Journal of Hydrology 553 (2017) 88–104

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

Performance of SMAP, AMSR-E and LAI for weekly agricultural drought forecasting over continental United States

Di Liu [a,b,*], Ashok K. Mishra [b,*], Zhongbo Yu [a], Chuanguo Yang [a], Goutam Konapala [b], Tue Vu [b]

[a] State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, HoHai University, Nanjing 210098, Jiangsu, China
[b] Glenn Department of Civil Engineering, Clemson University, 202 Lowry Hall, Clemson, SC 29634, USA

ARTICLE INFO

Article history:
Received 14 February 2017
Received in revised form 13 July 2017
Accepted 25 July 2017
Available online 27 July 2017
This manuscript was handled by A. Bardossy, Editor-in-Chief, with the assistance of Hamid Moradkhani, Associate Editor

Keywords:
Soil water deficit index
Data driven technique
Remotely sensed products
Data assimilation

ABSTRACT

This study applied support vector machines (SVMs) and data assimilation (DA) methods to investigate the performance of in-situ and remotely sensed products (i.e., leaf area index (LAI), AMSR-E and SMAP soil moisture retrievals) for near-real time agricultural drought forecasting for in-situ stations located in continental United States (CONUS). The agricultural drought was quantified using soil water deficit index (SWDI) derived based on available soil moisture and basic soil water parameters. It is observed that SVMs or SVM-DA with limited meteorological variables as inputs able to forecast SWDI at most of the in-situ stations up to 1–2-week lead time. Addition of remotely sensed products (i.e., LAI, AMSR-E, SMAP) either individually or simultaneously as inputs to SVMs can able to improve SWDI forecast at most of the stations where the strong relationship exists between LAI (and/or AMSR-E, SMAP) with SWDI. Such improvement can persist up to 2–4-week lead time at some of the stations. But the efficiency tends to decrease with the increase in lead time. The addition of both LAI and SMAP (AMSR-E) performs better than the independent addition of LAI or SMAP (AMSR-E). Typically, the performance of drought prediction varies at local scales; therefore it is difficult to generalize our findings at regional scale.

1. Introduction

Soil moisture plays an important role in the global water and energy cycle. The anomalies of soil moisture may exert great impact on the subsequent climate variables, thus, leading to the evolution of climate extremes (e.g., flood, drought and heat wave) by linking hydroclimatic fluxes at different spatio-temporal scales (e.g., Jaeger and Seneviratne, 2011; Liu et al., 2014a, b). Among the three commonly recognized categories of drought (i.e., meteorological, hydrological and agricultural) (Mishra and Singh, 2010), agricultural drought has a more direct and immediate impact on food security. Generally, agricultural drought is considered to begin when the soil moisture availability to plants drops to a level, which would adversely affects the crop yield and, hence, agricultural production (Panu and Sharma, 2002; Mishra et al., 2015). Therefore, it is reasonable to consider soil moisture databases as one potential resources for agricultural drought monitoring.

Currently, the soil moisture information at a coarser resolution is available for different sectorial applications with the development of remote sensing techniques, such as, the Advanced Microwave Scanning Radiometer-Earth Observing System sensor (AMSR-E) (Njoku et al., 2003), the Advanced Scatterometer (ASCAT) soil moisture retrievals (Wagner et al., 1999), the Soil Moisture and Ocean Salinity (SMOS) Soil Moisture Level 2 User Data Product (SMUDP2 file) (Kerr et al., 2010) and the newly launched Soil Moisture Active Passive (SMAP) mission (Entekhabi et al., 2010). These remotely sensed SM retrievals can provide an unprecedented spatial and temporal resolution of SM data across a range of scales, but is limited in terms of sensing at different depth. There is an ongoing effort to retrieve (develop) root zone SM profile using the surface SM values (Li et al., 2012; Tran et al., 2013; Ridler et al., 2014).

The growing of multiple SM databases activate the proposition of SM related drought indices in the agricultural drought monitoring (e.g., Hunt et al., 2009; Martínez-Fernández et al., 2015, 2016; Mishra et al., 2015; Keshavarz et al., 2014; Scaini et al., 2015; Sridhar et al., 2008). For example, Sridhar et al. (2008) and Hunt et al. (2009) proposed and revised the soil moisture index (SMI) to characterize the water deficit in the soil for drought. Mishra et al. (2015) applied the Standardized Soil Moisture Index (SSMI) at different soil layers and the Standardized Soil Water Availability Index (SSWI) to diagnose the agricultural drought severity based

* Corresponding authors at: State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, HoHai University, Nanjing 210098, Jiangsu, China (D. Liu); Glenn Department of Civil Engineering, Clemson University, 202 Lowry Hall, Clemson, SC 29634, USA (A.K. Mishra).

E-mail addresses: liudi@hhu.edu.cn, lenaliudi@163.com (D. Liu), ashokm@g.clemson.edu (A.K. Mishra).

http://dx.doi.org/10.1016/j.jhydrol.2017.07.049

on crop modeling. Scaini et al. (2015) used SMOS-derived SM anomalies for drought monitoring assessment. Martínez-Fernández et al. (2015) defined the soil water deficit index (SWDI) using soil moisture series solely for agricultural drought monitoring based on the percentile method proposed by Hunt et al. (2009). However, unlike the SMI defined in (Hunt et al., 2009; Sridhar et al., 2008), the deficit threshold for SWDI is fixed by the field capacity instead of the wilting point. Martínez-Fernández et al. (2016) further assessed the SWDI for agricultural drought monitoring with SMOS soil moisture retrievals.

However, some previous literatures stated that the water supplied to crops depends on the available SM in the root zone. Therefore, agricultural drought should not be treated equal for all crops, instead, it should be calculated based on the root zone SM rather than at a fixed soil layer depth (Mishra et al., 2015). Among the proposed SM related drought index, the SWDI proposed by Martínez-Fernández et al. (2015) directly include the available SM at the root zone for the crop growth. SWDI considered to be an effective agricultural drought index as it is based on soil moisture and basic soil water parameters, including field capacity and wilting point (Martínez-Fernández et al., 2015, 2016). Further, the SWDI can provide accurate status of agricultural drought because it is related to the soil moisture suction capacity of different type of crops. SWDI is a simple and useful way for the drought monitoring over large area with the increasing availability of soil moisture databases worldwide. The root zone SM predictions are important for agricultural drought monitoring and forecasting and to improve water resources management. But it is still a challenge to accurately predict the SM that varies in depth, space and time.

According to previous studies, the data-driven techniques, such as support vector machines (SVMs), are able to estimate or predict SM at root zones with antecedent meteorological variables (i.e., surface air temperature, precipitation, solar radiation, relative humidity) (e.g., Gill and McKee, 2007; Kaheil et al., 2008; Liu et al., 2010; Yu et al., 2012; Liu et al., 2016). In addition to that, data assimilation (DA) is proved to be a promising technique to optimally estimate land surface's variables (i.e., SM) by merging observed information into models (e.g., Kumar et al., 2008; Das et al., 2011; Li et al., 2012; Tran et al., 2013; Han et al., 2014; Kornelsen and Coulibaly, 2014; Yin et al., 2015; Montzka et al., 2011). Ensemble Kalman Filter (EnKF) is a kind of Monte Carlo approximation of sequential Bayesian filtering process, which alternates between an ensemble forecast step and a state variable update step (Reichle et al., 2002). EnKF is a popular DA technique used in hydrology (Gill and McKee, 2007; Liu et al., 2010; Yin et al., 2014; Mishra et al., 2015; Yin et al., 2015). The dual EnKF technique requires two separate state-space representation for the state and parameters through two interactive filters by updating model parameters and model states, which makes it more efficient in the improvement of estimations (i.e., Liu et al., 2016; Moradkhani et al., 2005).

This study investigates the potential of individual SVMs or coupled with dual EnKF DA technique (SVM-DA) to predict agricultural drought using in-situ meteorological variables and the remotely sensed products (i.e., soil moisture, leaf area index). Although the remote sensing products are applied for drought forecasting, however, the application of SMAP is limited. This study can provide insight to the potential improvements that can be achieved in agricultural drought forecasting by addition of a variety of remote sensing products including SMAP. The main objectives are: (1) To investigate the efficiency of meteorological variables as input of SVM for the agricultural drought forecast at subsequent weekly time scales; (2) To assess whether the addition of remotely sensed products as forcing variables can improve the performance of drought forecasting; and (3) To evaluate the performance of coupled SVM-DA technique to forecast agricultural drought in multiple in-situ stations located in CONUS.

The remaining structure of the paper is organized as follows: data description is given in Section 2; the model setup, experimental design and evaluation methodologies are introduced in Section 3; the results are analyzed in Section 4 and Section 5 concludes the main findings.

2. Data description

2.1. In-situ and remotely sensed data sets

The in-situ observed hydrometeorological variables, including air temperature (T2m), solar radiation (SR), precipitation (P), relative humidity (RH), and soil moisture (SM) at different soil depth from 5 cm to 100 cm, obtained from the United States Climate Reference Network (USCRN) stations with automated measured instrument are adopted in this study. The hourly, daily and monthly data are available at USCRN website (https://www.ncdc.noaa.gov/crn/qcdatasets.html). The daily time series during October 1st 2009 and March 31st 2016 are used in this study.

The AMSR-E Global daily 25 km soil moisture data (http://nsidc.org/data/docs/daac/ae_land3_l3_soil_moisture.gd.html) (Njoku et al., 2003) during October 1st 2009 and September 30th 2011, and the SMAP L3 Radiometer Global daily 36 km EASE-Grid soil moisture Version 1 (SPL3SMP) during April 1st 2015 to March 31st 2016 are selected to provide the surface SM data. To be convenient, the remotely sensed soil moisture, including AMSR-E and SMAP, are shortened as RSM in the following.

Compared to AMSR-E data, the SMAP satellite mission was newly launched by NASA in January 2015 (Entekhabi et al., 2010) to retrieve global soil moisture information via measuring brightness temperature through geophysical inversion. SMAP products are defined at different levels: Level 2 for half orbit based, Level 3 for daily composites and Level 4 for model assimilation (Brown et al., 2013). Pan et al. (2016) pointed out that the Level 2 and Level 3 are essentially no difference lying in the mid-latitudes where the neighboring swaths do not overlap. SMAP L3 is an estimate of surface soil moisture within the top 5 cm of the soil column in conjunction with a forward model of brightness temperature and ancillary products (O'Neill et al., 2016). In this study, the passive L-band radiometer 36 km resolution product is selected for the drought forecasting, because the active L-band radar failed after two months in orbit. The details about the retrieved and processed approach for SMAP L3 Passive data can be found in O'Neill et al. (2015).

The MCD15A3H version 6 MODIS Level 4 Leaf Area Index (LAI) products at 4-day composite set with 500 meter pixel size (Myneni et al., 2015) is selected to provide the LAI information to complement the in-situ observed meteorological variables.

The "near-neighbor" method is used to select the remotely sensed data for each study site over CONUS. The "near-neighbor" area is set within 0.25 degree of each site. For example, if the longitude and latitude of one site is [−86.0E 34.3N], then the remotely sensed data (i.e., AMSR-E, SMAP, LAI) within longitude [−86.25E −85.75E], latitude [34.05N 34.55N] are selected and averaged for use in this study.

2.2. Selected experimental locations

Considering the consistency in data availability for in-situ meteorological data and remote sensing data, two time periods are selected for the analysis: (a) 1st period based on available AMSR-E data between October 1st 2009 and September 30th 2011; and (b) 2nd period based on available SMAP data between

April 1st 2015 and March 31st 2016. The selected study stations are plotted in Fig. 1 (details shown in Table S1). To better describe the position, hydrologic characteristic and main land use type of each site, the 18 hydrologic unit (shorten as H01 to H18, defined in Table 2) and 8 subregions (Table 3) defined in Notaro et al. (2006) (Fig. S1) are provided to classify each site. There are total 14 study stations during the 1st time period; while 59 study stations during the 2nd time period, which includes the 14 study stations at the 1st time period. The overlapped 14 stations studied at both the 1st and 2nd time period mainly located over the Midwest (S4, S5, S7, S10, S14), Southeast (S1, S2, S3, S8, S12), Northern Great Plains (S6, S9, S11,) and Southern Great Plains areas (S13).

2.3. Soil water deficit index

The SWDI is calculated based on in-situ soil moisture information available in soil layers from 5 cm to 50 cm depth using Eqs. (1)–(3) according to the methods used in Martínez-Fernández et al. (2015, 2016):

$$SM_{0-50} = \frac{5SM_5 + 5SM_{10} + 10SM_{20} + 30SM_{50}}{50} \tag{1}$$

$$SWDI = \left(\frac{\theta - \theta_{FC}}{\theta_{AWC}}\right) \tag{2}$$

$$\theta_{AWC} = \theta_{FC} - \theta_{WP} \tag{3}$$

where θ is the soil water content, which equals to SM_{0-50}. θ_{FC} denotes field capacity and θ_{WP} is the wilting point. θ_{AWC} is the available water content, which is the difference between θ_{FC} and θ_{WP}. The wilting point (WP) and field capacity (FC) are calculated based on the percentile method proposed by Hunt et al. (2009), which considers 5th percentile of the observed SM time series during the growing season as θ_{WP} and 95th percentile as θ_{FC}. Here, the 5th and 95th percentile is derived from the accumulated SM_{0-50} time series during the crop growing season (April 1 to October 31) in each year from 2009 to 2016.

As illustrated in Allen et al. (1998), Martínez-Fernández et al. (2015) and Martínez-Fernández et al. (2016), the soils have an excess water when the SWDI is positive; while the soil is at the field capacity when it equals zero; and negative SWDI values indicate agricultural drought. When the SWDI ≤ -1, the water deficit is absolute (wilting point), and at this point, the soil water content is below the lower limit of available water for plants (Savage et al., 1996).

3. Experimental design and evaluation

3.1. Experimental design

The SVMs and dual EnKF technique used in (Liu et al., 2016) is adopted for the near real time SWDI forecast. Consistently, for SVMs, it has training and predicting processes, while the SVM-DA approach include one more step (i.e., updating process). For example, consider the SVM-DA framework (Fig. S2), which suggests if the observed data are available at the predicted time, then the initial predictions (Y_p) and observed data are used to update the

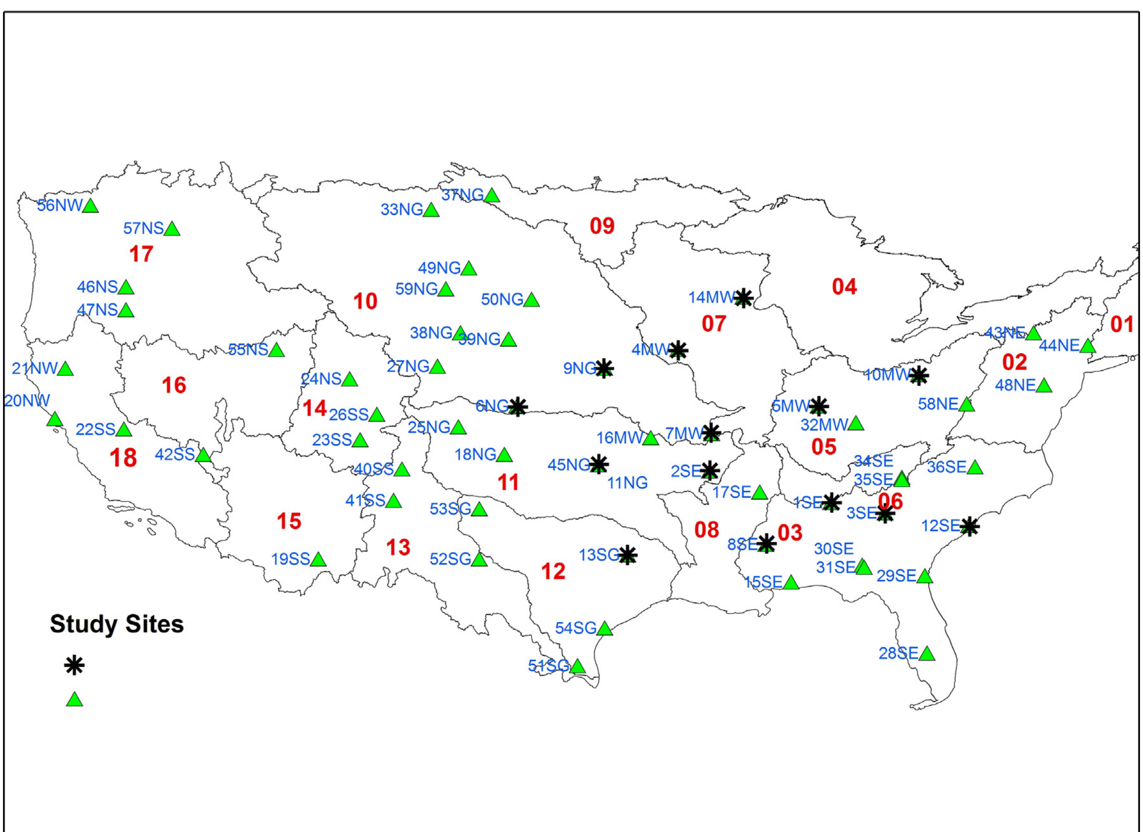

Fig. 1. Location of in-situ Stations used in the analysis. All the in-situ Stations (Black Star and Green Triangle) are used for SMAP data analysis, whereas Black Star represent the location of each Station for both AMSR-E and SMAP data analysis. Data length for AMSR-E is during October 1st 2009 to September 30th, 2011 (named as the 1st time period); for SMAP is during April 1st 2015 to March 31st 2016 (named as the 2nd time period). Red number for the 18 hydrologic regions defined in Table 2 over CONUS. Blue character for the number of the selected 59 Stations and their land use type categorized based on the 8 subregions defined in Table 3 (e.g., 1SE for the 1st Station and located within Southeast).

Table 1
Experimental design for weekly soil water deficit index (SWDI) forecast at subsequent 1–4 weeks lead time.

Cases	Abbreviation	Input	Output	Kernel function	Kernel Parameter	Length of Training Set	Tuning	Regulariza-tion Parameter	Cost Function	Model Update	Predict
Case1	C1	A1	B	K	S1	50/25	T1	G1	C1	M1	F1
Case2	C2	A2	B	K	S1	50/25	T1	G1	C1	M1	F1
Case3	C3	A3	B	K	S1	50/25	T1	G1	C1	M1	F1
Case4	C4	A4	B	K	S1	50/25	T1	G1	C1	M1	F1
Case5	C5	A1	B	K	S1	50/25	T1	G1	C1	M2	F1

Abbreviation:
T2m: average daily air temperature (Celsius); **SR:** total daily solar energy (WJ/m^2);
P: daily total precipitation (mm); **RH:** average relative humidity (%);
LAI: leaf area index; **RSM:** bias corrected remote sensing soil moisture (AMSR-E, SMAP);
SWDI: soil water deficit index within 50 cm soil layer depth.
Baseline:
A1 (T2m, SR, P, RH); **A2** (T2m, SR, P, RH, LAI)
A3 (T2m, SR, P, RH, RSM); **A4** (T2m, SR, P, RH, LAI, RSM)
B (SWDI)
F1: Antecedent 3 weeks averaged data as input to predict SWDI at the subsequent 1–4 week lead time;
K = {RBF-kernel}; **S1** = {auto selected}; **T1** = {crossvalidation}; **G1** = {auto selected}; **C1** = {mse, error = 5};
M1: {no data assimilation}; **M2**: {assimilated with in-situ derived SWDI at weekly time scale}.
Training and Predicting Ensembles:
1st study time period: October 1st 2009 to September 30th 2011 (total 98 weeks) with previous 50 weeks as the training ensemble while the following for the prediction test.
2nd study time period: April 1st 2015 to March 31st 2016 (total 45 weeks) with previous 25 weeks as the training ensemble while the following for the prediction test.
Length of Training Set: (50/25) means the 50 weeks and 25 weeks for the training ensemble size for the 1st and 2nd study time period, respectively.

Table 2
Description of the major geographic areas or regions over the continental United States based on the first level classification of Hydrologic Unit.

Code	Name	Code	Name	Code	Name
01	New England	07	Upper Mississippi	13	Rio Grande
02	Mid-Atlantic	08	Lower Mississippi	14	Upper Colorado
03	South Atlantic-Gulf	09	Souris-Red-Rainy	15	Lower Colorado
04	Great Lakes	10	Missouri	16	Great Basin
05	Ohio	11	Arkansas-White-Red	17	Pacific Northwest
06	Tennessee	12	Texas-Gulf	18	California

**Note:* USGS divides the United States into some smaller hydrological units successively which are classified into four levels: regions, sub-regions, accounting units, and cataloging units (Seaber, et al., 1987), named as "hydrological unit code" (HUC). The first-level of hydrological units, region, comprises either the drainage area of a major river (e.g. Missouri region) or the combined drainage areas of a series of rivers (e.g. Texas-Gulf region). Every region is named by a unique two-digital number by USGS, i.e. HUC2.

Table 3
Land use type over the 8 sub regions defined in Notaro et al. (2006).

Subregion	Latitude	Longitude	Abbreviation	Main Crop Type
Northern Great Plains	34.4–49N	105–96W	NG	Grassland, Cropland
Southern Great Plains	25–34.4N	105–96W	SG	Grassland, Cropland, Shrubland
Northern Shrubland	40–49N	119.9–105W	NS	Shrubland, Evergreen Forest
Southern Shrubland	30.8–40N	119.9–105W	SS	Shrubland
Midwest	37–45N	96–80W	MW	Cropland
Southeast	25–37N	96–75W	SE	Evergreen Forest, Cropland
Northwest	38–49N	124–120W	NW	Evergreen Forest, Shrubland
Northeast	37–47.5N	80–67W	NE	Deciduous Forest

model state and provide forcing data with the EnKF technique; the predictions from this updated model state and input forcing data are named as the 1st update (Y_d). Then the observed data is used to update the posterior Y_d with EnKF technique to get the final prediction (Y_{du}) named as the 2nd update. The predictions obtained at the 2nd update are the final results from the SVM-DA. The observed error and modeled error are assumed to be Gaussian distribution. In this study, the input variable error is assumed to be Gaussian distribution with 1% of mean variables of each input variable (e.g., T2m, SR, P, RH, LAI, RSM) in the initial training ensemble. The SWDI observation error is assumed to be with 1% variance of observed SWDI in the initial training ensemble. The modeled (or predicted SWDI) error is assumed to be with the same observed error. As stated in (Liu et al., 2016; Yin et al., 2015), the EnKF technique can reach its maximum efficiency when the updating ensemble size approaches around 15, therefore, the updated ensemble size is set to be the same size of training ensemble. The details about SVM and the procedure of the dual EnKF can be found in (Samuel et al., 2014; Liu et al., 2016). The experiments conducted for the SWDI predictions are given in Table 1 and described as follows:

Case 1 is conducted with limited accessible in-situ meteorological data (T2m, SR, P, RH) as input for SVMs to predict the weekly SWDI at subsequent 1–4 weeks lead time. This Case aims to investigate whether the limited in-situ meteorological data can be used for the weekly SWDI predictions at subsequent shortly time period and to provide a benchmark for the comparison with Cases 2 to Case 5.

Case 2 incorporated the LAI data to combine with the input meteorological variables in Case 1 to predict the weekly SWDI.

As indicated in Ines et al. (2013) and Mishra et al. (2015), LAI has a vital role in simulating agricultural crop yields and agricultural drought indices. Therefore, the addition of LAI is expected to improve the SWDI predictions. Similarly, Case 3 incorporates the RSM data instead of LAI data and combines with the input meteorological variables in Case 1 to predict the SWDI. Case 4 added both the LAI and RSM to combine with the input meteorological variables in Case 1 to predict the SWDI. Case 2–4 aims to assess whether the addition of remotely sensed LAI or SM data individually or simultaneously with in-situ meteorological data in Case 1 can improve the forecasting results.

Case 5 applies coupled SVM-DA model with the same input variables in Case 1 to predict the SWDI with the purpose to investigate the efficiency of SVM-DA for the improvement of the SWDI predictions with SVMs solely. The trial and error method is prior used to set up the SVMs prediction model. The details of experimental design can be found in Liu et al. (2016).

3.2. Evaluation criteria

The goodness of fit (GOF) statistics: correlation coefficient (R), root mean square error ($RMSE$) and mean absolute error (MAE) between the predicted SWDI and observed SWDI derived from in-situ SM data (named as observed SWDI) are used to evaluate the performance of the proposed modeling framework. The calculation about these criteria can be found in (Brocca et al., 2011; Yu et al., 2012; Liu et al., 2016; Cheng et al., 2016). As indicated, the smaller the $RMSE$ and MAE are and the larger R is, the better the model performs (Liu et al., 2010; Yu et al., 2012; Liu et al., 2016).

In addition, one recently proposed criteria, normalized information contribution (NIC) (Kumar et al., 2014; Yan et al., 2017) is selected to better evaluate the efficiency of remotely sensed products for the SWDI predictions at subsequent 1–4 weeks. NIC is similar to the assimilation efficiency (AE) index used in Yin et al. (2014) and Liu and Mishra (2017). The calculation procedure for NIC is given in Eq. (4) as following:

$$NIC = \frac{RMSE_{CTR} - RMSE_{UP}}{RMSE_{CTR}} \quad (4)$$

where, $RMSECTR$ means the $RMSE$ from the control run, which is derived from Case 1 in this study; while $RMSEUP$ means the $RMSE$ derived from the updated experiment, which belongs to Case 2 to Case 5 in this study. If $NIC > 0$, it indicates that the updated Case improves the predictions compared to the control run (Case 1 with only meteorological variables as input in this study); if $NIC = 0$, the updated Case does not add any skill; if $NIC < 0$, the updated Case degrades the prediction skill compared to the control run; and if $NIC = 1$, the updated Case achieve the maximum skill.

3.3. Evaluation of AMSR-E, SMAP with In-situ SM

Due to the difference in sensing depth and spatial resolution between satellite and in-situ sensors (Escorihuela et al., 2010), the bias correction approach based on Cumulative Density Function (CDF) is applied to remove systematic differences between satellite and in-situ SM data. This CDF-matching method is initially proposed by Reichle and Koster (2004), and later adopted in many SM bias correction studies (e.g., Drusch et al., 2005; Brocca et al., 2011, Yan et al., 2017). The mean AMSR-E data derived from ascending and descending products are used in this study. The statistical properties based on raw and bias-corrected remotely sensed SM with respect to the in-situ SM at the surface soil layer (SM1 at the 5 cm depth) is provided (Figs. S3.1 and S3.2). The statistical results indicate that the bias-corrected remotely sensed SM compares well with the in-situ observed SM estimated by the magnitude of mean and standard deviation, higher R, combined with lower $RMSE$ and MAE at each study site during each study period. Therefore, in this study, the bias-corrected remotely sensed SM is applied for the SWDI prediction.

All the data sets, including in-situ observed hydrometeorological data and remotely sensed data, used in this study are averaged from daily into weekly time series.

3.4. Relationship between meteorological (remotely sensed) variables with SWDI

The relationship between meteorological variables and remotely sensed products with SWDI over the CONUS are investigated (Fig. 2). It is observed that the R between T2 m and SWDI (named as T2m-SWDI) are negative at most study stations, though some stations located at the upper Northern Great Plains within H10 (S49, S59, S50, S39), Northern Shrubland (S24NS, S55NS), and lower Southern Great Plains (S51, S54) and Southeast (S28) along the Gulf of Mexico show positive signals. Among the negative signals, the T2m-SWDI correlation coefficient is stronger ($R < -0.5$) at the stations located at typical hydrologic regions, such as, the Northeast (H02), Southeast (H03, H06), Southern Great Plains (H12), lower Northern Great Plains (H11), lower Southern shrubland (H13, H15, H18), Northwest and Northern Shrubland (H17, H18). The SR-SWDI presents similar spatial pattern with T2m-SWDI, also implying negative signals at most stations located at the Hydrological Regions close to the Ocean (i.e., H02, H06, H11 to H13, H17, H18), while positive signals at the upper Northern Great Plains within H10, Northern Shrubland and upper Southern Shrubland within H16 and H14.

Compared to T2m or SR, the P-SWDI implies weak ($R < 0.5$) but positive signals at most study stations over CONUS with an exception of Southern and Northern Shrubland within H13, H14, and H15. The RH-SWDI shares similar spatial pattern and resembles similar amplitude with P-SWDI over CONUS, with the only obvious difference located at the Northeast and Upper Southeast within H02, H03 and H08 and upper Northern Great Plains within H10 where the RH-SWDI presents negative signals while P-SWDI is positive.

The LAI-SWDI correlation coefficient almost agree with T2m-SWDI and/or SR-SWDI both in the spatial pattern and amplitude over CONUS providing strong negative correlation (magnitude lower than -0.5) at most stations over the regions, such as, Northeast (H02), upper Southeast (H03, H06), Southern Shrubland (H13, H15), part of Northwest and Northern Shrubland within H17 and H18; while positive signals over the Upper Northern Great Plains within H10 and Central Northern Shrubland within H14, H16; and weak negative signals for the stations at the remaining subregions.

The bias corrected RSM presents significant positive relationship with SWDI ($R > 0.5$) at most study stations. Additionally, the stations located at the Upper Northern and Lower Southern Great Plains and Shrubland within H10 to H14 have slightly lower magnitude of R than the remaining stations.

Among the correlation coefficient between hydrometeorological variables (T2m, SR, P, RH, LAI, RSM) with SWDI, most stations located at Northeast within H02, Southeast within H03 and H06, Southern Great Plains within H12-H13 (not including S51, S54), and part of Northwest within H17 and H18, have strong negative T2m-SWDI combined with strong negative SR-SWDI, LAI-SWDI, and strong positive ($R > 0.75$) RSM-SWDI correlation coefficient, while the stations located at the Northern Great Plains within H10, Shrublands within H14 and lower Southern Great Plains within H12 present weak T2m-SWDI correlation coefficient (R within $[-0.5\ 0.5]$) combined with weak SR-SWDI, LAI-SWDI and lower RSM-SWDI correlation coefficient. The weak correlation coefficient between meteorological variables with SWDI over Great

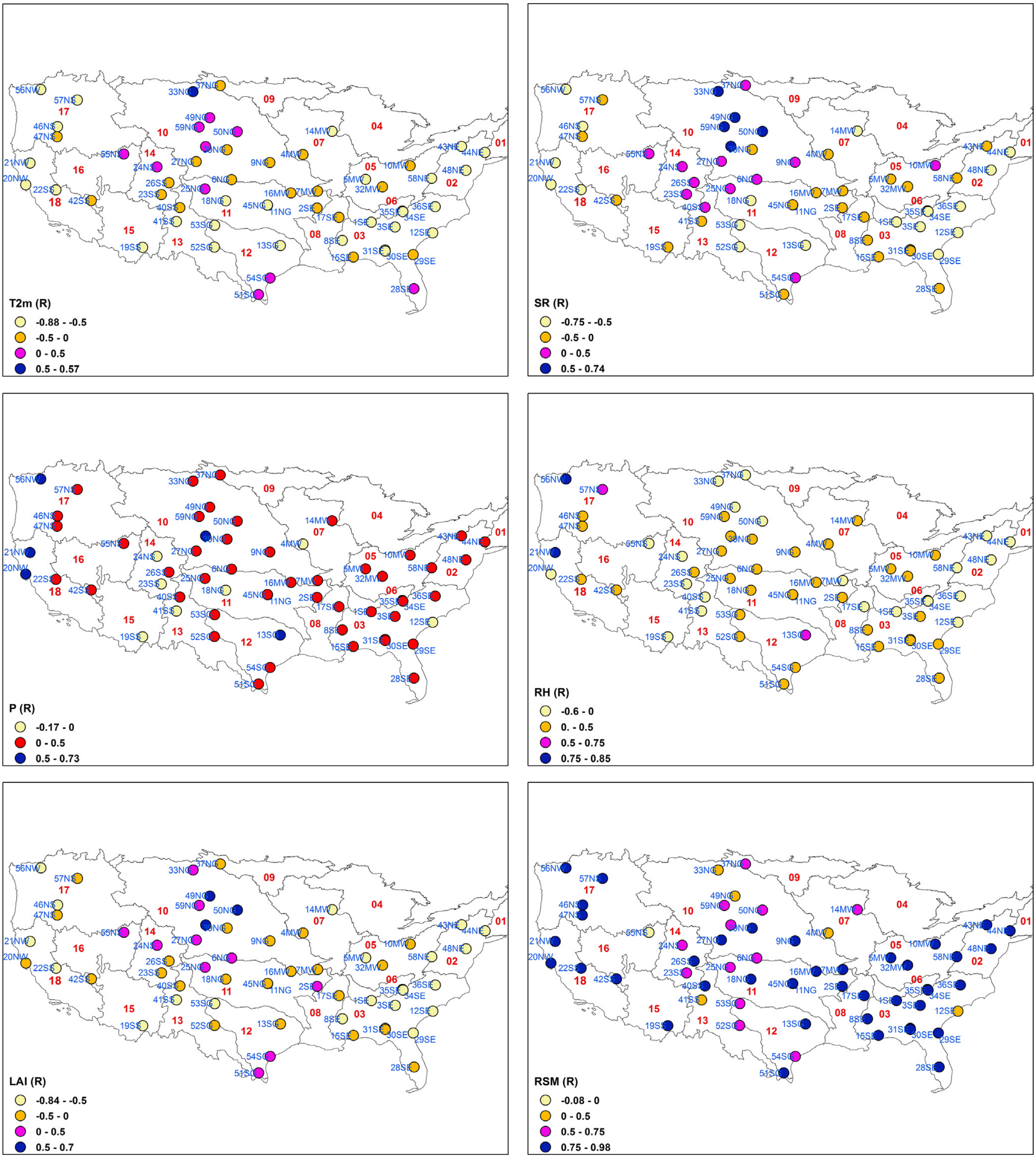

Fig. 2. Correlation coefficient between hydroclimatic variables (i.e., T2m, SR, P, RH, LAI, RSM) and SWDI over the selected 59 study Stations at the 2nd study time period during April 1st 2015 to March 31st 2016. The red numbers (01 to 18) for the HUC2 regions. The blue character for the selected 59 Stations and their subregion defined in Table 3.

Plains and Shrublands may be due to the local climate regime and land-atmosphere interactions. As stated in previous studies (e.g., Koster et al., 2006; Mei and Wang, 2011, 2012), the hot spots of the strong synoptic-scale coupling for SM-P or SM-T2m mainly appear in the central Great Plains of North America (Fig. S4). The positive T2 m-SWDI, SR-SWDI signals (i.e., the central to upper Northern Great Plains, part of Southern Great Plains and Southeast along the Gulf of Mexico) may be due to the local cold weather regime or the strong land-atmosphere coupling strength.

In this study, the antecedent meteorological variables and remotely sensed products (i.e., AMSR-E, SMAP, LAI) are used to predict the subsequent weekly SWDI. The cross correlation coefficient approach is adopted to evaluate the lagged time between SWDI and each input variables. It is observed that the lagged time for each selected local variables (T2m, SR, P, RH, LAI, RSM) with SWDI are almost within the antecedent 5 weeks at most of the 59 stations selected over CONUS (Fig. S5.1) at the 2nd time period with available SMAP. In addition, the lagged correlation coefficient between selected input variables at antecedent 1–5 weeks with current weekly SWDI tends to decrease with lagged time increase, especially after 4–5 weeks, at most study stations (Fig. S5.2). At the 1st time period over the 14 stations with available AMSR-E, the lagged time and lagged correlation coefficient presents similar results (Figs. S6.1 and S6.2) with that at the 2nd time period. It is observed that the stations located at the Northern Great Plains (i.e., S6, S9, S11) and Midwest (i.e., S4, S5) presents slightly weaker T2m-SWDI, SR-SWDI, LAI-SWDI, RSM-SWDI (AMSR-E) signals compared to the remaining stations. The lagged correlation tends to be decreased with lagged time increased, especially after 4 weeks length.

Therefore, to keep consistency, the variables, including T2m, SR, P, RH, LAI, and RSM, at the antecedent 3 weeks are used as input to predict the SWDI results at the subsequent 1–4 weeks. The component of T2m, SR, P and RH are selected as the benchmark input variables in Case 1 for the SWDI predictions at subsequent 1–4 weeks lead time. All the input and output variables of SVMs during each study period are standardized prior to assign equal weights with a range of values different for each variable. The predicted SWDI are transferred to the actual data type for the performance evaluation. The SVMs are validated with different initial training length and different component of meteorological variables (results not shown) based on the "trail-and-error" method.

4. Results analysis

Due to the non-overlapping of AMSR-E and SMAP data, we selected two time periods for analysis: (a) the 1st period for AMSR-E (October 1st 2009 to September 30th 2011); and (b) the 2nd period for SMAP (April 1st 2015 to March 31st 2016). The results are discussed in two sections: Firstly, we evaluated the performance of AMSR-E, SMAP and LAI for the drought forecasting at two different time periods at the 14 selected stations; Secondly, generalize our findings at the selected 59 stations for COUNS with a focus on recently launched SMAP product during the 2nd time period.

4.1. Evaluation of AMSR-E, SMAP and LAI

The GOF criteria (*R, RMSE, MAE*) between predicted SWDI at the subsequent 1st week with the observed SWDI at the 14 selected stations based on the two time periods are provided in Fig. 3. It is observed that the predicted SWDI at the subsequent 1st week from Case 1 (black circle) capture well the observed SWDI based on the significant correlation coefficient ($R > 0.5$) combined with low *RMSE* ($RMSE < 0.3$) and *MAE* ($MAE < 0.3$) at most of the stations (e.g., 1st period: S1-S3, S5, S7, S13; 2nd period: S1, S6-S8, S9, S10, S13, S14) with strong linkage between meteorological variables (i.e., T2 m, SR, P. RH) and SWDI (1st period: Fig. S5.1; 2nd period: Fig. S6.1). Such findings indicate the efficiency of antecedent meteorological variables for the SWDI prediction at subsequent 1st week. The strong linkage between meteorological variables and SWDI may contribute to the better performance of data-driven technique for the drought forecasting.

The addition of remotely sensed products (i.e., LAI and RSM (1st time period: AMSR-E; 2nd time period: SMAP)) as inputs for SVMs (Case 2 to Case 4) slightly improved the SWDI predictions at some of the stations (e.g., 1st time period: S3, S6-S8, S10-S13; 2nd time period: S1, S2, S7-S10, S12, S14). The relationship between LAI, RSM and in-situ SWDI varies between stations, which also decide improvement achieved in drought forecasting. For example, Case

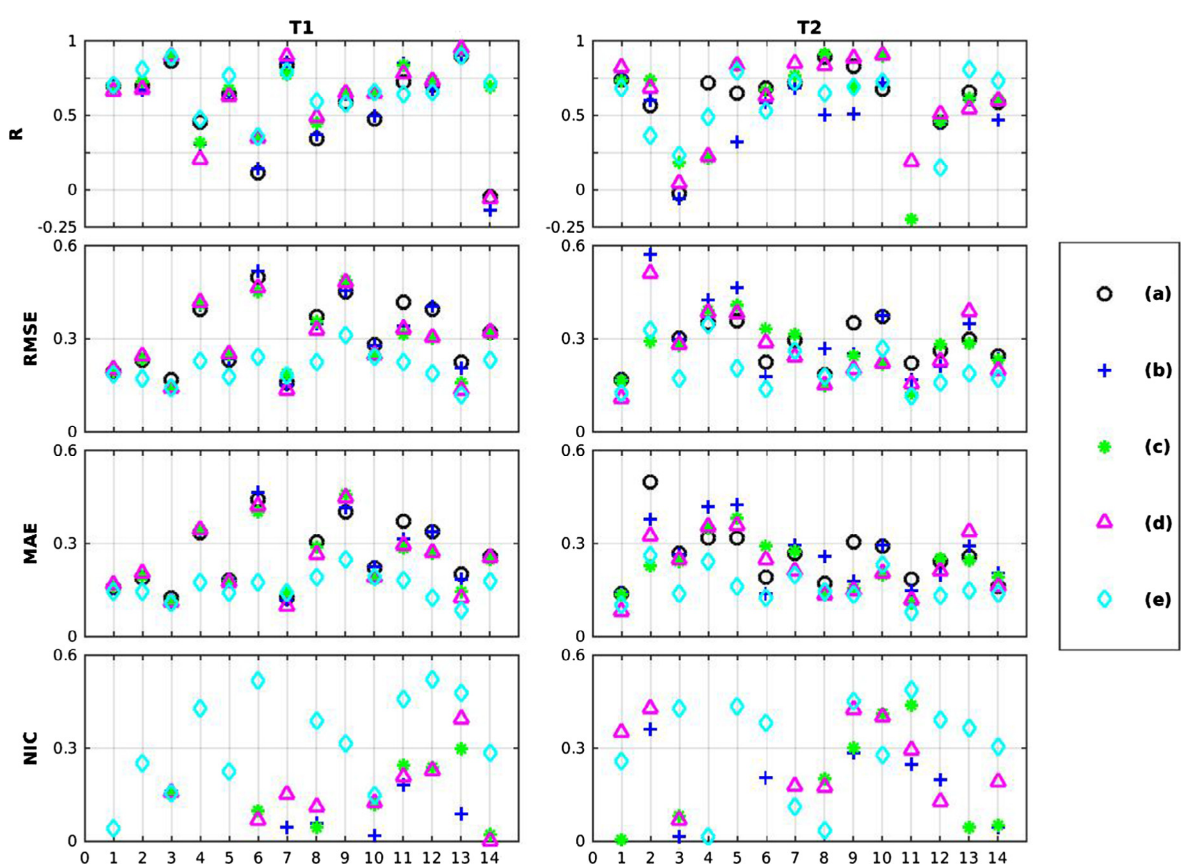

Fig. 3. GOF comparison among Case 1 to Case 5 for the selected 14 study stations at both study time period: T1 (1st study time period during October 1st 2009 to September 30th, 2011), T2 (2nd study time period during April 1st 2015 to March 31st 2016). (a) Case 1; (b) Case 2; (c) Case 3; (d) Case 4; (e) Case 5. X-axis for the 14 study stations.

2 with LAI added as input slightly improved the SWDI predictions at stations with significant LAI-SWDI relationship (e.g., 1st time period: S7, S10, S13; 2nd time period: S12) or weak meteorological relationship (e.g., 2nd time period: S2, S6, S9) based on the decreased *RMSE* or *MAE* magnitude with respect to Case 1 (Fig. 3). Consistently, Case 3 with RSM added as input has better performance at stations (e.g., 1st time period: S3, S10-S13; 2nd time period: S1, S8-S10, S13, S14) based on good RSM-SWDI correlation coefficient. Case 4 with both RSM and LAI added simultaneously as input performs better at stations (e.g., 1st time period: S3, S7, S8, S10-S13; 2nd time period: S1, S7-S10, S12, S14) based on either significant LAI-SWDI or RSM-SWDI linkage. Obviously, the addition of both LAI and RSM at the same time in Case 4 performs better than the individual addition of LAI in Case 2 or RSM in Case 3 based on the signals of GOF results and the improved performance at more number of stations.

Compared to the solely performance of SVMs in Case 1 to Case 4, Case 5 with SVM-DA coupling model lead to lower *RMSE* and *MAE* magnitude, while slightly higher or similar range of *R* magnitude at most stations at each study time period. Moreover, the *NIC* values for Case 5 with respect to Case 1 are greater than zero at most of the stations for both the 1st and 2nd study time period. Such information indicate the improvement of SWDI predictions at the subsequent 1st week with SVM-DA coupling model to the solely SVMs.

The time series of the SWDI predictions at the subsequent 1st week for five stations (S7MW, S8 SE, S10MW, S11 NG, S13 SG) with different land use type are selected to further evaluate the performance of in-situ meteorological variable and remotely sensed products added as inputs of SVMs (Fig. 4.1 and Fig. 4.2 for the 1st and 2nd time period, respectively). Consistent with GOF results, the predicted SWDI at the subsequent 1st week for each Case at selected site can capture the trend of observed SWDI (Fig. 4.1, Fig. 4.2). The addition of remotely sensed products (Case 2 to Case 4) resemble similar trend and close magnitude of SWDI predictions with Case 1 at most of the times with only a little improvement. The SVM-DA coupling model has better performance compared to the predictions from Cases with solely SVM (Case 1 to Case 4) based on the smaller bias with respect to the observed SWDI at most of the predicted time in each site (cyan diamond). However, the time series of relative error (*RE*) (Fig. 5.1 and Fig. 5.2 for the 1st and 2nd time period, respectively) indicate obvious improvement of Case 2 to Case 4 with respect to Case 1 according to the lower RE signals at some predicted times at each site. In addition, Case 4 suggests higher improvement at each site based on the lower *RE* values at most predicted time period among Case 2, Case 3 and Case 4. Moreover, consistent with the GOF results and time series analysis, Case 5 with the SVM-DA coupling model (cyan diamond) has the best performance based on the lower *RE* compared to the SVM solely predictions.

To assess the persistence of SVMs or SVM-DA coupling model for the SWDI predictions, the *R* and *NIC* values at the subsequent 1st to 4th week for each Case are evaluated to do a comparison (Fig. 6.1 and Fig. 6.2 for the 1st and 2nd time period, respectively). It is observed that the *R* and *NIC* signals tends to be weak based on the decreased *R* magnitude at most stations for both study period with SVMs solely (Case 1 to Case 4), implying the reduced efficiency of SVMs for the SWDI predictions at longer lead time. However, there are few stations maintain consistency (e.g., S1, S3, S5, S7, S10-S13 at the 1st time period (Fig. 6.1); S5-S7, S9-S10 at the

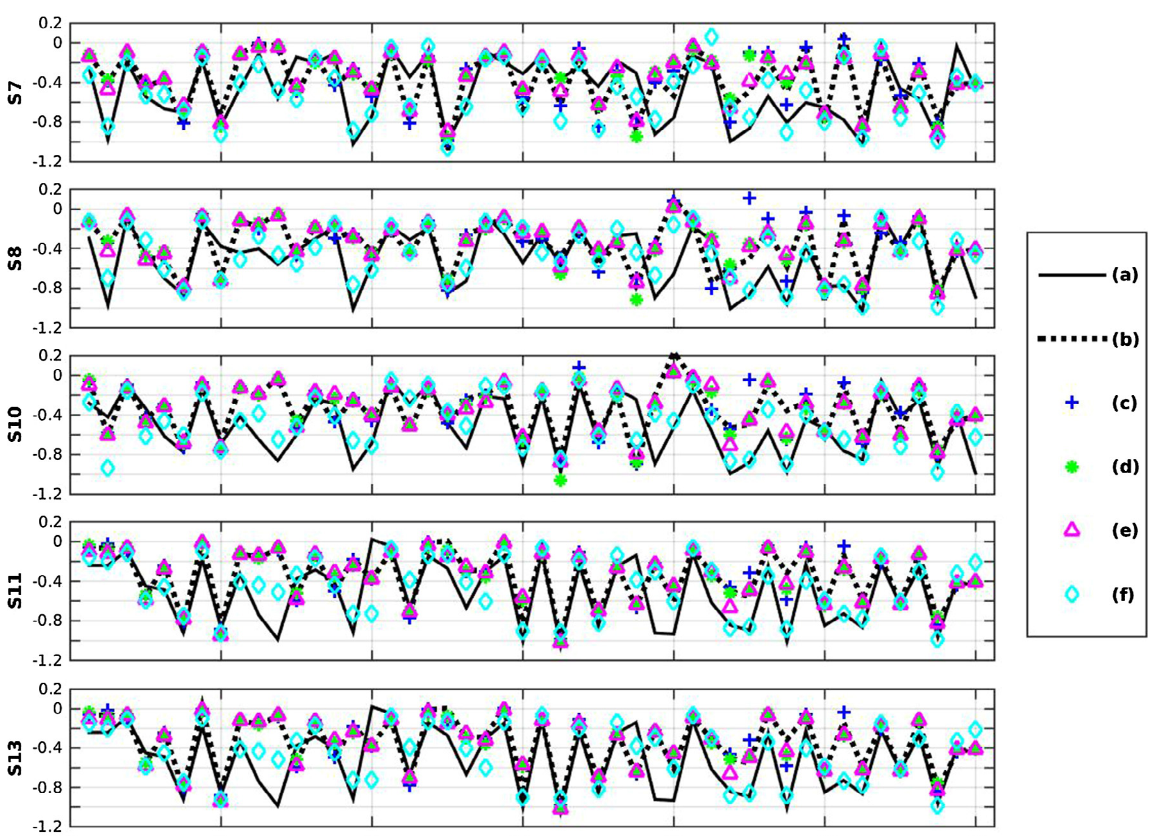

Fig. 4.1. Time series of SWDI prediction at subsequent 1st week at 5 selected study stations (S7, S8, S9, S10, S11, S13) with NIC value from Case 2 to Case 5 greater than zero compared to Case 1 at the 1st study time period during October 1st 2009 to September 30th, 2011. (a) observed SWDI; (b) Case 1; (c) Case 2; (d) Case 3; (e) Case 4; (f) Case 5. X-axis for the predicted time (Unit: Week).

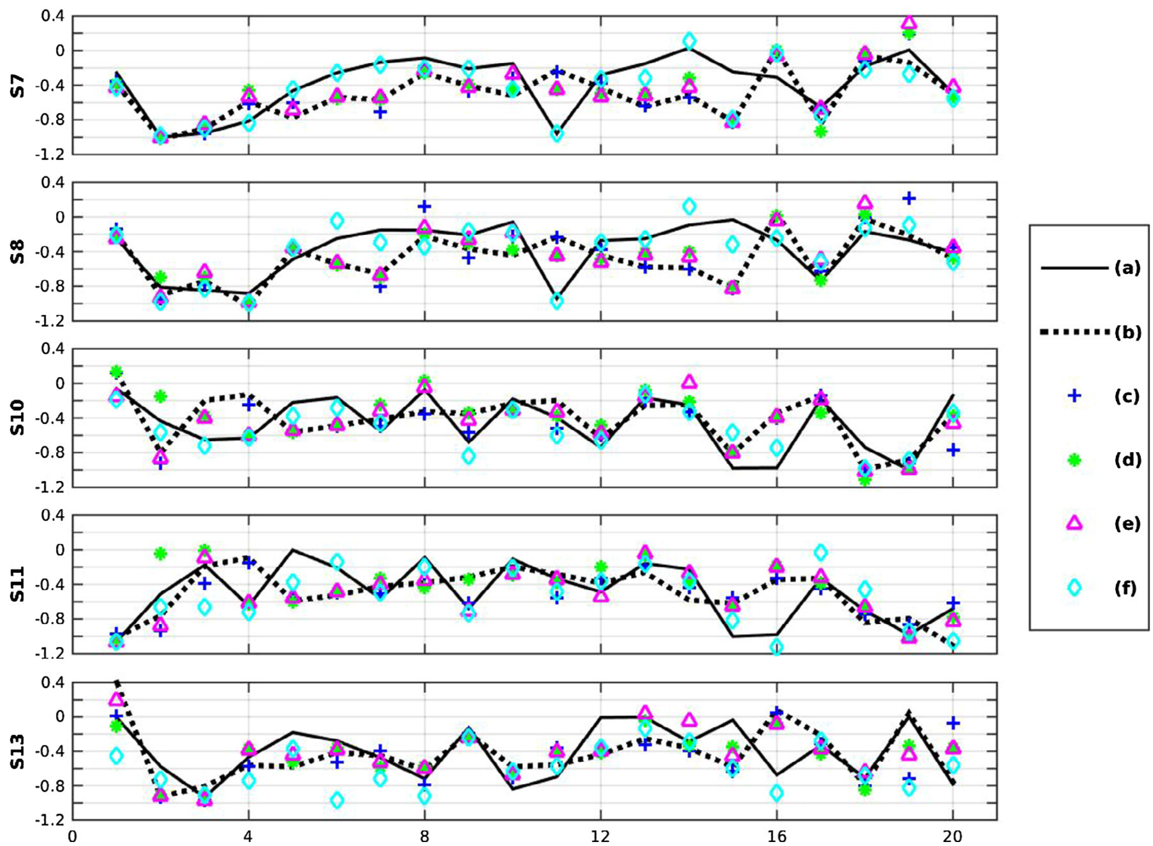

Fig. 4.2. Same as Fig. 4.1, but for the 2nd study time period during April 1st 2015 to March 31st 2016.

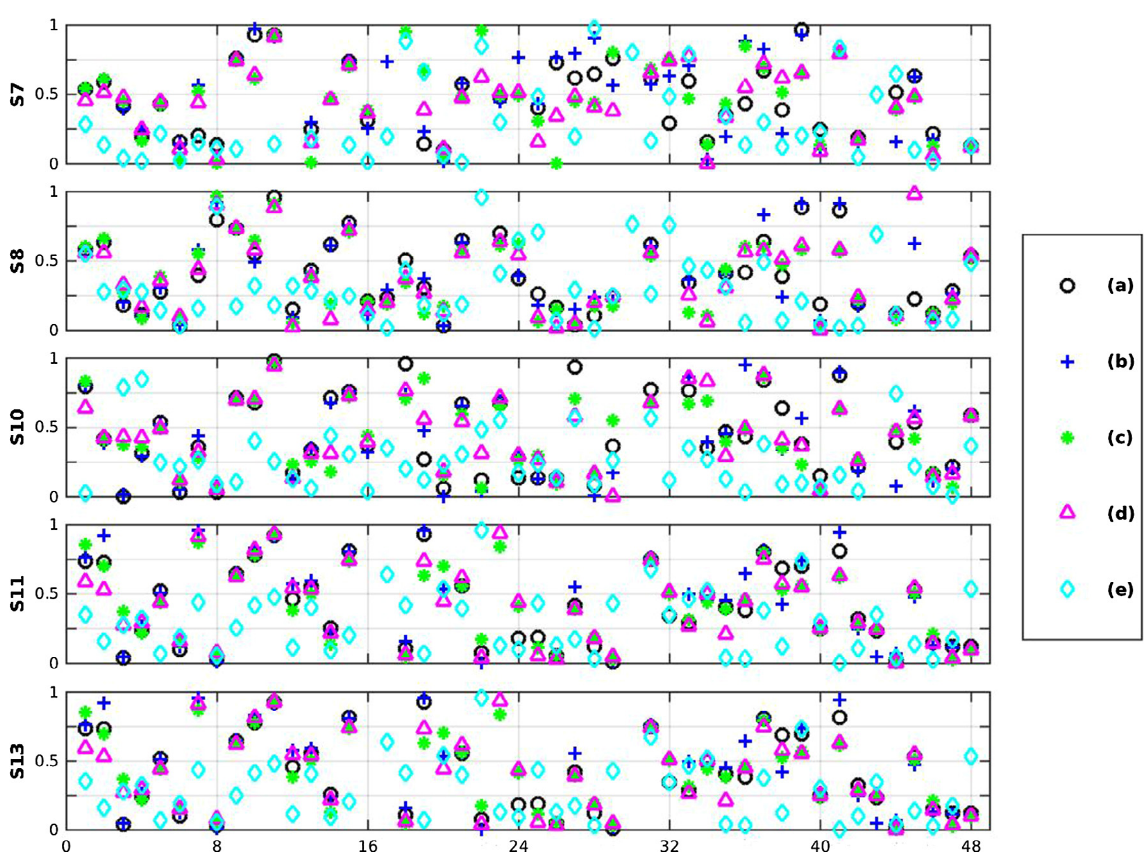

Fig. 5.1. Time series of relative error for predicted SWDI versus observed SWDI at subsequent 1st week at 5 selected study stations at the 1st study time period during October 1st 2009 to September 30th, 2011 (upper). (a) Observed SWDI; (b) Case 1; (c) Case 2; (d) Case 3; (e) Case 4; (f) Case 5. X-axis for the predicted time (Unit: Week).

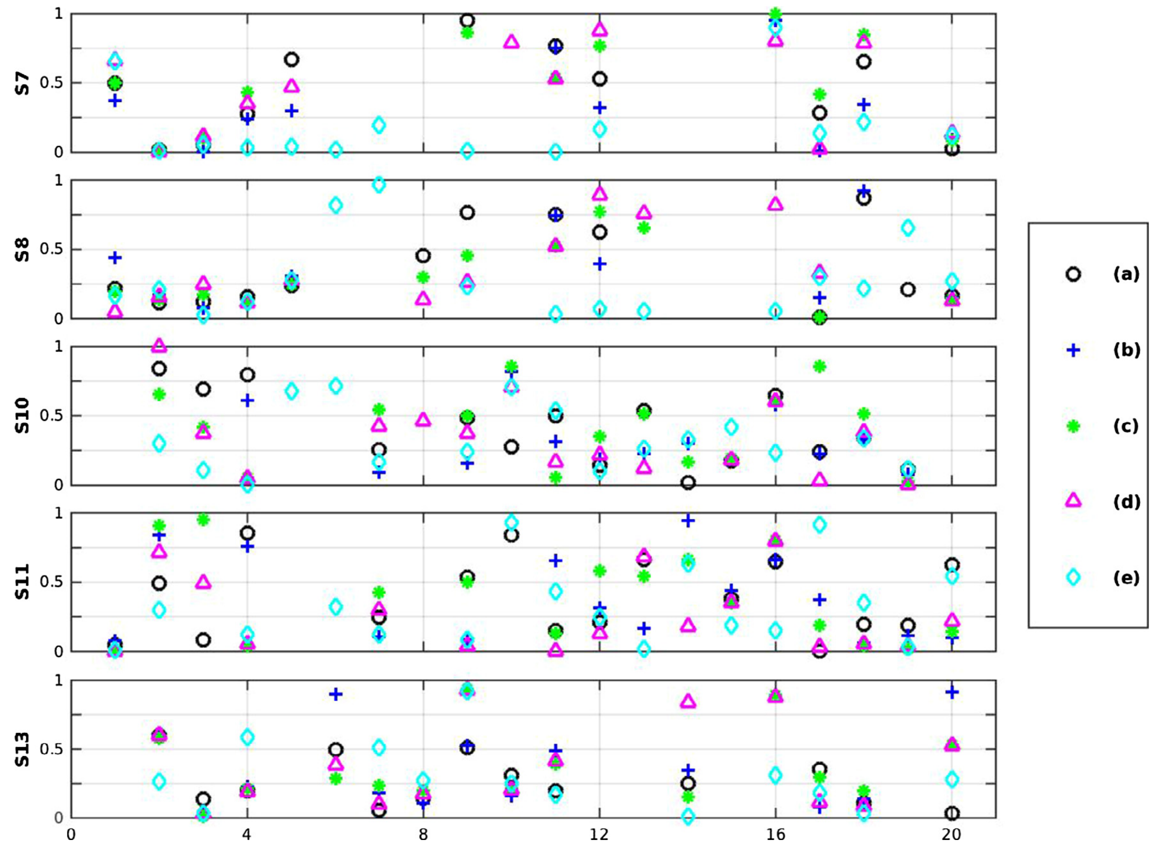

Fig. 5.2. Same as Fig. 5.1, but for the 2nd study time period during April 1st 2015 to March 31st 2016.

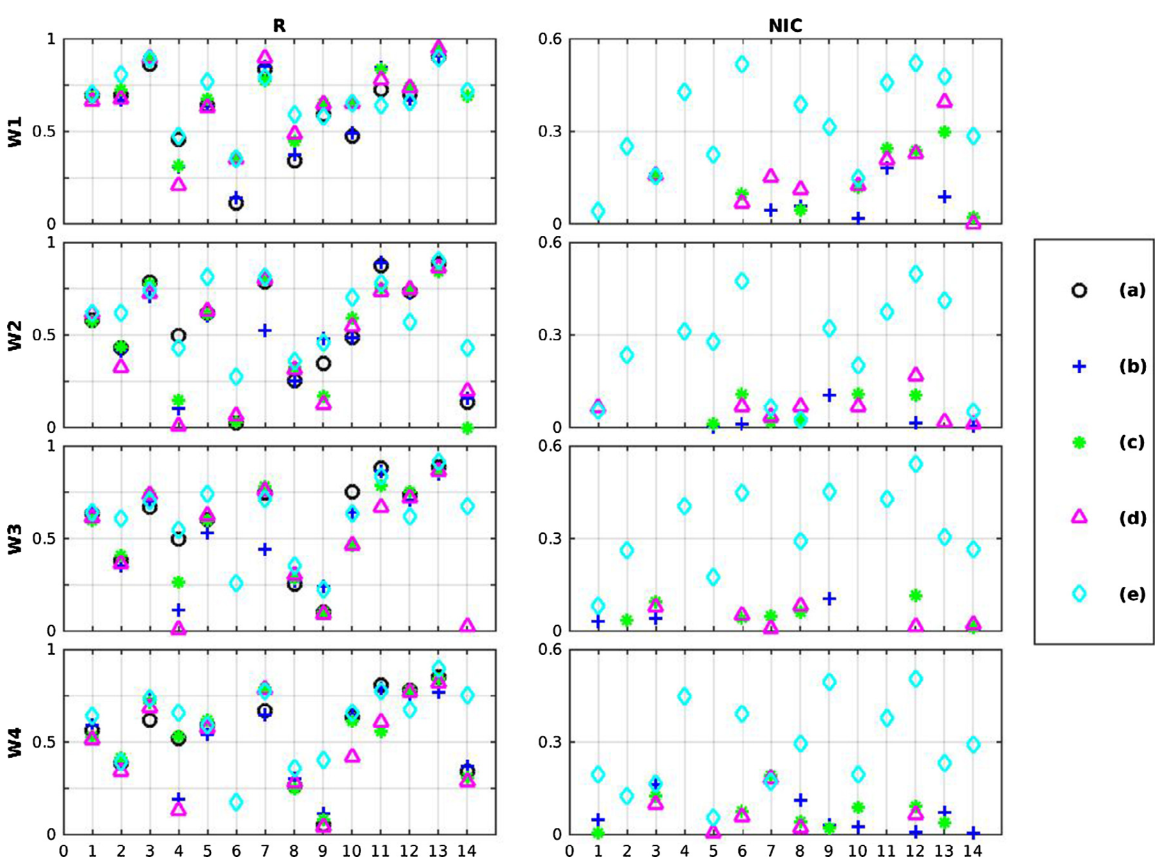

Fig. 6.1. R and NIC comparison at the subsequent 1–4 weeks lead time for the selected 14 study stations at the 1st study time period during October 1st 2009 to September 30th, 2011. (a) Case 2; (b) Case 3; (c) Case 4; (d) Case 5. (Only the results above zero are shown). X-axis for the 14 stations.

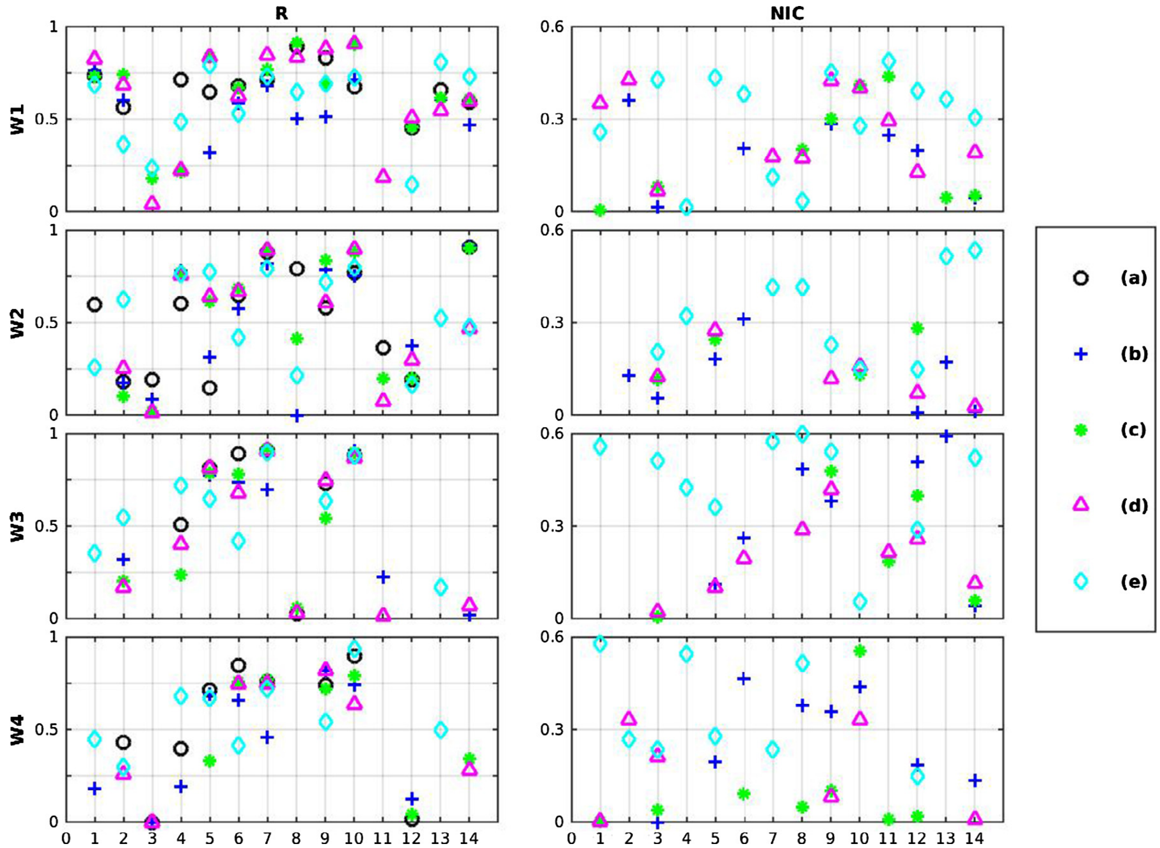

Fig. 6.2. Same as Fig. 6.1, but for the 2nd study time period during April 1st 2015 to March 31st 2016.

2nd time period (Fig. 6.2)) based on the higher *R* signals between predicted SWDI and observed SWDI. The addition of remotely sensed products, especially with LAI and RSM simultaneously at Case 4, attribute to improve the predictions at some stations (e.g., 1st time period: S7, S12, S13 (Fig. 6.1); 2nd time period: S9, S10 (Fig. 6.2)) based on the positive *NIC* value. However, such improvement tends to decrease with the increase in predicted time according to the decreased *R* magnitude (lower than 0.5) and *NIC* values (close to zero or tends to be negative) at most stations. Unlike SVMs solely, the SVM-DA seem to have better persistence for the SWDI predictions at longer predicted time (2nd to 4th week) based on the persistent significant correlation coefficient and the positive higher NIC values at most stations at each time period.

4.2. Evaluation of SMAP over the continental United States

Overall, 59 stations are selected to evaluate the performance of SMAP for drought prediction. The GOF criteria, including *R*, *RMSE*, and *MAE*, between predicted SWDI at the subsequent 1st week from each Case (Case 1 to Case 5) with the observed SWDI are plotted in Fig. 7.1, Fig. 7.2, and Fig. 7.3, respectively. As it is shown, the results from Case 1 present significant positive correlation coefficient signals ($R > 0.5$) (Fig. 7.1) combined with low *RMSE* (Fig. 7.2) and *MAE* (Fig. 7.3) magnitude (within 0.3) at most stations with strong linkage between meteorological variables (i.e., T2m, SR) with SWDI, such as Northeast within H02 (e.g., S44, S48, S58N), Southeast within H03 (e.g., S3, S36, S28, S29, S15), Lower Southern Great Plains within H12 (e.g., S52, S54), Southern and lower Northern Shrublands and lower Northwest within H13-H15 and H18 (e.g., S21-S23, S41-S42), indicating good performance of SVMs. Besides, some stations with weak relationship between meteorological variables with SWDI but located at the regions with strong land-atmosphere feedback (e.g., Koster et al., 2006; Dirmeyer et al., 2009; Mei and Wang, 2012), such as, Midwest within H05, H07 and H11 (e.g., S5, S10, S14, S16), Northern Great Plains within H10 and H11 (e.g., S18, S38, S39, S49), also present good performance of drought forecasting with SVMs. Such findings demonstrate the contribution of strong linkage between meteorological variables with SWDI and the possible influence of the land-atmosphere feedback for the drought forecasting with data-driven technique.

The addition of SMAP SM in Case 3 and Case 4 as SVMs input have slightly better performance with respect to Case 1 at some stations located at Southern Shrubland within H18 (e.g., S20, S22), and Southeast within H03 (e.g., S31, S8, S15), based on the higher magnitude of *R* values combined with low *RMSE* and *MAE* magnitude.

Based on the constraint of *NIC* magnitude greater than zero and *R* magnitude greater than 0.5, the stations are identified to further investigate the improvement in SWDI forecasting with the addition of remotely sensed products as inputs in Case 2 to Case 4 and the SVM-DA coupling model in Case 5 (Fig. 8). The LAI added in Case 2 provide some improvement at the stations mainly located at Southeast, lower Southern Great Plains within H03 and H12, lower Northern Great Plains and Shrublands within H11, H13, H14. Most of these stations have weak T2m-SWDI and SR-SWDI, but strong LAI-SWDI correlation coefficient (Fig. 2). Compared to Case 2, the addition of SMAP SM in Case 3 provide more improvement at the stations located at Northern Great Plains within H10-H11, part of Midwest, and Shrublands that has good RSM-SWDI correlation coefficient but weak SWDI relationship with local meteorological variables (i.e., T2m, SR, P, RH) (Fig. 2). Among Case 2 to Case 4 with addition of remotely sensed products, Case 4 added with both LAI

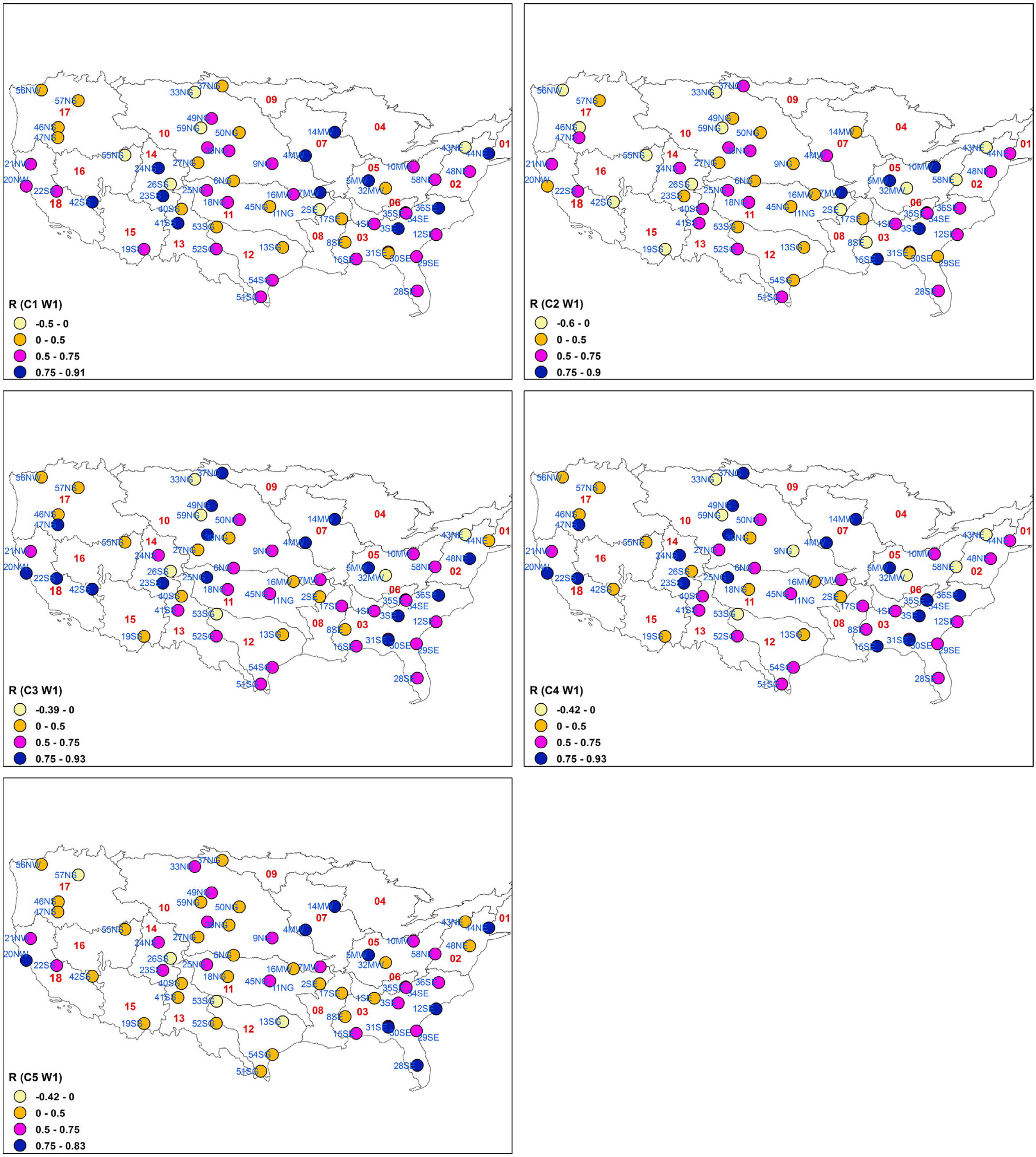

Fig. 7.1. Correlation Coefficient comparison among Case 1 to Case 5 at the selected study sites at the 2nd study time period during April 1st 2015 to March 31st 2016. The legend in each subplot presents the criteria for different Cases (C1 to C5) at 1 week lead time (W1), e.g., R (C1 W1) for correlation coefficient (R) of Case 1 (C1) at 1 week lead time (W1).

and SMAP SM comparatively performs best for the SWDI predictions with respect to Case 1 according to the improved performance with more stations have positive *NIC* values and significant *R* signals. By comparison, most of the stations improved in Case 2 and/or Case 3 are included in Case 4 with an exception over the Shrublands (e.g., S22, S40, S41, S42) and part of Great Plains (e.g., S51, S25, S18, S9). There are few stations where the addition of LAI and SMAP SM as input slightly reduce the prediction performance. Most of these stations present two kinds of linkage between meteorological (remotely sensed) variables with SWDI: (a) weak linkage of meteorological variables with SWDI and weak relationship between remotely sensed products (LAI, SMAP SM) and SWDI (e.g., S40, S41, S42, S51, S9, S25); (b) strong linkage between meteorological variables with SWDI (e.g., S22, S18). The addition of LAI and SMAP SM simultaneously may conflict with each other for the SWDI predictions at few stations.

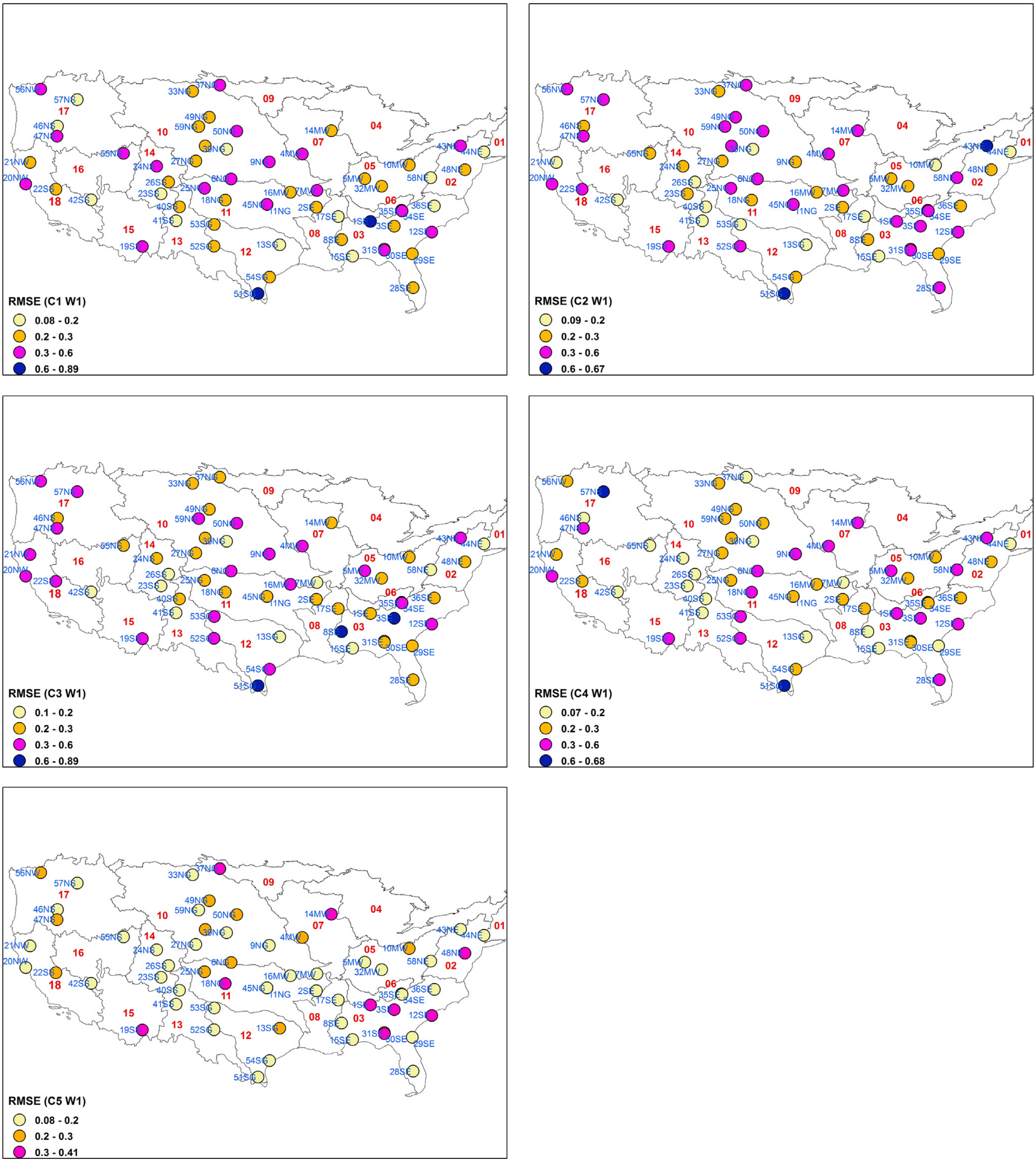

Fig. 7.2. Same as Fig. 7.1, but for the results of RMSE comparison among Case 1 to Case 5 at the selected study sites at the 2nd study time period during April 1st 2015 to March 31st 2016.

The SVM-DA coupling model in Case 5 present some improvement at the stations mainly located at the Southeast, Midwest, Northern Great Plains and part of middle Shrublands.

The addition of remotely sensed products (Case 2 to Case 4) and SVM-DA coupling model (Case 5) help to decrease the bias with observed SWDI at most stations based on the number of total stations with positive *NIC* values (Table 4). Case 4 seems to be more efficient for the improvement of SWDI predictions than Case 2 and Case 3 based on the more stations with significant *R* (40), *NIC* (38) and both significant *R* and *NIC* (28) criteria. The SVM-DA coupling performs best in reducing the bias with observed SWDI according to the total number of stations (47) with *NIC* values greater than zero. Among Case 3 to Case 5, there are greater than 20 stations present both significant correlation coefficient ($R > 0.5$) and positive *NIC* values ($NIC > 0$) between predicted SWDI with observed SWDI. Most of these improved stations located at

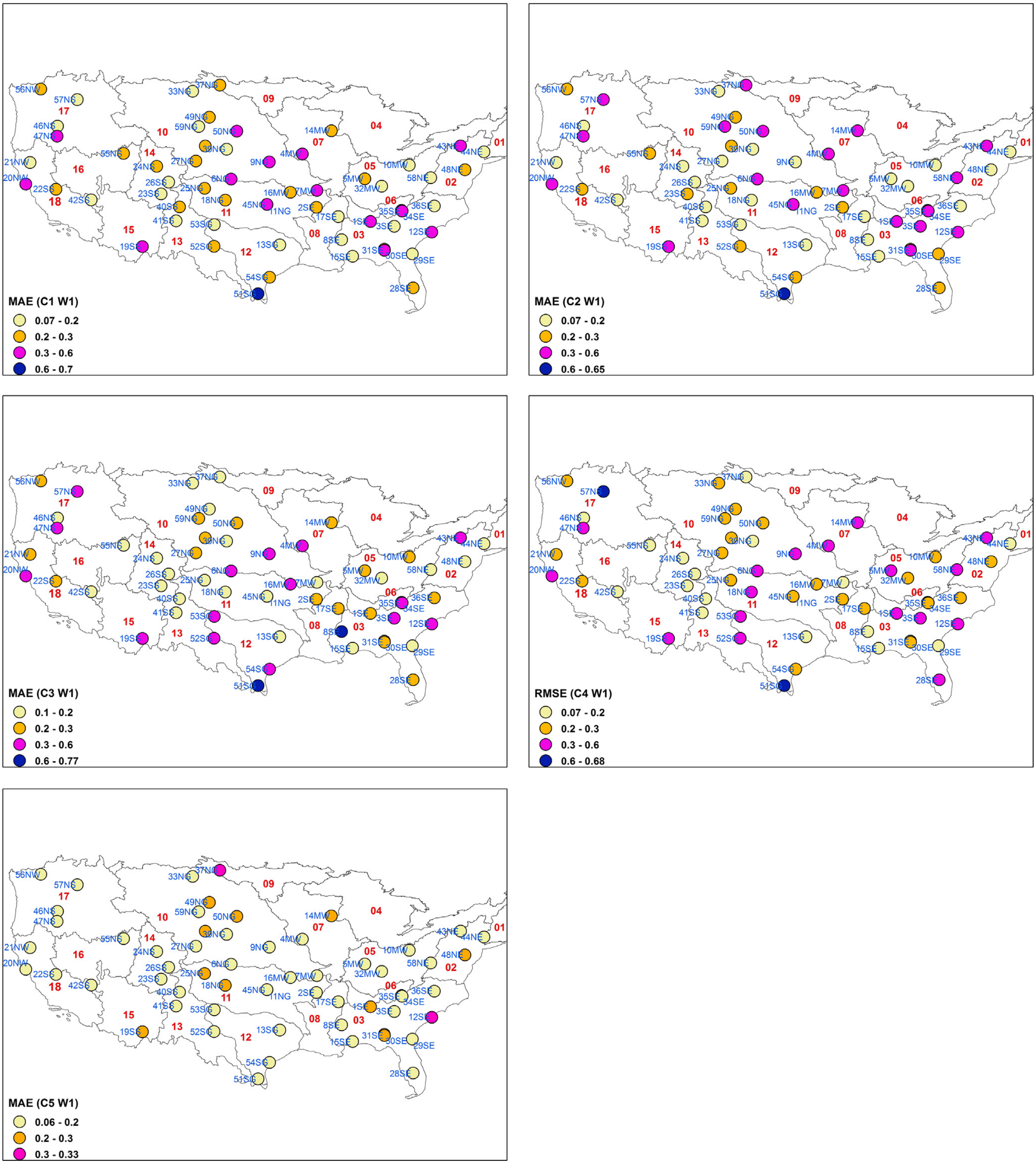

Fig. 7.3. Same as Fig. 7.1, but for the results of MAE comparison among Case 1 to Case 5 at the selected study sites at the 2nd study time period during April 1st 2015 to March 31st 2016.

the Great Plains (NG, SG), Shrublands (SS, NS) and Southeast (SE). These information demonstrates the efficiency of the remotely sensed products (i.e., LAI, SMAP SM) for the improvement of drought monitoring. The DA technique is able to reduce the predicted bias at most stations.

The time series of relative error (*RE*) between predicted SWDI and observed SWDI at several stations (S15, S23, S30, S35, S38, S48, S49, S54) with *R* magnitude greater than 0.5 and *NIC* values above zero from Case 2 to Case 5 (Fig. S7) further demonstrate the improvement of SWDI predictions in Cases with addition of remotely sensed products (Case 2 to Case 4) or with SVM-DA coupling model (Case5) based on the lower *RE* magnitude at most predicted time at most selected stations (e.g., S30, S38, S48, S49). Among Case 2 to Case 5, Case 4 has better performance compared to Case 2 or Case 3, while Case 5 with the SVM-DA coupling model has the lowest *RE* at most predicted time steps.

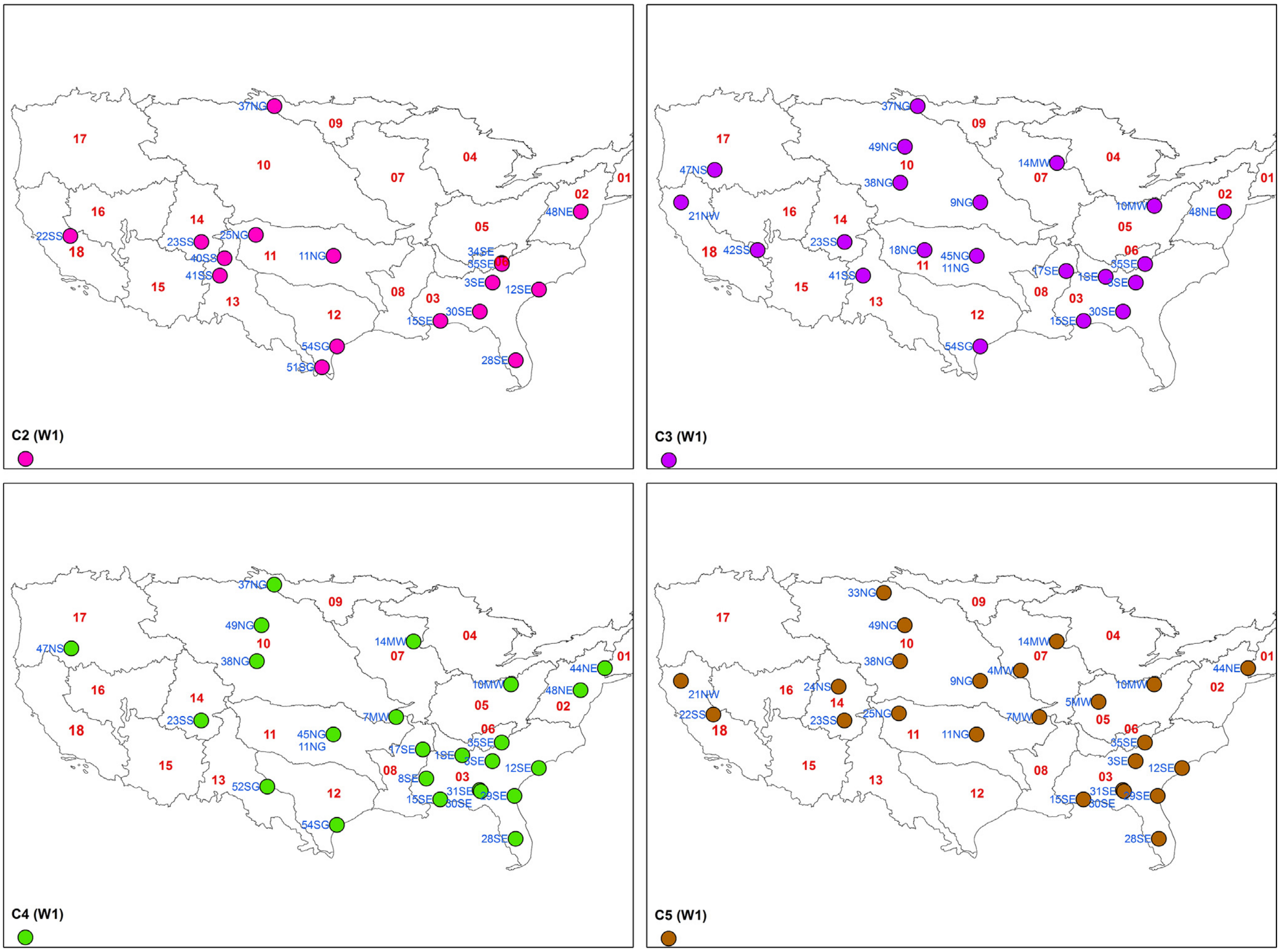

Fig. 8. Stations (solid colored circle) with R magnitude greater than 0.5 and NIC values greater than zero at the subsequent 1st week at the 2nd study time period during April 1st 2015 to March 31st 2016 from Case 2 to Case 5, respectively. The legend in each subplot presents the NIC for different Cases (C2 to C5) at 1 week lead time, e.g., C2 (W1) for NIC of Case 2 (C2) at 1 week lead time (W1).

Table 4
Number of stations with positive NIC, and both significant correlation coefficient ($R \geq 0.5$) and positive NIC for each Case over the whole CONUS and eight subregions. The number of stations within each subregion are given in the parenthesis, for example, NG(13) means there are 13 stations located at NG.

Category	Case	Total	NG(13)	SG(4)	NS(6)	SS(7)	MW(7)	SE(15)	NW(3)	NE(4)
NIC > 0	C2	30	6	3	3	3	5	7	2	1
	C3	32	9	1	3	4	3	9	1	2
	C4	38	10	2	5	5	2	11	1	2
	C5	47	10	4	5	6	5	13	3	1
$R \geq 0.5$ & NIC > 0	C2	14	2	1	1	1	4	3	1	1
	C3	26	9	1	2	2	3	8	0	1
	C4	28	9	2	2	3	2	9	0	1
	C5	24	6	0	2	2	3	9	2	0

In all, we found that the stations located at the Northeast within H02 (e.g., S44, S48), Southeast within H03 (e.g., S1, S36, S12, S28, S29), Midwest within H05, H07 and H11 (e.g., S5, S10, S4, S7), Lower Southern Great Plains within H12 (e.g., S54), Northern Great Plains within H10 (e.g., S18, S25, S49), Southern and lower Northern Shrublands and lower Northwest within H13-H15 and H18 (e.g., S21-S24, S40) have better SWDI predictions at most Cases with SVM solely or SVM-DA coupling model based on the GOF criteria (positive *R* signals with magnitude greater than 0.5 combined with *RMSE* and/or *MAE* lower than 0.3). The addition of remotely sensed products are proved to be efficiency to improve the SWDI predictions at stations mainly located at Southeast (Case 2, Case 3, Case 4), Northeast and Midwest (Case 3 and Case 4), central Shrublands (Case 2 and Case 3), Northern Great Plains (Case 3 and Case 4). Among Case 2 to Case 4, the addition of RSM and LAI simultaneously in Case 4 have better performance for the SWDI predictions, especially over the South East and North East CONUS. The SVM-DA coupling model (Case 5) reduced the bias between predicted SWDI with observations based on the lower *RMSE* and/or *MAE* magnitude (lower than 0.3) combined with positive *NIC* values at most study stations. The improved stations with significant correlation coefficient and positive *NIC* values mainly located at the Southeast, Midwest, Northern Great Plains and part of central Shrublands (Case 5).

The persistence of SVMs solely (Case 1 to Case 4) over the whole CONUS at the 2nd time period is weakened at most stations based on the reduced R magnitude (lower than 0.5) with the predicted lead time prolonged from 1-week to 4-week (Fig. S8). However, the addition of remotely sensed products seems to be able to keep good performance at longer lead times (1 week to 4 weeks) at some stations based on the high magnitude of R ($R \geq 0.5$) (Fig. S8) and positive *NIC* values (Fig. S9). The coupled SVM-DA for the SWDI predictions in Case 5 can improve drought forecast at longer lead times (up to 4 weeks) at most of the stations compared to the SVMs solely (Case 1 to Case 4) based on the consistent significant R magnitude and positive *NIC* values.

5. Conclusions

This paper investigates the performance of limited in-situ meteorological variables (i.e., T2m, P, SR and RH) and remotely sensed products (including LAI, AMSR-E and SMAP) for the near-real time agricultural drought prediction using data-driven technique SVMs, solely or coupled with DA technique (SVM-DA). Due to limited time length of meteorological data and remotely sensed products, two time periods are selected in this study: 1st study time period during October 1st 2009 to September 30th 2011 with available AMSR-E data; the 2nd period during April 1st 2015 to March 31st 2016 with available SMAP data. The major conclusions are drawn from this study as follows:

(1). The limited meteorological variables are able to predict the SWDI at most study stations over the CONUS at one week lead time with SVMs solely or SVM-DA coupling model based on the GOF criteria and time series analysis. The best performance was observed for stations mainly located in the regions with strong linkage between meteorological variables with SWDI or regions with strong land-atmosphere feedback including Northern Great Plains and Midwest.

(2). The addition of remotely sensed products (i.e., LAI, AMSR-E, SMAP) slightly improve the SWDI predictions at most of the stations, where the strong relationship between LAI (and/or AMSR-E, SMAP) with SWDI exists. Typically, the improvements take place at local scales; therefore it is difficult to generalize the improvement at regional scale. In general, the main improvement with addition of LAI is located at Southeast and central Shrublands. Based on the limited number of stations, the improvement with addition of SMAP (AMSR-E) mainly located at Southeast, Northeast, Midwest, and Northern Great Plains. The improvement with addition of both LAI and SMAP (AMSR-E) at the same time performs better than the independent addition of LAI or SMAP (AMSR-E).

(3). Coupled SVM-DA model are more efficient in the SWDI predictions based on the lower bias with observed SWDI at most study stations, though the correlation coefficient signals decreased at some stations. The improved stations with significant correlation coefficient and positive *NIC* values mainly located at the Southeast, Midwest, and Northern Great Plains.

(4). The performance for the SWDI predictions with solely SVMs is reasonable up to 1–2 weeks lead time at most stations. The addition of remotely sensed products can slightly improve the predictions at 2–4 weeks lead time at some stations, but the efficiency tends to decrease with the increase in lead time. The SVM-DA coupling model have good performance at longer lead time (up to 4 weeks) at most stations based on the consistent good GOF criteria signals.

(5). The strong linkage between meteorological variables with SWDI may contribute to the good performance of drought forecasting. However, the local land-atmosphere feedback may also influence the drought forecasting. At few stations, the addition of remotely sensed products (i.e., LAI, AMSR-E, SMAP) as input may not perform well due to the contrasting nature of association between SWDI and meteorological (Remotely sensed) products. For example, few stations either have weak (strong) linkage between meteorological (remotely sensed) variables with SWDI or vice versa.

Conflict of interest

Authors do not have COI for their manuscript.

Acknowledgement

This study was supported by the United States Department of Agriculture (USDA) award 2015-68007-23210, the Fundamental Research Funds for the Central Universities of China (Grant No. 2015B00214), the National Natural Science Foundation of China (Grant Nos. 41471016; 41323001; 51539003), the National Key Research and Development Program of China (Grant Nos. 2016YFC0402706; 2016YFC0402710), the Special Fund of State Key Laboratory of Hydrology-Water Resources (Grant No. 20145027312) and the open foundation of State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, HoHai University (Grant No. 2015490311).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jhydrol.2017.07.049.

References

Allen, R.G., Pereira, L.S., Raes, D., et al., 1998. Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements. Irrig. Drain. Pap. 56, FAO, Rome.
Brocca, L., Hasenauer, S., Lacava, T., et al., 2011. Soil moisture estimation through ASCAT and AMSR-E sensors: an intercomparison and validation study across Europe. Remote Sens. Environ. 115, 3390–3408.
Brown, M.E., Escobar, V., Moran, S., et al., 2013. NASA's soil moisture active passive (SMAP) mission and opportunities for applications users. Bull. Amer. Meteor. Soc. 94 (8), 1125–1128.
Cheng, K.S., Lien, Y.T., Wu, Y.C., Su, Y.F., 2016. On the criteria of model performance evaluation for real-time flood forecasting. Stoch. Environ. Res. Risk Assess., 1–24
Das, N., Entekhabi, D., Njoku, E., 2011. Algorithm for merging SMAP radiometer and radar data for high resolution soil moisture retrieval. IEEE Trans. Geosci. Remote Sens. 49, 1504–1512.
Dirmeyer, P.A., Schlosser, C.A., Brubaker, K.L., 2009. Precipitation, recycling, and land memory: An integrated analysis. J. Hydrometeorol. 10 (1), 278–288.
Drusch, M., Wood, E.F., Gao, H., 2005. Observation operators for the direct assimilation of TRMM microwave imager retrieves soil moisture. Geophys. Res. Lett. 32, L15403.
Entekhabi, D., et al., 2010. The Soil Moisture Active Passive (SMAP) mission. Proc. IEEE 98, pp. 704–716.
Escorihuela, M.J., Chanzy, A., Wigneron, J.P., et al., 2010. Effective soil moisture sampling depth of L-band radiometry: A case study. Remote Sens. Environ. 114, 995–1001.
Gill, M.K., McKee, M., 2007. Soil moisture data assimilation using support vector machines and ensemble Kalman filter. J. Am. Water Resour. Assoc. 43, 1004–1015.
Han, X., Franssen, H.-J.H., Montzka, C., et al., 2014. Soil moisture and soil properties estimation in the Community Land Model with synthetic brightness temperature observations. Water Resour. Res. 50, 6081–6105.
Hunt, E.D., Hubbard, K.G., Wilhite, D.A., Arkebauer, T.J., Dutcher, A.L., 2009. The development and evaluation of a soil moisture index. Int. J. Climatol. 29, 747–759.
Ines, A.V.M., Das, N.N., Hansen, J.W., et al., 2013. Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. Remote Sens. Environ. 138, 149–164.

Jaeger, E.B., Seneviratne, S.I., 2011. Impact of soil moisture-atmosphere coupling on European climate extremes and trends in a regional climate model. Clim. Dyn. 36, 1919–1939.
Kaheil, Y.H., Gill, M.K., McKee, M., et al., 2008. Downscaling and assimilation of surface soil moisture using ground truth measurements. IEEE Trans. Geosci. Remote Sens. 46, 1375–1384.
Kerr, Y.H., Waldteufel, P., Wigneron, J.P., Delwart, S., Cabot, F., Boutin, J., Escorihuela, M.J., Font, J., Reul, N., Gruhier, C., Juglea, S.E., 2010. The SMOS mission: New tool for monitoring key elements ofthe global water cycle. Proc. IEEE 98 (5), 666–687.
Keshavarz, M.R., Vazifedoust, M., Alizadeh, A., 2014. Drought monitoring using a soil wetness deficit index (SWDI) derived from MODIS satellite data. Agric. Water Manage. 132, 37–45.
Kornelsen, K.C., Coulibaly, P., 2014. Root-zone soil moisture estimation using data-driven methods. Water Resour. Res. 50, 2946–2962.
Koster, R.D., Guo, Z.C., Dirmeyer, P.A., et al., 2006. GLACE: The global land-atmosphere coupling experiment Part I: Overview. J. Hydrometeorol. 7, 590–610.
Kumar, S.V., Riechle, R.H., Peters-Lidard, C.D., et al., 2008. A land surface data assimilation framework using the land information system: Description and applications. Adv. Water Resour. 31, 1419–1432.
Kumar, S.V., Peters-Lidard, C.D., Mocko, D., Reichle, R., Liu, Y., Arsenault, K.R., Xia, Y., Ek, M., Riggs, G., Livneh, B., Cosh, M., 2014. Assimilation of remotely sensed soil moisture and snow depth retrievals for drought estimation. J. Hydrometeorol. 15, 2446–2469.
Li, B., Toll, D., Zhan, X., et al., 2012. Improving estimated soil moisture fields through assimilation of AMSR-E soil moisture retrievals with an ensemble Kalman filter and a mass conservation constraint. Hydrol. Earth Syst. Sci. 16, 105–119.
Liu, D., Mishra, A.K., Yu, Z., 2016. Evaluating uncertainties in multi-layer soil moisture estimation with support vector machines and ensemble Kalman filtering. J. Hydrol. 538, 243–255.
Liu, D., Mishra, A.K., 2017. Performance of AMSR_E soil moisture data assimilation in CLM4. 5 model for monitoring hydrologic fluxes at global scale. J. Hydrol. 547, 67–79.
Liu, D., Wang, G.L., Mei, R., et al., 2014a. Impact of initial soil moisture anomalies on climate mean and extremes over Asia. J. Geophys. Res. 119, 1–14.
Liu, D., Wang, G.L., Mei, R., et al., 2014b. Diagnosing the strength of land-atmosphere coupling at sub-seasonal to seasonal time scales in Asia. J. Hydrometeorol. 15, 320–339.
Liu, D., Yu, Z., Lv, H.S., 2010. Data assimilation using support vector machines and ensemble Kalman filte for multi-layer soil moisture prediction. Water Sci. Eng. 3, 361–377.
Martínez-Fernández, J., González-Zamora, A., Sánchez, N., et al., 2015. A soil water based index as a suitable agricultural drought indicator. J. Hydrol. 522, 265–273.
Martínez-Fernández, J., González-Zamora, A., Sánchez, N., et al., 2016. Satellite soil moisture for agricultural drought monitoring: assessment of the SMOS derived Soil Water Deficit Index. Remote Sens. Environ. 177, 277–286.
Mei, R., Wang, G., 2011. Impact of sea surface temperature and soil moisture on summer precipitation in the United States based on observational data. J. Hydrometeor. 12, 1086–1099.
Mei, R., Wang, G., 2012. Summer land–atmosphere coupling strength in the United States: Comparison among observations, reanalysis data, and numerical models. J. Hydrometeorol. 13 (3), 1010–1022.
Mishra, A.K., Ines, A.V.M., Das, N.N., et al., 2015. Anatomy of a local-scale drought: application of assimilated remote sensing products, crop model, and statistical methods to an agricultural drought study. J. Hydrol. 526, 15–29.
Mishra, A.K., Singh, V.P., 2010. A review of drought concepts. J. Hydrol. 391, 202–216.
Montzka, C., Moradkhani, H., Weihermüller, L., et al., 2011. Hydraulic parameter estimation by remotely-sensed top soil moisture observations with the particle filter. J. Hydrol. 399, 410–421.
Moradkhani, H., Sorooshian, S., Gupta, H.V., et al., 2005. Dual state–parameter estimation of hydrological models using ensemble Kalman filter. Adv. Water Resour. 28, 135–147.
Myneni, R., Knyazikhin, Y., Park, T., 2015. MCD15A2H MODIS/Terra + Aqua Leaf Area Index/FPAR 8-day L4 Global 500 m SIN Grid V006. NASA EOSDIS Land Processes DAAC.
Njoku, E.G., Jackson, T.J., Lakshmi, V., et al., 2003. Soil moisture retrieval from AMSR-E. IEEE Trans. Geosci. Remote Sens. 41, 215–229.
Notaro, M., Liu, Z., Williams, J.W., 2006. Observed vegetation–climate feedbacks in the United States. J. Climate 19 (5), 763–786.
O'Neill, P.E., Chen, S., Njoku, E.G., et al., 2015. SMAP L3 radiometer global daily 36 km EASE-grid soil moisture, version 1.
O'Neill, P.E., Chan, S., Njoku, E.G., et al., 2016. SMAP L3 radiometer global daily 36 km EASE-grid soil moisture, version 2.
Pan, M., Cai, X., Chaney, N.W., et al., 2016. An initial assessment of SMAP soil moisture retrievals using high-resolution model simulations and in situ observations. Geophys. Res. Lett. 43 (18), 9662–9668.
Panu, U.S., Sharma, T.C., 2002. Challenges in drought research: some perspectives and future directions. Hydrol. Sci. J. 47, S19–S30.
Reichle, R.H., Mclaughlin, D.B., Entekhabi, D., 2002. Hydrologic data assimilation with the ensemble Kalman filter. Mon. Weather Rev. 130, 103–114.
Reichle, R.H., Koster, R.D., 2004. Bias reduction in short records of satellite soil moisture. Geophys. Res. Lett. 31 (19).
Ridler, M.-E., Madsen, H., Stisen, S., et al., 2014. Assimilation of SMOS-derived soil moisture in a fully integrated hydrological and soil-vegetationatmosphere transfer model in Western Denmark. Water Resour. Res. 50, 8962–8981.
Samuel, J., Coulibaly, P., Dumedah, G., et al., 2014. Assessing model state and forecasts variation in hydrologic data assimilation. J. Hydrol. 513, 127–141.
Savage, M.J., Ritchie, J.T., Bland, W.L., et al., 1996. Lower limit of soil water availability. Agron. J. 88, 644–651.
Scaini, A., Sánchez, N., Vicente-Serrano, S.M., et al., 2015. SMOS-derived soil moisture anomalies and drought indices: a comparative analysis using in situ measurements. Hydrol. Process. 29, 373–383.
Seaber, P.R., Kapinos, F.P., Knapp, G.L., 1987. Hydrologic Unit Maps. U.S. 822 Geological Survey Water-Supply Paper 2294.
Sridhar, V., Hubbard, K.G., You, J., et al., 2008. Development of the soil moisture index to quantify agricultural drought and its "user friendliness" in severity-area-duration assessment. J. Hydrometeorol. 9 (4), 660–676.
Tran, A.P., Vanclooster, M., Lambot, S., 2013. Improving soil moisture profile prediction from ground-penetrating radar data: A maximum likelihood ensemble filter approach. Hydrol. Earth Syst. Sci. 17, 2543–2556.
Wagner, W., Lemoine, G., Rott, H., 1999. A method for estimating soil moisture from ERS scatterometer and soil data. Remote Sens. Environ. 70 (2), 191–207.
Yan, H., Moradkhani, H., Zarekarizi, M., 2017. A probabilistic drought forecasting framework: a combined dynamical and statistical approach. J. Hydro. 548, 291–304.
Yin, J., Zhan, X., Zheng, Y., et al., 2015. Optimal ensemble size of ensemble Kalman filter in sequential soil moisture data assimilation. Geophys. Res. Lett. 42.
Yin, J., Zhan, X., Zheng, Y., et al., 2014. Impact of quality control of satellite soil moisture data on their assimilation into land surface model. Geophys. Res. Lett. 41, 7159–7166.
Yu, Z., Liu, D., Lü, H., et al., 2012. A multi-layer soil moisture data assimilation using support vector machines and ensemble particle filter. J. Hydrol. 475, 53–64.
Entekhabi, D., Njoku, E.G., O'Neill, P.E., et al., 2010. The soil moisture active passive (SMAP) mission. Proc. IEEE 98 (5), 704–716.