Remote Sensing of Environment xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Estimating surface soil moisture from SMAP observations using a Neural Network technique

J. Kolassa[*,a,b], R.H. Reichle[b], Q. Liu[c,b], S.H. Alemohammad[d], P. Gentine[d], K. Aida[e], J. Asanuma[e], S. Bircher[l], T. Caldwell[m], A. Colliander[f], M. Cosh[g], C. Holifield Collins[n], T.J. Jackson[g], J. Martínez-Fernández[h], H. McNairn[i], A. Pacheco[i], M. Thibeault[j], J.P. Walker[k]

[a] Universities Space Research Association/NPP, Columbia, MD, USA
[b] Global Modelling and Assimilation Office, NASA Goddard Spaceflight Center, Greenbelt, MD, USA
[c] Science Systems and Applications Inc., Lanham, MD, USA
[d] Columbia University, New York, NY, USA
[e] University of Tsukuba, Tsukuba, Japan
[f] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
[g] USDA ARS Hydrology and Remote Sensing Laboratory, Beltsville, MD, USA
[h] Instituto Hispano Luso de Investigaciones Agrarias (CIALE), Universidad de Salamanca, Salamanca, Spain
[i] Agriculture and Agri-food Canada, Ottawa, Ontario, Canada
[j] Comisiòn Nacional de Actividades Espaciales (CONAE), Buenos Aires, Argentina
[k] Department of Civil Engineering, Monash University, Clayton, Victoria, Australia
[l] Centre d'Etudes Spatiales de la BIOsphère, (CESBIO-CNES, CNRS, IRD, Université Toulouse III), Toulouse, France
[m] University of Texas at Austin, TX, USA
[n] USDA ARS Southwest Watershed Research Center, Tucson, AZ, USA

ARTICLE INFO

Keywords:
Soil moisture remote sensing
SMAP
Data assimilation
Microwave radiometer

ABSTRACT

A Neural Network (NN) algorithm was developed to estimate global surface soil moisture for April 2015 to March 2017 with a 2–3 day repeat frequency using passive microwave observations from the Soil Moisture Active Passive (SMAP) satellite, surface soil temperatures from the NASA Goddard Earth Observing System Model version 5 (GEOS-5) land modeling system, and Moderate Resolution Imaging Spectroradiometer-based vegetation water content. The NN was trained on GEOS-5 soil moisture target data, making the NN estimates consistent with the GEOS-5 climatology, such that they may ultimately be assimilated into this model without further bias correction. Evaluated against in situ soil moisture measurements, the average unbiased root mean square error (ubRMSE), correlation and anomaly correlation of the NN retrievals were 0.037 m^3m^{-3}, 0.70 and 0.66, respectively, against SMAP core validation site measurements and 0.026 m^3m^{-3}, 0.58 and 0.48, respectively, against International Soil Moisture Network (ISMN) measurements. At the core validation sites, the NN retrievals have a significantly higher skill than the GEOS-5 model estimates and a slightly lower correlation skill than the SMAP Level-2 Passive (L2P) product. The feasibility of the NN method was reflected by a lower ubRMSE compared to the L2P retrievals as well as a higher skill when ancillary parameters in physically-based retrievals were uncertain. Against ISMN measurements, the skill of the two retrieval products was more comparable. A triple collocation analysis against Advanced Microwave Scanning Radiometer 2 (AMSR2) and Advanced Scatterometer (ASCAT) soil moisture retrievals showed that the NN and L2P retrieval errors have a similar spatial distribution, but the NN retrieval errors are generally lower in densely vegetated regions and transition zones.

1. Introduction

Soil moisture is a key variable for many surface and boundary layer processes, such as the coupling of the water and energy cycles (Seneviratne et al., 2006; Gentine et al., 2011; Bateni and Entekhabi, 2012) or the partitioning of precipitation into runoff and infiltration (Philip, 1957, Corradini et al., 1998, Assouline, 2013). Soil moisture is also a key determinant of the carbon cycle (McDowell, 2011; Sevanto

* Corresponding author.
E-mail address: jana.kolassa@nasa.gov (J. Kolassa).

https://doi.org/10.1016/j.rse.2017.10.045
Received 21 April 2017; Received in revised form 24 October 2017; Accepted 30 October 2017
0034-4257/ Published by Elsevier Inc.

et al., 2014; Jung et al., 2017). The importance of soil moisture has been recognized by the World Meteorological Organization by naming it an Essential Climate Variable (GCOS, 2009) and thus encouraging efforts to obtain better soil moisture observations, which is challenging because of its high variability both in space and time.

One avenue to obtain observations of soil moisture is through satellite instruments that provide global observations with a relatively short revisit period of 2–3 days. In particular, L-band (1.4 GHz) microwave instruments exhibit a high sensitivity to soil moisture in the top ~5 cm of the soil in sparsely to moderately vegetated areas. This has led to the launch of two L-band satellite missions to observe soil moisture, the European Soil Moisture and Ocean Salinity (SMOS) mission in 2009 (Kerr et al., 2010) and the NASA Soil Moisture Active Passive (SMAP) mission (Entekhabi et al., 2010) in 2015.

Traditionally, satellite soil moisture retrievals from L-band (and other) sensors are implemented through the inversion of Radiative Transfer Models (RTMs) (e.g. Owe et al., 2001; Kerr et al., 2012; O'Neill et al., 2015), which explicitly formulate the physical relationships linking surface soil moisture to satellite brightness temperature observations. The RTM inversion technique is used to produce the official SMOS and SMAP retrieval products, and is able to provide high quality soil moisture estimates (Al Bitar et al., 2012; Chan et al., 2016b; Colliander et al., 2017) with a typical latency of 12 to 24 h. However, this approach requires accurate knowledge of the physical relationships between the surface state and the satellite observations as well as their associated parameters, which are often empirically estimated and thus uncertain. Moreover, RTM inversions also require explicit information on other surface states, including surface soil temperature and vegetation, and are thus typically ill-posed problems. Additionally, for time critical applications, such as near real time flood prediction or soil moisture assimilation into weather prediction models, retrieval products with a shorter latency are required.

Data assimilation provides another option to generate improved soil moisture estimates through the merging of satellite and model information, and can yield soil moisture estimates that are of higher quality than estimates from satellite observations or models alone (e.g. Entekhabi et al., 1994; Walker and Houser, 2001; Liu et al., 2011; Lahoz and De Lannoy, 2014). For soil moisture assimilation, the observations and model estimates have to be unbiased with respect to each other, which is typically achieved by locally matching the mean and variability of the satellite observations to those of the model (Reichle and Koster, 2004). While this satisfies the requirements of the assimilation system, it has the side effect of removing some independent information in the satellite observations. Given the high quality of soil moisture observations from SMOS and SMAP this is undesirable.

As an alternative to RTM inversions, statistical Neural Network (NN) retrieval algorithms have been successfully implemented for a number of sensors in recent years (Aires et al., 2005; Chai et al., 2009; Kolassa et al., 2013, 2016; Rodríguez-Fernández et al., 2015; Santi et al., 2016). Instead of explicitly formulating physical relationships, NNs are calibrated on a sample of satellite observations and corresponding soil moisture estimates (the target data) to model the global statistical relationship between the satellite observations and surface soil moisture. As a result, NN retrievals can offer several general advantages over traditional RTM inversions. First, they do not require an explicit parameterization of physical relationships and are thus not affected by errors in our knowledge of these relationships or their parameters. Second, after a one-time calibration, NNs are computationally extremely efficient and can provide soil moisture estimates almost immediately after arrival of the instrument data, thereby shortening the latency. Third, training a NN non-locally on target data from a model, yields NN retrievals that are globally unbiased with respect to the model, with spatial and temporal patterns that are driven by the satellite observations (e.g. Alemohammad et al. (2017), Jimenez et al. (2013), Kolassa et al. (2016)). This may reduce the need for bias correction prior to an assimilation and at the same time retain more of the independent information contained in the spatial and temporal patterns of the satellite observations.

In this study, we develop the first NN algorithm to retrieve global surface soil moisture from SMAP observations. The motivation for this work is twofold. First, we investigate statistical retrieval techniques as a possible alternative or supplement to the existing physically-based SMAP retrieval algorithms. Since statistical techniques require less ancillary data and are subject to different algorithm-related errors than physically-based retrievals, NN retrievals may provide useful information where and when RTMs are known to be uncertain. For SMOS, the NN technique has been successfully implemented (Rodríguez-Fernández et al., 2015). However, it is not obvious that a NN for SMAP will work equally well, given the differences between SMOS and SMAP in the observing geometry (multiple vs. single incidence angle) and instrument error characteristics (De Lannoy et al., 2015). Second, we aim to investigate the potential of statistical techniques to generate a soil moisture product with characteristics beneficial to SMAP soil moisture assimilation. The NN algorithm retrieves soil moisture in the climatology of the target model and thus may reduce the need for bias correction prior to data assimilation. In a follow-on study, we will investigate whether this results in a more efficient use of SMAP observations during data assimilation.

The NN retrieval algorithm is trained with SMAP brightness temperatures and two ancillary datasets as inputs, and with target data from the NASA Goddard Earth Observing System version 5 (GEOS-5) model (Section 2). Using the trained NN, we compute global estimates of volumetric surface soil moisture for the period April 2015 to March 2017 and evaluate them using a number of different metrics and techniques (Section 3). We compare the SMAP NN soil moisture estimates to the target GEOS-5 model soil moisture to identify the independent information provided by the SMAP observations that can potentially inform the model during data assimilation (Section 4.1). Next, we assess the SMAP NN retrievals against independent in situ measurements and compare their skill to that of the SMAP Level-2 passive (L2P) retrieval product and the GEOS-5 model soil moisture (Section 4.2). Finally, we assess the global error distributions of the SMAP NN, GEOS-5 and SMAP L2P products using a triple collocation (TC) analysis in conjunction with soil moisture retrievals based on observations from the Advanced Microwave Scanning Radiometer 2 (AMSR2) and the Advanced Scatterometer (ASCAT), which have independent errors with respect to the SMAP and GEOS-5 products (Section 4.3).

2. Datasets

2.1. Neural Network inputs and target datasets

2.1.1. SMAP observations

The main input to the NN soil moisture retrieval algorithm are the SMAP brightness temperatures. SMAP was launched in January 2015 and is equipped with an L-band (1.4 GHz) radiometer observing on four different channels, horizontal and vertical polarization as well as the 3rd and 4th Stokes' parameter. SMAP is in a sun-synchronous, near-circular orbit with equator crossings at 6 AM and 6 PM local time and a revisit time of 2–3 days (Entekhabi et al., 2010). Brightness temperature data have been collected since 31 March 2015.

For our NN retrieval product we use SMAP Level-1C brightness temperatures (Chan et al., 2016) for the April 2015 to March 2017 period. The data are provided on the 36-km resolution Equal-Area Scalable Earth version 2 (EASEv2) grid (Brodzik et al., 2012) as daily half-orbit files. We only use observations from the 6 AM overpass, in order to minimize observation errors due to Faraday rotation and the difference between the soil and canopy temperatures (Entekhabi et al., 2010; O'Neill et al., 2015). A test of different input combinations indicated that using data from all four SMAP channels as inputs to the retrieval algorithm yielded the best NN retrieval performance (not