https://doi.org/10.5194/essd-2026-126
Preprint. Discussion started: 13 April 2026

Open Access Earth System Science Data Discussions

European daily dataset of soil moisture from *in situ* meteorological observations

Gerard van der Schrier[1], Milan Fischer[2,3], Martin Mulder[4], Jos van Dam[4], Jan Řehoř[2,5], and Miroslav Trnka[2,3]

[1]Royal Netherlands Meteorological Institute, De Bilt, the Netherlands
[2]Global Change Research Institute, Czech Academy of Sciences, Brno, Czech Republic
[3]Department of Agrosystems and Bioclimatology, Mendel University in Brno, Brno, Czech Republic
[4]Department of Environmental Sciences, Wageningen University and Research Centre, Wageningen, The Netherlands
[5]Institute of Geography, Masaryk University, Brno, Czech Republic

Correspondence: G. van der Schrier (schrier@knmi.nl)

Abstract. This paper documents the development and validation of an *in situ* dataset for daily soil moisture based on observed meteorological variables from the European Climate Assessment & Dataset (ECA&D). These data are input to the SoilClim model which provides soil water contents values for the upper 10 cm, the root zone and the deeper layer of the soil. Soil characteristics are taken from the widely used SoilGrids system. The meteorological input data are daily precipitation totals, daily averaged temperature, and global radiation sums. The pan-European dataset for soil water content results in a set of 5758 stations for which daily soil water content values are produced, for the surface zone (upper 10 cm) and for the root zone (upper 40 cm), for the period for which all input data is available. Model results are validated against observations of actual evapotranspiration (ET_a) from the FluxNet experiment. Calculated soil water content are compared against a selection of the observations of the International Soil Moisture Network (ISMN). In the comparisons against observations, calculated values of ET_a and soil water content (SWC) from the state-of-the-art soil dynamics model SWAP, which is more advanced than SoilClim, are used as well. The comparisons show that SoilClim and SWAP show generally the same behaviour albeit that SoilClim saturates quicker than SWAP. On average, the similarity between the models and observation for ET_a is higher (correlations 0.77 and 0.78 for SWAP and SoilClim respectively) than for the observed soil water content, where average correlations for the surface and rootzone vary between 0.52 and 0.62.

Based on the rootzone soil water content, four climate indices are calculated. These are the Soil Moisture Index (SMI) which is a metric to compare soil water content values across the diverse climatic zones which are present in Europe. This index relates to water content levels that are above or below the 50% level of the field capacity. In addition, an index is introduced that simply counts the number of days where the soil water content drops below the 50% level of the field capacity. In addition, two indices assess more extreme soil water content values: the number days where the soil water content is either nearly at field capacity or close to the wilting point. The 1991-2020 spring and summer climatology is provided for this dataset in terms of these indices. The relatively recent 2018 drought is put in a historical perspective by a comparison against the 1976 drought and a trend analysis of summer desiccation since 1979 is provided.

Summary of Comments on 20260901_essd-2026-126.pdf

Page: 1

Number	Author	Subject	Date	Comment
Number: 1	Author: Willem Vervoort	Subject: Inserted Text	Date: 25/09/2026 2:07:16 PM	simulated
Number: 2	Author: Willem Vervoort	Subject: Inserted Text	Date: 25/09/2026 2:08:25 PM	s
Number: 3	Author: Willem Vervoort	Subject: Cross-Out	Date: 25/09/2026 2:09:15 PM	
Number: 4	Author: Willem Vervoort	Subject: Cross-Out	Date: 25/09/2026 2:09:12 PM	
Number: 5	Author: Willem Vervoort	Subject: Cross-Out	Date: 25/09/2026 2:09:04 PM	
Number: 6	Author: Willem Vervoort	Subject: Cross-Out	Date: 25/09/2026 2:09:09 PM	
Number: 7	Author: Willem Vervoort	Subject: Highlight	Date: 25/09/2026 2:10:10 PM	"Nearly" and "close" that does not seem very precise...
Number: 8	Author: Willem Vervoort	Subject: Inserted Text	Date: 25/09/2026 2:10:32 PM	.
Number: 9	Author: Willem Vervoort	Subject: Inserted Text	Date: 25/09/2026 2:10:40 PM	In addition,

https://doi.org/10.5194/essd-2026-126
Preprint. Discussion started: 13 April 2026

Open Access Earth System Science Data Discussions

The use of *in situ* observations, the daily resolution and the climate indices included in the described dataset make it suitable as validation of other soil moisture datasets and the assessment of changes in the occurrence of extremes in soil moisture over the European domain. The dataset of daily soil moisture from *in situ* meteorological observations is available at https://doi.org/10.21944/phk8-at71 (van der Schrier et al., 2026).

1 Introduction

Northwest Europe has recently seen three consecutive springs and summers that have been dry and sunny (2018, 2019, 2020). In between these seasons, the autumn and winters failed to compensate for the shortage of water. This has led to widespread water availability issues. Van der Wiel et al. (2023) indicate that such multi-year droughts are expected to occur more often in a warming climate, with the resulting feed-back on heatwaves as we have recently seen (Lemus-Canovas et al., 2024). An early example where successive dry and (for spring and summer) sunny seasons have led to a severe drought is the 1921 drought (van der Schrier et al., 2021), which resulted in a drought that is still the severest on record (van der Schrier et al., 2006). While droughts are among the costliest climatic extremes, excessively wet periods can be equally problematic from the perspective of a farmer when fields are too wet for harvesting and crops perish. Especially in the northern half of Europe, where the annual sum of precipitation is increasing, and wet autumns have occurred more often in recent decades (Kendon et al., 2024). The 12-month period until the end of the 2024 spring stands out in this respect as it has been the wettest 12-month period since records began in parts of Central Europe (Tagesspiegel, 2024; KNMI, 2024). Recurring droughts and pluvials are aspects of climatic change that have a large societal relevance but our ability to monitor the impacts of these climatic events is limited. The common approach is to use standard drought indices like SPI, SPEI and scPDSI which are all based on atmospheric parameters only and, for the SPEI and scPDSI, make attempts to capture the loss of moisture from the soils to atmosphere too. A comprehensive review of drought indices can be found in Heim (2002). All these indices are used as proxies for soil water content but the dynamics of soil water content is not or only very rudimentary captured in these indices. The time scale at which these drought indices are usually calculated - monthly - is too coarse to monitor short term variations in the soil water balance. It also ignores the climate change-related shift in the distribution of daily precipitation towards longer dry spells and more extreme precipitation (Kyselý, 2009; Zolina et al., 2010; Cavanaugh et al., 2015). Extreme precipitation produces larger runoff and smaller infiltration of moisture into the soil, which further decouples precipitation amount from soil moisture content. The shift in the precipitation distribution impacts soil moisture which is not captured when using monthly mean values. It is the soil water balance that needs to be resolved to address short-term fluctuations, like flash droughts (spells of extreme heat that are related to a lack of soil moisture) and floods that occur when strong rains are combined with saturated soils.

The importance of soil moisture in climate dynamics contrasts with the lack of *in situ* observational datasets for soil water content. There are several consequences of this omission. One is that alternative sources for soil moisture information, like the ERA5 Land (Muñoz Sabater et al., 2021) and the C3S Soil Moisture satellite dataset (Dorigo et al., 2019), lack a continent-wide validation dataset. For the assessment of the annual soil moisture anomaly in the 2019 European State of the Climate (C3S, 2020), which had a summer where many national heat records were broken, this produced a view on soil moisture anomalies

Page: 2

Number: 1	Author: Willem Vervoort	Subject: Inserted Text	Date: 25/09/2026 2:11:06 PM

simulated

Number: 2	Author: Willem Vervoort	Subject: Inserted Text	Date: 25/09/2026 2:11:28 PM

experienced

Number: 3	Author: Willem Vervoort	Subject: Highlight	Date: 25/09/2026 2:13:00 PM

This is particularly problematic

Number: 4	Author: Willem Vervoort	Subject: Inserted Text	Date: 25/09/2026 2:12:14 PM

if

Number: 5	Author: Willem Vervoort	Subject: Cross-Out	Date: 25/09/2026 2:13:17 PM

Number: 6	Author: Willem Vervoort	Subject: Cross-Out	Date: 25/09/2026 2:13:36 PM

Number: 7	Author: Willem Vervoort	Subject: Cross-Out	Date: 25/09/2026 2:15:29 PM

Number: 8	Author: Willem Vervoort	Subject: Inserted Text	Date: 25/09/2026 2:15:54 PM

only very rudimentary or not at all

Number: 9	Author: Willem Vervoort	Subject: Highlight	Date: 25/09/2026 2:16:50 PM

repeat of the above??

https://doi.org/10.5194/essd-2026-126
Preprint. Discussion started: 13 April 2026

that was not entirely consistent between these two products. Focusing on the Mediterranean region, a lack of a uniform and consistent validation dataset for soil water content interferes with that region's vulnerability to soil moisture deficits. Recently Ardilouze et al. (2022) argued that a model analysis of the soil moisture - precipitation feedback in the Mediterranean is hampered by a 'wide array of uncertainties', pointing to soil moisture and its representation in climate models.

Observational sites for soil moisture are costly to set-up and to operate. The consequence is that only a few ground-based soil moisture measurement networks are available (Robock et al., 2000). Except of a few long-running stations, like those in the Ukraine (Robock et al., 2005), the records are usually short. The poor availability of ground observations for soil moisture which is limiting progress in these disciplines (Seneviratne et al., 2010), it has been the motivation to construct an observation-based dataset for soil water content. Here, we document the construction and validation of a dataset of daily values of soil moisture for Europe based on observed *in situ* daily values of precipitation, temperature, global radiation and observed soil characteristics which force the soil moisture model SoilClim (Hlavinka et al., 2011).

The network of meteorological stations used in this dataset is that of the European Climate Assessment & Dataset (ECA&D, Klein Tank et al. (2002)) and the meteorological data are owned and provided by the European National Meteorological Services (NMSs). The use of these quality controlled meteorological observed quantities has the benefit that it makes use of the dense network and the full historical records of these observations. In addition, the World Meteorological Organization (WMO, WMO (2023)) prescribes that these stations are situated over short-clipped grass and should be representative for a wider region. This has the benefit that a certain standardisation in drought monitoring is in place. The drawback is that soil moisture information for other land uses, like forests and agricultural crops, is missing in this dataset.

Specific for soil moisture monitoring purposes, a set of soil moisture indices is used based on the rootzone soil moisture content, which aims to provide a historical perspective of soil water content variations. These indices are taken either from the literature, like the Soil Moisture Index proposed by Hunt et al. (2009) which draws upon earlier suggestions (Betts, 2004). Inspired by the climate impact indices of the WMO Commission for Climatology's Expert Team on Sector-specific Climate Indices (ET-SCI), soil water content indices are added that monitor the occurrence of more extreme conditions. These relate to the number of daily occurrences of moisture stressed, very moisture stressed or very moisture saturated conditions.

The calculation of soil water content for this dataset is based on the SoilClim model (Hlavinka et al., 2011). There are more complex models for water balance estimates in a soil–crop system available, like the SWAP model (Kroes et al., 2017; Heinen et al., 2024), but that requires detailed or specialized input data (e.g. water retention function, hydraulic conductivity function, root density and other inputs). The motivation to use SoilClim is that it can be used for a wide range of climatic conditions, it requires a modest amount of input data and it has low computational costs. In this study, the SoilClim is validated against observations of actual evapotranspiration, a selected set of soil moisture observations, and it is compared against SWAP model results where both soil moisture models are forced with similar input data.

Page: 3

Number: 1 Author: Willem Vervoort Subject: Cross-Out Date: 25/09/2026 2:18:22 PM

Number: 2 Author: Willem Vervoort Subject: Cross-Out Date: 25/09/2026 2:18:22 PM

Number: 3 Author: Willem Vervoort Subject: Cross-Out Date: 25/09/2026 2:18:25 PM

Number: 4 Author: Willem Vervoort Subject: Inserted Text Date: 25/09/2026 2:18:36 PM
This

Number: 5 Author: Willem Vervoort Subject: Highlight Date: 25/09/2026 2:21:00 PM
I don't think this really argues why a simulated soil moisture dataset is acceptable in the absence of a observed soil moisture dataset. I think this needs to be made explicit. Under which conditions are simulated data fine relative to observed data and why is the "semi-observed" dataset here better than the satellite data?

Number: 6 Author: Willem Vervoort Subject: Inserted Text Date: 25/09/2026 2:18:52 PM
simulated

Number: 7 Author: Willem Vervoort Subject: Cross-Out Date: 25/09/2026 2:22:31 PM

Number: 8 Author: Willem Vervoort Subject: Inserted Text Date: 25/09/2026 2:22:44 PM
is used

Number: 9 Author: Willem Vervoort Subject: Inserted Text Date: 25/09/2026 2:23:11 PM
ese

https://doi.org/10.5194/essd-2026-126
Preprint. Discussion started: 13 April 2026

2 Data and Methods

2.1 ECA&D

The European Climate Assessment & Dataset (ECA&D) is a repository of meteorological station records for Europe with the purpose providing a platform with easy access to high quality *in situ* observations. In addition, derived data products are produced to monitor climatic change and variations. The long records in ECA&D give the historical perspective to recent climatic extremes. The data collated by the ECA&D initiative are mostly supplied directly by the European NMSs. For the countries outside the European Union the number of stations available to ECA&D may be a subset of the complete network available to the NMSs. This is due to restrictions concerning the exchange of data. Other sources of data are repositories that, similar to ECA&D, aim to provide unified access to a large set of data. The original source of these data is the NMSs as well.

ECA&D is the backbone of the Climate Data node for the WMO Regional Climate Centre for Europe and the Middle East. The domain covered is referred to as Regional Association (RA) VI.

ECA&D collaborates with regional meteorological services, like the Meteorological Service of Catalonia in Spain, environmental services, like the ARPA's (Regional Environmental Protection Agencies) in Italy or river authorities like those in Flanders and Walloon. Universities and observatories share their data with ECA&D as well, like the Armagh Observatory in Northern Ireland, the Radcliffe Meteorological Station of Oxford University in England or the data gathered by the Wageningen University in the Netherlands. In some countries, data are supplied by both the NMS and from an complementary network through the ministry of agriculture, like in Luxembourg and Portugal.

Figure 1 shows the spatial and temporal availability of rain gauges, stations that provide daily averaged temperature and stations providing global radiation. These are the meteorological variables that are required in the calculation of soil moisture content. The figure shows the large amount of station data in the western half of the continent, with the number of rain gauges outnumbering by far stations that provide daily averaged temperature. The number of stations that provide global radiation values is clearly the smallest and the availability of the latter variable is the bottleneck in further expanding the dataset with soil moisture content.

Figure 2 illustrates the availability of stations for which soil moisture data can be provided. The total number of stations in this map is 5758. The number of stations provided through other sources than directly through the NMS varies depending on the variables. For daily precipitation and average temperature, the percentages of data not directly provided by the NMSs are 5.05% and 6.06% respectively. In contrast, for global radiation this percentage is 13.28%. The strong contribution of the European NMSs to ECA&D parallels their increasing willingness to share data more abundantly. This is fuelled by efforts of the World Meteorological Organization to urge member states to share data (WMO, 2022) and EU regulations 2019/1024 and 2023/138 on Open Data and re-use of public sector information and, more specifically, on Meteorological data (EU Commission, 2022). Nevertheless, the efforts of organisations to build large and consistent repositories with *in situ* data have been used

Page: 4

Number: 1	Author: Willem Vervoort	Subject: Inserted Text	Date: 25/09/2026 2:23:45 PM
of			
Number: 2	Author: Willem Vervoort	Subject: Cross-Out	Date: 25/09/2026 2:24:01 PM
Number: 3	Author: Willem Vervoort	Subject: Cross-Out	Date: 25/09/2026 2:25:28 PM
Number: 4	Author: Willem Vervoort	Subject: Inserted Text	Date: 25/09/2026 2:25:55 PM
minimum			

https://doi.org/10.5194/essd-2026-126
Preprint. Discussion started: 13 April 2026

to supplement data for global radiation, where the data from the European domain of the World Radiation Data Centre[1] and the Baseline Surface Radiation Network (BSRN)[2] (Driemel et al., 2018) are shared with ECA&D.

Any gaps in the time series provided to ECA&D are filled with data from neighbouring series to form more temporally complete series (ECA&D Project Team, 2012; van der Schrier et al., 2013). To update the series to near real time when updates of the NMSs are not available, data are used from synoptic messages (SYNOP) distributed via the Global Telecommunications System WMO (2007). The SYNOP data are replaced when validated data become available from the NMSs (van den Besselaar et al., 2012).

A range of different methods is used to calculate the daily values, and this is often dependent upon the convention used by a given data provider. For example, daily mean temperatures may be the average of maximum and minimum temperatures, or may be the average of hourly readings, or another frequency such as three hourly. In addition, the start and end times of the 24h period over which elements are measured are not standardized across Europe and the date attached to each observation may sometimes be the end of the observation period rather than the start. Although no attempts are made to adjust for this inhomogeneity across Europe, ECA&D staff apply a day shift when necessary to make sure that the observed period coincides as much as possible with the date attached to the observation (ECA&D Project Team, 2012).

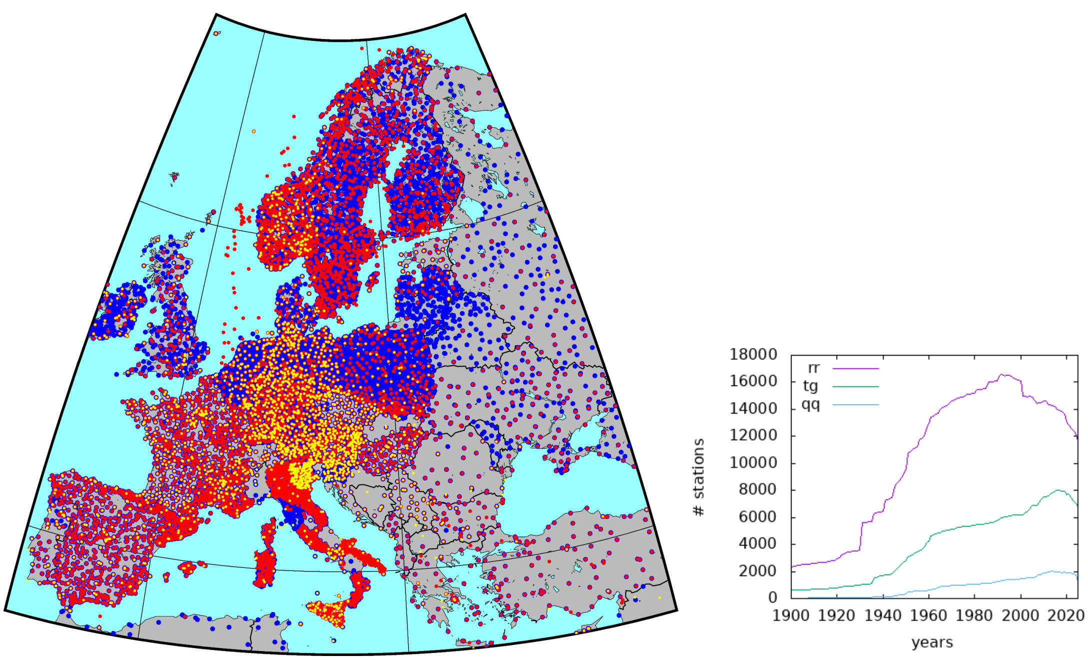

Figure 1. Map of the locations of rain gauges (blue) and stations recording daily average temperature (red) and global radiation (yellow) in ECA&D (left). The coloured dots are overlain with varying sizes to make clear that many stations provide all three variables. The right plot shows the temporal evolution of the number of rain gauges versus the number of global radiation stations since 1900.

[1]http://wrdc.mgo.rssi.ru
[2]https://bsrn.awi.de

Page: 5

Number: 1 Author: Willem Vervoort Subject: Highlight Date: 25/09/2026 2:28:47 PM

what is iT, qq and tg in this plot? This is not defined in the caption

https://doi.org/10.5194/essd-2026-126
Preprint. Discussion started: 13 April 2026

Open Access Earth System Science Data Discussions

AW10p	number of days when the Soil Water Content < 10% of maximum Available Water Content
	Let AW_{ij} be the daily soil water content on day i in period j and let AW10
	be the 10% value of the maximum Available Water Capacity.
	The index in the period j is determined by counting days where $\mathrm{AW}_{ij} < \mathrm{AW10}$.
AW90p	number of days when Soil Water Content > 90% of maximum Available Water Content
	Let AW_{ij} be the daily soil water content on day i in period j and let AW90
	be the 90% value of the maximum Available Water Capacity.
	The index in the period j is determined by counting days where $\mathrm{AW}_{ij} > \mathrm{AW90}$.
SMI	Soil Moisture Index
	Based on actual water content and a known field capacity and wilting point, the SMI
	is a continuous function and is scaled from 5.0 to -5.0, with 5.0 representing water
	content at field capacity and -5.0 representing water content at wilting point.
	SMI represents the time mean value.
SMI0	number of days with moisture stress
	Let SMI_{ij} be the daily Soil Moisture Index on day i in period j. The number
	of days in period j is determined by counting the days where $\mathrm{SMI}_{ij} < 0$.

Table 2. Definition of climate indices for soil water content.

2.5 Climate indices

An index specifically designed for soil moisture monitoring is the Soil Moisture Index (SMI) (Hunt et al., 2009) This index can identify a quick onset of agricultural drought by comparing the observed dryness of a soil relative to the plant's ability to extract water. Hunt et al. (2009) explain the rationale behind the SMI with the observation that actual evapotranspiration became limiting halfway between field capacity and wilting point. The SMI is defined such that positive values correspond to the situation that soil moisture is above the 50% threshold of total available water and negative SMI values correspond to the situation that soil moisture is below that threshold. An SMI value of 0 separates the stress versus the non-stress situations. For the current purpose, SMI is calculated for every day and aggregated values are presented. To highlight the frequency of days where soil moisture stress kicks in, the number of days per period where the daily SMI values drop below 0 is calculated and presented as index SMI0.

Indices that focus more on the extreme cases of moisture stress or moisture saturated situations are the AW10p and AW90p indices, respectively. These indices calculate the number of days in each period where soil moisture is close to the wilting point or close to the field capacity. A more detailed definition is given in table 2 and the indices are schematically illustrated in fig. 4. In the phase space of soil water content values and the fineness of soil texture, possible values of soil water content are limited by the wilting point and the field capacity. When soil water content is in purple or green areas in the plot, indices AW90p and AW10p respectively become non-zero. Index SMI0 gets larger the more days soil water content is in the blue area. SMI indicates where, on average over a specified period, soil water content is found in this diagram.

Page: 11

Number: 1	Author: Willem Vervoort	Subject: Inserted Text	Date: 25/09/2026 9:05:50 PM
highlighting			
Number: 2	Author: Willem Vervoort	Subject: Inserted Text	Date: 25/09/2026 9:06:00 PM
omes			
Number: 3	Author: Willem Vervoort	Subject: Cross-Out	Date: 25/09/2026 9:06:25 PM
Number: 4	Author: Willem Vervoort	Subject: Cross-Out	Date: 25/09/2026 9:06:13 PM
Number: 5	Author: Willem Vervoort	Subject: Cross-Out	Date: 25/09/2026 9:06:15 PM
Number: 6	Author: Willem Vervoort	Subject: Cross-Out	Date: 25/09/2026 9:06:18 PM
Number: 7	Author: Willem Vervoort	Subject: Cross-Out	Date: 25/09/2026 9:06:20 PM
Number: 8	Author: Willem Vervoort	Subject: Cross-Out	Date: 25/09/2026 9:06:28 PM
Number: 9	Author: Willem Vervoort	Subject: Inserted Text	Date: 25/09/2026 9:07:28 PM
paper			
Number: 10	Author: Willem Vervoort	Subject: Cross-Out	Date: 25/09/2026 9:07:39 PM

https://doi.org/10.5194/essd-2026-126
Preprint. Discussion started: 13 April 2026

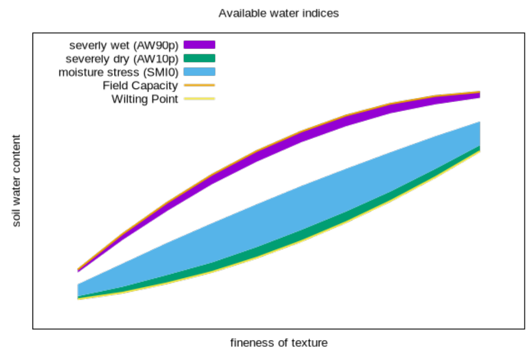

Figure 4. Schematic relationship between soil moisture and texture, with indications as to what the indices AW10p, AW90p and SMI0 relate to. Figure inspired by Barker et al. (2005, fig 1-6).

3 The operational embedding

3.1 The selection of ECA&D stations

Soil moisture estimates require ~~input data for~~ the amount of precipitation and the reference evapotranspiration ET_0. The stations available in ECA&D may not have all required variables present at each location as networks with only rain gauges are included as well or if only part of the available variables are shared with ECA&D. The mismatch in density of the rain gauge network and those of temperature and radiation (Figure 1) did motivate us to use the approach described below. The basis for the calculations are the blended series as described in Sect. 2.1.

Starting with the full network of stations that provide daily precipitation, the closest stations in the neighbourhood (within 25km) that provides global radiation (QQ) data and a station that provides daily mean temperature (TG) are located. For the station that provides temperature, the additional requirement is added that the elevation difference with the rain gauge should be less than 250 m. These latter two stations often coincide, but that is not always the case. If these stations can be found, a search is made for the nearest station that provides sea-level pressure (SLP). If no SLP-providing station can be found in the neighbourhood, the standard atmospheric pressure of 1030hPa is used as the influence of air pressure on ET_0 values is very small. With the daily values of QQ, TG and SLP, daily values of the Makkink reference evapotranspiration ET_0 can be calculated (Sect. 2.3).

The limiting factor in terms of the network density for which soil moisture can be calculated is the number of global radiation stations. The collection of global radiation records and the transformation of sunshine duration to global radiation records, the former often go further back in time. The transformation from sunshine duration series to global radiation series is done following the approach of Suehrcke et al. (2013). To illustrate this point, the number of rain gauges available in ECA&D (irrespective of stop and start dates) is 20210. For daily mean temperature this value is 9100 while for global radiation the number is 2307 (status summer 2025). The argument to start the selection of ECA&D stations with the precipitation station

Page: 12

Number: 1	Author: Willem Vervoort	Subject: Highlight	Date: 25/09/2026 9:09:43 PM

Do you mean increasing clay content? Also, axes labels are too small

Number: 2	Author: Willem Vervoort	Subject: Cross-Out	Date: 25/09/2026 9:10:31 PM

https://doi.org/10.5194/essd-2026-126
Preprint. Discussion started: 13 April 2026

is that precipitation has a much shorter decorrelation length than daily mean temperature or daily values of global radiation which are much more spatially homogeneous in comparison to precipitation.

3.2 The selection of the soil characteristics

For the subset of precipitation stations that can be linked to neighbouring stations that provide the input for the reference evapotranspiration ET_0, the coordinates are used to extract the soil characteristics for the grid square that corresponds to that location. The soil hydraulic functions are derived following Wösten et al. (1999) where the Mualem - van Genuchten parameters ThetaR, ThetaS, Alpha, n, Tort and Ksat are determined from sand, silt, clay, bulk density, organic carbon content values sourced from the SoilGrids system (Hengl et al., 2017).

The Available Water Content (AWC) needs to be corrected for the amount of stones in the soils. This is done by

$$AWC = (1 - gravel)(FC - WP) \tag{2}$$

where $gravel$ is the volume fraction of stones in the soils and FC and WP are the Field Capacity and Wilting Point.

In the calculations of soil water content in the soils, the actual groundwater table is ignored. Although Fan et al. (2013) present global observations of water table depth compiled from government archives and literature, a pan-European view on the actual groundwater table, updated in near-real time as is required for the current dataset, is not available. Therefore we apply both in SoilClim and SWAP free drainage. For SWAP, the model is initialized with the pressure head of each compartment in hydrostatic equilibrium with the initial groundwater level, which is set at -200 cm below soil surface.

3.3 Surface and root-zone soil moisture content

The WMO Essential Climate Variables for soil moisture are 'Surface soil moisture' and 'Root-zone soil moisture' (?) which is the motivation to include these two variables in the dataset. The output of the SoilClim model of the upper layer represents the soil moisture of the 0–10 cm layer. This is taken to be representative of the surface soil moisture. The study of Jackson et al. (1996, their fig.1) shows that 83% of the root biomass of temperate grassland is in the upper 30 cm of the soil. This motivates us to relate the soil moisture in the upper two layers of SoilClim (0–40 cm) to the root-zone soil moisture.

An additional motivation to provide both the surface and root-zone soil moisture content relates to the temporal variability of these two layers. Under dry conditions, the surface layer gradually decouples from the lower layer (Hirschi et al., 2014) and in in-situ observations show that the surface layer will then have a lower soil moisture content and variability compared to the root-zone layer.

Number: 1 Author: Willem Vervoort Subject: Highlight Date: 26/09/2026 10:21:24 AM
I think you need to be clearer that the results are from "non-calibrated models". as far as I understand there was no local fine-tuning of the models, the models are purely driven by the available data.

Number: 2 Author: Willem Vervoort Subject: Highlight Date: 25/09/2026 9:23:08 PM
THis is a repeat from earlier. .Maybe use: as indicated....

Number: 3 Author: Willem Vervoort Subject: Highlight Date: 25/09/2026 9:23:32 PM
missing reference??

https://doi.org/10.5194/essd-2026-126
Preprint. Discussion started: 13 April 2026

siteid	name	lat	lon	elev	start	stop	surface
AT-Neu	Neustift	47.1167	11.3175	970	2002	2012	grass
BE-Lon	Lonzee	50.5516	4.7462	167	2004	2014	crop
CH-Oe1	Oensingen grassland	47.2858	7.7319	450	2002	2008	grass
CZ-BK2	Bily Kriz grassland	49.4944	18.5429	855	2004	2012	grass
DE-Geb	Gebesee	51.0997	10.9146	162	2001	2014	crop
DE-Gri	Grillenburg	50.9500	13.5126	385	2004	2014	grass
DE-Kli	Klingenberg	50.8931	13.5224	478	2004	2014	crop
ES-LJu	Llano de los Juanes	36.9266	-2.7521	1600	2004	2013	crop
ES-LgS	Laguna Seca	37.0979	-2.9658	2267	2007	2009	shrub
ES-Ln2	Lanjaron-Salvage	36.9695	-3.4758	2249	2009	2009	shrub
FI-Jok	Jokioinen	60.8986	23.5134	109	2000	2003	grass
FI-Let	Lettosuo	60.6418	23.9595	111	2009	2012	crop
FR-Gri	Grignon	48.8442	1.9519	125	2004	2014	crop
IT-MBo	Monte Bondone	46.0147	11.0458	1550	2003	2013	grass
NL-Hor	Horstermeer	52.2403	5.0713	2.2	2004	2011	grass

Table 3. Names and start & stop dates of FluxNet records. Surface coverage information is from Teuling et al. (2010) and from the ESA landsurface classification.

4 Verification

4.1 Verification against FluxNet

The SWAP and SoilClim models output actual evapotranspiration ET_a data when calculating the soil moisture profile. These data are generated for a selection of stations from the FluxNet initiative (Pastorello et al., 2020) as reference, using meteorological input data from the nearest ECA&D stations.

As the focus of the present study is on soil moisture records based on the meteorological records of the European NMHSs, the FluxNet stations that measure over grassland and rainfed crops are selected for this comparison as this is the prescribed surface coverage for WMO meteorological stations. The selection is partly based on earlier studies like the study of Graf et al. (2022), who analysed the energy partitioning across terrestrial ecosystems during the 2018 drought in Europe. This difference in response to heatwave in terms of heat balance between grass and forest land was described earlier by Teuling et al. (2010). The distinction in FluxNet stations for forest, grass, crop and marshes which is made by Graf et al. (2022), is used here as well. Table 3 shows the FluxNet stations over grass- and cropland which are used here for the validation of the soil moisture stations.

The variable used from the FluxNet dataset is latent heat with gapfilling (LE_F_MDS).

For the location of each selected FluxNet station, the nearest station which provides global radiation and temperature is located, the nearest rain gauge is located and the nearest station that provides SLP. These three locations can point to the same station, but as the networks for radiation, precipitation and SLP are very different in terms of density, this need not be the case. The SLP data is converted to station level pressure and is used for the psychrometric constant (Allen et al., 1994, their eq. 1.4). For some stations, this approach turns out to be rather unlucky. For instance the Czech Fluxnet station Bily Kriz is

Page: 14

Number	Author	Subject	Date
Number: 1	Author: Willem Vervoort	Subject: Highlight	Date: 25/09/2026 9:25:51 PM

Again a repeat from earlier

Number	Author	Subject	Date
Number: 2	Author: Willem Vervoort	Subject: Inserted Text	Date: 25/09/2026 9:25:37 PM

located in

Number	Author	Subject	Date
Number: 3	Author: Willem Vervoort	Subject: Inserted Text	Date: 25/09/2026 9:26:28 PM

between

Number	Author	Subject	Date
Number: 4	Author: Willem Vervoort	Subject: Inserted Text	Date: 25/09/2026 9:26:37 PM

the

https://doi.org/10.5194/essd-2026-126
Preprint. Discussion started: 13 April 2026

coupled to the meteorological station Lysa Hora but the latter station is situated on the top of a mountain at 1323 m, it is an extremely leeward spot with precipitation of ca. 1500 mm / year and therefore likely to be not too representative of the wider surroundings.

Figure 5 shows scatter diagrams comparing the FluxNet data and the calculated SWAP and SoilClim data. Table 4 provides the (Pearson) correlation coefficient, the RMSE and the relative RMSE of the comparison. In terms of correlation, the modelled ET_a is not too far apart from the observed values; both SWAP and SoilClim correlate at 0.74. The difference between the models is shown in the RMSE and rRMSE values which are for SoilClim consistently higher than for SWAP. This indicates that SWAP calculates ET_a more closely than SoilClim for these stations.

The models overestimate the ET_a values in most cases, particularly for CZ-BK2 and FI-Jok. For station NL-Hor, some underestimation for the higher values of measured ET_a is present. For station IT-MBo, the complex topography and the difficulty in finding meteorological stations close to the observational site have resulted in a fairly large scatter.

Unfortunately, a comparison between observed and modelled ET_a for the three Spanish stations was not possible. The timeseries of ET_a for these stations showed a distinctly different behaviour in the dry season than what was modelled. The much higher observed values of ET_a in comparison to the modelled figures indicate an additional water source, like irrigation, than what is provided as rain in the nearby rain gauges.

The resemblance between the observed ET_a data for the selected Fluxnet stations and the modelled data using the nearest meteorological station is generally fair. The correlation is high, and where the correlation between observed and modelled ET_a is low, the similarity between the SWAP and SoilClim output is strong. This suggests that the poor correlation might be attributed to meteorological data that are not sufficiently representative for the FluxNet site.

Figure 6 shows the average seasonal cycle, the median and the range of values between the 25th and 75th percentiles of ET_a for four selected FluxNet stations and the SWAP and SoilClim simulations. The daily values of the median and percentiles are determined by combining all available data and by using a 31-day window centered around each calendar day to further increase sample size and smooth the seasonal cycle. These plots show that the seasonal cycle is reasonably well captured in the Austrian station, although there is some overestimation in the early part of the year. The miss-representation of the amplitude in summer for the Czech station is obvious in Figure 5 as well and is related to the awkward location of the meteorological station used for this comparison. Nevertheless, these figures confirm that both soil moisture models produce similar seasonal cycles in ET_a although differences with observed values exist.

4.2 Verification against ISMN

The International Soil Moisture Network (ISMN, Robock et al. (2000)) maintains a global in-situ soil moisture database. This initiative started as a community effort, funded by the European Space Agency, to serve as a centralised data hosting facility for globally available *in situ* soil moisture measurements (Dorigo et al., 2021, 2011). Similar to the comparison against FluxNet stations, ISMN stations for use in the validation are selected that are located in grassland or rainfed crop. In addition, the time series should have at least five years of data.

Page: 15

Number	Author	Subject	Date	Text
Number: 1	Author: Willem Vervoort	Subject: Cross-Out	Date: 25/09/2026 9:27:12 PM	
Number: 2	Author: Willem Vervoort	Subject: Inserted Text	Date: 25/09/2026 9:27:09 PM	not
Number: 3	Author: Willem Vervoort	Subject: Inserted Text	Date: 25/09/2026 9:27:41 PM	the correlation for both
Number: 4	Author: Willem Vervoort	Subject: Inserted Text	Date: 25/09/2026 9:27:44 PM	is
Number: 5	Author: Willem Vervoort	Subject: Cross-Out	Date: 25/09/2026 9:27:51 PM	
Number: 6	Author: Willem Vervoort	Subject: Inserted Text	Date: 25/09/2026 9:27:55 PM	are

https://doi.org/10.5194/essd-2026-126
Preprint. Discussion started: 13 April 2026

station	SWAP corr/RMSE/rRMSE	SoilClim corr/RMSE/rRMSE
AT-Neu	0.85/0.76/0.52	0.84/1.02/0.68
Patscherkofel (AT) / Patscherkofel (AT)		
BE-Lon	0.74/0.84/0.78	0.81/1.01/0.93
Maastricht (NL) / Perwez (BE))		
CH-Oe1	0.86/0.69/0.41	0.85/0.82/0.49
Basel Binningen (CH) / Basel Binningen (CH)		
CZ-BK2	0.77/1.19/1.88	0.75/1.54/2.43
Lysa Hora (CZ) / Lysa Hora (CZ)		
DE-Geb	0.77/0.81/0.74	0.77/0.91/0.83
Erfurt-Bindersleben (DE) / Erfurt-Bindersleben (DE)		
DE-Gri	0.83/0.81/0.76	0.85/1.07/1.00
Nossen (DE) / Tharandt-Grillenburg (DE)		
FI-Jok	0.76/0.98/1.28	0.78/1.41/1.84
Jokioinen Jokioisten Observatorio (FI) / Jokioinen Jokioisten Observatorio (FI)		
FI-Let	0.74/0.83/0.67	0.76/1.04/0.83
Jokioinen Jokioisten Observatorio (FI) / Jokioinen Jokioisten Observatorio (FI)		
FR-Gri	0.68/0.85/0.61	0.69/0.95/0.68
Trappes (FR) / Trappes (FR)		
IT-MBo	0.59/1.19/0.76	0.57/1.27/0.81
Obergurgl (AT) / Aldeno (IT)		
NL-Hor	0.50/1.24/0.96	0.50/1.36/1.05
De Bilt (NL) / Loenen aan de Vecht (NL)		

Table 4. Pearson correlation coefficients, RMSE (in [mm/day]) and rRMSE values between ET_a data from selected FluxNet stations and simulated ET_a values from SWAP and SoilClim. Each row has a line added with the nearest meteorological station (providing temperature and global radiation) and the nearest rain gauge respectively.

The depth levels for which soil moisture is provided is not standardized in the ISMN repository. Motivated by this, we matched the calculated soil moisture levels at the surface and rootzone levels from SWAP and SoilClim to a range of levels which were provided by the selected ISMN stations. Measured soil moisture at the upper 10 cm in the soils (and in a few cases in the upper 20 cm) are compared against the calculated surface soil moisture. When observations are available until a depth of 50 cm, then these are related to the root zone.

Table A1 shows the correlation, RMSE and rRMSE values of the comparison of observed soil moisture values at surface, rootzone and deep levels from the ISMN stations with the simulations of SWAP and SoilClim forced by the observations from the nearest ECA&D station. The general picture that emerges from this table is that correlations between modelled and observed soil moisture are on average around 0.55. For the surface layer, the average correlations between the ISMN series and the SWAP or SoilClim output are 0.55 and 0.52 respectively. For the root zone, these numbers are 0.51 and 0.64 for SWAP and SoilClim respectively. When the correlation is poor between ISMN observed soil moisture and a model, it will be poor for the other model as well. This suggests that for the poorly correlating stations, the selection of the nearest meteorological

Page: 16

Number: 1	Author: Willem Vervoort	Subject: Highlight	Date: 25/09/2026 9:39:59 PM

This is a bit of an awkward table to read. Why don't you split these columns up?

Number: 2	Author: Willem Vervoort	Subject: Inserted Text	Date: 25/09/2026 9:42:37 PM

W

https://doi.org/10.5194/essd-2026-126
Preprint. Discussion started: 13 April 2026

Open Access Earth System Science Data Discussions

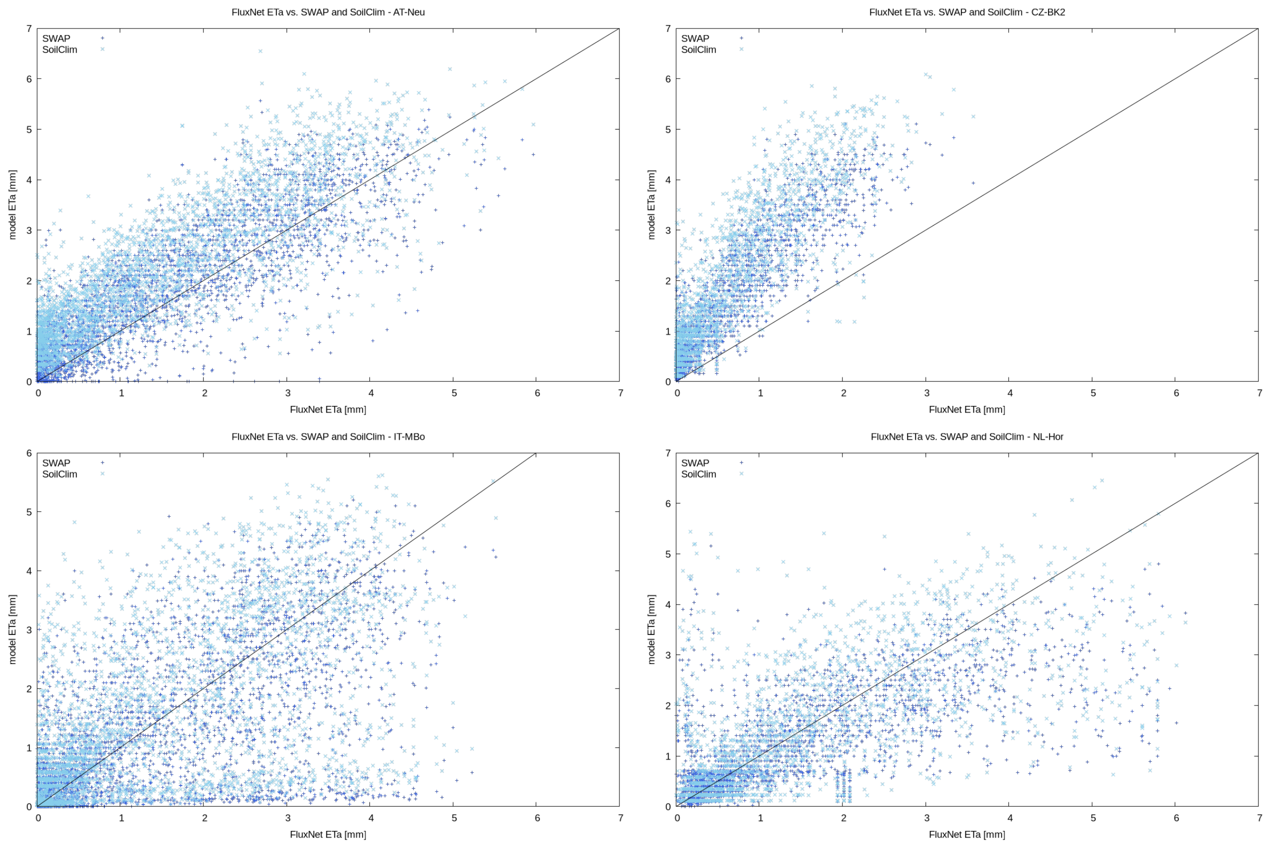

Figure 5. Scatterplots with oberved ET_a values from the FluxNet station on the horizontal axis and modelled ET_a values from SWAP (dark blue) and SoilClim (light blue), for FluxNet stations AT-Neu, CZ-BK2, IT-MBo and NL-Hor.

station is not optimal. For the surface and root-zone layers, the SWAP model has a consistently lower rRMSE than SoilClim which mirrors the observation from the comparison with ET_a of the FluxNet stations that SWAP models the vertical motion of moisture in and out of the soil more realistically. However, for the deep layer, the SoilClim model has a lower rRMSE.

Figure 7 shows examples from ISMN stations San Pietro Capofiume, Hoal_01 and Berzeme showing scatter diagrams between observed ISMN values and modelled SWAP and SoilClim values for the surface layer and root zone. The clustering of point at the top-end of the diagrams shows that the SoilClim model tends to saturate earlier than what is observed and what is modelled by SWAP. For the Berzeme station, the SoilClim shows a quicker depletion of soil water from the surface layer.

The average seasonal cycle in soil water content for the rootzone is shown in fig. 8. For station San Pietro Capofiume, the two models simulations agree fairly well but the saturation in soil moisture in late winter and early spring by SoilClim is obvious. Both models show a stronger depletion in soil moisture in summer than the observations. For station Hoal_01, the medians in the seasonal cycles are not too far apart - except for SoilClim's saturation in late winter, early spring and a decrease in SWAP's soil moisture for mid-spring. The resemblance in seasonal cycle for station Berzeme between observations and model results

Number: 1 Author: Willem Vervoort Subject: Highlight Date: 25/09/2026 9:40:39 PM
Figure labels are too small, axis titles, axis labels etc

Author: Willem Vervoort Subject: Sticky Note Date: 25/09/2026 9:41:32 PM
symbols are also too small and why both blue? Make one red and one blue? Think about colour blind people

https://doi.org/10.5194/essd-2026-126
Preprint. Discussion started: 13 April 2026

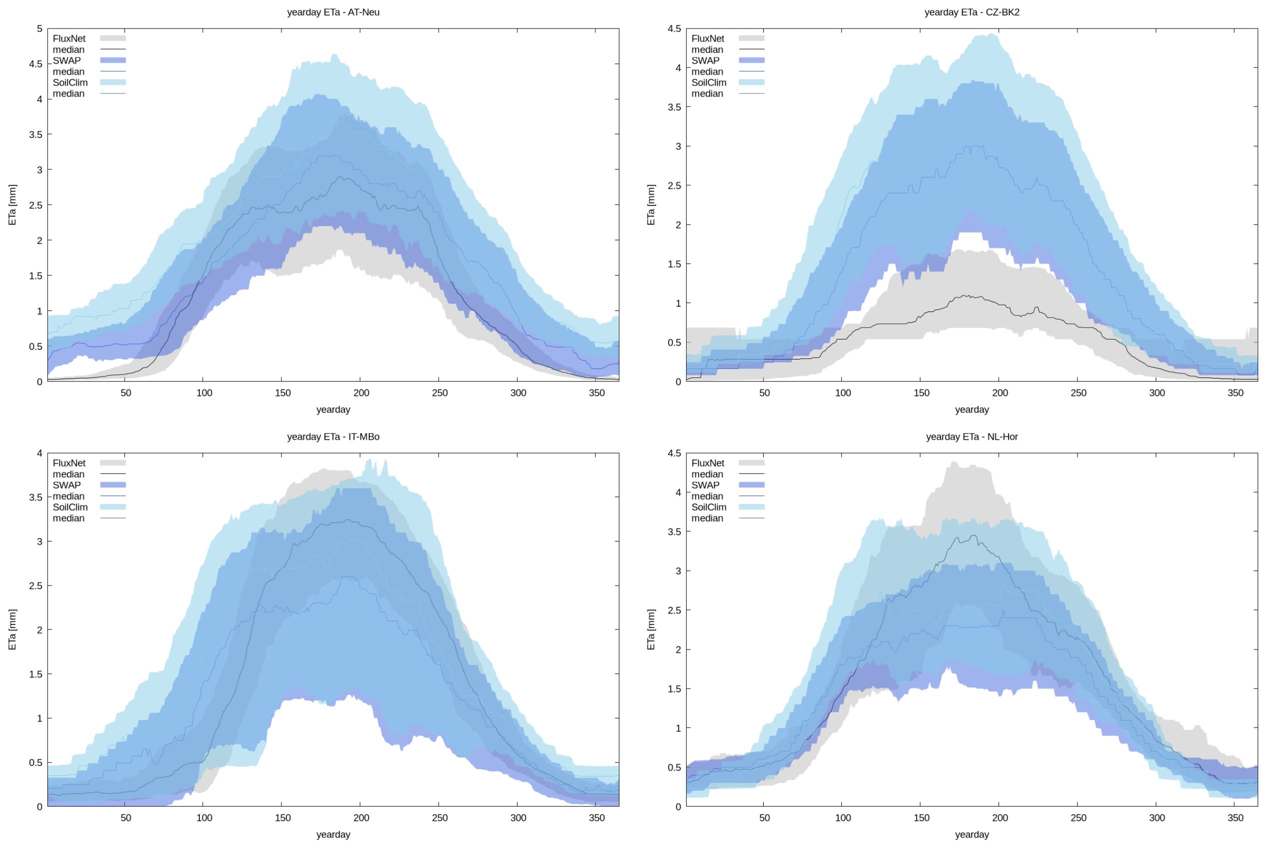

Figure 6. Comparison of the average seasonal cycle of ET_a as provided by the observation (gray), SWAP (dark blue) and SoilClim (light blue), for FluxNet stations AT-Neu, CZ-BK2, IT-MBo and NL-Hor. The lines represent the median values and the shaded areas give the range between the 25th and 75th percentiles.

is clearly the worst of the subset shown in this figure. While the model results agree fairly well, they overestimate the observed soil water content.

5 Climate Assessment

5.1 The historical perspective

The length of the records in ECA&D allows for putting recent extreme events into a fairly long historical perspective without losing the pan-European scope. An example is the relatively recent drought of 2018 which is often compared in terms of severity with the drought of 1976. Figure 9 shows the number of moisture stressed days (SMI0) in summer for both these years. It shows that for both years the epicentre of the drought was in central Europe with the complete season experiencing moisture stress. But the 1976 drought had a larger extent to the west, over the Netherlands into England, to the south-west over

Number: 1 Author: Willem Vervoort Subject: Highlight Date: 25/09/2026 9:44:12 PM

See comment with other figures on the design of the figure

https://doi.org/10.5194/essd-2026-126
Preprint. Discussion started: 13 April 2026

Open Access Earth System Science Data Discussions

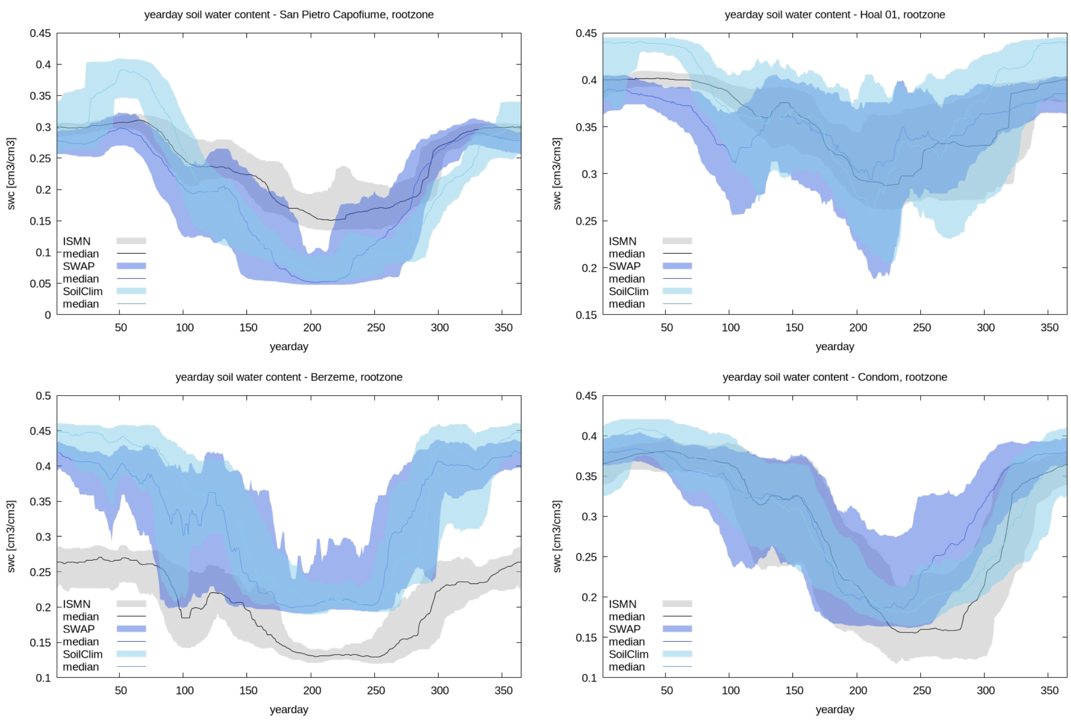

Figure 8. Comparison of the average seasonal cycle of soil water content in the rootzone as provided by the observation (gray), SWAP (dark blue) and SoilClim (light blue), for ISMN stations San Pietro Capofiume, Hoal_01, Berzeme and Condom. The lines represent the median values and the shaded areas give the range between the 25th and 75th percentiles.

reduced germination. Soils that are too wet are impossible to access, disrupting the harvesting schedule (Wigley and Atkinson, 1977). A recent example of a very wet spring was 2024 in western Europe (C3S, 2024). Figure 10 (upper panel) shows the number of very moisture saturated days in De Bilt and Brno in spring. It shows that 2024 is indeed high, but ranks third since 1961 (after 1983 and 1977). This figure also shows that such moisture saturated days in spring are rare in Brno.

This contrasts with a time series of very moisture stressed days (AW10p). Figure 10 (lower panel) shows that most of the springs in Brno, and nearly all in De Bilt, lack such very dry situations. However, the recent decades in Brno show an increase in springs with very moisture stressed days, although none reach the level of the 1974 spring. In the period since 1960, De Bilt experienced 10 very moisture stressed days in the 2002 spring.

5.2 First assessment of climatology and trends

In an interesting 20-year old publication, Robock et al. (2005) analysed 141 soil moisture records from the Ukraine spanning the period from 1958 to 2002. In that publication, no summer desiccation was (yet) found which was the initial hypothesis given that the climate projections available at the time indicated gradually more droughty summer conditions. Instead the observations indicated an overall positive (wetting) trend over this period with the trend leveling off in the last two decades of

Page: 20

Number: 1 Author: Willem Vervoort Subject: Highlight Date: 26/09/2026 10:30:19 AM

I am not sure this should be part of a "data paper".

https://doi.org/10.5194/essd-2026-126
Preprint. Discussion started: 13 April 2026

Open Access Earth System Science Data Discussions

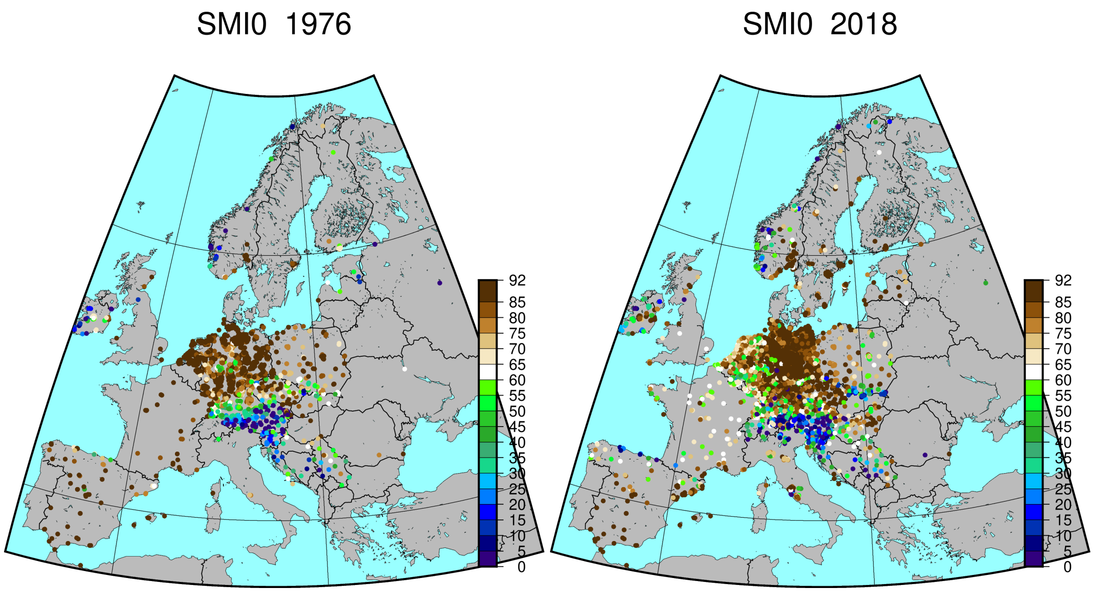

Figure 9. Comparison of the SMI0 index for summer 1976 (left) and 2018 (right) which were both very dry summers in NW Europe.

the period they observed. In this section, a first brief look at the climatology of the four proposed indices is taken and an initial trend analysis is made to assess if summer desiccation is observed over Europe.

Figure 11 shows the spring and summer 1991-2020 climatology of the indices SMI and SMI0. The climatological SMI spring values are without exception positive, meaning that for an average spring and on average over the season, moisture stress will not be an issue. It is interesting that in Ireland, Norway and northern Scandinavia the SMI values are relatively close to zero. In summer, this picture changes considerably. The areas that remain (for an average summer and on average over the season) free of moisture stress are Scandinavia, Ireland and northern UK, the low countries, the greater Alpine area and the Caucasus. A few stations in the central parts of the Balkan peninsula also manage to stay green. A clear gradient to the south-west into Spain and into the east toward Poland of more severe moisture stress is seen.

The spring climatology of the number of days with moisture stress (SMI0) shows that no or hardly any days are observed for much of the domain, but Spain, southern France, the Balkan and Poland show areas where already a significant number of days fall below the SMI=0 threshold. For summer, the area with a large number of days experiencing moisture stress is quite large and only the highest parts of the Alps, Ireland, the northern UK and Norway have a more modest number of moisture stress days.

Figure 12 shows the trends over the spring and summer season in the period 1979-2023 for the number of moisture stress days (SMI0) and the Soil Moisture Index (SMI) for spring and summer respectively. The trends are calculated using the Theil-Sen regression and the significance of the trend is determined from the Mann-Kendall test. The trends that pass the 95% significance level are plotted as black-rimmed circles. The trend map for spring SMI0 generally shows an increase in the number of moisture stressed days over Europe, with the strongest trends in northern Spain, France, Central Europe and

Number: 1 Author: Willem Vervoort Subject: Highlight Date: 26/09/2026 10:26:34 AM

Why do you call this "climatology"? I think it is just a spatial map of soil moisture stress by season?

https://doi.org/10.5194/essd-2026-126
Preprint. Discussion started: 13 April 2026

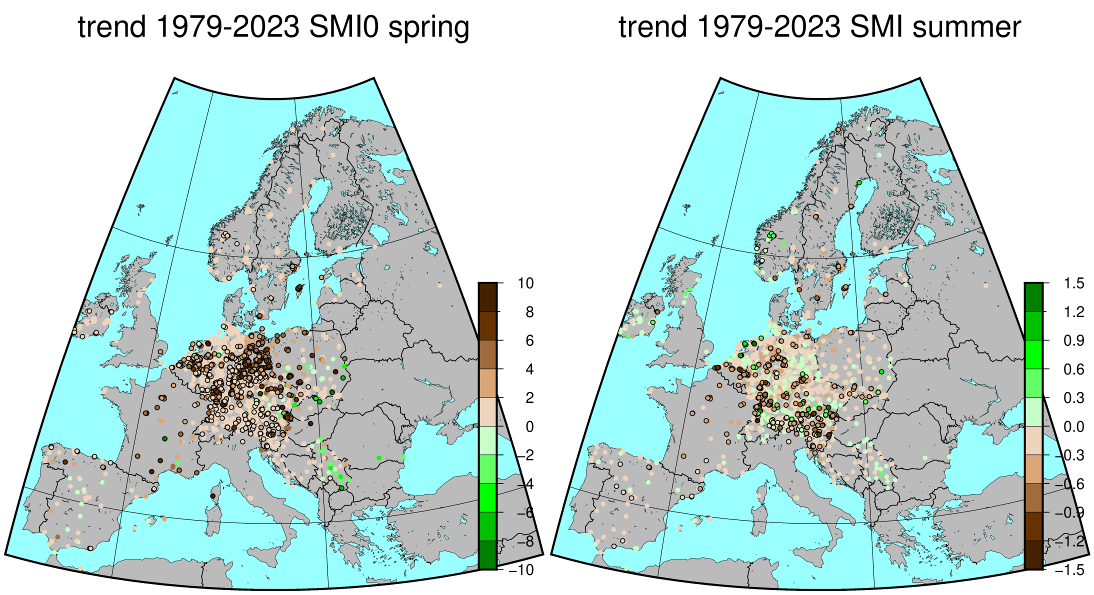

Figure 12. Trends over the 1979-2023 period for spring SMI0 (left) and summer SMI (right). The unit of the trends is index/decade for SMI and days/decade for SMI0. Trends that reach the 95% significance level are plotted with a black rimmed circle.

causing earlier soil moisture decrease during spring with a higher probability of early soil moisture deficit, which is another example of an important soil moisture - temperature feedback (Ionita et al., 2017; Řehoř et al., 2024).

In the dataset presented here, the challenge put forward by Seneviratne et al. (2010) to optimize the use of existing observed information for soil moisture - climate research is taken up. A pan-European dataset with *in situ* observations of meteorological parameters (ECA&D) is used to calculate daily soil moisture values for the surface and the rootzone, using a state-of-the art soil dynamics model. This has resulted in a dataset of nearly 6000 stations (status July 2025) for which a set of four climate impact indices is calculated for monitoring purposes. The aim is to update the dataset regularly along with the regular updates of ECA&D. However, the bottleneck of the current set-up are the lack of *in situ* global radiation and/or sunshine duration data. With over 20,000 rain gauges in the dataset, there is some potential to increase the coverage in some places and make the coverage denser in others. Increasing the number of long-running sunshine duration records - there are now about 400 stations that provide sunshine duration data back to 1950 provides the potential to extend the historical perspective further back in time.

An important motivation to focus on short-term variations in soil water content is the recent insight that flash droughts are more common than previously thought - also in Europe (Shah et al., 2022; Yuan et al., 2023). Flash droughts are a recently recognized type of extreme event distinguished by sudden onset and rapid intensification of drought conditions with severe impacts (Lesinger and Tian, 2022). They unfold on sub-seasonal-to-seasonal timescales (weeks to months) (Pendergrass et al., 2023; Yuan et al., 2023) and can develop into severe droughts in a few weeks. In addition to precipitation deficits, flash droughts are also caused by abnormally high evapotranspiration that rapidly depletes the soil water. The intermittent nature and rapid

Page: 24

Number: 1 Author: Willem Vervoort Subject: Highlight Date: 26/09/2026 10:32:16 AM

Do you really need all this in your data paper? Seems not really related to the data, the quality of the data or the integration of the data