Daily root-zone soil moisture from multimodal foundation-model embeddings and meteorology: estimation and interpretation

Prajwal Khanal[1], Yijian Zeng[1], Joost Beckers[2], Zhongbo (Bob) Su[1]

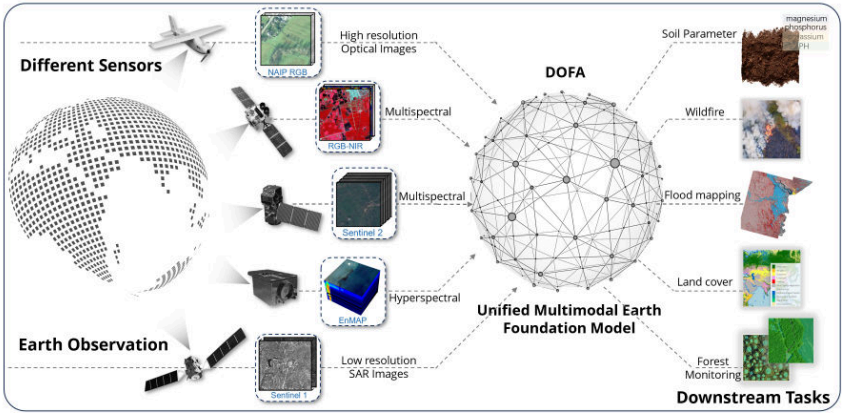
[1]Faculty ITC, University of Twente, Enschede, The Netherlands [2]Planet Labs, Haarlem, The Netherlands

2026 Workshop on Machine Learning for Land and Hydrology

ECMWF, Reading

3 November 2026

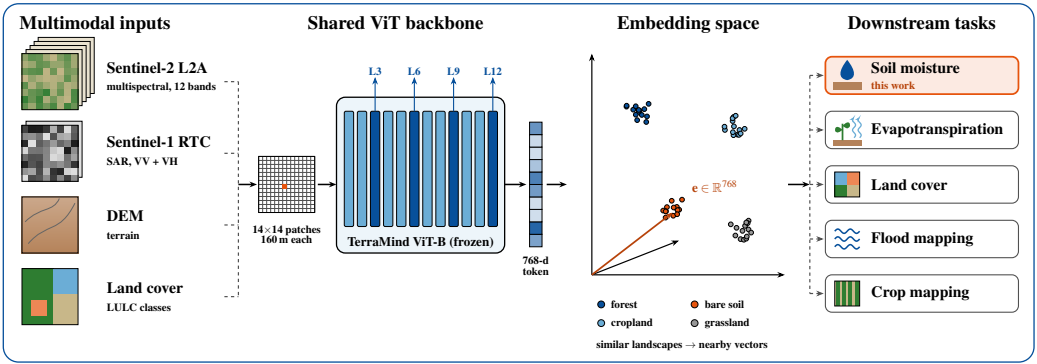
Motivation Geospatial foundation models (GFMs)



DOFA, a multimodal geospatial foundation model transferred to many downstream tasks (Xiong et al., 2024).

Foundation models pre-trained on unlabelled multimodal imagery provide embeddings that transfer to many downstream tasks.

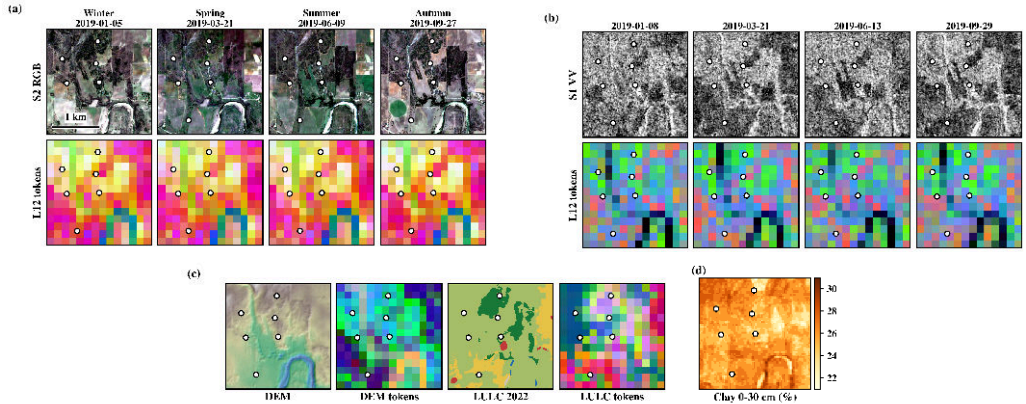
Motivation TerraMind embeddings for soil moisture



Embedding extraction. Sentinel-2, Sentinel-1, elevation and land-cover images pass through the frozen TerraMind encoder (ViT-B), which returns one 768-dimensional embedding per 160 m patch.

In this study, TerraMind embeddings are used as predictors of daily soil moisture.

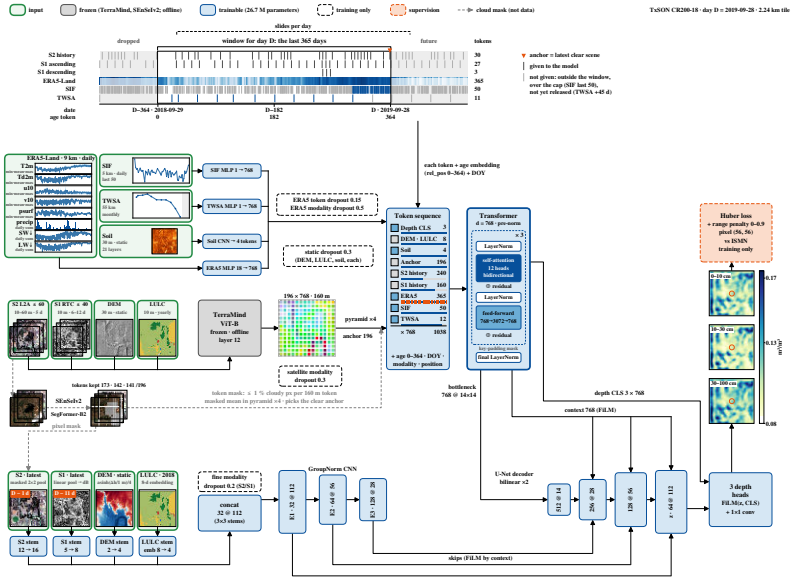
Motivation Example: TerraMind embeddings at one site (TxSON, 2019)



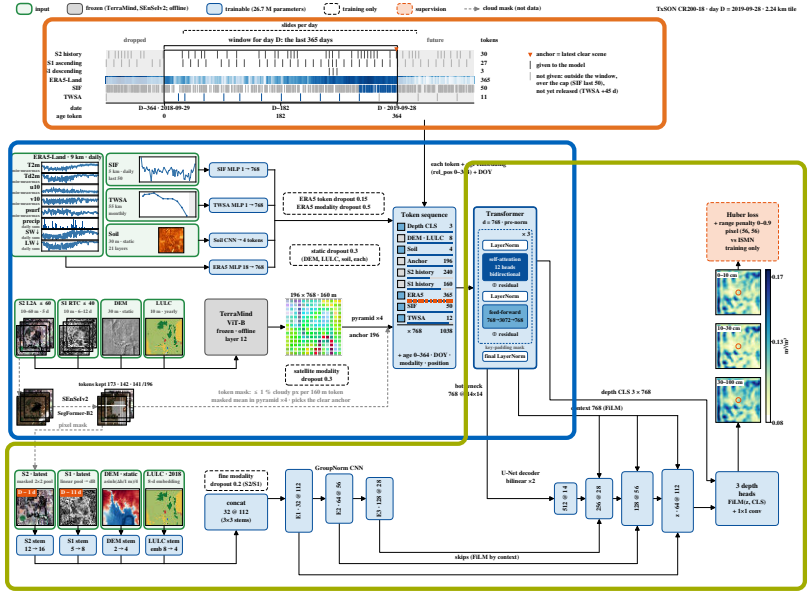
TerraMind embeddings at TxSON, 2019. (a) Sentinel-2 in four seasons, (b) Sentinel-1 VV, (c) elevation and land cover, each above its embeddings; (d) clay at 0-30 cm. Colours: first three principal components; dots: stations.

Each 160 m patch is represented by one 768-dimensional embedding that changes with season and sensor.

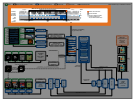
Method Model architecture

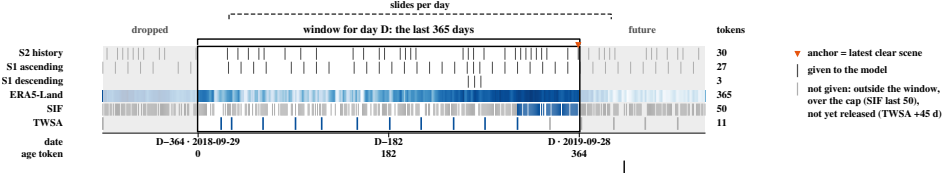


Method Model architecture

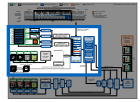


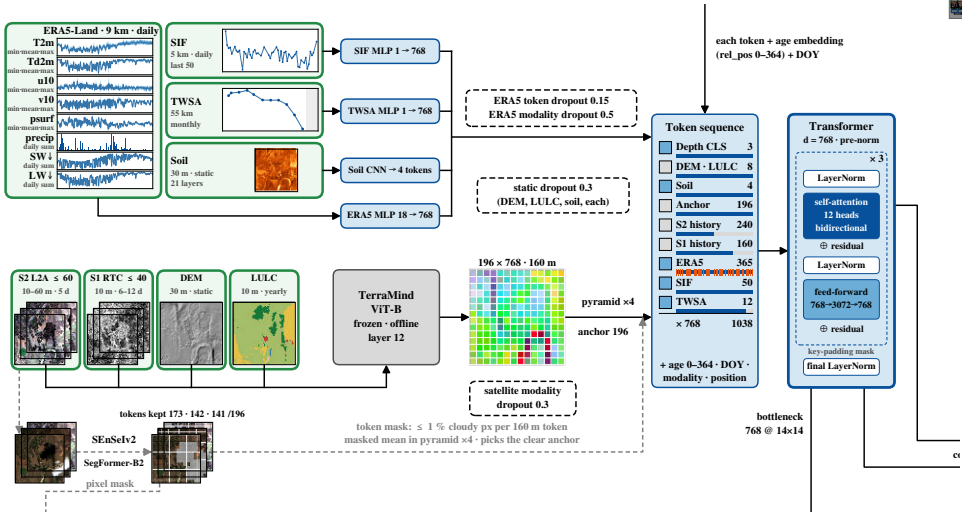
Method Inputs: what the model sees for day D



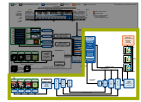


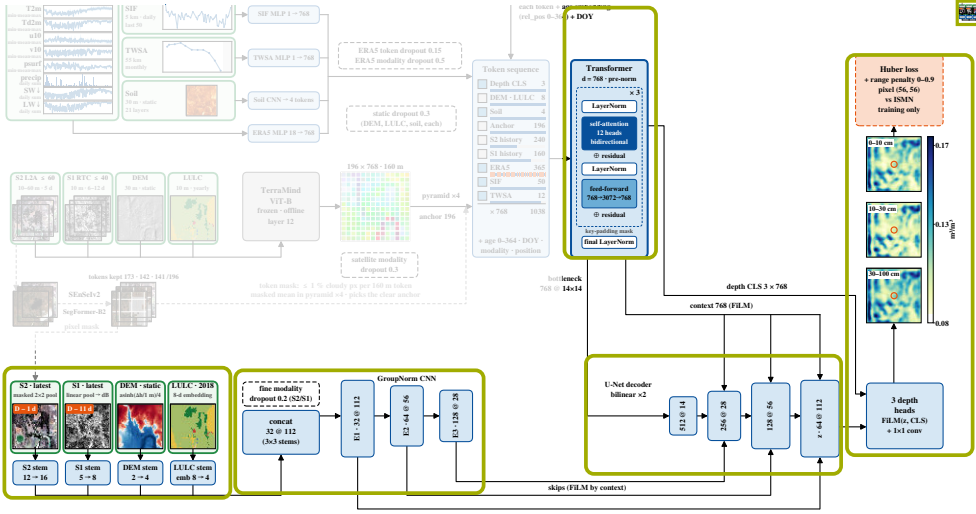
Method Coarse path: tokens → transformer



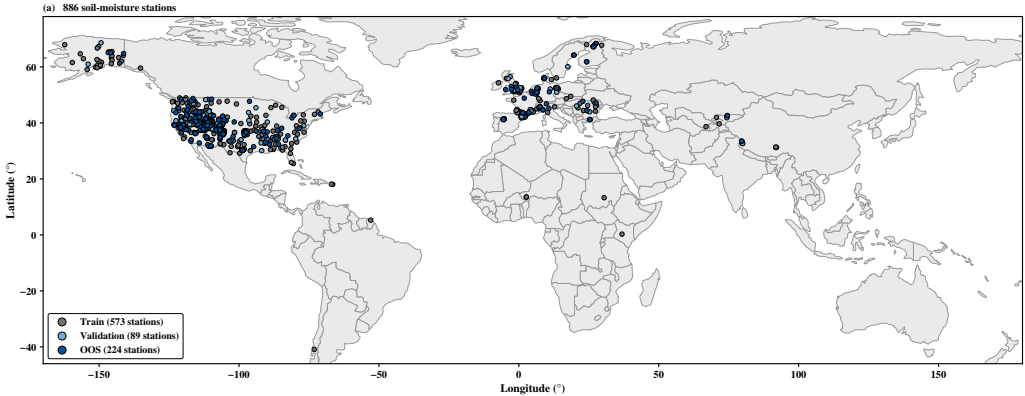


Method Fine path: 20 m CNN → U-Net decoder → depth heads





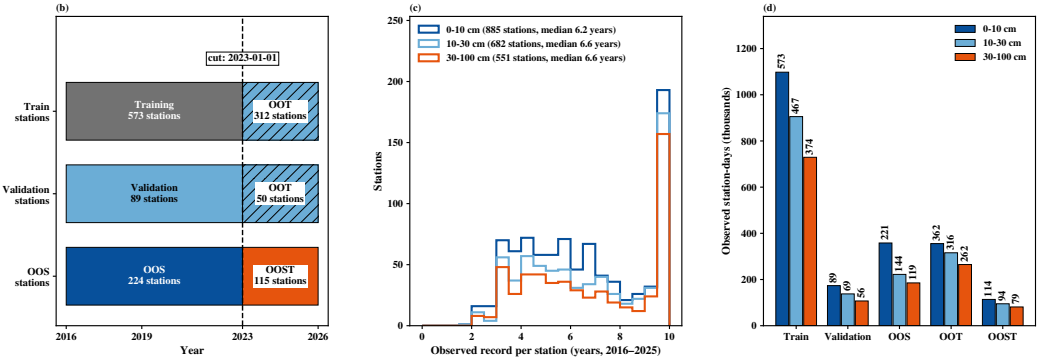
Data In-situ soil moisture: global coverage



In-situ soil-moisture stations from ISMN and flux networks (n = 886): 573 training, 89 validation and 224 spatial hold-out stations.

Station coverage is dense in North America and Europe and sparse in the tropics and the Southern Hemisphere.

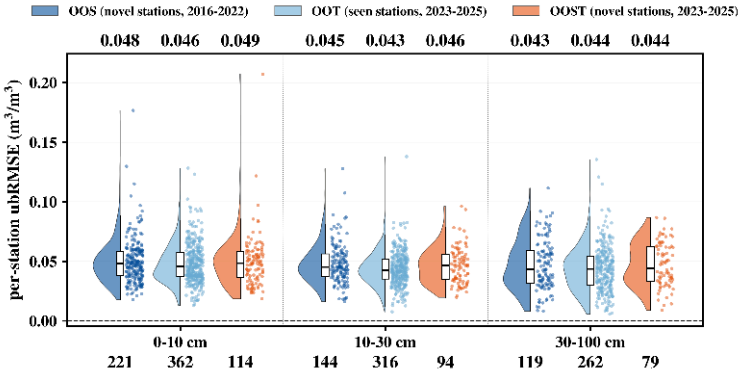
Data Depth coverage, record length and splits



Station records. (b) Training, validation and spatial hold-out stations and the temporal split on 1 January 2023; (c) length of the observed record per station; (d) number of observed station-days.

Observations cover 885, 682 and 551 stations at 0-10, 10-30 and 30-100 cm, with median records of 6.2, 6.6 and 6.6 years.

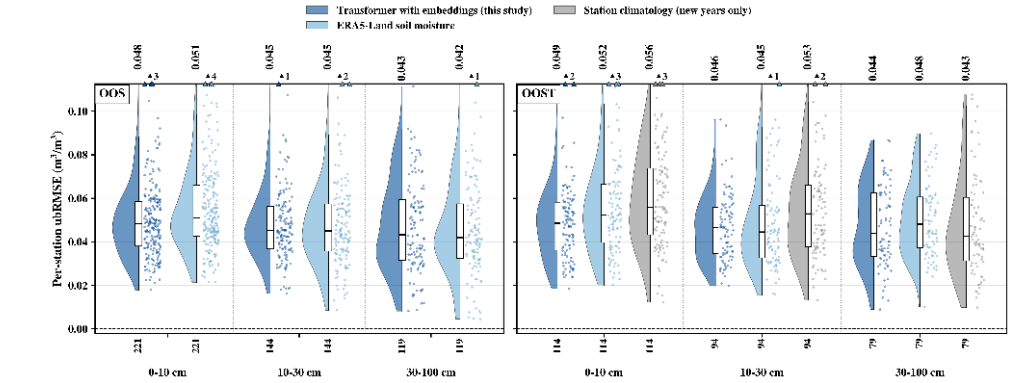
Results Per-station ubRMSE by depth



Per-station ubRMSE for the spatial (OOS), temporal (OOT, 2023-2025) and spatio-temporal (OOST) hold-out sets: density, median and interquartile range, one dot per station; medians above, station counts below.

The median ubRMSE was 0.046, 0.044 and 0.047 m^3/m^3 in the spatial, temporal and spatio-temporal hold-out sets.

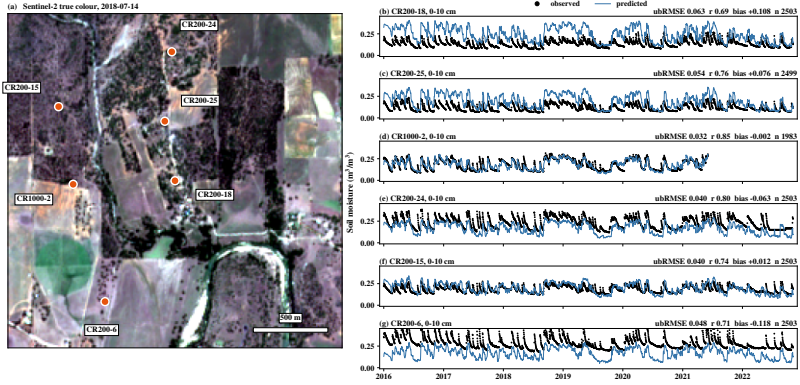
Results Comparison with reference estimates



Per-station ubRMSE of the model, ERA5-Land soil moisture and station climatology (OOST only, as it needs the station record) in the spatial and spatio-temporal hold-out sets; medians above, counts below.

The model had a lower ubRMSE than ERA5-Land at 0-10 cm and than station climatology in new years.

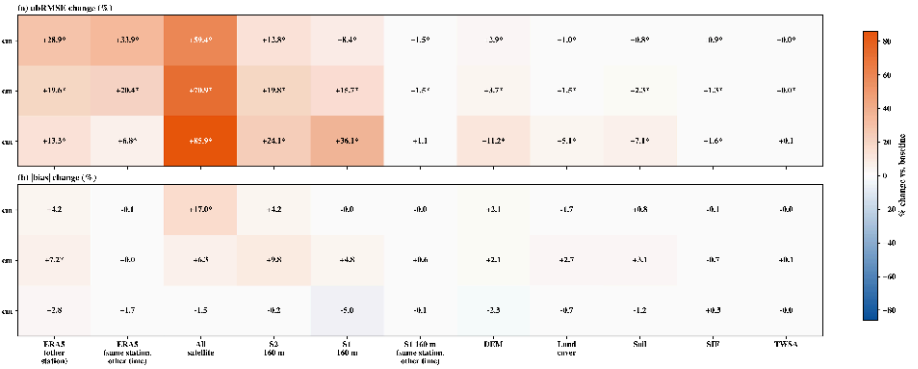
Results TxSON: six stations inside one 2.2 km tile



Six stations within one 2.24 km input tile (TxSON, Texas). (a) Sentinel-2 true colour with the station locations; (b-g) estimated and observed soil moisture at 0-10 cm.

The dynamics were reproduced at all six stations (r 0.69-0.85; ubRMSE 0.032-0.063 m^3/m^3), but biases of −0.12 to +0.11 m^3/m^3 remained.

Results Input exchange: which inputs does the model use?



Median per-station change in (a) ubRMSE and (b) absolute bias at the spatial hold-out stations when one predictor group is taken from another station (or from another time). **:* $p < 0.05$ *(station bootstrap of the median,* $n = 2000$*).*

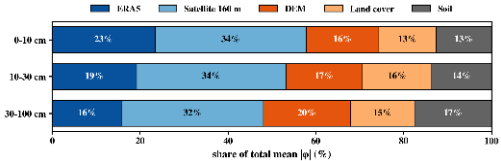
Satellite inputs from another station increased the ubRMSE by 59-86 % (rising with depth), meteorology by 13-29 % (falling with depth).

Interpretation Which predictors drive the estimate? Exact group Shapley values

Players: 5 predictor groups meteorology, satellite embeddings, elevation, land cover, soil.

Groups, not single inputs: one Shapley value per input would require 2^N model evaluations; with five groups all $2^5 = 32$ coalitions are evaluated, so the values are exact, not sampled. Based on SHAPformer (Hertel et al., 2026).

Additive, per day and depth: estimate $= f(\varnothing) + \sum_i \varphi_i$ (m^3/m^3).



Share of the total mean absolute Shapley value ($|\varphi|$) of each predictor group at the spatial hold-out stations, at 0-10, 10-30 and 30-100 cm.

Satellite embeddings received the largest share (32-34 %); meteorology fell with depth (23 to 16 %), elevation and soil rose (16 to 20 % and 13 to 18 %).

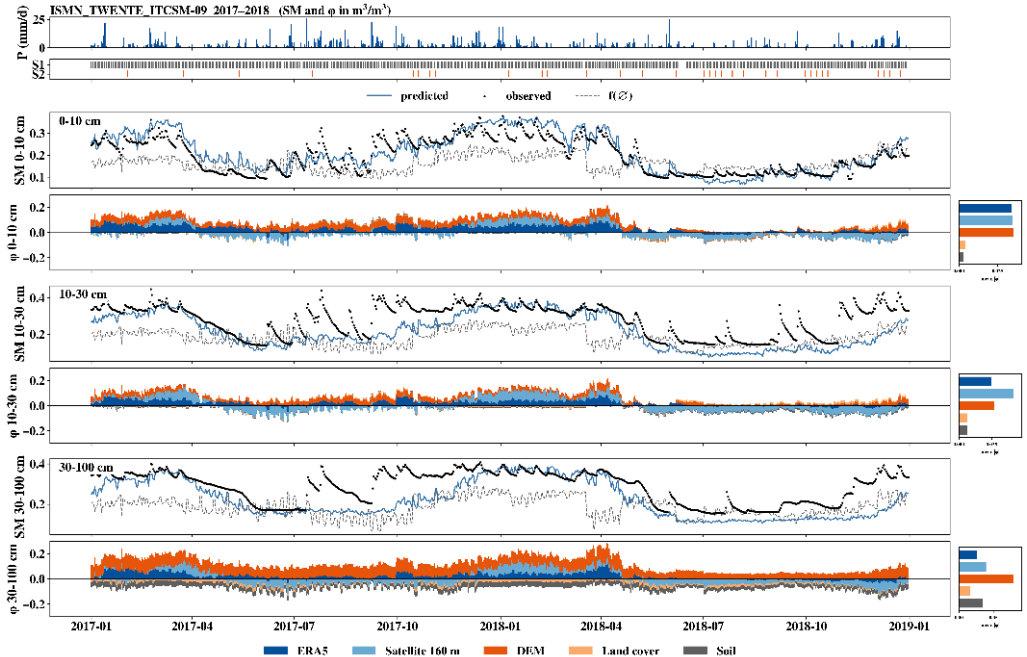

ISMN_TWENTE_ITCSM-09_2017–2018 (SM and φ in m³/m³)
P (mm/d)
predicted
observed
f(∅)
0-10 cm
SM 0-10 cm
φ 0-10 cm
10-30 cm
SM 10-30 cm
φ 10-30 cm
30-100 cm
SM 30-100 cm
φ 30-100 cm
2017-01
2017-04
2017-07
2017-10
2018-01
2018-04
2018-07
2018-10
2019-01
ERA5
Satellite 160 m
DEM
Land cover
Soil

Table 1. *Median correlation over the spatial hold-out stations between Shapley value anomalies (φ'; met: meteorology, sat: satellite embeddings, static: elevation, land cover and soil) and observed soil-moisture anomalies (θ'), i.e. deviations from the station climatology (31-day running mean of day-of-year means). Static inputs are time-invariant, so φ'_{static} arises from interactions only; $r(\cdot \mid \cdot)$: partial correlation. *: $p < 0.05$ (station bootstrap, $n = 2000$).*

Correlation	0-10 cm	10-30 cm	30-100 cm
Meteorology: $r(\varphi'_{\text{met}}, \theta')$	0.49*	0.42*	0.28*
$r(\varphi'_{\text{met}}, \theta' \mid \varphi'_{\text{static}}, \varphi'_{\text{sat}})$	0.47*	0.41*	0.27*
Satellite embeddings: $r(\varphi'_{\text{sat}}, \theta')$	0.26*	0.16*	0.09*
$r(\varphi'_{\text{sat}}, \theta' \mid \varphi'_{\text{static}})$	0.24*	0.17*	0.15*
$r(\varphi'_{\text{sat}}, \theta' \mid \varphi'_{\text{static}}, \varphi'_{\text{met}})$	0.24*	0.15*	0.14*

The satellite embeddings carry a time-varying signal that is weaker than that of meteorology and strongest near the surface (partial r 0.24, 0.15 and 0.14).

Conclusions

RQ1 How accurate are daily estimates at 0-10, 10-30 and 30-100 cm at new stations and in new years, compared with ERA5-Land and station climatology?

- The median ubRMSE at 0-10, 10-30 and 30-100 cm was 0.048, 0.045 and 0.043 m^3/m^3 at new stations and 0.046, 0.043 and 0.044 m^3/m^3 in new years, so **new stations were about as accurate as new years**.
- The ubRMSE was **lower than that of ERA5-Land** at 0-10 cm (0.048 vs 0.051 m^3/m^3 at new stations) and **lower than that of station climatology in new years** at all depths (0.046, 0.043 and 0.044 vs 0.058, 0.052 and 0.046 m^3/m^3).

RQ2 Which predictor groups drive the estimates, and how does this change with depth?

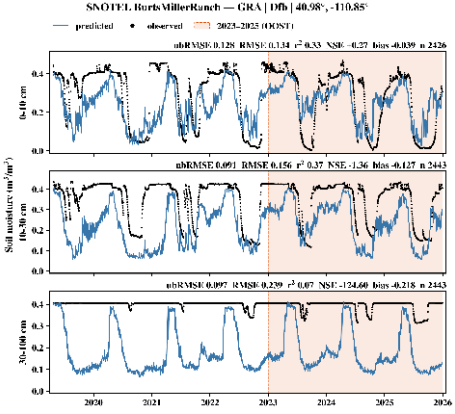
- The satellite embeddings received **the largest share of the attribution at every depth** (32-34 %).
- The **share of meteorology decreased with depth**, from 23 % at 0-10 cm to 16 % at 30-100 cm, whereas those of **elevation and soil increased**, from 16 % to 20 % and from 13 % to 18 %.

RQ3 Do the embeddings carry a static (site) or a time-varying signal?

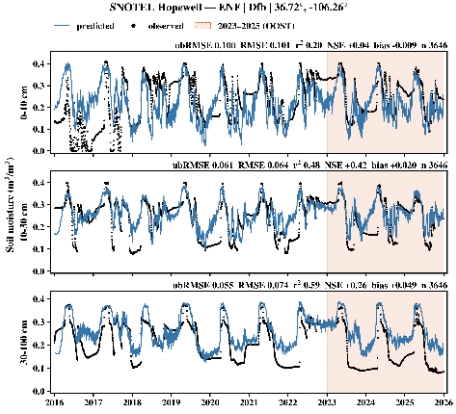
- **Mainly static:** another station's embeddings raised the ubRMSE by 59-86 %, whereas the station's own Sentinel-1 embeddings from another time changed it by at most 1.5 %.
- **A weaker time-varying signal:** the partial correlation between the Shapley value of the satellite embeddings and soil moisture, both as anomalies and after controlling for the static groups and meteorology, was 0.24, 0.15 and 0.14 at 0-10, 10-30 and 30-100 cm.

Backup slides

Results Sites where the time series have some issues



SNOTEL Burts Miller Ranch, Utah



SNOTEL Hopewell, New Mexico

Spatial hold-out stations with the highest ubRMSE at all three depths; shaded: spatio-temporal hold-out (2023-2025).

At 30-100 cm the estimates deviate in summer: too dry at Burts Miller Ranch and too wet at Hopewell.

Players: five predictor groups

meteorology (ERA5-Land), satellite embeddings (S1/S2 history), elevation, land cover, soil

Value of a coalition S: the estimate $f(S)$ at one depth with only the groups in S present

$$\varphi_i = \sum_{S \subseteq K \setminus \{i\}} \frac{|S|!\,(|K| - |S| - 1)!}{|K|!} \left[f(S \cup \{i\}) - f(S) \right], \; |K| = 5$$

Additive: estimate $= f(\varnothing) + \sum_i \varphi_i$, every day and depth

$f(\varnothing)$ keeps SIF, water storage and the 20 m image inputs

SHAPformer (Hertel et al., 2026)

a removed group is masked from attention; all coalitions evaluated exactly, no sampling, no background data

Our changes

- Players are groups of many inputs, not single variables.
- An absent group is removed exactly as in the training dropout, so all $2^5 = 32$ coalitions are inputs the model was trained on.
- Shapley values are computed separately for the three depths.

Backup **Gradient-boosted trees: embeddings vs site maps**

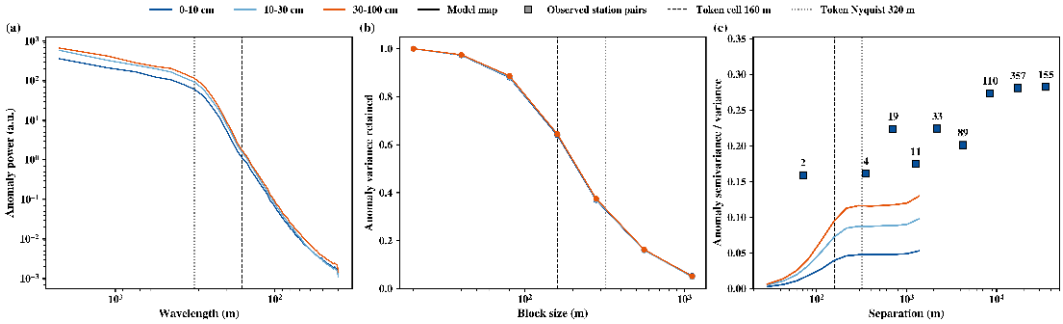
Median ubRMSE (m^3/m^3) at 0-10 / 10-30 / 30-100 cm, new stations (OOS; 221 / 144 / 119 stations)

	Without embeddings	With embeddings
Without site maps	0.049 / 0.047 / 0.046	0.046* / 0.044* / 0.039*
With site maps	0.044* / 0.042* / 0.039*	0.044 / 0.043 / 0.039

*All models use the same dynamic predictors (meteorology, SIF and terrestrial water storage); embeddings are station means before 2023, reduced to 20 principal components per modality. *: different from the model without that input (left, or above), $p < 0.05$ (two-sided station bootstrap of the median difference, $n = 2000$).*

Station-mean embeddings replaced the site maps but added no skill to them.

Backup Spatial resolution of the estimated maps (TxSON)



Daily anomaly maps of the 30 TxSON tiles at spatial hold-out stations. (a) Radial power spectrum; (b) anomaly variance retained after block averaging; (c) normalised semivariance of the maps (lines) and of the observed station anomalies at 0-10 cm (squares; numbers: station pairs). Dashed: 160 m embedding cell; dotted: 320 m.

The maps contain spatial structure down to about 250 m, which the few stations sharing a tile could neither confirm nor rule out.