Abstract

High spatial and temporal resolution surface soil moisture is required for most hydrological and agricultural applications. The recently developed Disaggregation based on Physical and Theoretical Scale Change (DisPATCh) algorithm provides 1-km-resolution surface soil moisture by downscaling the 40-km Soil Moisture Ocean Salinity (SMOS) soil moisture using Moderate Resolution Imaging Spectroradiometer (MODIS) data. However, the temporal resolution of DisPATCh data is constrained by the temporal resolution of SMOS (a global coverage every 3 days) and further limited by gaps in MODIS images due to cloud cover. This paper proposes an approach to overcome these limitations based on the assimilation of the 1-km-resolution DisPATCh data into a simple dynamic soil model forced by (inaccurate) precipitation data. The performance of the approach was assessed using ground measurements of surface soil moisture in the Yanco area in Australia and the Tensift-Haouz region in Morocco during 2014. It was found that the analyzed daily 1-km-resolution surface soil moisture compared slightly better to in situ data for all sites than the original disaggregated soil moisture products. Over the entire year, assimilation increased the correlation coefficient between estimated soil moisture and ground measurements from 0.53 to 0.70, whereas the mean unbiased RMSE (ubRMSE) slightly decreased from 0.07 to 0.06 m$^3$ m$^{-3}$ compared to the open-loop force–restore model. The proposed assimilation scheme has significant potential for large-scale applications over semiarid areas, since the method is based on data available at the global scale together with a parsimonious land surface model.

1. Introduction

Soil moisture is an important variable of the terrestrial hydrosphere. Whereas precipitation provides the amount of available water at the surface, soil moisture impacts the partitioning of rainfall into runoff, evaporation, and infiltration. Moreover, soil moisture is highly variable in space and time, as a result of 1) the alternation between wetting and drying events and 2) the heterogeneity in land cover, topography, and soil properties. An accurate and continuous description of soil moisture in space and time is therefore critical for understanding the continental water cycle and for achieving efficient and sustainable water management (Entekhabi 1995; Gao et al. 2014; Rodriguez-Iturbe 2000).

Satellite remote sensing is often the most practical and effective method to observe the land surface soil moisture over large geographical areas. The recent Soil Moisture and Ocean Salinity (SMOS) mission, launched in 2009, operates at L band (the optimal microwave band to estimate soil moisture; Kerr 2007; Njoku and Entekhabi 1996) and provides near-surface soil moisture (SSM) with a resolution of about 40 km (Kerr et al. 2012). This mission has been complemented by the Soil Moisture Active Passive (SMAP) satellite mission launched in 2015, ensuring the continuity of L-band passive microwave data for global SSM monitoring (Entekhabi et al. 2010b). Recent studies, based on the temporal stability of soil moisture (Vachaud et al. 1985), have shown that even coarse-scale satellite soil moisture can add a benefit in hydrological modeling (Pauwels et al. 2001; Draper et al. 2011; Brocca et al. 2012; Alvarez-Garreton et al. 2015; Chen et al. 2014; Massari et al. 2015; Lievens et al. 2015). Nevertheless, the
current spatial resolution of microwave radiometers is too coarse for most hydrological and agricultural applications. Therefore, downscaling methodologies have been developed to improve the spatial resolution of passive microwave–derived SSM data (Das et al. 2014; Fang et al. 2013; Kim and Hogue 2012; Merlin et al. 2008a; Piles et al. 2011; Sánchez-Ruiz et al. 2014; Srivastava et al. 2013). For example, Disaggregation based on Physical and Theoretical Scale Change (DisPATCh) estimates the SSM variability within a 40-km resolution SMOS pixel at 1-km resolution using Moderate Resolution Imaging Spectroradiometer (MODIS) data (Merlin et al. 2012, 2013). However, the temporal resolution of DisPATCh data based on SMOS and MODIS data is limited by 1) gaps in MODIS images due to cloud cover and 2) the 2–3-day temporal resolution of global SMOS coverage (Djamai et al. 2016).

A land surface model (LSM) forced by uncertain meteorological inputs and constrained with discontinuous disaggregated soil moisture through data assimilation could both address the issue of discontinuity in the soil moisture products and as well improve the LSM estimate. Several studies have been undertaken to assimilate the observed satellite brightness temperature directly (Crow and Wood 2003; Dumdedah et al. 2011; Margulis et al. 2002; Reichle et al. 2007; Lievens et al. 2016, 2017) and/or the satellite SSM retrieval (Reichle et al. 2008; Draper et al. 2011; Brocca et al. 2012; Dumdedah and Walker 2014; Ridler et al. 2014; Kumar et al. 2014; Wanders et al. 2014; Lievens et al. 2015; Leroux et al. 2016) into LSMS. Others studies have assimilated coarse-scale SSM into fine land surface models to produce fine-scale model predictions and consistently improve soil moisture and other land surface variables (Reichle et al. 2001b, 2010; Parada and Liang 2004; Pan et al. 2009; Pan and Wood 2009; De Lannoy et al. 2010, 2012; De Lannoy and Reichle 2016; Sahoo et al. 2013; Lievens et al. 2016, 2017). These approaches are based on spatial error correlations that are modeled within the assimilation system. Moreover, Djamai et al. (2016) estimated SSM at 1-km resolution during cloudy days by combining DisPATCh data and the Canadian Land Surface Scheme (CLASS), forced by a 30-km atmospheric reanalysis. However, the SSM DisPATCh estimates were not improved by the combination of DisPATCh and CLASS when compared to in situ measurements of the SMAP Validation Experiments dataset in 2012 over Winnipeg, Canada. In a similar context, Dumdedah et al. (2015) assimilated DisPATCh data into the Joint U.K. Land Environment Simulator (JULES) to estimate root-zone soil moisture over the Yanco area in Australia. The assimilation of DisPATCh data into the JULES model had a limited positive impact on the SSM estimation accuracy compared to DisPATCh and open-loop JULES simulation.

These results demonstrate that data assimilation remains one of the most promising approaches to link satellite-based SSM with LSMS, while accounting for uncertainties in the observation data and the simulated output from the model (Calvet et al. 1998; Entekhabi et al. 1994; Jackson et al. 1981; Reichle et al. 2001a; Sabater et al. 2007). However, assimilation strategies still need to be improved. Two aspects should be addressed when assimilating downscaled SSM data into an LSM: 1) the number of state variables in the LSM should be consistent with the available observations in order to eliminate equifinality (Beven 1989; Franks et al. 1997) and 2) the accuracy in forcing data at the application scale. Most of the surface models developed since the 1980s ( Sellers et al. 1986; Noilhan and Planton 1989) have a large number of variables that cannot be directly measured at the model application scale (Demaria et al. 2007; Franks et al. 1997). As overparameterization is the main limitation for implementation of such complex models in an operational context, there is a need to develop simplified modeling approaches that are forced by available remote sensing and meteorological data (Allen et al. 1998). A number of studies have shown the potential of this approach (Alberge et al. 2008; Ceballos et al. 2005; Pellarin et al. 2006; Wagner et al. 1999) for representing components of the surface water budget. One of the main issues is that large-scale datasets of meteorological variables are currently unavailable at 1 km (or higher) spatial resolution. Nevertheless, a disaggregation/assimilation coupling scheme is potentially capable of compensating errors in atmospheric (mainly precipitation) forcing data available at a coarse scale only (Merlin et al. 2006).

Within this context, the objective of this study was to develop a new methodology based on an assimilation scheme for interpolating DisPATCh SSM in a suboptimal manner using global (meteorological and soil map) datasets. Since DisPATCh is a physically based method to provide native SSM at 1-km resolution using 1-km resolution MODIS data, the native resolution of the DisPATCh SSM products developed is 1-km resolution. The approach was tested using ground measurements of soil moisture and precipitation over two semiarid sites: 1) the Yanco area in the Murrumbidgee River catchment, Australia, and 2) the Tensift-Haouz basin located in central Morocco.

2. Sites description

a. Yanco: Murrumbidgee catchment (Australia)

The Murrumbidgee catchment, located in the southeastern part of Australia, covers about 82,000 km²
(34°–37°S, 143°–150°E) and is a part of the Murray–Darling basin. The Yanco study site is a 55 km x 55 km area located in the center of the Murrumbidgee western plains where the topography is flat, with very few geological outcroppings. The soil texture is predominantly sandy loam. The climate is semiarid, with an average annual precipitation of about 300 mm, while evaporative demand is about 1200 mm yr⁻¹, according to the reference evapotranspiration (ET₀), derived from the Food and Agriculture Organization (FAO) Penman–Monteith equation (Allen et al. 1998). The land use in the western part of the site comprises irrigation, while elsewhere land use is composed of rain-fed crops and native pasture with scattered trees.

The Yanco region has been intensively monitored for remote sensing studies since 2001 (Smith et al. 2012). This area has been selected as a core site for the calibration/validation of the SMOS (Peischl et al. 2012), SMAP (Panciera et al. 2014), and Global Change Observation Mission–Water (GCOM-W1; Mladenova et al. 2011) missions and has also been the focus of field experiments dedicated to algorithm development studies for the SMOS and SMAP missions: National Airborne Field Experiment 2006 (NAFE06; Merlin et al. 2008b), Australian Airborne Cal/Val Experiments for SMOS (AACES-1, -2; Peischl et al. 2012), and Soil Moisture Active Passive Experiments (SMAPex-1, -2, -3; Panciera et al. 2014). To assess the European Centre for Medium-Range Weather Forecasts (ECMWF) interim reanalysis (ERA-Interim) precipitation product, OzNet ground-based precipitation measurements using tipping-bucket rain gauges were used (Smith et al. 2012). These data are available online (http://www.oznet.org.au). Seven sites presenting the best data quality and continuity were selected for this study (Yanco 1, 2, 8, 9, 10, 12, and 13). Table 1 displays the site characteristics, and their locations are shown in Fig. 1. These sites are representative of the three main land uses of the region (Fig. 1): irrigated crops (Yanco 9), rain-fed crops (Yanco 1 and 11; typically wheat and fallow), and grazing (Yanco 2, 8, 10, 13; typically perennial grass-type vegetation).

### b. Tensift-Haouz basin (Morocco)

The Tensift-Haouz basin covers about 24 000 km² (30.75°–32.40°N, 7.05°E–9.9°W) around the city of Marrakech, in central Morocco (Fig. 2). The climate is semiarid, typically Mediterranean, with an average annual precipitation of about 250 mm (Chehbouni et al. 2008) concentrated between November and April over the Haouz plain, where the study site is located. Evaporative demand is about 1600 mm yr⁻¹.

In the Tensift-Haouz basin, the Sidi Rahal monitoring station was installed on a rain-fed wheat field (Fig. 2) in December 2013, in the framework of the Joint International Laboratory TREMA (a French acronym for Remote Sensing and Water Resources in the Semiarid Mediterranean; http://trema.ucam.ac.ma; Jarlan et al. 2015). It is equipped with micrometeorological instruments to estimate latent and sensible heat fluxes at the soil–vegetation–atmosphere interface and probes for the measurements of soil water content at different depths. The automatic meteorological station installed in the vicinity was equipped with sensors for the measurement of rainfall, global radiation, temperature, relative humidity, and wind speed at a half-hourly time step. The soil texture is predominantly loams. Information about the monitoring stations is provided in Table 1 and Fig. 2.

### 3. Materials and method

#### a. Globally available data

1) **SMOS soil moisture data**

The SMOS level-3 (L3) one-day global SSM (MIRCLF31A/D, version 7.72) product in reprocessing mode RE04 is used in this study as input to the DisPATCH algorithm and assimilation scheme. These products are presented in NetCDF format on the Equal-Area Scalable Earth Grid (EASE-Grid), with a grid spacing of ~25 km x 25 km in cylindrical projection. Note that L3 data are a 25-km grid representation of the 40-km data.
Details on the processing algorithms can be found in Jacquette et al. (2013) and in Kerr et al. (2014). For comparison purposes, the assimilation scheme was applied at 1-km resolution using the nondisaggregated (25 km) SMOS L3 data by oversampling the 25-km product.

2) DisPATCh Soil Moisture Data

DisPATCh provided 1-km-resolution SSM data from 40-km SMOS SSM and 1-km MODIS land surface temperature (LST), normalized difference vegetation index (NDVI), and GTOPO30 digital elevation model (DEM) data. MODIS-derived soil temperature was used to estimate soil evaporative efficiency (SEE), which is known to be relatively constant during the day in clear-sky conditions (Merlin et al. 2012). MODIS-derived, 1-km-resolution SEE was finally used as a proxy for SSM variability within the low-resolution pixel using a first-order series expansion around the SMOS observation. The disaggregated SSM products are expressed in meters cubed per meters cubed. The current version of the DisPATCh methodology is fully described in Molero et al. (2016). Note that only the ascending SMOS overpass (0600 local time) was used in this paper.

The DisPATCh product was derived from the average of an output ensemble for each SMOS overpass time. This output ensemble was obtained by applying DisPATCh to 1) four SMOS resampling grids by taking advantage of the L3 SMOS data oversampling, 2) three MODIS overpass dates by taking into account the MODIS data collected within ±1 day around the SMOS overpass, and 3) two daily MODIS observations aboard Terra and Aqua. The number of elements used to compute this average (a maximum of 24 elements per SMOS overpass) is called the DisPATCh count. Note that the DisPATCh count is often smaller than 24 because of gaps in MODIS data associated with cloud cover and/or limited overlap with the SMOS swath. The error of the DisPATCh product is taken as the standard deviation from the output ensemble computing. This error accounts for the downscaling and retrieval errors (more details in Merlin et al. 2012; Malbéeau et al. 2016).

DisPATCh outputs have been validated mostly in semiarid conditions where SEE is well constrained by the SSM: the Murrumbidgee catchment in Australia (Bandara et al. 2015; Malbéeau et al. 2016; Molero et al. 2016), the Little Washita watershed in Oklahoma, Walnut Gulch in Arizona (Molero et al. 2016), the

FIG. 1. The experimental Yanco area located in southeastern Australia showing the SMOS L3 grid corner (red crosses), DisPATCh grid (black crosses), the selected OzNet stations, and the irrigated area.
Tensift-Haouz basin in central Morocco (Merlin et al. 2015), and the Lleida area in Spain (Escorihuela and Quintana-Seguí 2016; Merlin et al. 2013).

3) VEGETATION INDEX

To estimate evapotranspiration, the vegetation cover $f_v$ was derived from the 1-km-resolution MODIS NDVI data. The NDVI dataset was extracted from the version-5 Terra MODIS vegetation indices’ 16-day L3 global 1-km grid product (MOD13A2). Fractional $f_v$ was computed using the linear relationship between NDVI of the fully covered vegetation and NDVI of the bare soil proposed by Gutman and Ignatov (1998).

4) METEOROLOGICAL DATASET

ERA-Interim (Dee et al. 2011) was used for meteorological (relative humidity, air temperature, wind speed, pressure, shortwave and longwave radiation, and precipitation) forcing. ERA-Interim is produced at the highest resolution of about 0.125° with a 3-hourly time step covering the period from January 1979 to present, with product updates at approximately 1 month behind real time. This study used the ERA-Interim datasets provided daily at 0.125° spatial resolution. Note that the product is generated at a much coarser resolution (a spectral T255 horizontal resolution, which corresponds to approximately 79 km spacing grid) and then mapped to 0.125°. The ERA-Interim atmospheric reanalysis is built upon a consistent assimilation of an extensive set of observations distributed worldwide from satellite remote sensing, in situ measurements, and radio soundings. ERA-Interim datasets are free of charge and available online (www.ecmwf.int/en/research/climate-reanalysis/era-interim). The environmental parameters simulated by ERA-Interim have been widely validated by in situ and remote sensing observations at different spatiotemporal scales (Balsamo et al. 2015; Bao and Zhang 2013; Boisvert et al. 2015; Mooney et al. 2011; Su et al. 2013; Szczypta et al. 2011; Wang and Zeng 2012). Several studies (Belo-Pereira et al. 2011; Pfeifroth et al. 2013; Szczypta et al. 2011; Zhang et al. 2013) have reported an overestimation of ECMWF precipitation data, but Balsamo et al. (2010) have shown that the original ERA-Interim products have reasonable skill for land applications at time scales from daily to annual over the conterminous United States. The total annual amount and daily distribution of ECMWF precipitation is compared to meteorological stations in this study for the two test sites.
5) Global Soil Texture

The relative amounts of bound and free water are influenced by the soil texture (sand, clay, and silt fractions) and bulk density. The map used for this study was a 0.01°-resolution combination of the soil maps (Kim 2013) from the FAO and the Harmonized World Soil Database (HWSD) and the regional datasets 1) State Soil Geographic Database (STATSGO), 2) National Soil Database Canada (NSDC), and 3) Australian Soil Resources Information System (ASRIS). Note that this soil texture map is used by both SMOS (Kerr et al. 2012) and SMAP (Entekhabi et al. 2010) level-2 SSM retrieval algorithms.

b. Land surface model

In an effort to reduce as much as possible the number of model parameters, while attempting to preserve the representation of the physics that controls the SSM dynamics, the LSM used in this study was based on the force–restore method developed by Deardorff (1977). This scheme is used in many LSMs, including Interactions between Soil, Biosphere, and Atmosphere (ISBA; Noilhan and Planton 1989). The force–restore method appears to be a good trade-off between realism (physics) and complexity (number of parameters) for calibration over large areas. In this semiphysical model, the dynamics of soil moisture is described within two layers: the SSM (noted \( \Theta_1 \)) and the root-zone soil moisture (noted \( \Theta_2 \)). In this study, only the SSM dynamics were simulated with the root-zone soil moisture taken as a buffer variable to minimize possible biases between DisPATCh SSM and the force–restore prediction for compensating errors in meteorological (mainly precipitation and irrigation) forcing data. The equation for SSM is

\[
\frac{\partial \Theta_1}{\partial t} = \frac{C_1}{\rho_w d_1} (P - E_g) - \frac{C_2}{\tau} (\Theta_1 - \Theta_{eq}),
\]

where \( \Theta_{eq} \) is the equilibrium soil moisture, \( P \) is the ERA-Interim precipitation reaching the soil surface, \( E_g \) is the evaporation at the soil surface, \( \rho_w \) is the density of liquid water, \( \tau \) is the time constant taken as 1 day, and \( d_1 \) is an arbitrary normalization depth of 10 cm. Parameters \( C_1 \) and \( C_2 \) are empirical parameters named force and restore coefficients, respectively, representing the process of mass exchange between the soil and the atmosphere, and the surface and the root-zone layer, respectively. The force and restore coefficients \( C_1 \) and \( C_2 \) are dimensionless and highly dependent upon both the soil moisture content and the soil texture. Note that coefficients \( C_1 \) and \( C_2 \) are spatially distributed based on Noilhan and Mahfouf (1996) and vary over time. They were calibrated against a multilayer soil moisture model (Noilhan and Mahfouf 1996) such that

\[
C_1 = C_{1\text{sat}} \left( \frac{\Theta_1}{\Theta_{sat}} \right)^{\left( \frac{\beta}{2} + 1 \right)} \quad \text{and} \quad (2)
\]

\[
C_2 = C_{2\text{ref}} \left( \frac{\Theta_2}{\Theta_{sat} - \Theta_1 + \Theta_{eq}} \right), \quad (3)
\]

where \( \Theta_{sat} \) is the saturated soil moisture for a given texture, \( \beta \) is the slope of the retention curve, \( C_{1\text{sat}} \) and \( C_{2\text{ref}} \) are hydraulic parameters, and \( \Theta_1 \) is a small numerical value equal to 0.001. Each parameter was estimated from clay/sand fractions and default empirical parameters [equations are detailed in Noilhan and Mahfouf (1996)]. Parameter \( E_g \) in Eq. (1) is expressed as in Allen (2000) and Allen et al. (2005) by

\[
E_g = \text{ET}_0 \times K_e, \quad (4)
\]

with \( \text{ET}_0 \) being the reference evapotranspiration estimated according to the FAO Penman–Monteith equation (Allen et al. 1998) and the ERA-Interim meteorological forcing data (relative humidity, air temperature, wind speed, pressure, shortwave and longwave radiation). Parameter \( K_e \) is the soil evaporation coefficient, computed from

\[
K_e = (1 - f_v)K_r, \quad (5)
\]

where \( K_r \) is the soil evaporation reduction coefficient derived from the SSM, and \( f_v \) is the vegetation cover [see section 3a(3)]. Soil evaporation from the exposed soil was assumed to take place in two stages: an energy-limiting stage and a falling-rate stage. After rain, evaporation was only determined by the energy available for evaporation, and thus \( K_r \) was set to 1; then when the soil surface dried out, \( K_r \) decreased linearly and evaporation was reduced. Parameter \( K_r \) was equal to zero when no significant water was left for evaporation, being when SSM was smaller than \( (1/2)\Theta_{wp} \) (where \( \Theta_{wp} \) is the soil moisture at wilting point), as reported by Allen et al. (1998).

c. Assimilation scheme: A combined 2DVAR and sequential approach

The purpose of assimilating DisPATCh data into a LSM was to combine the downscaled snapshots of DisPATCh SSM with the continuous LSM predictions, in order to obtain the best estimate of the SSM at 1 km every day. The simplified two-dimensional variational (2DVAR) method developed by Balsamo et al. (2004) to analyze the root-zone soil moisture (as a buffer variable) was combined with a simplified Kalman filter approach to update the SSM state. The relation between
surface and root-zone soil moisture is not physically based with the force–restore scheme. For that reason, a linear variational algorithm may not be well suited for updating surface soil moisture by contrast with the root zone. Moreover, the sequential approach is able to update the potentially rapid changes related to irrigation that are not represented by the LSM but that are observed in DisPATCh data. Thus, the two-scheme procedure has the advantage of considering the two temporal dynamics, these being the (rapid) surface and (slow) root-zone soil moisture.

The 2DVAR method was initially designed to analyze the root-zone soil moisture using 2-m air temperature and humidity observations (Balsamo et al. 2004). It has been adapted by Sabater et al. (2007) to analyze the root-zone soil moisture from SSM observations and to the analysis of both aboveground biomass and root-zone soil moisture by Sabater et al. (2008). The simplified 2DVAR method has also been applied to the analysis of aboveground biomass from satellite-derived leaf area index products over West Africa (Jarlan et al. 2008). In the present study, \( Q^b \) was taken as a buffer variable without any dynamic equation. Stated differently, this variable was left free to adjust the model prediction to DisPATCh SSM through the simplified 2DVAR approach. This first step of the assimilation algorithm was necessary to represent SSM dynamics with consistency to the restore parameter. The analyzed state is given by

\[
\Theta^a = \Theta^b + K(y - H\Theta^b),
\]

where the superscripts \( a \) and \( b \) indicate the analysis and background, respectively; \( y \) is the DisPATCh SSM; and \( H \) is the observation operator that allows the projection of the state vector in the observation space. In the 2DVAR approach, \( H \) is computed from a one side finite difference, while \( H \) is equal to 1 in the sequential approach. The SSM update step is close to that of the Kalman filter, but the propagation of the background error matrix was avoided here for simplicity purpose. Parameter \( K \) is called the gain and is calculated as

\[
K = BH^T(\text{H}B^T + R)^{-1},
\]

where \( B \) and \( R \) are the covariance matrices of the background and SSM observations errors, respectively. Matrix \( R \) has scalar values equal to \( \sigma_{\text{obs}} \) (DisPATCh error), and \( B \) is calculated as

\[
B = \begin{pmatrix} \sigma_{\Theta_1} & 0 \\ 0 & \sigma_{\Theta_2} \end{pmatrix},
\]

where \( \sigma_{\Theta_1} \) is \( \Theta_1 \) background error and \( \sigma_{\Theta_2} \) is \( \Theta_2 \) background error.

Considering a 1-day assimilation window, \( H \) equals

\[
H = \begin{bmatrix} \Delta\Theta_1(t) \\ \Delta\Theta_2(t-1) \\ 0 \end{bmatrix},
\]

IMPLEMENTING AND EVALUATING THE DATA ASSIMILATION ALGORITHM

(i) Background error covariance matrix

The parameters \( B \) and \( R \) determine the relative weight given to the background and observation covariance, respectively, while \( \sigma_{\text{obs}} \) corresponds to the observation (DisPATCh) error [see section 3a(2)]. Observation errors are correlated in space. An accurate estimation of the background error is likely to be the most difficult task in the error prescription (Bouttier 1994; Reichle et al. 2002). Thus, a sensitivity analysis to
background error on SSM and root-zone soil moisture was carried out; a set of $\sigma_Q^1$ and $\sigma_Q^2$ were compared in order to estimate both background errors since there is no propagation equation of the background error covariance matrix using variational assimilation. In practice, an ensemble of 10 perturbations from 0.02 to 0.1 m$^3$ m$^{-2}$ was built for both the background error terms, and the global statistics [correlation coefficient $r$, root-mean-square error (RMSE), and mean bias] were computed based on the analyzed and in situ SSM comparison. Results of the sensitivity study are displayed in Fig. 3. The optimal choices obtained from this sensitivity study were about 0.04 and 0.09 m$^3$ m$^{-2}$ for $\sigma_Q^1$ and $\sigma_Q^2$, respectively. Note that the same sensitivity study has been performed at 25 km, and the optimal choices obtained are 0.05 and 0.06 m$^3$ m$^{-2}$ for $\sigma_Q^1$ and $\sigma_Q^2$, respectively. Nevertheless, the range of bias and RMSE were low (about 0.009 m$^3$ m$^{-2}$) for the whole range of

![Fig. 4. Cumulative daily precipitation (mm) for all sites. The blue lines are the ERA-Interim precipitation at 0.125° spatial resolution distributed by the ECMWF, and the red lines are the in situ precipitation. Note that in situ data are not available for Yanco 8.](image)
potential values. This means that the sensitivity analysis for both background errors presented limited choices. Interestingly, a Θ1 background error lower than that of Θ2 also seems consistent with the objective of the study, since Θ2 was considered as a buffer variable to minimize biases on Θ1. Finally, this quite low value of background error on Θ1 was also certainly to be attributed to the good quality of ERA-Interim data, which were the main forcing of the Θ1 dynamics. Based on this analysis, the suboptimal values of background error were chosen for the implementation of the data assimilation algorithm.

(ii) Statistical metrics

It was important to assess the performance of the method, not only in terms of linear dependency and error, but also in terms of relative variability of the original and updated dataset. Therefore, r, RMSE, unbiased RMSE (ubRMSE), and the mean bias were used to fully assess the accuracy of SSM (Entekhabi et al. 2010a). Moreover, a new metric called the gain of downscaling (GDOWN), introduced by Merlin et al. (2015), was also used. The gain is a measure of the statistical improvement dedicated to disaggregated SSM products. The gain can range from −1 to 1, where positive values indicate better correspondence with in situ than low-resolution products such as SMOS data. One key advantage of GDOWN, with regard to other performance metrics, is to provide an estimate of the overall improvement in soil moisture data with a single value.

4. Results and discussion

The DisPATCh/assimilation approach has been run over the entire year 2014 for both areas (Yanco in Australia and Tensift-Haouz in Morocco). First, ERA-Interim precipitation products were assessed and validated using ground measurements. After, the analyzed SSM was evaluated at the time of DisPATCh availability. Finally, the analyzed SSM was assessed for the entire year datasets. a. ERA-Interim precipitation assessment

Although the assimilation scheme can compensate for error in precipitation input data, a good agreement of ERA-Interim with ground rainfall in terms of frequency (instead of quantity) is preferable to update the SSM state on a daily basis. A preliminary comparison between ERA-Interim precipitation and the station data showed that ERA-Interim presented too frequent low-rainfall events (between 0.1 and 3 mm day⁻¹). This has already been observed by Ibrahim et al. (2012) and Diaconescu et al. (2015) over another semiarid region in the West African Sahel. The general overestimation of wet days is due to the fact that precipitation in analyses is mainly model generated, and therefore highly related to forecast–model physical parameterizations (surface pressure, temperature, and wind). In this study, the precipitation values during low-rainfall events (<3 mm day⁻¹) were set to zero (Ibrahim et al. 2012; Diaconescu et al. 2015). After this preprocessing, ERA-Interim precipitation was in better agreement with local station data (not shown). The daily ERA-Interim precipitation values were compared to the in situ data using 24-h accumulation from the raw 30-min observations. Figure 4 and Table 2 reported the annual amounts and differences between the two precipitation datasets for each site. With an average bias of 27 mm yr⁻¹ and an r of 0.48, ERA-Interim annual amounts matched quite well with the in situ observations, considering the large resolution of ERA-Interim data and the high spatial variability of precipitation in semiarid regions. Apart from sites Yanco 10 and Sidi Rahal, biases remained below 40 mm yr⁻¹. Figure 4 and Table 2 also showed that timing was well reproduced at ±1 day, based on the correlation coefficient value when using a 3-day accumulated precipitation. For instance, daily ERA-Interim precipitation amounts at the Sidi Rahal site were really well correlated (r of 0.93) with the ground measurements. Regarding the Yanco 10 site, the daily r was low; however, it increased greatly using the 3-day accumulated
precipitation (from 0.18 to 0.51). These results were similar to results found in Balsamo et al. (2010). In particular, none of the big storm events recorded by the local stations were missed by ERA-Interim. Both timing and event amount were particularly well reproduced for Yanco 1 and 2. Two anomalies were noted at the Sidi Rahal site: the ERA-Interim precipitation was underestimated compared with in situ observations, whereas the ERA-Interim precipitation was overestimated at the Yanco 10 station. For both sites, all events were well detected but the amounts of water were under- and overestimated for the Sidi Rahal and Yanco 10 sites, respectively. On average, ERA-Interim precipitation data compared quite well with in situ stations apart from moderate biases. The ERA-Interim dataset was thus used in the data assimilation algorithm in order to evaluate the performance of the approach when precipitation data are inaccurate, which is a very likely situation when no meteorological station is available.

b. Assimilation results

Herein, the performance of the approach was assessed by comparing SMOS, DisPaTCh, open-loop, and analyzed SSM with in situ measurements at the time of DisPaTCh availability, in order to check if the analyzed SSM shows an improvement with regard to disaggregated SSM. All statistics were estimated on a yearly basis to evaluate the capability of a dynamical model to interpolate and, potentially, to improve DisPaTCh SSM data. Table 3 showed the yearly statistics for each monitoring station and the number of comparison days. In this section, the number of days used was strongly dependent on the number of SMOS overpasses and cloud coverage. When comparing the statistics obtained over both areas, it was observed that the disaggregation and the assimilation scheme reduced bias by approximately 0.02 m$^3$ m$^{-2}$, while $r$ was systematically higher after data assimilation. The disaggregation at 1 km (DisPaTCh) has the advantage to produce SSM at a spatial resolution closer to the representativeness scale of ground measurements; it has been shown that this reduces the bias (Malbétanuet al. 2016). The mean $r$ over the Yanco area rose from 0.62 to 0.77 after disaggregation and up to 0.80 after application of the assimilation scheme. Regarding Yanco 2, $r$ improved from 0.69 to 0.79, bias reduced from 0.03 to 0.01 m$^3$ m$^{-3}$, and ubRMSE decreased from 0.07 to 0.05 m$^3$ m$^{-3}$. Similar results were observed for Sidi Rahal as $r$ increased from 0.82 to 0.87 after assimilation while bias was closer to 0, and ubRMSE decreased slightly from 0.05 to 0.04 m$^3$ m$^{-3}$. Across all sites, the assimilation scheme does not significantly improve SSM in terms of statistics when using a limited (DisPaTCh) time series for validation. Differences can be appreciated more easily through qualitative

<table>
<thead>
<tr>
<th>Country</th>
<th>Station</th>
<th>SMOS</th>
<th>DepATCH</th>
<th>Ec</th>
<th>Analyzed</th>
<th>SMOS</th>
<th>DepATCH</th>
<th>Ec</th>
<th>Analyzed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morocco</td>
<td>Sidi Rahal</td>
<td>104</td>
<td>0.64 (0.10)</td>
<td>0.01</td>
<td>0.03 (0.01)</td>
<td>0.01</td>
<td>0.03 (0.01)</td>
<td>0.01</td>
<td>0.03 (0.01)</td>
</tr>
<tr>
<td>Australia</td>
<td>Yanco 1</td>
<td>104</td>
<td>0.67 (0.10)</td>
<td>0.01</td>
<td>0.03 (0.01)</td>
<td>0.01</td>
<td>0.03 (0.01)</td>
<td>0.01</td>
<td>0.03 (0.01)</td>
</tr>
<tr>
<td>Australia</td>
<td>Yanco 2</td>
<td>111</td>
<td>0.67 (0.14)</td>
<td>0.01</td>
<td>0.03 (0.01)</td>
<td>0.01</td>
<td>0.03 (0.01)</td>
<td>0.01</td>
<td>0.03 (0.01)</td>
</tr>
<tr>
<td>Australia</td>
<td>Yanco 10</td>
<td>114</td>
<td>0.65 (0.10)</td>
<td>0.01</td>
<td>0.03 (0.01)</td>
<td>0.01</td>
<td>0.03 (0.01)</td>
<td>0.01</td>
<td>0.03 (0.01)</td>
</tr>
<tr>
<td>Australia</td>
<td>Yanco 12</td>
<td>79</td>
<td>0.65 (0.13)</td>
<td>0.01</td>
<td>0.03 (0.01)</td>
<td>0.01</td>
<td>0.03 (0.01)</td>
<td>0.01</td>
<td>0.03 (0.01)</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>69</td>
<td>0.65 (0.13)</td>
<td>0.01</td>
<td>0.03 (0.01)</td>
<td>0.01</td>
<td>0.03 (0.01)</td>
<td>0.01</td>
<td>0.03 (0.01)</td>
</tr>
</tbody>
</table>
inspection of scatterplots. Sidi Rahal (Fig. 5b), Yanco 9 (Fig. 6b), and Yanco 10 (Fig. 7b) illustrate the Table 3 results. In these three plots, the analyzed distribution appears closer and more symmetric around the 1:1 line than for both DisPATCh and open loop.

As illustrated in Table 4, GDOWN was approximately equal for both the disaggregated and the analyzed data. Moreover, the only site with negative values (meaning that SMOS shows better results) was Yanco 12, because DisPATCh and the analyzed SSM had a larger mean bias than SMOS (Table 3). However, \( r \) was slightly improved after assimilation, which means that the SSM dynamics were better represented. Yanco 2 was the site with the best enhanced GDOWN, with values going from 0.35 to 0.53 before and after assimilation, respectively. This was mainly due to a large improvement of \( r \) (from 0.47 to 0.79) and of the bias (from \(-0.03\) to \(-0.01\) m\(^3\) m\(^{-3}\)). The assimilation scheme was also compared to the open-loop estimates using the same statistics (Table 3). The assimilation clearly outperformed the open-loop prediction at the time of DisPATCh availability.

As a summary, this new approach improved the \( r \) values and also reduced the ubRMSE with regard to either the satellite observations or the model open loop, indicating that the assimilation has the capability to improve the SSM estimates over the model results or DisPATCh alone.

c. Soil moisture analysis

To assess the potential of using the force–restore model and the assimilation scheme for infilling disaggregated SSM, a conventional validation strategy was adopted, by comparing the analyzed and in situ SSM datasets for the full time series. This strategy was useful to characterize the overall quality of the analyzed SSM over both areas. The open-loop estimate of SSM estimates determined from the force–restore model forced by ERA-Interim were also computed for comparison purposes, along with the assimilation scheme applied to the original SMOS L3 product. This comparison evaluated the contribution of fine (instead of coarse) resolution soil moisture data. Table 5 displays temporal statistics for open-loop, analyzed 25-km, and analyzed 1-km SSM estimates for all in situ stations. Overall results showed that \( r \) was about 0.7, while mean bias was equal to 0.03 m\(^3\) m\(^{-3}\) and ubRMSE was 0.06 m\(^3\) m\(^{-3}\) for the analyzed SSM estimates. The \( r \) values were found to be systematically higher after assimilation, whereas bias, RMSE, and ubRMSE were equivalent for both datasets.
Regarding the Sidi Rahal station (Fig. 5, Table 5), $r$ after data assimilation was about 0.83, while bias was close to 0.01 m$^3$ m$^{-3}$ and the ubRMSE was around 0.06 m$^3$ m$^{-3}$. The time series exhibited the dominant seasonal cycle very well and showed a similar dynamical response to precipitation events. Comparison between analyzed 25-km and 1-km statistics showed that DisPATCh SSM improved results for all sites, even though the original SMOS L3 data had a larger temporal repetition. In others words, the spatial information provided by DisPATCh provides superior assimilation results despite the data gaps associated with cloud coverage over the study sites. Figures 6 and 7 illustrate time series for two sites in the Yanco area. Regarding Yanco 10 (Fig. 7a), data assimilation enhanced $r$ from 0.47 to 0.70, whereas the bias was larger by approximately 0.01 m$^3$ m$^{-3}$. The open-loop SSM estimates showed a threshold for dry conditions at around 0.1 m$^3$ m$^{-3}$ due to $\Theta_2$ being forced to 0.1 m$^3$ m$^{-3}$ by lack of information. The analyzed SSM was not constrained any more by this artificial threshold. This demonstrated the importance of the analysis of $\Theta_2$ for a correct estimate of $\Theta_1$. Interestingly, the best improvement was observed for the irrigated site Yanco 9 (Fig. 6a) where precipitation was supplemented by irrigation inputs that were not taken into account in the model run. Consequently, the assimilation of DisPATCh data improved $r$ (from 0.42 to 0.74), while bias, RMSE, and ubRMSE were similar compared to open-loop results. The time series in Fig. 6a showed water input events at day 140 and day 325 (for example) that were certainly due to irrigation. Thus, this approach could be used to detect and retrieve irrigation information that is very difficult to obtain over large areas on a daily basis. This information is requested by managers to monitor and control irrigation, especially for the monitoring of groundwater (Le Page et al. 2012).

The coupled scheme has the advantage of combining the spatial (but static) information provided by DisPATCh data with the temporal (but monodimensional) information provided by the force–restore scheme, in order to get SSM estimates every day at 1 km (Merlin et al. 2006). Figures 8 and 9 showed the temporal average of SSM during one year over the Yanco area and the Tensift-Haouz basin, respectively. The irrigated areas are indicated for comparison purposes. Regarding the Australian case study, the Murrumbidgee riverbanks and irrigated areas appeared wetter than the dry grassland. The wet area located in the south of the study area is the floodplain of the Yanco Creek system, which is a tributary of the Murrumbidgee River downstream of Narrandera, flowing southwest. Over the Tensift-Haouz basin, the wetter pixels were mainly located in the irrigated areas. Nevertheless, it is important to keep in mind that the irrigated areas indicated may be inaccurate since many boreholes have been dug since the beginning of the 2000s, and they are neither registered nor monitored. A wet zone was also shown in the southeast of the study area. This region corresponds to the Atlas Mountains and their piedmont. A large amount of water in the piedmont is provided by the water from Atlas snowmelt (Boudhar et al. 2009). However, it is necessary to note that DisPATCh data may be unreliable in mountainous areas as the illumination effect on LST can be significant in steep-sided valleys (Malbéteau et al. 2017), and no correction for such effects has been included in DisPATCh yet (Molero et al. 2016). Note that the disaggregation

### Table 4. GDOWN results.

<table>
<thead>
<tr>
<th>Country</th>
<th>Site</th>
<th>DisPATCh</th>
<th>Analyzed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morocco</td>
<td>Sidi Rahal</td>
<td>0.232</td>
<td>0.330</td>
</tr>
<tr>
<td>Australia</td>
<td>Yanco 1</td>
<td>0.119</td>
<td>0.112</td>
</tr>
<tr>
<td></td>
<td>Yanco 2</td>
<td>0.352</td>
<td>0.530</td>
</tr>
<tr>
<td></td>
<td>Yanco 8</td>
<td>0.571</td>
<td>0.314</td>
</tr>
<tr>
<td></td>
<td>Yanco 9</td>
<td>0.014</td>
<td>0.067</td>
</tr>
<tr>
<td></td>
<td>Yanco 10</td>
<td>0.108</td>
<td>0.235</td>
</tr>
<tr>
<td></td>
<td>Yanco 12</td>
<td>-0.111</td>
<td>-0.066</td>
</tr>
<tr>
<td></td>
<td>Yanco 13</td>
<td>0.282</td>
<td>0.220</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>0.196</td>
<td>0.218</td>
</tr>
</tbody>
</table>
images (Figs. 8, 9) present a slight boxy artifact at low (SMOS) resolution. Such an artifact is typical of downscaling methods like DisPATCh that apply a conservation law at low resolution (meaning that the average of disaggregated SSM at SMOS resolution is set to the SMOS observation).

As a summary, the proposed downscaling/assimilation scheme showed systematically higher $r$ values with regard to the open loop and with regard to DisPATCh alone, indicating that the dynamic of the SSM at a daily time scale has been improved. The maps of yearly average SSM were consistent with the main hydrological characteristics of both catchments (rivers, wetlands, and irrigated areas). This opens perspectives for the retrieval of irrigation water inputs.

5. **Conclusions**

The DisPATCh algorithm has been developed to improve the spatial resolution of readily available passive microwave–derived SSM data that is too coarse for many hydrological and agricultural applications. However, the temporal resolution of DisPATCh data based on SMOS and MODIS data is limited by the data gaps in MODIS images due to cloud cover and by the temporal resolution of SMOS. This paper evaluated the potential of assimilating DisPATCh data into the force–restore soil moisture model, forced by the ERA-Interim precipitation data in order to obtain daily SSM at 1 km resolution. A variational scheme was used for root-zone soil moisture analysis taken as a buffer variable, together with a sequential approach for the update of SSM. The approach was tested during a 1-yr period (2014) over two semiarid regions: 1) the Yanco zone in Australia and 2) the Tensift-Haouz basin in Morocco.
This study opens perspectives for developing new remote sensing–based methods in order to retrieve irrigation water inputs at 1-km resolution and/or to improve precipitation estimates. In particular, several studies have been undertaken to estimate and/or improve precipitation estimates based on remotely sensed coarse-scale SSM (Brocca et al. 2013, 2014; Pellarin et al. 2008, 2013). Continuous SSM data in space and time could allow the disaggregation of coarse-scale precipitation data from reanalysis datasets at 1-km resolution for hydrological and agronomical applications. Likewise, a 1-km daily irrigation input dataset could help improve knowledge on how water is used for irrigation purposes.

**Acknowledgments.** This study was supported by the MIXMOD-E project (ANR-13-JS06-0003-01), funded by the French agency ANR (Agence Nationale de la Recherche), as well the REC project (RISE-2014-645642-REC), funded by the European H2020 program. Initial setup and maintenance of the Murrumbidgee monitoring network used in this study was funded by the Australian Research Council (DP0343778, DP0557543, DP0879212, DP0984586, DP140100572) and by the CRC for Catchment Hydrology. Sidi Rahal station was funded by the MISTRAL METASIM/SICMED project, and it is managed in the framework of the Joint International Laboratory TREMA (http://trema.ucam.ac.ma).

**REFERENCES**


Balsamo, G., F. Bouyssel, and J. Noilhan, 2004: A simplified bi-dimensional variational analysis of soil moisture from screen-level observations in a mesoscale numerical weather-prediction...


