Chapter 12
Soil Moisture Dynamics Estimated from MODIS Time Series Images

Thomas Gumbricht

Abstract The annual cyclic phenomena of soil surface wetness influences for instance vegetation growth, drought, flooding, and soil properties. This study presents an attempt to define metrics relevant for capturing the soil moisture dynamics from an annual series of wetness estimates derived from global Moderate-resolution imaging spectroradiometer (MODIS) images. Different algorithms for both smoothing and gap-filling the time series are tested with the results compared to in-situ data. Neither the smoothing nor the gap-filling improve the capturing of the surface wetness phenology compared to using unsmoothed time series data. The smoothing, however removes the effects of erratic rainfall events and noise, and the smoothed time series was considered more robust for identification of wetness phenology. Metrics capturing the global surface wetness phenology for 2011, extracted after smoothing using a simplified locally weighted scatterplot smoothing (LOWESS) model, are presented at a spatial resolution of 500 m for the calendar year 2011.

12.1 Introduction

Soil moisture content and its annual variation are of key interest for understanding e.g. vegetation production, rainfall to runoff response and flooding, drought and fire risk, and soil formation including the occurrence of wetlands. Hitherto, time-series of continental to global scale soil surface wetness have been restricted to coarse scale (~25–50 km) estimates derived from passive microwave sensors and global hydrological models (Haddeland et al. 2011; Ochsner et al. 2013; Papa et al. 2013; Reichle et al. 2004; Wagner et al. 2003). The successful launch of the Soil Moisture Active Passive (SMAP, smap.jpl.nasa.gov) mission in January 2015 now produces refined estimates, even though one the radar instruments stopped operating in July 2015. But also the SMAP data is at a too coarse scale for satisfying the needs in many applications. Elsewhere (Gumbricht 2015) I introduce an attempt to

T. Gumbricht
Karttur AB, Stockholm, Sweden
e-mail: thomas.gumbricht@gmail.com

Y. Ban (ed.), *Multitemporal Remote Sensing*, Remote Sensing and Digital Image Processing 20, DOI 10.1007/978-3-319-47037-5_12

develop a quantitative estimation of soil wetness retrieved from optical image data, the Transformed Wetness Index (TWI). TWI is a non-linear normalized difference (ND) index that uses biophysical feature vectors representing the soil line and wetness as input. TWI can in principle be derived from any multi-spectral image data source, but soil moisture content is better captured with Short-Wave Infra-Red (SWIR, 1100–2500 nm) compared to visible and near-infrared (VNIR, 400–1100 nm) wavelengths. Adopting Bidirectional Reflection Distribution Function (BRDF) corrected Moderate-resolution Image Spectroradiometer (MODIS) data for calculating TWI has several advantages for soil moisture mapping; the product includes three SWIR bands, the data are reflectance corrected and easily accessible for the whole earth, atmospheric attenuations including cloud contaminations are negligible, and the annual cycle of wetness can be captured from discrete images over an annual cycle, each representing a 16-day period. A disadvantage is that the exact date for each observation is not known, only the 16-day interval within which it falls is.

This study briefly describes TWI, and focuses on techniques for retrieving metrics for capturing the global annual soil moisture dynamics. While several studies validate the temporal performance of microwave derived soil-moisture estimates compared to in-situ probes measuring soil moisture (Draper et al. 2009; Jackson et al. 2010; Reichle et al. 2007), relatively few studies have explored the metrics of seasonal and annual variations in soil-moisture (Cheema et al. 2011). To remove noise and outliers, and fill in data-gaps, different smoothing algorithms are tested, and model results compared to in-situ probed wetness. The study presents phenological metrics depicting the global soil moisture dynamics for the calendar year 2011.

12.1.1 Phenological Characterization

Characterization of cyclic annual or seasonal natural phenomena (phenology) has been widely adopted for studying vegetation dynamics extracted from time-series of satellite images (Heumann et al. 2007; Jones et al. 2011; Tan et al. 2011). Phenological characterization includes both value based and temporal metrics. The most common value metrics include mean (MEAN), minimum (MIN), maximum (MAX), and seasonal integration (INT) (e.g. accumulated vegetation growth). The most widely used temporal metrics include time of start of (growing) season (SS), end of season (ES), length of season (LS) and timing of peak season (PS). Some vegetation phenological studies also explore the derivative of the vegetation cycle during green-up and brown-down to identify both timing and rate of change representing the periods of maximum growth and maximum senescence.

Most vegetation phenology studies use a per-pixel definition of seasons, either by analyzing derivates or by threshold(s). Seasonal separation is usually dynamic,

derived from the statistical moments of an annual cycle, or by using multi-year statistical moments. The latter requires that the data show some persistency, and the more advanced models used in e.g. the widely adopted TIMESAT software (Jönsson and Eklundh 2004) for instance require a minimum of 3 years of data. Studies capturing the phenology of for instance snow cover or ice formation (Kang et al. 2012), instead rely on absolute thresholds.

The soil moisture regimes in many landscapes vary more erratically and less predictable than vegetation growth and density (or snow/ice phenology). While vegetation growth is dependent on soil moisture, the biological processes driving vegetation growth and senescence moderate the variability and change rates of the vegetation density compared to soil moisture. Further, the surface wetness cycle can vary on short distances, dependent on e.g. topography, hydrology and, not least, human management. Hydrological recharge ("uphill") and discharge ("downhill") areas usually have different wetness conditions, with e.g. ridges next to floodplains able to show extreme differences both in moisture content and timing. These differences between the phenology of vegetation and surface wetness prompt different approaching for both time-series smoothing and definition of relevant metrics for characterization of phenology.

12.1.1.1 Smoothing

Reducing bias and smoothing noise is commonly regarded as a prerequisite when retrieving phenology from satellite derived time-series data (Atkinson et al. 2012; de Beurs and Henebry 2010; Hird and McDermid 2009; Jönsson and Eklundh 2004). Noise reduction and gap-filling techniques are either based on model fitting or smoothing algorithms. The best technique depends both on the satellite derived index, as well as the objective of the study and the phenology metrics explored. For indexes with a known bias (e.g. the Normalized Difference Vegetation Index, NDVI), methods adjusting the smoothing considering this bias perform better; Hird and McDermid (2009), for instance, found that asymmetric Gaussian (Jönsson and Eklundh 2004) and double logistic (Beck et al. 2006) model fitting outperformed smoothing algorithms when cleaning time series derived from NDVI. These models, however, performed less well in a comparative study using the (unbiased) Medium Resolution Imaging Spectrometer (MERIS) Terrestrial Chlorophyll Index (MTCI) (Atkinson et al. 2012).

More advanced models for smoothing vegetation indexes implicitly rely on rates and restrictions in biological processes, and are also criticized for being over-parameterized (de Beurs and Henebry 2010). Advanced models are further less suited for studies covering different ecological or climatic regions or landscapes, and most can not be adopted without a-priori assumptions on the annual seasonality. Also, other routine algorithms used for smoothing time-series, including Fourier-analysis and harmonics, can not be adopted for smoothing irregular and asymmetrical time-series with an unknown number of annual cycles.

Assuming that the TWI soil moisture estimates are unbiased, the smoothing should aim at identifying the local mean. The most widely approach for identifying the local mean is moving average. Seeking a more flexible approach, this study adopted a modified variant of the robust locally weighted scatterplot smoothing (LOWESS) (Cleveland 1979; Cleveland and Devlin 1988) for smoothing time-series of soil moisture. LOWESS is flexible both through allowing weighting, and the selection of polynomial functions for fitting local regressions to an arbitrary sized filter window. LOWESS can be applied without any a-priori assumptions on seasonality (smoothness) and at the same time adapts well to data variations. Outlier removal in LOWESS most commonly uses an iterative process and the statistical moments of the time-series itself to decide whether any single observation should be omitted or not. As discussed above, changes in soil surface moisture can be fast, but are not per se linked to the statistical moments of the time-series at large. An alternative LOWESS approach was thus developed in this study, using absolute thresholds for discarding outliers. For regular interval time-series data, the filter size in effect determines the length of the period influencing any observation, and in this study also the LOWESS filter size was set to absolute values.

12.1.1.2 Metrics

For vegetation, the local variations in production can be directly captured from a vegetation index and used for setting both dynamic and local thresholds for e.g. growing seasons. Relevant thresholds for soil wetness instead relate to biophysical characteristics, including e.g. wilting point, field capacity and water content at fully saturated soils. These thresholds vary both locally and with soil type (Brady and Weil 2007), and ideally soil wetness phenology should relate to local thresholds. There is, however at present no technique available for directly capturing these thresholds from remote sensing data. The TWI soil moisture estimates used in this study are given as volume water over total volume, and converted to percent. Assuming a soil pore volume of 50 %, a soil moisture estimate of 50 thus represent a fully water saturated soil.

In this study, the phenological extraction was done for four arbitrarily fixed soil moisture thresholds: Flooding Seasons (FS) (soil moisture >50 %), Soaking Season (SS) (soil moisture >37.5 %), Wet Season (WS) (soil moisture >25 %), and non-Dry Season (DS) (soil moisture >12.5 %). The thresholds are loosely set to represent: FS, fully water saturated soil conditions, SS: soil conditions favoring rapid rainfall to runoff conditions; WS: soil wetness at field capacity (threshold for drainage of soil water), and DS: representing soil moisture at the wilting point. The labeling should be regarded as one of convenience. Both the length of the season with soil moisture exceeding each threshold, as well as the annual integrated soil moisture above each of these thresholds are calculated as phenological metrics (Table 12.1). Additionally, the start and end dates of up to three seasons of each threshold are also calculated.

Table 12.1 Definition of metrics for capturing the annual soil surface moisture phenology

Abb.	Full label	Definition
Value metrics		
MEAN	Mean soil wetness	Mean annual soil wetness (%)
SD	Standard deviation of soil wetness	Variation in annual soil wetness (%)
MIN	Minimum soil wetness	Lowest recorded wetness (%)
MAX	Maximum soil wetness	Highest recorded wetness (%)
IFS	Integration of flood season wetness	Integration of soil moisture >50 %
ISS	Integration of soaking season wetness	Integration of soil moisture >37.5 %
IWS	Integration of wet season wetness	Integration of soil moisture >25 %
IDS	Integration of non-dry season wetness	Integration of soil moisture >12.5 %
Temporal metrics		
LFS	Length of flooding season	Length of season with soil moisture >50 %
LSS	Length of soaking season	Length of season with soil moisture >37.5 %
LWS	Length of wet season	Length of season with soil moisture >25 %
LDS	Length of non-dry season	Length of season with soil moisture >12.5 %
PWS	Peak wet season	Day of year with highest recorded wetness
PDS	Peak dry season	Day of year with lowest recorded wetness

12.2 Objective

The primary aim of the study was to create global maps of annual soil moisture dynamics at moderate spatial resolution, to be used as support for other mapping efforts, including mapping of tree cover, soil organic carbon and wetlands (Gumbricht 2015). The main objective was to define metrics for capturing the surface wetness phenology, and to identify a smoothing algorithm that both removes/reduces outliers and smoothes soil moisture time-series noise.

12.3 Data

This study was based on a global annual time-series of 16-day composited MODIS BRDF data (MCD43A4) for the calendar year 2011. To allow Inverse Distance Weighting (IDW) of data gaps, and smoothing at the beginning and end of the calendar year 2011, the two last MODIS tile-dates for 2010 and the two first for 2012 were also included. If data for IDW were lacking, the 2011 MODIS TWI time-series was filled by extrapolation (all data used for model development could be filled by IDW, but the production of the global maps required extrapolation over some densely clouded regions). For the equatorial tropics (MODIS vertical tiles 8 and 9) the complete time series for 2010, 2011 and 2012 were combined to fill in data gaps over heavily clouded regions.

Ground probed soil moisture data were taken from all networks and stations available from the International Soil Moisture Network (ISMN) (Dorigo et al. 2011; Ochsner et al. 2013). For each station only the top most recording (usually within 5 cm from the soil surface) was used. In total 745 stations were used in this study. 451 stations with data covering at least 6 months in 2011 were used for testing the time series smoothing and divided into two random sub-sets, for model calibration (281 stations) and model validation (170 stations). The land cover for each ISMN station was extracted from the MODIS land cover product (MCD12Q1 version 051) for 2011 using the International Geosphere-Biosphere Program (IGBP) classification scheme.

The performance validation of TWI compared to in-situ data was done solely using the 2011 MODIS data, and all ISMN stations with available data for 2011 (disregarding the number of observations coinciding with the MODIS imagery).

12.4 Definition of the Transformed Wetness Index (TWI)

At its core, TWI is a normalized difference (ND) index, but rather than using original satellite image bands as inputs, the ND algorithm in TWI uses data obtained after a linear transformation of the image bands. The transformation is achieved by a fixed orthogonal matrix optimized to separate wet and dry pixels. The first transformation component aligns from dark soil reflectance to light soil reflectance, representing the soil-line (Baret et al. 1993) brightness (Eq. 12.1). The second and third components represent photosynthetic and non-photosynthetic vegetation, while the fourth represents open water (Eq. 12.2). Omitting the vegetation components and using the soil line and wetness components in an ND approach has two distinct advantages; the vegetation influence is reduced, and the index can be adjusted for local soil conditions. The calculations of MODIS TWI used in here retain the reflectance value factors (reflectance * 10^5) of the MODIS MCD43A4 product, and the soil line (sl) and wetness (w) components are calculated as:

$$\begin{aligned} sl = {} & 0.3148(RL - 563) + 0.3209(NIR - 1008) + 0.3595(BL - 147) \\ & +0.3364(GL - 507) + 0.2498(SWIRa - 1531) \\ & +0.6573(SWIRb - 1836) + 0.2471(SWIRc - 1699) \end{aligned} \tag{12.1}$$

$$\begin{aligned} w = {} & 0.1882(RL - 563) + 0.0384(NIR - 1008) + 0.4940(BL - 147) \\ & +0.3501(GL - 507) - 0.3581(SWIRa - 1531) \\ & -0.1731(SWIRb - 1836) - 0.6621(SWIRc - 1699) \end{aligned} \tag{12.2}$$

with band order given as in the MODIS reflectance products (RL = red, NIR = near infrared, BL = blue, GL = green, SWIR = short wave IR).

The TWI ND algorithm is defined by a reference line of iso-wetness and applied using a trigonometric, scale preserving, rotation combined with a re-scaling factor (R) and a calibration factor (C) allowing for non-linear reflectance mixing between soil and water:

$$TWI = R * \frac{sin(\beta + 45)(sl + w + a) + cos(\beta + 45)(-sl + w + a)}{sin(\beta + 45)(sl - w - a) + cos(\beta + 45)(sl + w + a) + C} \quad (12.3)$$

where a is the reference iso-wetness line intercept with the soil-line and β the reference iso-wetness line slope. The global values for the reference iso-wetness line and the calibration factor were determined from reference sites sampled by the author in Botswana, Uganda, Kenya and Indonesia: a (2080), β (−57°) and C (7000). Setting the re-scaling factor R to 5942 scales the TWI range to approximately −4300 for dry soil, 2000 for water saturated dark soil and 3500 for deep open water. MODIS TWI is converted to actual soil moisture, Θ_{TWI} (volume/volume) by a linear-power function:

$$\Theta_{TWI} = (TWI + 4300)/430 + 1.067^{(TWI+4300)*0.0086} \quad (12.4)$$

TWI performance was evaluated directly comparing Θ_{TWI} to in-situ data and after assimilation of Θ_{TWI} to fit the statistical moments (mean and variance) of each local in-situ time-series (Reichle et al. 2004, 2007). The assimilation was done against the in-situ data representing the same 16-day period as each Θ_{TWI} estimate. Each 24 h cycle of in-situ data, adjusted for the local noon where first calculated separately, and then averaged. The smoothing algorithms and parameterizations were evaluated by aggregating all local results, and comparing model results to in-situ data by the coefficient of determination (r^2), the Random Mean Square Error (RMSE), and model efficiency (E) (Nash and Sutcliffe 1970):

$$E = 1 - \frac{\sum (\Theta_o - \Theta_m)^2}{\sum (\Theta_o - \overline{\Theta}_o)^2} \quad (12.5)$$

where $\overline{\Theta}_o$ is the mean of observed soil moisture, and Θ_m is modeled and Θ_o observed soil moisture for matching data pairs.

12.5 Estimating Soil Moisture Dynamics from TWI

To retrieve metrics of the global annual soil moisture dynamics, the TWI estimates (Θ_{TWI}) derived for each 16-day cycle in 2011 were used. Soil moisture metrics for the annual phenology were extracted from the original TWI time-series, and after applying different smoothing procedures. Adopted smoothing models include a weighted moving average (WMA) model, and a modified LOWESS model assigning lower weights to gap-filled data points. In the latter, the weighting (w_i) combines the standard LOWESS tri-cube weight function with lower weights for gap-filled points:

$$w_i = \left(1 - \left|\frac{x - x_i}{d(x)}\right|^3\right)^3 * wf_i \tag{12.6}$$

where x is the point to be estimated, wf_i is the weight assigned for each data value i falling within the filter span of x, and $d(x)$ the maximum distance (i.e. number of days) between x and the predictors x_i. The weights for gap-filled points (wf) were varied between 0.25 and unity (step = 0.25), whereas original points were assigned a wf of unity. The maximum filter span ($d(x)$) was set to predefined values of 24 days (3 data points), 40 days (5 data points), and 56 days (7 data points). The LOWESS local polynomial fitting was done using linear regression (first degree polynomial). The weighting for the WMA used wf directly, and the same filter spans.

Different approaches were validated for (1) general smoothing, (2) reducing or removing outliers, and (3) gap-filling. The influence of outliers was tested by an iterative approach, identifying outliers after an initial smoothing by comparing the original and smoothed time-series and assigning points varying more than between 1 to 10 % units of soil moisture (step = 1 % unit) as outliers. The smoothing algorithm was then iterated but with the outlier values altered (set either to the smoothed value, or replaced by IDW gap-filling) or removed. Gap-filling was tested by first filling all data time-series using IDW and then applying the smoothing algorithm. In the combined tests of outlier smoothing and gap-filling, the IDW for gap-filling was iterated after the outlier replacement to reflect the initially smoothed outliers. Trials were made using varying weights also for outliers, and by using separate weights for outliers and gap-filled points. None of these more complex models, however improved model performances, and hence they are not presented. Tested smoothing approaches and models are summarized in Table 12.2.

All smoothing models operate on the original Θ_{TWI} time-series, and assimilation to fit in-situ data (see Sect. 12.4) was done as a post-processing step. The results of the smoothing models were then evaluated compared to the original Θ_{TWI} estimate, with all Θ_{TWI} time-series assimilated to the in-situ data. The effects of smoothing outliers and gap-filling was also separately examined using r^2 and RMSE compared to coinciding observations in the in-situ data.

Table 12.2 Summary of smoothing algorithms tested for extracting annual phenology from soil moisture time-series data

Model code	Gap-filling	Outliers	Smoothing algorithm	Smoothed data
GF	IDW	No action	None (orignal model)	None
GF_{WMA}	IDW	WMA smoothed	WMA	All
GF_{LOWESS}	IDW	LOWESS smoothed	LOWESS	All
GFO_{LOWESS}	IDW	LOWESS smoothed	LOWESS	Outliers only
$GFRO_{LOWESS}$	IDW	Replaced by IDW	LOWESS	All
T	None	No action	None (orignal model)	None
T_{LOWESS}	None	LOWESS smoothed	LOWESS	All
TO_{LOWESS}	None	LOWESS smoothed	LOWESS	Outliers only

For in-situ sites with data covering at least 10 months, the phenological performance was evaluated modified after Hird and McDermid (2009). Phenology was calculated for both assimilated Θ_{TWI} and in-situ data representing; value metrics: MIN, MAX and INT (integration of soil moisture during wet season); and temporal metrics: LS, PWS and PDS. To allow all local time-series to be evaluated in the model development, the threshold between wet and dry seasons was set to the mean wetness of each local time-series (with Θ_{TWI} assimilated to fit the in-situ data, the mean is always the same for both time-series). The score for each phenological metric (PS_{m}) was calculated as:

$$PS_m = 1 - \frac{|P_m^M - P_i^M|}{P_{max}} \tag{12.7}$$

where P^{M} is the derived phenological metric of the in-situ (i) and modeled (m) time-series of soil moisture, and P_{max} is the maximum range for each metric, simply pre-defined to a soil moisture value of 50 %, and a maximum temporal value of 365 days. To convert INT to a maximum range of 50, INT was divided by LWS. Phenological metrics for in-situ data was calculated using the full in-situ time series. The phenological performance was evaluated for each local (pixel-wise) time-series individually, and using the overall mean of all local performances for model evaluation.

Model performances were evaluated using all the criteria presented above, but primarily ranked from the smoothing model abilities to capture the soil-moisture phenology compared to the in-situ data. Model performances for the best phenological parameter settings for each smoothing model were tested against the validation data set using the same smoothing and rescaling approach as in the calibration. Exploring the results of the smoothing algorithms, the global phenology for 2011 was calculated using a LOWESS smoothing without any gap-filling (and thus no weighting) or outlier adjustment, and a filter size of 5 data points (or less if gaps occur within the filter) with a maximum distance span ($d(x)$) set to 35 days.

Figure 12.1 illustrates some of the smoothing algorithms and the effects of varying filter sizes and how the adjustment of outliers affect the smoothing results. The figure also shows the simplified LOWESS weighing algorithms selected for generating the global surface wetness phenology for 2011.

12.6 Results and Discussions

12.6.1 Transformed Wetness Index

Compared to 745 in-situ ground sites Θ_{TWI} has a bias of 2.5 % and a global RMSE of 14.0 %, which reduces to 8.5 % when assimilating the mean and variance to fit local in-situ data. The assimilated RMSE for cosmic-ray soil moisture probes (Zreda

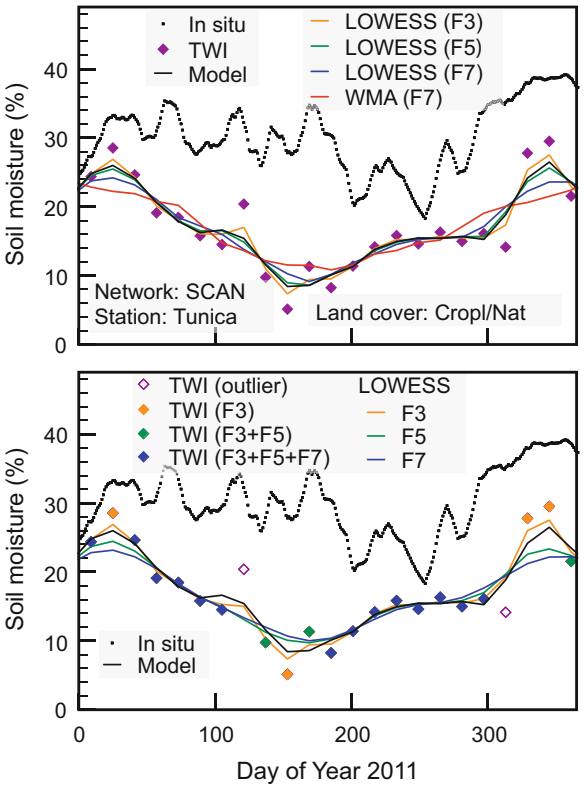

Fig. 12.1 Illustration of the effects of different smoothing parameterizations. The *upper panel* shows the results for LOWESS smoothing with filter sizes (F) of 3, 5 and 7 elements, and the results from a WMA smoothing using 7 elements. The *lower panel* shows how the smoothing is affected by identified outliers, using LOWES smoothing with filter sizes of 3, 5 and 7 elements. The identified outliers are replaced by the initial smoothing value, and the smoothing algorithm applied in a second iteration which generated the finally smoothed time-series. Both panels also show the simplified LOWESS model selected for calculating the global phenology for 2011 (see text). The example represent a real time-series without any data gaps (see Fig. 12.2)

et al. 2008), with a foot-print comparable to the resolution of MODIS images, is 5.3 % (with a bias of 3.7 %). The assimilated model efficiency for the cosmic-ray probes is 0.79, compared to an overall global model efficiency of 0.42.

Θ_{TWI} overestimates surface wetness for regions with dark surfaces, including basaltic outcrops and vertisols. Dense stands of e.g. reeds and papyrus leads to underestimations of soil moisture content over wetlands, whereas the soil moisture is overestimated in dense and moist forests (in particular over evergreen needleleaf forests). The model performance for non forested areas is hence better, with an estimated bias of −0.4 % and a global RMSE of 11.6 % (574 stations), reduced to 8.0 % for assimilated time-series. Θ_{TWI} in general underestimates the variations in soil moisture, which is probably related both to the inability of optical sensors to

capture surface conditions under clouds (i.e. during precipitation events and rainy seasons), and to differences in soil moisture variations at point scale compared to the 500 m foot-print of the MODIS images.

12.6.2 Phenological Characterization

The differences in overall smoothing model performance, whether using statistical measures or scores relating to phenological metrics, are small (Tables 12.3 and 12.4). No particular model can be identified as performing better than the others. Indeed, the unsmoothed time-series performs equally well, or better, compared to the smoothing algorithms. Further, the results of the calibration tests for each individual model only show minimal variances in performance between different parameter settings (not explicitly shown, but indicated from the results in Table 12.3). The parameters presented in Table 12.3 represent the parameterizations that best captures the phenological performances compared to in-situ data. Using other criteria (e.g. model RMSE, model efficiency E, or outlier RMSE) for selecting the optimal model lead to different parameterizations.

The largest variations between different models relate to the smoothing of outliers. Models smoothing the complete time-series (GF_{WMA}, GF_{LOWESS} and T_{LOWESS}) all indicate that a thresholds of 6 % soil moisture is optimal for identifying outliers (i.e. peaks and troughs in soil moisture larger than 6 % units over a period of approximately one to two months are likely to be erroneous). But only very few such peaks and troughs are identified (Tables 12.3 and 12.4), and the RMSE of the identified outliers themselves increase in the smoothing except for in the T_{LOWESS} model. Applying smoothing to only outliers (models GFO_{LOWESS} and TO_{LOWESS}), instead identifies a very large number of outliers as optimal. Non-outliers are not smoothed in these models, and the outlier smoothing hence replaces the general smoothing, reducing both the overall RMSE and the RMSE of the outliers themselves. The model replacing the outliers with IDW filled value ($GFRO_{\mathrm{LOWESS}}$) is the worst performing. These results indicate that in general the Θ_{TWI} time-series data have no problems related to outliers. Visual inspection of the data (Fig. 12.2), reveals that most peaks and troughs are no artifacts, but discernible also in the in-situ data. MODIS TWI rather tends to miss peaks with short duration. For any 16-day interval the MODIS data always represent cloud free conditions, more likely to represent drier ground conditions compared to cloudy conditions.

The RMSE and r^2 of the gap-filled data points show a low fit to the in-situ data (Table 12.3), and gap-filled models do not perform better compared to those with no gap-filling. With rainfall events linked to cloud cover, it is more likely that peaks in surface wetness are missed by the MODIS optical sensors, indicated in Fig. 12.2. As the gap-filling interpolates adjacent observations, gaps due to cloudy conditions with associated precipitation can not be properly estimated. The LOWESS models used in this study utilize the trends adjacent to gaps for the gap-filling, but the LOWESS filling of gaps only performs slightly better than the filling by IDW.

Table 12.3 Smoothing model calibration results, showing the performances of different smoothing algorithms compared to in-situ data. The overall results compare the fully smoothed and then assimilated time-series with the in-situ data, with performance reported as the random mean square error (RMSE), coefficient of determination (r^2), model efficiency (E), and the mean score for phenological metrics agreement ($\overline{PS}_m$). Results for the gap-filling are reported as the RMSE and r^2 compared to in-situ data coinciding with the gap-filled data points. Outlier smoothing results are reported in the same manner as for gap-filling, but also comparing the unsmoothed RMSE and r^2 for the smoothed outliers. For both gap-filling and outlier smoothing the number of validated points (and in the parenthesis, the total number of affected points) are given (n). The model results shown represent the best performing parameterzation for $\overline{PS}_m$ for each model. For the parameter settings; F = filter size, T = threshold for identifying outliers (%) and *wf* = weight for gap filled data points

	Overall results				Gap-filling			Outlier smoothing					Parameters		
								Orignal		Smoothed					
Model	RMSE	r^2	E	$\overline{PS}_m$	RMSE	r^2	n	RMSE	r^2	RMSE	r^2	n	F	T	*wf*
GF	7.84	0.53	0.46	0.85	8.34	0.29	320(1429)	–	–	–	–	–	–	–	–
GF_{WMA}	7.89	0.53	0.45	0.86	8.56	0.26	320(1429)	6.64	0.71	6.26	0.74	21(99)	3	6	0.25
GF_{LOWESS}	7.89	0.53	0.45	0.86	8.47	0.27	320(1429)	5.87	0.80	6.92	0.72	15(63)	5	6	0.25
GFO_{LOWESS}	7.87	0.53	0.45	0.86	8.41	0.28	320(1429)	8.14	0.46	8.01	0.47	1093(3299)	5	1	0.75
$GFRO_{LOWESS}$	7.90	0.52	0.45	0.86	8.75	0.22	320(1429)	8.28	0.44	8.78	0.36	616(1837)	3	1	0.75
T	7.82	0.53	0.46	0.85	–	–	–	–	–	–	–	–	–	–	–
T_{LOWESS}	7.88	0.53	0.45	0.86	–	–	–	5.93	0.77	3.23	0.93	13(60)	5	6	–
TO_{LOWESS}	7.85	0.53	0.46	0.86	–	–	–	8.06	0.48	7.91	0.50	1029(3069)	5	1	–

Table 12.4 Smoothing model validation results. See Table 12.3 for explanations

	Overall results				Gap-filling			Outlier smoothing					Parameters		
								Orignal		Smoothed					
Model	RMSE	r^2	E	$\overline{PS}_m$	RMSE	r^2	n	RMSE	r^2	RMSE	r^2	n	F	T	wf
GF	7.90	0.49	0.40	0.85	8.56	0.28	256(1260)	–	–	–	–	–	–	–	–
GF_{WMA}	8.00	0.48	0.38	0.85	8.75	0.25	256(1260)	7.49	0.42	7.07	0.49	44(144)	3	6	0.25
GF_{LOWESS}	8.00	0.48	0.40	0.85	8.75	0.25	256(1260)	7.80	0.36	7.57	0.40	30(96)	5	6	0.25
GFO_{LOWESS}	7.98	0.48	0.38	0.85	8.68	0.26	256(1260)	7.98	0.48	7.83	0.50	942(3015)	5	1	0.75
$GFRO_{\text{LOWESS}}$	8.06	0.47	0.37	0.85	8.71	0.26	256(1260)	8.12	0.47	7.62	0.47	499(1704)	3	1	0.75
T	7.90	0.49	0.40	0.85	–	–	–	–	–	–	–	–	–	–	–
T_{LOWESS}	8.0	0.48	0.38	0.85	–	–	–	7.80	0.36	7.50	0.41	30(94)	5	6	–
TO_{LOWESS}	7.96	0.48	0.39	0.85	–	–	–	7.93	0.49	7.74	0.52	895(2817)	5	1	–

Table 12.5 Production model results (including both the calibration and validation data sets), reported for both the original (not rescaled) MODIS TWI model (with wet season set to the average wetness of the in-situ data for each local site) and with each MODIS TWI local time-series assimilated to fit the mean and standard deviation of the in-situ data. Results are reported for the global dataset, and separately for cosmic-ray probes, and for forests and non-forest (as identified from the MODIS land cover product MCD12Q1.v005). See Table 12.3 for explanations. Phenological scores (Eq. 12.7) for applicable metrics are separately reported

	Overall results				Phenological scores ($\mathbf{PS_m}$)						
Regionalization	RMSE	r^2	E	$\overline{PS}_m$	MIN	PDS	MAX	PWS	LWS	IWS	MEAN
	Orginal time-series										
Global	11.1	0.02	−0.50	0.74	0.80	0.74	0.80	0.61	0.57	0.86	0.84
Cosmic-ray probes	11.3	0.06	−0.51	0.79	0.84	0.85	0.86	0.69	0.51	0.90	0.82
Non-forested sites	10.9	0.06	−0.07	0.75	0.84	0.75	0.81	0.63	0.56	0.87	0.86
Forested sites	11.3	0.0	−2.1	0.69	0.65	0.70	0.71	0.56	0.61	0.81	0.75
	Assimilated time-series										
Global	7.9	0.51	0.42	0.85	0.94	0.74	0.93	0.61	0.88	0.92	–
Cosmic-ray probes	5.0	0.82	0.80	0.89	0.96	0.85	0.95	0.69	0.89	0.95	–
Non-forested sites	7.4	0.56	0.50	0.86	0.94	0.75	0.94	0.63	0.89	0.92	–
Forested sites	9.4	0.32	0.13	0.82	0.93	0.70	0.91	0.56	0.84	0.91	–

Accepting that neither outlier reduction nor gap-filling improve the phenological extraction, but that smoothing in general removes small temporal variations (noise), a simpler LOWESS model was used for calculating global wetness phenology. The selected model smoothing filter size was set to five elements, with a maximum date span of 35 days. The advantage of five elements is primarily that, compared to a filter with only three elements, it bridges single data gaps. Reducing the maximum date span from 40 to 35 days, however reduces the weight of the more distant dates. With the identification of outliers omitted and no gap-filling (i.e. no weighting), the LOWESS weighting algorithm (Eq. 12.6) can be replaced by a simple weighting vector ($\boldsymbol{w}$) applied to each element of the original time-series:

$$\boldsymbol{w} = [0.08 \;\; 0.74 \;\; 1.0 \;\; 0.74 \;\; 0.08] \tag{12.8}$$

The results using $\boldsymbol{w}$ for weighting, and a first degree polynomial for estimating the smoothed values are summarized in Table 12.5. In general the smoothing reduces the variance, but otherwise does not affect model performance. The global phenology value metrics (% soil moisture) for mean (MEAN) and standard deviation (SD) are shown in Fig. 12.3, with additional global maps shown as Figs. 12.5 and 12.6 in the Appendix.

The smoothing algorithm adopted for calculating wetness phenology reduces the variance compared to the original time-series data. The Θ_{TWI} original model underestimates the variance compared to point scale in-situ data, and this underestimation is further accentuated by the smoothing. Comparing the smoothing results

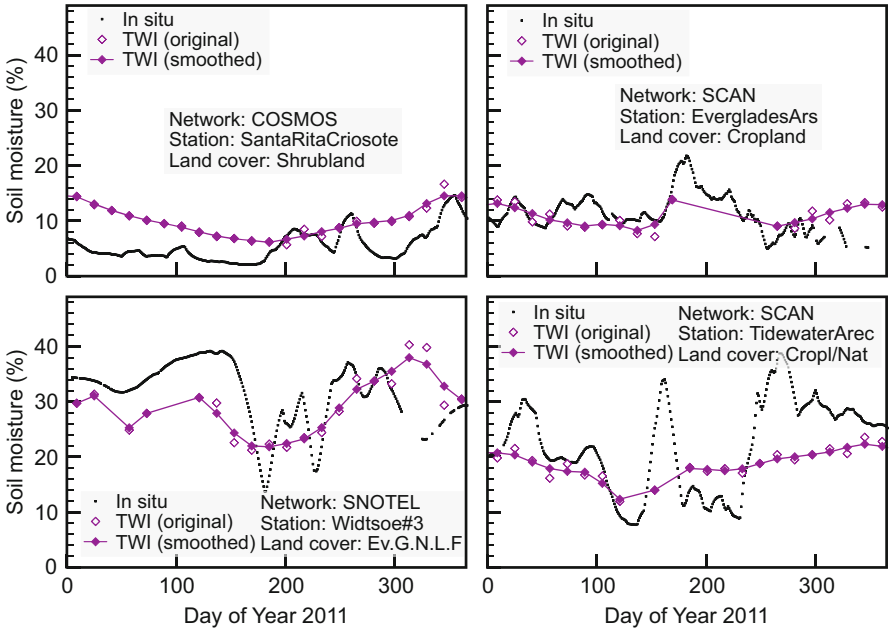

Fig. 12.2 Comparison of soil moisture observations from in-situ station data and soil moisture estimated by the MODIS Transformed Wetness Index (TWI) (expressed as volumetric soil moisture – Θ_{TWI}). Each panel shows the MODIS TWI original (not rescaled) time-series, and the same time-series smoothed using the simplified LOWESS weighting algorithm selected for calculating the global phenology for 2011 (see text). The four panels represent different networks and land cover (from the MODIS land cover product MCD12Q1.v051) as indicated in each graph (Ev.G.N.L.F. = evergreen needleleaf forest; Cropl/Nat = cropland/natural vegetation mosaic). Details about each station are available in the ISMN dataset

with cosmic-ray probes (measuring soil moisture at a spatial scale equalling that of the MODIS pixel size) the variances of the Θ_{TWI} time-series are adjusted to closer matching the variances captured by the cosmic-ray probes. Arguably, the results for the cosmic-ray probes more correctly represent the ability of Θ_{TWI} to capture the spatially integrated soil moisture regime.

For non-forested sites the Θ_{TWI} estimated soil moisture bias is negligible (−0.4 %) and does only marginally affect extraction of the wetness phenology. The phenology is also better captured for non-forested sites compared to forested sites (Table 12.5). Large positive bias over primarily evergreen needleleaf forests (16.2 %), but also other forests (10 %), cause over-estimation of both temporal metrics and value metrics when retrieving soil moisture. The Okavango Swamps in Botswana (Fig. 12.4 and Appendix Fig. 12.7) are less affected. The maps over the Okavango Swamps clearly capture both the wetter sites, but also separates the more permanent (central swamps) from the distal floodplains.

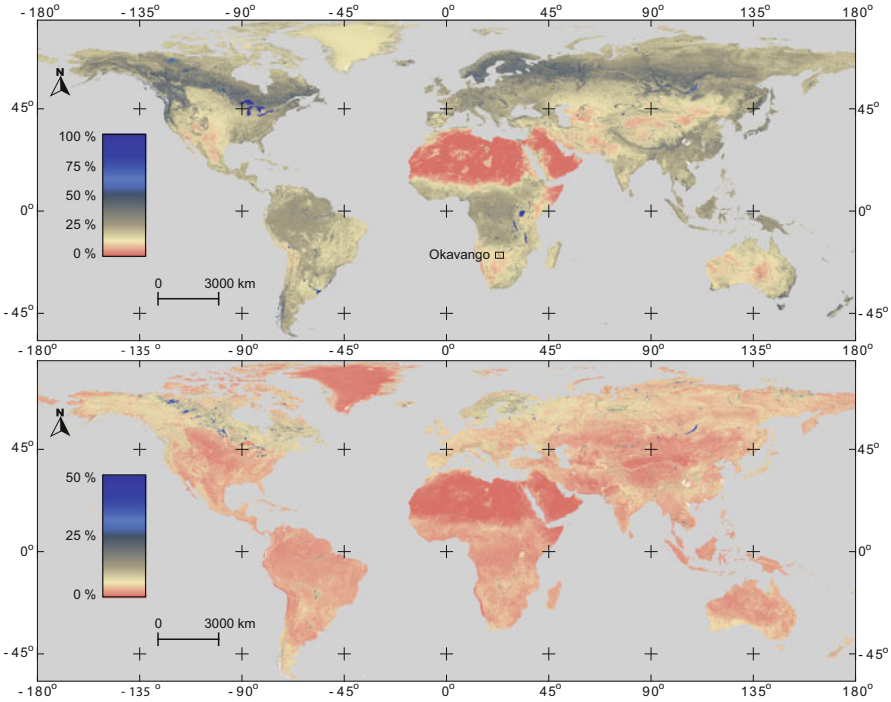

Fig. 12.3 Mean (MEAN) (*top panel*) and standard deviation (SD) (*bottom panel*) soil surface wetness (%) metrics for 2011 estimated from the Moderate-resolution imaging spectroradiometer (MODIS) Transformed Wetness Index (TWI). The wetness phenology metrics is calculated after applying a locally weighted scatterplot smoothing (LOWESS) to the original MODIS TWI time-series (expressed as volumetric soil moisture – Θ_{TWI}). For the tropical region (between 20 degrees latitudes) the map represents the average conditions for 2010, 2011 and 2012, whereas for other regions it represents the calendar year 2011. Note that the scaling is different for MEAN and SD. White areas have too few recordings for extracting phenological metrics

12.7 Conclusions

This study has attempted to define metrics and identify smoothing algorithms for capturing the annual soil moisture phenology from Θ_{TWI}, a soil wetness index derived from MODIS data. While most approaches for retrieving time-series of e.g. vegetation, snow or ice cover are based on the assumption that proximity in time or space is the key for cleaning and smoothing, this assumption is challenged when mapping variable properties with unknown or unpredictable scales of spatial and temporal variation.

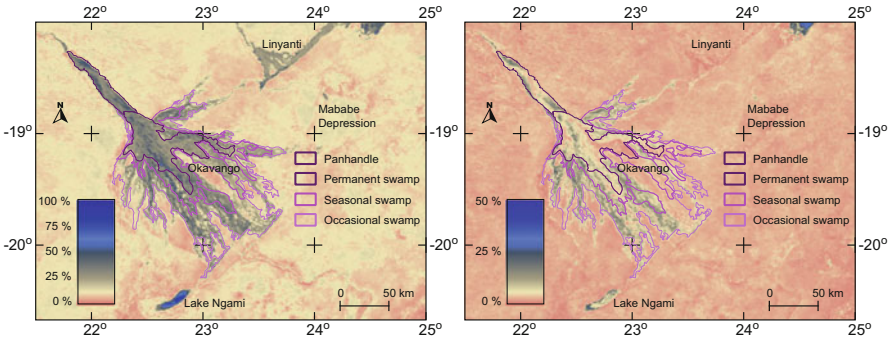

Fig. 12.4 Mean (MEAN) (*left panel*) and standard deviation (SD) (*right panel*) phenological soil moisture metrics for the Okavango Swamps, Botswana for 2011 (see Fig. 12.3). The wetland mask is modified after McCarthy et al. (2003) and divided into four regions: (1) the Panhandle tectonically bound entry channel, (2) the central permanent wetlands, (3) the seasonally inundated wetlands, and (4) the most distal areas only temporally flooded. Okavango is situated on a tectonically active region and the swamp naturally oscillates across the underlaying alluvial fan. Compared to the estimated area of wetlands between 1972 and 2000, the maps indicate that the north-eastern parts of the swamps are drying out and the south-western parts receive more water. Further, the maps reveal that Lake Ngami was filled during 2011, whereas it contained no or little water from 1972 to 2000

Metrics for the global surface wetness phenology have been developed at a spatial scale of 500 m. A large bias and high RMSE for ecosystems dominated by dense forests causes overestimated wetness conditions in particular for higher latitude evergreen needleleaf forests. The phenology metrics for non forested landscapes, and large (non forested) wetlands are better captured. Phenology is better captured when compared to in-situ data captured by cosmic-ray probes, integrating the soil moisture over a spatial scale closely matching the MODIS image resolution.

The LOWESS smoothing algorithm used in this study is a compromise, but does have several advantages. It is flexible by allowing selection of both filter size (i.e. defining the temporal range that influences the soil wetness conditions at any particular date), and customized weighting (and also by selecting different fitting functions, but this was not tested in this study). But primarily LOWESS can handle irregular time-series data, and thus removes the need for pre-defining e.g. rates of changes and seasonality commonly required by more advanced smoothing algorithms and model fitting routines. As the MODIS product used in this study is delivered with regular interval time-steps (16 days), the LOWESS weighing function could be replaced by a fixed vector, also decreasing the computational processing time manyfold.

Acknowledgements This work was supported by USAID (Grant No. EEM-G-00-04-00010-00) as part of the CGIAR research programme on Forests, Trees and Agroforestry.

Appendix

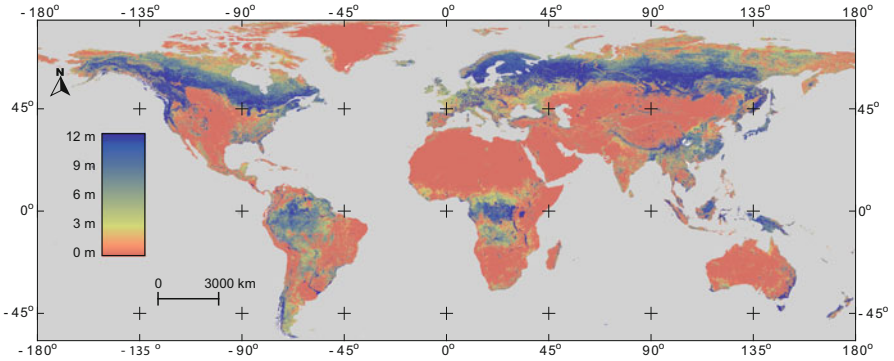

Fig. 12.5 Lengths of wet season (LWS) (months (m) with soil wetness >25 % in 2011). The overestimation of soil moisture content under forests, in particular evergreen needleleaf forests causes a large positive bias in temperate forests

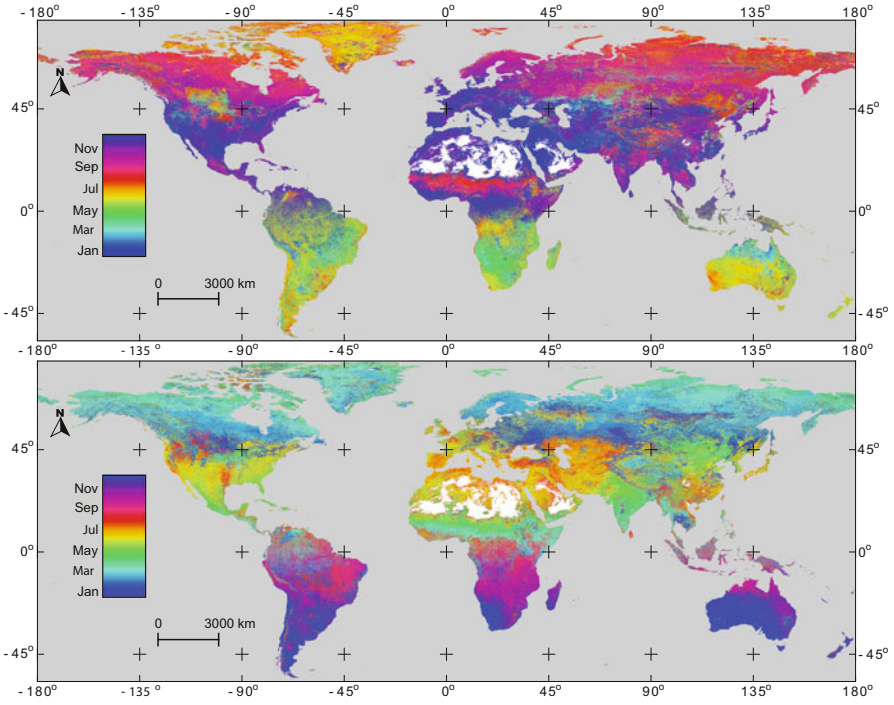

Fig. 12.6 Day of year with highest (PWS – *top panel*) and lowest (PDS – *bottom panel*) recorded wetness 2011. The map clearly shows the spatially mirrored wetness phenology across the equator. Large parts of the Sahara Desert have no discernible seasonality, and no dates for peak wet and dry seasons can be retrieved

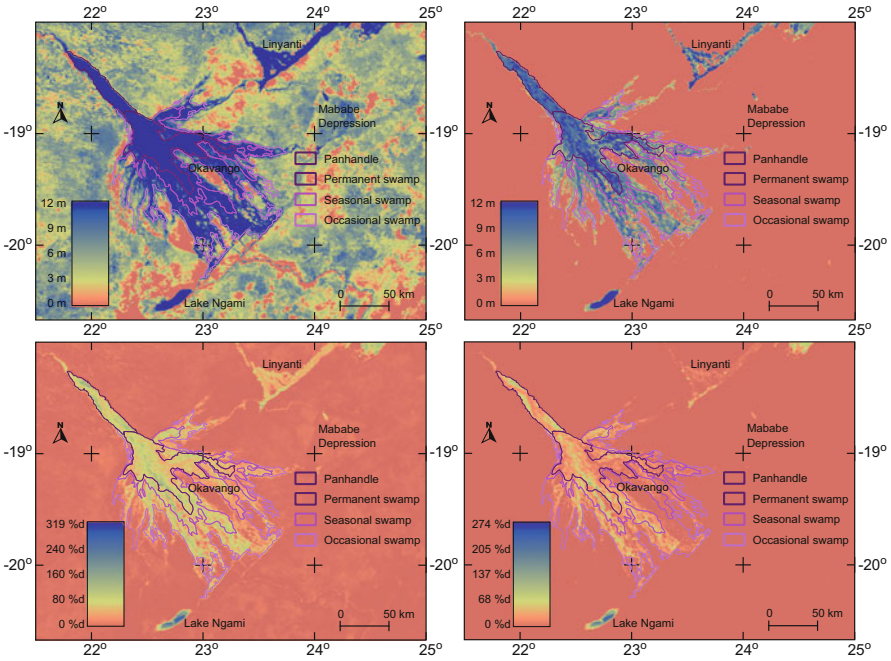

Fig. 12.7 Illustration of soil moisture phenological metrics for the Okavango Swamps in Botswana. The *upper panels* show the Lengths of non-dry season (LDS) and Wet season (LWS), while the *lower panels* show the integrated non-dry season (IDS) and integrated wet season (IWS). All panels represent the calendar year 2011. The location of the Okavango swamps is indicated in Fig. 12.3

References

Atkinson PM, Jeganathan C, Dash J, Atzberger C (2012) Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. Remote Sens Environ 123:400–417. doi:10.1016/j.rse.2012.04.001

Baret F, Jacquemoud S, Hanocq JF (1993) The soil line concept in remote sensing. Remote Sens Rev 7(1):65–82. doi:10.1080/02757259309532166

Beck PSA, Atzberger C, Hogda KA, Johansen B, Skidmore AK (2006) Improved monitoring of vegetation dynamics at very high latitudes: a new method using MODIS NDVI. Remote Sens Environ 10:321–334

Brady NC, Weil RR (2007) The nature and properties of soils, 14th edn. Prentice Hall, Upper Saddle River

Cheema M, Bastiaanssen W, Rutten M (2011) Validation of surface soil moisture from AMSR-E using auxiliary spatial data in the transboundary Indus Basin. J Hydrol 405(1–2):137–149. doi:10.1016/j.jhydrol.2011.05.016

Cleveland WS (1979) Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc 74:829–826

Cleveland WS, Devlin SJ (1988) Locally weighted regression: an approach to regression analysis by local fitting. J Am Stat Assoc 83(403):596–610

de Beurs KM, Henebry GM (2010) Spatio-temporal statistical methods for modeling land surface phenology, chap 9. In: Hudson IL, Keatley MR (eds) Phenological research. Springer Netherlands, Dordrecht, pp 177–208. doi:10.1007/978-90-481-3335-2

Dorigo WA, Wagner W, Hohensinn R, Hahn S, Paulik C, Xaver A, Gruber A, Drusch M, Mecklenburg S, van Oevelen P, Robock A, Jackson T (2011) The international soil moisture network: a data hosting facility for global in situ soil moisture measurements. Hydrol Earth Syst Sci 15(5):1675–1698. doi:10.5194/hess-15-1675-2011

Draper CS, Walker JP, Steinle PJ, de Jeu RAM, Holmes TRH (2009) An evaluation of AMSR–E derived soil moisture over Australia. Remote Sens Environ 113(4):703–710. doi:10.1016/j.rse.2008.11.011

Gumbricht T (2015) Hybrid mapping of pantropical wetlands from optical satellite images, hydrology, and geomorphology, chap 20. In: Tiner R, Lang MW, Klemas VV (eds) Wetland classification and mapping. CRC Press/Taylor and Francis, pp 435–455. doi:10.1201/b18210-24

Haddeland I, Clark DB, Franssen W, Ludwig F, VoßF, Arnell NW, Bertrand N, Best M, Folwell S, Gerten D, Gomes S, Gosling SN, Hagemann S, Hanasaki N, Harding R, Heinke J, Kabat P, Koirala S, Oki T, Polcher J, Stacke T, Viterbo P, Weedon GP, Yeh P (2011) Multimodel estimate of the global terrestrial water balance: setup and first results. J Hydrometeorol 12(5):869–884. doi:10.1175/2011JHM1324.1

Heumann B, Seaquist J, Eklundh L, Jönsson P (2007) AVHRR derived phenological change in the Sahel and Soudan, Africa 1982–2005. Remote Sens Environ 108(4):385–392. doi:10.1016/j.rse.2006.11.025

Hird JN, McDermid GJ (2009) Noise reduction of NDVI time series: an empirical comparison of selected techniques. Remote Sens Environ 113(1):248–258. doi:10.1016/j.rse.2008.09.003

Jackson TJ, Cosh MH, Bindlish R, Starks PJ, Bosch DD, Seyfried M, Goodrich DC, Moran MS, Du J (2010) Validation of advanced microwave scanning radiometer soil moisture products. IEEE Trans Geosci Remote Sens 48(12):4256–4272. doi:10.1109/TGRS.2010.2051035

Jones MO, Jones LA, Kimball JS, Mcdonald KC (2011) Satellite passive microwave remote sensing for monitoring global land surface phenology. Remote Sens Environ 115(4):1102–1114. doi:10.1016/j.rse.2010.12.015

Jönsson P, Eklundh L (2004) TIMESAT – a program for analyzing timeseries of satellite sensor data. Comput Geosci 30:833–845

Kang K, Duguay CR, Howell SEL (2012) Estimating ice phenology on large northern lakes from AMSR-E: algorithm development and application to Great Bear Lake and Great Slave Lake, Canada. The Crysophere 6:235–254. doi:10.5194/tc-6-235-2012

McCarthy JM, Gumbricht T, McCarthy T, Frost P, Wessels K, Seidel F (2003) Flooding patterns of the Okavango Wetland in Botswana between 1972 and 2000. AMBIO 32(7):453–457. doi:10.1579/0044-7447-32.7.453

Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models part I – a discussion of principles. J Hydrol 10(3):282–290. doi:http://dx.doi.org/10.1016/0022-1694(70)90255-6

Ochsner TE, Cosh MH, Cuenca RH, Dorigo WA, Draper CS, Hagimoto Y, Kerr YH, Njoku EG, Small EE, Zreda M (2013) State of the art in large-scale soil moisture monitoring. Soil Sci Soc Am J 77(6):1888. doi:10.2136/sssaj2013.03.0093

Papa F, Prigent C, Aires F, Jimenez C, Rossow WB, Matthews E (2010) Interannual variability of surface water extent at the global scale, 1993–2004. J Geophys Res 115(D12):D12111. doi:10.1029/2009JD012674

Reichle RH, Koster RD, Dong J, Berg AA (2004) Global soil moisture from satellite observations, land surface models, and ground data: implications for data assimilation. J Hydrometeorol 5(3):430–442

Reichle RH, Koster RD, Liu P, Mahanama SPP, Njoku EG, Owe M (2007) Comparison and assimilation of global soil moisture retrievals from the advanced microwave scanning radiometer for the Earth observing system (AMSR-E) and the scanning multichannel microwave radiometer (SMMR). J Geophys Res Atmos 112(D9):D09,108. doi:10.1029/2006JD008033

Tan B, Morisette JT, Wolfe RE, Gao F, Ederer GA, Nightingale J, Pedelty JA (2011) An enhanced TIMESAT algorithm for estimating vegetation phenology metrics from MODIS data. IEEE J Select Top Appl Earth Observ Remote Sens 4(2):361–371. doi:10.1109/JSTARS.2010.2075916

Wagner W, Scipal K, Pathe C, Gerten D, Lucht W, Rudolf B (2003) Evaluation of the agreement between the first global remotely sensed soil moisture data with model and precipitation data. J Geophys Res Atmos 108(D19):4611. doi:10.1029/2003JD003663

Zreda M, Desilets D, Ferré TPA, Scott RL (2008) Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic-ray neutrons. Geophys Res Lett 35(21):L21,402. doi:10.1029/2008GL035655