levy@ceh.ac.uk]PeterLevy

Daily soil moisture maps for the UK (2016-2023) at 2 km resolution

[p[1]
[1]Centre for Ecology and Hydrology, Bush Estate, Penicuik, Midlothian, EH26 0QB, United Kingdom
Correspondence: Peter Levy (plevy@ceh.ac.uk)

1 Abstract

Soil moisture is important in many hydrological and ecological processes. However, data sets which are currently available have issues with accuracy and resolution. To translate remotely-sensed data to an absolute measure of soil moisture requires mapped estimates of soil hydrological properties and estimates of vegetation properties, and this introduces considerable uncertainty. We present an alternative methodology for producing daily maps of soil moisture over the UK at 2-km resolution ("SMUK"). The method is based on a simple empirical model, calibrated with five years of daily data from cosmic-ray neutron sensors at ~40 sites across the country. The model is driven by precipitation, humidity, a remotely-sensed "soil water index" satellite product, and soil porosity. The model explains around 70 % of the variance in the daily observations. The spatial variation in the parameter describing the soil water retention (and thereby the response to precipitation) was estimated using daily water balance data from ~1200 catchments with good coverage across the country. The model parameters were estimated by Bayesian calibration using a Markov chain Monte Carlo method, so as to characterise the posterior uncertainty in the parameters and predictions. We found that the simple model could emulate the behaviour of a more complex process-based model. Given the high resolution of the inputs in time and space, the model can predict the very detailed variation in soil moisture which arises from the sporadic nature of precipitation events, including the small-scale and short-term variations associated with orographic and convective rainfall. Predictions over the period 2016 to 2023 demonstrated realistic patterns following the passage of weather fronts and prolonged droughts. The model has negligible computation time, and inputs and predictions are updated daily, lagging approximately one week behind real time.

2 Introduction

Soil moisture is an important controlling variable in many hydrological and ecological processes. In modelling flood risk, the soil saturation status is important in predicting runoff and the catchment response to precipitation (Ahlmer et al., 2018; Chifflard et al., 2018). It affects plant growth, crop yield, and irrigation needs. It also determines the aerobic status of the soil, and thereby the balance between different microbial processes such as nitrification and denitrification, and methanogenesis and methanotrophy (and thus the emissions of the greenhouse gases nitrous oxide and methane, (Davidson, 1992; Zou et al., 2022). The availability of soil moisture data at high resolution in near-real time therefore has many potential applications in environmental science.

Here we describe data from a model for mapping soil moisture across the UK, based on the highly accurate COSMOS network data, but incorporating other data sources where these improve predictions. Soil moisture predictions carry considerable uncertainty, and we aim to include an appropriate characterisation of our predictive uncertainty.

3 Collection/generation methods

3.1 Data sources

Here we briefly describe the data sources used as inputs to the model.

3.1.1 COSMOS network

COSMOS-UK is a network of sites equipped with cosmic-ray soil moisture sensors (Evans et al., 2016; Stanley et al., 2019). The cosmic-ray measurement principle utilizes naturally occurring fast neutrons generated by cosmic rays. These neutrons interact with water molecules in soil, and the back-scattered flux of slow neutrons is proportional to the soil water content. The neutron detectors are installed just above the ground, so there is no disturbance to the soil structure. A single sensor measures a circular footprint with a radius of approximately 100-240 m, and is sensitive to soil moisture in the top 15-30 cm, decreasing exponentially with radial distance (Köhli et al., 2015). A full suite of meteorological measurements are also made at the sites, including all the variables necessary to calculate potential evapotranspiration. The network provided 61225 observations of daily mean soil moisture, after filtering for data quality and missing values.

3.1.2 Soils data

Soil porosity was estimated using different data sources for Great Britain (GB) and the rest of the domain (Ireland and the edge of continental Europe). For GB, the UK Countryside Survey has collected bulk density ρ and soil organic carbon measurements within several hundred 1-km squares since 2007 (Emmett, 2010). Soil organic carbon has been interpolated on to a 1-km grid covering GB using a generalised additive model (Thomas et al., 2020), and we used this to predict bulk density, using the relationship between these variables established from the sample data. Porosity was calculated as $1 - \rho/2.65$, where 2.65 is the standard value for particle density of the solid fraction in g cm^3. For the rest of the domain, the same procedure was used, but using mapped estimates of soil carbon from (de Brogniez et al., 2015). In principle, mapped estimates of bulk density from (Ballabio et al., 2016) could have been used to estimate porosity more directly, but these did not reflect the range in porosity seen in organic soils in the UK and Ireland, where values over 70 % are typical. Other variables were obtained from mapped estimates based on the European LUCAS data (Ballabio et al., 2016, 2019).

3.1.3 Meteorological data

The main observations of precipitation are ultimately derived from the Met Office NIMROD system (Met Office, 2003), which are assimilated into the Met Office UK Atmospheric High-Resolution Model for weather forecasting [NWP-UKV; Met Office (2016)]. The NIMROD system is based on a network of 15 C-band rainfall radars covering the UK (Harrison et al., 2000). This provides 2-km resolution composite data for precipitation rate every five minutes, from 2004 to the near-present. The assimilation system combines processed radar and satellite data, together with surface reports into the UK Met Office Numerical Weather Prediction (Milan et al., 2020). Bias corrections are well developed to improve absolute accuracy of precipitation totals, although there may be further scope for improvement (Yu et al., 2020). The system provides the best combination of accuracy and spatial and temporal resolution available. Other meteorological variables for model development were also taken from the Met Office UK Atmospheric High-Resolution Model for weather forecasting.

3.1.4 SCAT-SAR

The Copernicus Global Land Services Scatterometer-Synthetic Aperture Radar Soil Water Index (SCAT-SAR SWI) is a fusion of two satellite products (Bauer-Marschallinger et al., 2018): it uses high-resolution synthetic-aperture radar (SAR) surface soil moisture (SSM) data generated by the Sentinel-1 satellite mission, and combines it with the Advanced Scatterometer (ASCAT) SSM data from the MetOp satellites. The resulting product achieves both, high temporal and spatial resolution (daily, at 1 km over Europe), while providing improved reliability and accuracy, and is available from 2015 to the present day (Bauer-Marschallinger et al., 2018). In the SCATSAR-SWI algorithm, the ASCAT SSM and SAR SSM values are given a weighting determined by the signal to noise ratio, specific to the time series at each grid point. Exponential moving averaging is then applied, similar to that described above for precipitation, except that a slightly different formulation is used, and a range of λ values are used to vary the weight placed on current versus past data. We choose an intermediate value ($T = 15$ in their notation). The time series is normalised relative to the minimum and maximum SSM values recorded at each grid point; the scaling is therefore somewhat arbitrary and varies from location to location.

3.1.5 NRFA data

The UK National River Flow Archive (NRFA) contains daily data on river flow for 1212 gauging locations covering catchments across Great Britain, along with corresponding catchment precipitation. Records span the period from the 1970s to the present day, although this varies among gauging sites. The gauging sites are classified into ~100 hydrometric areas, representing either single catchments or neighbouring hydrologically similar catchments. The data were obtained via the R package rnrfa (Vitolo et al., 2016).

3.2 Model development

We developed a simple statistical model which explains the temporal and spatial variability in the COSMOS observations of soil moisture. We used a hierarchical modelling approach to account for the non-independence of data from the same sites (Bates et al., 2015). An exponential moving average (EMA) was applied to precipitation data as described in (Levy and COSMOS-UK, 2023). The model predicts volumetric soil moisture at time t and site s, based on EMA precipitation, vapour pressure deficit D and the SCAT-SAR Soil Wetness Index I as:

$$\begin{aligned}
\theta'_{ts} &= \beta_0 + \beta_{\mathrm{P}}\langle P\rangle_{ts} + \beta_{\mathrm{D}} D_{ts} + \beta_{\mathrm{I}} I_{ts} + b_s + \epsilon_{ts} \\
b_s &\sim N(0, \Psi) \\
\epsilon_{ts} &\sim N(0, \sigma^2) \\
\theta_{ts} &= \theta_{max,s} - \theta'_{ts}
\end{aligned} \tag{1}$$

where $\theta_{\max}$ is the soil moisture content at saturation and θ' is the deviation from this. β_{D} and β_{I} are the additional regression coefficients, and b_s is the local deviation at site s (a so-called "random intercept" term which accounts for the within-site correlation of residuals). The local deviations b are assumed to be independently drawn from a normal distribution with mean zero and variance Ψ.

To improve the representation of spatial variation in β_{P}, we used the NRFA data on river discharge and catchment precipitation. Whilst we cannot infer absolute soil moisture values in the catchment, we can estimate the *change* in soil moisture from these data by calculating the quantity:

$$\widehat{\Delta S} = P - Q - E_{\mathrm{pot}} \tag{2}$$

closely correlated with $\Delta\theta = \Delta S/z$, but with evapotranspiration estimated at its potential rate E_{pot} rather than the actual rate. The $\widehat{\text{hat}}$ symbol denotes that this is an estimator of ΔS, rather than the true value.

For each of the 1212 NRFA catchments, we calculated daily values of $\widehat{\Delta S}$, the exponential-moving-averaged precipitation $\langle P\rangle$, and its day-to-day difference $\Delta\langle P\rangle$. The EMA filter linearises the relationship so that we can derive the slope representing how the change in soil water storage $\widehat{\Delta S}$ responds to change in (EMA) precipitation $\Delta\langle P\rangle$. For each of the 1212 NRFA catchments, we calculated this slope by linear regression i.e. fitting the linear model:

$$\widehat{\Delta S} = c + m\Delta\langle P\rangle. \tag{3}$$

The slope parameter m is closely related to the β_{P} parameter in Eq. (1), as it relates the change in soil water storage to the change in precipitation. Absolute values are not comparable as the effective depth z is unknown, but the spatial patterns in both will be similar. We therefore use the spatial variation in m from the 1212 NRFA catchments to estimate the relative spatial variation in β_{P}. For each NRFA site s, we calculate the value:

$$\beta_{\mathrm{P},s} = \beta_{\mathrm{P}} \frac{m_s}{\bar{m}} \tag{4}$$

where β_{P} is the global (or "fixed effect") derived from Eq. (1), m_s is the slope derived from Eq. (3) for site s, and $\bar{m}$ is the mean value of m across all 1212 sites.

3.3 Spatial extrapolation of β_{P}

The above equations provide values of $\beta_{\text{P},s}$ for 1212 sites, but we want to make predictions on a 2-km grid covering the UK. This is a common problem in the area of geostatistics, where we have limited measurement locations and we need to estimate values at many unobserved locations. The standard approach is known as kriging, a form of weighted local averaging, where the estimates of values at unrecorded places are weighted averages of the observations. The kriging weights are calculated on the basis of the "semivariogram", which quantifies the form of the increasing variance (or decreasing covariance) between pairs of points as the distance between them increases. Various mathematical models are used to describe the covariance as a function of this distance; the Matérn model is a common choice for its flexibility (Pardo-Iguzquiza and Chica-Olmo, 2008) in which the covariance is given by

$$C_\nu(d) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{d}{\phi}\right)^\nu K_\nu \left(\frac{d}{\phi}\right) \tag{5}$$

where σ^2 is the variance, ϕ is a range or scale parameter, ν is a shape parameter, Γ is the gamma function, and K_ν is the modified Bessel function. Graphically, this shows the scale at which values are highly correlated, and how this changes with spatial scale. Prediction at a new location is based on all the observations (within some cut-off distance), each weighted according to the degree of correlation at that distance predicted by the semivariogram. Kriging has been shown to be optimal in the sense that it provides estimates with minimum variance and without bias (in the long-term statistical sense) (Cressie, 1990). Here, we applied the method with the Matérn covariance model, but in a Bayesian approach described below, using the geoR package for the R statistical software (Ribeiro Jr et al., 2020).

3.4 Quality control

Quality control of the COSMOS data, on which the model is calibrated, is described in (Stanley et al., 2019). In the modelling, we used a Bayesian approach via Hamiltonian Markov chain Monte Carlo (MCMC) sampling, an efficient iterative algorithm for calculating numerical approximations of multi-dimensional integrals (Hoffman and Gelman, 2014; Betancourt, 2017; Bürkner, 2018). This approach explicitly attempts to quantify the probability distributions of the parameters, given the data and any prior information. This has the advantages that it provides a robust means of estimating uncertainty on the parameters and predictions.

We also apply the Bayesian approach to the spatial extrapolation, so we recognise that the semivariogram is not a fixed, known quantity, but a geostatistical model with uncertain parameters. In brief, this means we account for the uncertainty in the variogram model, and represent each of the parameters as a probability distribution. Rather than assuming the variance is known, we calculate the posterior distribution of the parameters, given the observed data, and sample many realisations of these to represent the uncertainty. We used the Bayesian kriging algorithm available in the geoR package (Ribeiro Jr. and Diggle, 2001) as described above, with uniform (uninformative) priors for the σ, ϕ and ν parameters. The algorithm makes some simplifying assumptions to speed computation time, such as discretising the prior parameter distributions.

3.4.1 Missing data

The data set from the Met Office NWP-UKV model held on The Centre for Environmental Data Analysis (CEDA, https://archive.ceda.ac.uk/ has some periods with missing data. In these periods, we have not attempted to fill in the gaps, so the soil moisture data set is not complete. Files for the days where no soil moisture can be calculated have been removed from the deposited data set.

4 Details of data structure

The data are stored in TIF files with a raster data structure for a grid covering the UK and Ireland. The grid is in the OS GB coordinate system, which uses a Transverse Mercator projection. The grid has 704 rows and 548 columns, and has a 2-km horizontal resolution. The header information for an example file for the 1st July 2023 is shown below.

```
## class      : RasterLayer
## dimensions : 704, 548, 385792  (nrow, ncol, ncell)
## resolution : 2000, 2000  (x, y)
## extent     : -239000, 857000, -185000, 1223000  (xmin, xmax, ymin, ymax)
## crs        : +proj=tmerc +lat_0=49 +lon_0=-2 +k=0.9996012717 +x_0=400000 +y_0=-100000 +el
## source     : dt_smuk_2023-07-01.tif
## names      : dt_smuk_2023.07.01
## values     : 0.1574903, 0.8015409  (min, max)
```

5 Nature and Units of recorded values

The data values are volumetric soil moisture, in units of m^3 water per m^3 soil volume. Because the model is calibrated on the COSMOS data, the depth corresponds to a variable layer of the top soil which itself depends on the soil moisture - shallower in wet soil and deeper in dry soil, varying over a range of 15 to 30 cm, decreasing exponentially with radial distance (Köhli et al., 2015).

Competing interests.

References

Ahlmer, A.-K., Cavalli, M., Hansson, K., Koutsouris, A. J., Crema, S., and Kalantari, Z.: Soil Moisture Remote-Sensing Applications for Identification of Flood-Prone Areas along Transport Infrastructure, Environmental Earth Sciences, 77, 533, https://doi.org/10.1007/s12665-018-7704-z, 2018.

Ballabio, C., Panagos, P., and Monatanarella, L.: Mapping Topsoil Physical Properties at European Scale Using the LUCAS Database, Geoderma, 261, 110–123, https://doi.org/10.1016/j.geoderma.2015.07.006, 2016.

Ballabio, C., Lugato, E., Fernández-Ugalde, O., Orgiazzi, A., Jones, A., Borrelli, P., Montanarella, L., and Panagos, P.: Mapping LUCAS Topsoil Chemical Properties at European Scale Using Gaussian Process Regression, Geoderma, 355, 113 912, https://doi.org/10.1016/j.geoderma.2019.113912, 2019.

Bates, D., Mächler, M., Bolker, B., and Walker, S.: Fitting Linear Mixed-Effects Models Using Lme4, Journal of Statistical Software, Articles, 67, 1–48, https://doi.org/10.18637/jss.v067.i01, 2015.

Bauer-Marschallinger, B., Paulik, C., Hochstöger, S., Mistelbauer, T., Modanesi, S., Ciabatta, L., Massari, C., Brocca, L., and Wagner, W.: Soil Moisture from Fusion of Scatterometer and SAR: Closing the Scale Gap with Temporal Filtering, Remote Sensing, 10, 1030, https://doi.org/10.3390/rs10071030, 2018.

Betancourt, M.: A Conceptual Introduction to Hamiltonian Monte Carlo, arXiv:1701.02434 [stat], 2017.

Bürkner, P.-C.: Advanced Bayesian Multilevel Modeling with the R Package Brms, The R Journal, 10, 395–411, 2018.

Chifflard, P., Kranl, J., zur Strassen, G., and Zepp, H.: The Significance of Soil Moisture in Forecasting Characteristics of Flood Events. A Statistical Analysis in Two Nested Catchments, Journal of Hydrology and Hydromechanics, 66, 1–11, https://doi.org/10.1515/johh-2017-0037, 2018.

Cressie, N.: The Origins of Kriging, Mathematical Geology, 22, 239–252, https://doi.org/10.1007/BF00889887, 1990.

Davidson, E. A.: Sources of Nitric Oxide and Nitrous Oxide Following Wetting of Dry Soil, Soil Science Society of America Journal, 56, 95–102, https://doi.org/10.2136/sssaj1992.03615995005600010015x, 1992.

de Brogniez, D., Ballabio, C., Stevens, A., Jones, R. J. A., Montanarella, L., and van Wesemael, B.: A Map of the Topsoil Organic Carbon Content of Europe Generated by a Generalized Additive Model, European Journal of Soil Science, 66, 121–134, https://doi.org/10.1111/ejss.12193, 2015.

Emmett, B.: Countryside Survey: Soils Report from 2007. Technical Report No. 9/07. NERC/Centre for Ecology & Hydrology 192pp. (CEH Project Number: C03259)., Tech. rep., 2010.

Evans, J. G., Ward, H. C., Blake, J. R., Hewitt, E. J., Morrison, R., Fry, M., Ball, L. A., Doughty, L. C., Libre, J. W., Hitt, O. E., Rylett, D., Ellis, R. J., Warwick, A. C., Brooks, M., Parkes, M. A., Wright, G. M. H., Singer, A. C., Boorman, D. B., and Jenkins, A.: Soil Water Content in Southern England Derived from a Cosmic-Ray Soil Moisture Observing System – COSMOS-UK, Hydrological Processes, 30, 4987–4999, https://doi.org/10.1002/hyp.10929, 2016.

Harrison, D. L., Driscoll, S. J., and Kitchen, M.: Improving Precipitation Estimates from Weather Radar Using Quality Control and Correction Techniques, Meteorological Applications, 7, 135–144, https://doi.org/10.1017/S1350482700001468, 2000.

Hoffman, M. D. and Gelman, A.: The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo, Journal of Machine Learning Research, 15, 1593–1623, 2014.

Köhli, M., Schrön, M., Zreda, M., Schmidt, U., Dietrich, P., and Zacharias, S.: Footprint Characteristics Revised for Field-Scale Soil Moisture Monitoring with Cosmic-Ray Neutrons, Water Resources Research, 51, 5772–5790, https://doi.org/10.1002/2015WR017169, 2015.

Levy, P. and COSMOS-UK, t.: Mapping Soil Moisture across the UK: Assimilating Cosmic-Ray Neutron Sensors, Remotely-Sensed Indices, Rainfall Radar and Catchment Water Balance Data in a Bayesian Hierarchical Model, EGUsphere, pp. 1–24, 2023.

Met Office: 1-Km Resolution UK Composite Rainfall Data from the Met Office Nimrod System, https://catalogue.ceda.ac.uk/uuid/27dd6ffba67f667a18c62de5c3456350, 2003.

Met Office: NWP-UKV: Met Office UK Atmospheric High Resolution Model Data, https://catalogue.ceda.ac.uk/uuid/f47bc62786394626b665e23b658d38 2016.

Milan, M., Macpherson, B., Tubbs, R., Dow, G., Inverarity, G., Mittermaier, M., Halloran, G., Kelly, G., Li, D., Maycock, A., Payne, T., Piccolo, C., Stewart, L., and Wlasak, M.: Hourly 4D-Var in the Met Office UKV Operational Forecast Model, Quarterly Journal of the Royal Meteorological Society, 146, 1281–1301, https://doi.org/10.1002/qj.3737, 2020.

Pardo-Iguzquiza, E. and Chica-Olmo, M.: Geostatistics with the Matern Semivariogram Model: A Library of Computer Programs for Inference, Kriging and Simulation, Computers & Geosciences, 34, 1073–1079, https://doi.org/10.1016/j.cageo.2007.09.020, 2008.

Ribeiro Jr., P. and Diggle, P.: geoR: A Package for Geostatistical Analysis, R-NEWS, 1, 15–18, 2001.

Ribeiro Jr, P. J., Diggle, P. J., Schlather, M., Bivand, R., and Ripley, B.: geoR: Analysis of Geostatistical Data, 2020.

Stanley, S., Antoniou, V., Ball, L., Bennett, E., Blake, J., Boorman, D., Brooks, M., Clarke, M., Cooper, H., Cowan, N., Evans, J., Farrand, P., Fry, M., Hitt, O., Jenkins, A., Kral, F., Lord, W., Morrison, R., Nash, G., Rylett, D., Scarlett, P., Swain, O., Thornton, J., Trill, E., Warwick, A., and Winterbourn, J.: Daily and Sub-Daily Hydrometeorological and Soil Data (2013-2019) [COSMOS-UK], 2019.

Thomas, A., Cosby, B., Henrys, P., and Emmett, B.: Topsoil Carbon Concentration Estimates from the Countryside Survey of Great Britain, 2007 Using a Generalized Additive Model, 2020.

Vitolo, C., Fry, M., and Buytaert, W.: Rnrfa: An R Package to Retrieve, Filter and Visualize Data from the UK National River Flow Archive, The R Journal, 8, 102–116, 2016.

Yu, J., Li, X.-F., Lewis, E., Blenkinsop, S., and Fowler, H. J.: UKGrsHP: A UK High-Resolution Gauge–Radar–Satellite Merged Hourly Precipitation Analysis Dataset, Climate Dynamics, 54, 2919–2940, https://doi.org/10.1007/s00382-020-05144-2, 2020.

Zou, J., Ziegler, A. D., Chen, D., McNicol, G., Ciais, P., Jiang, X., Zheng, C., Wu, J., Wu, J., Lin, Z., He, X., Brown, L. E., Holden, J., Zhang, Z., Ramchunder, S. J., Chen, A., and Zeng, Z.: Rewetting Global Wetlands Effectively Reduces Major Greenhouse Gas Emissions, Nature Geoscience, 15, 627–632, https://doi.org/10.1038/s41561-022-00989-0, 2022.