

SOIL VISION BY WEENAT: ESTIMATING SOIL MOISTURE FOR EUROPEAN AGRICULTURE USING MACHINE LEARNING METHODS

SOIL VISION PAR WEENAT : ESTIMATION DE L'HUMIDITÉ DU SOL POUR L'AGRICULTURE EUROPÉENNE À L'AIDE DE MÉTHODES D'APPRENTISSAGE AUTOMATIQUE

E. Buisson, M. Sauveur, ¹P. Giquel, S. Coly, M. El Fakraoui, R. Vallée

ABSTRACT

In the context of growing water scarcity, optimising irrigation management is a major challenge for modern agriculture. Agriculture accounts for at least 70% of global freshwater withdrawals, and more than 50% of this water could be used more efficiently. One of the main operational obstacles lies in the difficulty of obtaining soil moisture data that are both accurate at the field scale and comprehensive enough to cover large territories for irrigation optimisation.

Weenat, a French company specialised in irrigation decision-support solutions since 2014, developed Soil Vision as part of the Meteorita and e-RASM research projects. This artificial intelligence system continuously estimates field-scale soil water content without requiring in-situ sensors on every field. The solution relies on a multi-source data fusion strategy integrating: a network of more than 20,000 in-situ soil sensors, a proprietary database of 2,000 physical soil samples, satellite data (NDVI, NDMI, EVI), and meteorological data spatialised at the square-kilometre resolution.

The core innovation consists of two machine learning models: the Soil Model, which characterises soil hydraulic properties from pedological data and laboratory measurements (XGBoost, MAE = 0.023 mm/mm volumetric water content, $R^2 = 92\%$), and the Soil Vision model, which estimates real-time soil water content from multi-source data. The system was trained and tested on several thousand fields and validated on 50 pilot farms across Europe (more than 200 fields). More than 150 operational proof-of-concept trials were conducted during the 2025 season. Results show a mean absolute error of approximately 1.5% in volumetric water content and a relative error of 5.5%, covering 95% of French soil types and climatic conditions. These performances enable agricultural professionals to make proactive decisions on irrigation, tillage, fertilisation and harvest, contributing to a more resilient, efficient and sustainable agriculture.

Keywords: Water scarcity; water content; soil moisture data; irrigation; tillage.

RESUME

Dans le contexte de la croissance de la pénurie des ressources en eau, l'optimisation de la gestion de l'irrigation pose un défi majeur en ce qui concerne l'agriculture moderne. L'agriculture représente au moins 70% des prélèvements mondiaux d'eau douce, et plus de 50% de cette ressource pourrait être utilisée plus efficacement. L'un des principaux obstacles opérationnels réside dans la difficulté d'obtenir des données sur l'humidité du sol à la fois précises à l'échelle du champ et suffisamment complètes pour couvrir de vastes territoires en vue de l'optimisation de l'irrigation.

Weenat, société française spécialisée dans les solutions d'aide à la décision en irrigation depuis 2014, a développé Soil Vision dans le cadre des projets de recherche Meteorita et e-RASM. Ce système d'intelligence artificielle évalue de façon continue la teneur en eau du sol à l'échelle du champ sans exiger des capteurs in situ sur chaque champ. La solution repose sur une stratégie de fusion de données

¹ WEENAT, Product & Innovation Department, Nantes, France — Email: emmanuel.buisson@weenat.com

multi-sources intégrant : un réseau de plus de 20 000 capteurs de sol in situ, une base de données propriétaire de 2 000 échantillons physiques de sols, des données satellites (NDVI, NDMI, EVI), et des données météorologiques spatialisées à la résolution du kilomètre carré.

L'innovation de base comprend deux modèles d'apprentissage automatique : le Soi Model, qui caractérise les propriétés hydrauliques du sol à partir de données pédologiques et de mesures de laboratoire (XGBoost, MAE = 0,023 mm/mm de teneur en eau volumétrique, $R^2 = 92\%$), et le modèle Soil Vision, qui estime la teneur en eau du sol en temps réel à partir de données provenant de sources multiples. Le système a été formé et testé sur plusieurs milliers de champs et validé sur 50 fermes pilotes à travers l'Europe (plus de 200 champs). Plus de 150 essais opérationnels de validation de concept ont été menés au cours de la saison 2025. Les résultats montrent une erreur absolue moyenne d'environ 1,5% sur la teneur en eau volumétrique et une erreur relative de 5,5%, couvrant 95% des types de sol français et des conditions climatiques. Ces performances permettent aux professionnels de l'agriculture de prendre des décisions proactives en matière d'irrigation, de travail du sol, de fertilisation et de récolte, contribuant ainsi à une agriculture plus résistante, efficace et durable.

Mots-clés : Pénurie d'eau; Teneur en eau; Données sur l'humidité du sol; Irrigation; Travail du sol.

INTRODUCTION

Water management in agriculture represents one of the most critical challenges of the 21st century. With at least 70% of global freshwater withdrawals destined for irrigation (FAO, 1997; Foley et al., 2011), optimising water use is both an environmental and an economic imperative. Current analyses indicate that more than 50% of this water could be employed more efficiently (Jägermeyr et al., 2015). Pressure on water resources is expected to intensify under the effects of climate change and increasing global food demand (Wada et al., 2011).

Soil moisture is a central variable in irrigation and fertilisation management. Yet its precise measurement at the field scale remains difficult to deploy at large scale. Existing methods suffer from a persistent precision-scale dilemma: in-situ sensors (tensiometers, capacitive probes) offer high accuracy but require significant investment and regular maintenance; satellite approaches provide broad spatial coverage but with limited resolution and accuracy, particularly in the presence of dense vegetation cover (Massari et al., 2021; Foster et al., 2020; Ma et al., 2022). Pedotransfer functions (Bastet, 1999; Saxton and Rawls, 2006; Abesh and Hubbart, 2023) allow estimation of soil hydraulic properties but do not provide temporal monitoring.

In this context, Weenat initiated in 2022 the Meteorita and e-RASM research projects with the objective of developing a solution capable of providing continuous, spatialised estimates of volumetric soil water content at the field scale, without requiring sensors on every field. This project led to the development of Soil Vision, an artificial intelligence system based on multi-source data fusion. This paper presents the technical architecture of Soil Vision, its two main components (Soil Model and the Soil Vision model itself), their respective performances and agricultural applications.

MATERIALS AND METHODS

Data Sources

Soil Vision relies on four categories of input data: (1) Soil data: pedological properties are extracted from the global Soil Grids database (250 m resolution, stratified by depth), providing textural and physico-chemical information. These public data are compared against a proprietary database of 2,000 physical soil samples collected and analysed in an in-house laboratory (WeeLab) on French agricultural fields in 2024 and 2025; (2) Meteorological data: spatialised indicators at the square-kilometer resolution including statistics (cumulative totals, means, extremes, etc.) over time windows of 1 to 28 days for daily parameters (precipitation, potential evapotranspiration, temperature, wind, relative humidity, radiation) as well as hourly indicators over 120 hours; (3) Satellite data: vegetation and moisture indices from Sentinel-2 and Landsat satellites (NDMI, NDVI, EVI) providing information on vegetation cover status

and surface conditions; and (4) Field data: a network of more than 20,000 in-situ soil sensors (tensiometers and capacitive probes) deployed by Weenat across Europe constitutes the training and validation soil moisture dataset.

SoilModel — Soil Hydrological Characterisation Model

The SoilModel is a machine learning model designed to estimate the hydraulic characteristics of a soil (available water capacity (AWC), volumetric water content at field capacity (fc), at permanent wilting point (pwp), at saturation and at residual state) from its pedological properties (texture, bulk density, coarse fragments, etc.). The training dataset comprises 2,000 physical soil samples covering the main French agricultural soil types. Each sample underwent tensio-gravimetric analyses to establish the water retention curve according to the Van Genuchten model.

Several algorithms were compared (Lasso, Ridge, kNN, Random Forest, XGBoost) in Python. An XGBoost model predicting all hydraulic thresholds of interest (fc, pwp, saturation, residual content and critical point [pF3]) was selected based on the quality of fit between predictions and measurements, consistency between hydraulic thresholds and operational practicality (deployment, maintenance, updates). A model including only textural variables (along with the threshold of interest and depth) was retained as a compromise between goodness of fit and parsimony. Best model selection was based on 5-fold cross-validation with a site-separation constraint. The final training of the best model followed a Leave-One-Out approach isolating each sampling site.

Table 1. Global SoilModel performance vs. individual models per hydraulic threshold (MAE in volumetric water content)

Parameter	Global SoilModel	Individual SoilModels
Global (theta)	0.026	0.029
Saturation (theta_s)	0.028	0.036
Field capacity (fc)	0.025	0.024
Permanent wilting point (pwp)	0.017	0.033
Residual (theta_r)	0.016	0.015
Available Water Capacity (AWC)	0.13 mm/cm	0.18 mm/cm

Table 2. SoilModel performance (MAE/RMSE in volumetric water content)

Parameter	MAE (valid.)	RMSE (valid.)	R ² (valid.)
All SoilGrids parameters	0.026	0.032	89%
Textural parameters only	0.023	0.029	92%

This model provides crucial information on a field's capacity to store water and also enables reconstruction of the entire water retention curve, thereby converting tensiometry measurements taken in the field into volumetric water content.

Soil Vision Model — Soil Water Content Estimation

The Soil Vision model aims to estimate real-time soil water content at the field scale, without in-situ sensors. The target variable is in fact the available water capacity filling ratio (ratio_awc), which is then converted back into volumetric water content via the Soil Model parameters.

The training database includes measurements from more than 2,000 Weenat tensiometers over the period 2021–2025, representing nearly 200,000 daily observations distributed across France, Spain,

Belgium and Germany. For each of these measurements, satellite, meteorological and soil data are extracted and input indicators calculated (see "Data Sources"). Among the 200 indicators considered, approximately fifty explanatory variables were selected after correlation clustering (threshold 0.9). Hourly data were excluded for operational feasibility reasons.

The test set selection uses a geographic strategy: sensors located within 14 km of 2025 pilot farms are reserved for testing (approximately 20% of the data), ensuring evaluation on spatially independent data. Among the algorithms compared (Lasso, Ridge, MLP, Random Forest, LightGBM, GradientBoosting, XGBoost), the XGBoost model optimised by RandomizedSearchCV was selected.

RESULTS

Soil Model Performance

The Soil Model, evaluated using a Leave-One-Out approach, achieves a global MAE of 0.026 in volumetric water content with an R^2 of 92%. Errors per hydraulic threshold are detailed in Table 1. Validation across the 14 textural classes of the French texture triangle (Jamagne classification) shows results consistent with bibliographic reference values. Clay content plays a key role in model results, with all textural parameters having a significant impact. Residual biases are identified for heavy clays (underestimated AWC) and soils with very high clay content (> 60%) that will be the subject of a supplementary sampling campaign in 2026.



Figure 1. Wee Lab soil analysis laboratory

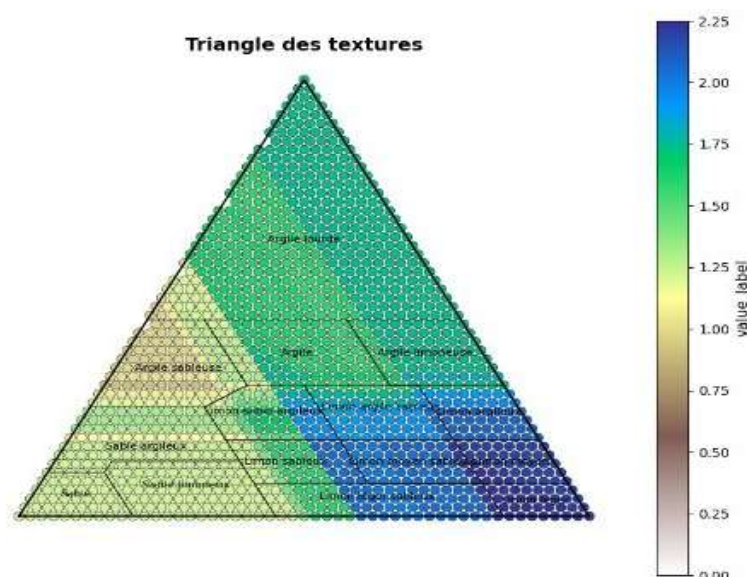


Figure 2. Estimated mean available water capacity values on the texture triangle

Soil Vision Model Performance

Operational tests demonstrated good fit of the water content dynamics (ability to reproduce drying and rewetting cycles) and good spatial variability. The overall mean absolute error on the validation network is approximately 1.5% in volumetric moisture, with a relative error of approximately 5.5%. These results cover 95% of soil types and climatic conditions characteristic of French agricultural zones.

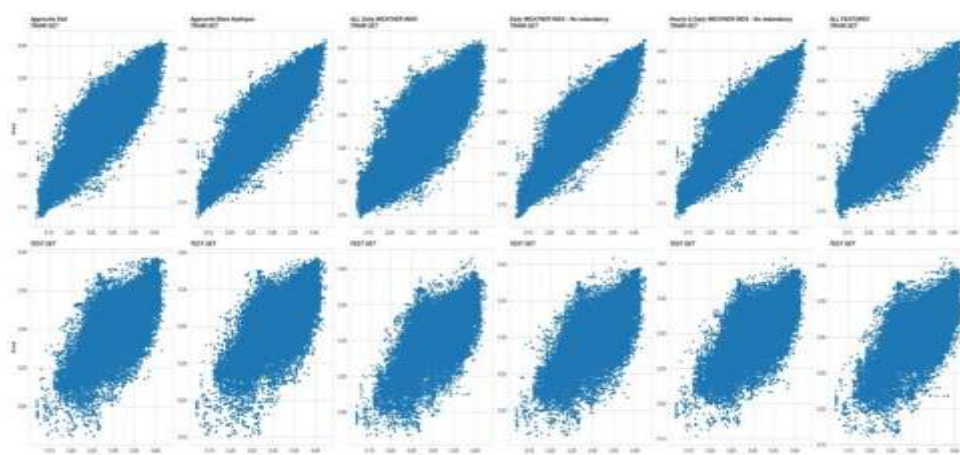


Figure 3. Comparison between measured water content and Soil Model results for the 6 tested variable sets (Row 1: training set, Row 2: validation set / One column per variable set)

Analysis of the most determining variables (SHAP method on the main features retained by XGBoost) identifies as priorities: cumulative water balance indicators (difference between water inputs and potential evapotranspiration), an aridity index over the past month, the number of days without significant precipitation, and satellite indices. Soil properties (particularly sand and clay contents) occupy intermediate ranks but constitute unavoidable structural factors. Comparison on concrete examples (carrot, sugar beet and cereal crops) confirms the model's ability to track water content variations in response to irrigation and precipitation events, with satisfactory calibration against the magnitudes measured by tensiometers.

DISCUSSION AND APPLICATIONS

The combination of the Soil Model and Soil Vision models forms the core of the irrigation management solution, providing agricultural professionals with real-time field-scale soil moisture estimates without dedicated equipment on each field. This approach overcomes the precision-scale dilemma that characterises existing solutions, relying exclusively on standardised data sources (satellite data, meteorological data spatialised at km², global Soil Grids database, etc.), which guarantees the reproducibility and geographic scalability of the system.

The operational applications of Soil Vision in agriculture are multiple. In terms of irrigation management, real-time knowledge of soil water content enables adjustment of irrigation dose and frequency to keep the crop within a comfortable water range, avoiding excesses (leaching, root asphyxia) and severe deficits (water stress, yield losses). Such optimisation can contribute to reducing water inputs by 10 to 30% depending on the cropping context. In terms of agronomic decision support, soil moisture data enable optimisation of the calendar of field operations (sowing, tillage, fertilisation, harvest) according to actual conditions. Monitoring of water stress levels throughout the growing season also enables refinement of yield forecasts. Finally, precise soil moisture data constitute an objective criterion for evaluating the performance of varieties, biostimulants or cultural practices within agronomic trials.

Validation on more than 200 fields across 50 pilot farms in Europe, complemented by more than 150 operational proof-of-concept trials in 2025, confirms the model's robustness for different pedo-climatic contexts. Limitations have however been identified: very clayey soils (> 60% clay) and very sandy soils are under-represented in the current training database, resulting in residual biases for these textural classes. The estimation depth is currently limited to the 0–30 cm horizon, which may be insufficient for some deep-rooted crops.

CONCLUSIONS AND PROSPECTS

Soil Vision represents a significant advance in the democratisation of soil moisture information for European agriculture. By combining multi-source data via advanced machine learning algorithms, the system overcomes the traditional dilemma between accuracy and spatial coverage. The Soil Model enables characterisation of the hydraulic properties of any soil type, while the Soil Vision model provides an estimate of soil water content with a mean absolute error of 1.5% in volumetric content.

Short-term perspectives (Summer 2026) include extension of the model to the 30–60 cm horizon, improved consideration of the vegetative cycle through new features based on satellite data, and enrichment of the soil database through a 2026 sampling campaign targeting under-represented soil types. In the longer term, integration of radar satellite data (Sentinel-1), measurements from capacitive probes and geographic extension to other European countries are planned to consolidate the position of Soil Vision as a reference tool for sustainable water management in agriculture.

ACKNOWLEDGEMENTS

This work was carried out as part of the Meteoira & e-RASM research projects, supported by I-Nov 10 call (sponsored by ADEME & BPI France) and ESA/CNES. The authors wish to thank all the pilot farms and partner organisations who participated in the operational validation of the Soil Vision system across Europe.

REFERENCES

- Abesh, B.F.; Hubbart, J.A. A Comparison of Saturated Hydraulic Conductivity (Ksat) Estimations from Pedotransfer Functions (PTFs) and Field Observations in Riparian Seasonal Wetlands. *Water* 2023, 15, 2711.
- Bastet, G. Estimation des propriétés de rétention en eau des sols à l'aide de fonctions de pédotransfert: développement de nouvelles approches. Université d'Orléans, 1999.
- FAO. Improving water-use efficiency. FAO Irrigation and Drainage Paper, 1997.

- Foley, J.A., et al. Solutions for a cultivated planet. *Nature* 2011, 478, 337–342.
- Foster, T.; Mieno, T.; Brozovic, N. Satellite-based monitoring of irrigation water use: assessing measurement errors and their implications for agricultural water management policy. *Water Resour. Res.* 2020, 56, e2020WR028378.
- Jägermeyr, J., et al. Water savings potentials of irrigation systems: global simulation of processes and linkages. *Hydrol. Earth Syst. Sci.* 2015, 19, 3073–3091.
- Ma, C.; Johansen, K.; McCabe, M.F. Monitoring Irrigation Events and Crop Dynamics Using Sentinel-1 and Sentinel-2 Time Series. *Remote Sens.* 2022, 14, 1205.
- Massari, C., et al. A review of irrigation information retrievals from space and their utility for users. *Remote Sens.* 2021, 13, 4112.
- Saxton, K.E.; Rawls, W.J. Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic Solutions. *Soil Sci. Soc. Am. J.* 2006, 70, 1569–1578.
- Wada, Y.; van Beek, L.P.H.; Bierkens, M.F.P. Modelling global water stress of the recent past: on the relative importance of trends in water demand and climate variability. *Hydrol. Earth Syst. Sci.* 2011, 15, 3785–3808.