSAT-2. *Can J Remote Sens*. 2014;30(6):893–902.
50. Lasne Y, Paillou P, Freeman A, Farr T, McDonald KC, Ruffie G, Malezieux J-M, Chapman B, Demontoux F. Effect of salinity on the dielectric properties of geological materials: Implication for soil moisture detection by means of radar remote sensing. *IEEE Trans Geosci Remote Sens*. 2008;46(6):1674–1688.
51. Brus DJ, Kempen B, Heuvelink GBM. Sampling for validation of digital soil maps. *Eur J Soil Sci*. 2011;62(3):394–407.
52. Golestani M, Mosleh Ghahfarokhi Z, Esfandiarpour-Boroujeni I, Shirani H. Evaluating the spatiotemporal variations of soil salinity in Sirjan playa, Iran using sentinel-2A and Landsat-8 OLI imagery. *Catena*. 2023;231:107375.
53. Fernández-Buces N, Siebe C, Cram S, Palacio JL. Mapping soil salinity using a combined spectral response index for bare soil and vegetation: A case study in the former Lake Texcoco, Mexico. *J Arid Environ*. 2006;65(4):644–667.
54. Nguyen K-A, Liou Y-A, Tran H-P, Hoang P-P, Nguyen T-H. Soil salinity assessment by using near-infrared channel and vegetation soil salinity index derived from Landsat 8 OLI data: A case study in the Tra Vinh Province, Mekong Delta, Vietnam. *Prog Earth Planet Sci*. 2022;7:1.
55. Ding J, Yu D. Monitoring and evaluating spatial variability of soil salinity in dry and wet seasons in the Werigan–Kuqa oasis, China, using remote sensing and electromagnetic induction instruments. *Geoderma*. 2014;235-236:316–322.
56. Zhang Q, Li L, Sun R, Zhu D, Zhang C, Chen Q. Retrieval of the soil salinity from Sentinel-1 dual-polarized SAR data based on deep neural network regression. *IEEE Geosci Remote Sens Lett*. 2022;19:4006905.
57. Wang N, Peng J, Chen S, Huang J, Li H, Biswas A, He Y, Shi Z. Improving remote sensing of salinity on topsoil with crop residues using novel indices of optical and microwave bands. *Geoderma*. 2022;422:Article 115935.
58. Touzi R, Goze S, Le Toan T, Lopes A, Mougin E. Polarimetric discriminators for SAR images. *IEEE Trans Geosci Remote Sens*. 1992;30(5):973–980.
59. Nurmemet I, Sagan V, Ding J-L, Halik U, Abliz A, Yakup Z. A WFS-SVM model for soil salinity mapping in Keriya Oasis, northwestern China using polarimetric decomposition and fully PolSAR data. *Remote Sens (Basel)*. 2018;10(4):598–618.
60. Tripathi NK, Rai BK, Dwivedi P. Spatial modelling of soil alkalinity in GIS environment using IRS data. Paper presented at: Proceeding of the 18th Asian Conference in Remote Sensing; 1997 Oct 20–25; Kuala Lumpur, Malaysia.
61. Khan NM, Rastoskuev VV, Sato Y, Shiozawa S. Assessment of hydrosaline land degradation by using a simple approach of remote sensing indicators. *Agric Water Manag*. 2005;77(1–3):96–109.
62. Douaoui AEK, Nicolas H, Walter C. Detecting salinity hazards within a semiarid context by means of combining soil and remote-sensing data. *Geoderma*. 2006;134(1–2):217–230.
63. Bannari A, Guedon AM, El-Harti A, Cherkaoui FZ, El-Ghmari A. Characterization of slightly and moderately saline and sodic soils in irrigated agricultural land using simulated data of advanced land imaging (EO-1) sensor. *Commun Soil Sci Plant Anal*. 2008;39(19–20):2795–2811.
64. Major DJ, Baret F, Guyot G. A ratio vegetation index adjusted for soil brightness. *Int J Remote Sens*. 2007;11(5): 727–740.
65. Scudiero E, Skaggs TH, Corwin DL. Regional scale soil salinity evaluation using Landsat 7, western San Joaquin Valley, California, USA. *Geoderma Reg*. 2014;2–3:82–90.
66. Alhammadi MS, Glenn EP. Detecting date palm trees health and vegetation greenness change on the eastern coast of the United Arab Emirates using SAVI. *Int J Remote Sens*. 2008;29(6):1745–1765.
67. Foody GM, Cutler ME, McMorrow J, Pelz D, Tangki H, Boyd DS, Douglas I. Mapping the biomass of bornean tropical rain forest from remotely sensed data. *Glob Ecol Biogeogr*. 2001;10(4):379–387.
68. Crippen R. Calculating the vegetation index faster. *Remote Sens Environ*. 1990;34(1):71–73.
69. Wu W, Al-Shafie WM, Mhaimeed AS, Ziadat F, Nangia V, Payne WB. Soil salinity mapping by multiscale remote sensing in Mesopotamia, Iraq. *IEEE J Sel Top Appl Earth Obs Remote Sens*. 2014;7(11):4442–4452.
70. Gitelson AA, Kaufman YJ, Merzlyak MN. Use of a green channel in remote sensing of global vegetation from eos-modis. *Remote Sens Environ*. 1996;58(3):289–298.
71. Huete A, Didan K, Miura T, Rodriguez EP, Gao X, Ferreira LG. Overview of the radiometric and biophysical performance of the modis vegetation indices. *Remote Sens Environ*. 2002;83(1–2): 195–213.
72. Rondeaux G, Steven M, Baret F. Optimization of soil-adjusted vegetation indices. *Remote Sens Environ*. 1996;50(2):95–107.
73. Ceccato P, Gobron N, Flasse S, Pinty B, Tarantola S. Designing a spectral index to estimate vegetation water content from remote sensing data: Part 1: Theoretical approach. *Remote Sens Environ*. 2002;82(2–3):188–197.