This is an Accepted Manuscript of an article published in the Journal of the American Water Resources Association published in December 2022, available online:
https://doi.org/10.1111/1752-1688.12965

Sensitivity of Enhanced Vegetation Index to Satellite-Derived Hydrologic Predictors in the Colorado River Basin, 2001-2019

Linnea Saby, Kevin L. McKee, Prakrut Kansara, Jonathan L. Goodall, Lawrence E. Band, Venkataraman Lakshmi

Department of Engineering Systems and Environment (Saby, Kansara, Goodall, Lakshmi, Band), University of Virginia, Charlottesville, Virginia, USA; Department of Environmental Sciences (Band), University of Virginia, Charlottesville, Virginia, USA; Department of Statistics (McKee), Virginia Polytechnic Institute and State University, Blacksburg, VA, USA. (Correspondence to Saby: ls3en@virginia.edu).

Research Impact Statement: We compare spatially-distributed long time series precipitation, surface soil moisture, and root zone soil moisture datasets for prediction of vegetation in the semi-arid Colorado River Basin.

ABSTRACT:
In water-limited regions, it is important to understand the response of vegetation to hydrologic predictors for fire management and water resource planning. We present a novel spatially-distributed analysis of ecohydrological interactions in the semi-arid Colorado River Basin (CRB) over 18 years, 2001-2019. The hydrologic predictors used include precipitation from the Integrated Multi-satellitE Retrievals for GPM (IMERG), 0-10 cm depth Surface Soil Moisture (SSM) estimations from the National Land Data Assimilation System (NLDAS), and newly available 0-40-cm depth SoilMERGE root zone soil moisture (RZSM) estimations. These are evaluated using time-lagged correlations with the Enhanced Vegetation Index (EVI) from MODerate Resolution Imaging Spectroradiometer (MODIS). We demonstrate that EVI response is stronger and more immediate to the hydrologic predictors in the hot and dry southwestern CRB, with lags times of 0-32 days in this region. We found SSM to be the best predictor of EVI, with interquartile range correlations of 0.20-0.37 across the CRB. RZSM had slightly lower interquartile range correlations with EVI of 0.15-0.35, and precipitation was the least effective predictor of EVI with interquartile range correlations of 0.11-0.26. Plotting these descriptive cross-correlations provides an overview of temporal dependence between SSM, RZSM, precipitation, and EVI with publicly available datasets.

(**KEYWORDS**: Vegetation response, remote sensing, ecohydrology, soil moisture)

1 INTRODUCTION

In the Colorado River Basin (CRB) located in the semi-arid southwestern United States, hydrologic models are used to inform key management decisions for reservoir operations, irrigation scheduling, fire management, and consumptive use monitoring (Lukas and Payton, 2020). The accuracy of these models has immediate implications for water resource availability, food security, and frequency of forest fires for the 40 million people who depend on the river (Pulwarty et al., 2005). In the CRB, over 90% of water that falls as precipitation returns to the atmosphere through evapotranspiration (Lukas and Payton, 2020). Plant physiology, climate, and soil type all interact to dynamically control the rate of transpiration, adding complexity and error in hydrologic models (Mahmood and Hubbard, 2003; van Wijk, 2011). Desert-adapted succulents often have extensive shallow root systems that absorb precipitation immediately and efficiently, while forested areas have deeper root systems that take advantage of slower percolation of moisture into a deeper root zone (Schulze et al., 1996). Since 2000, the CRB has been experiencing an historic drought, which can change these patterns by increasing sensitivity to water availability and selecting for more adaptable vegetation types.

Parameterizing the heterogeneous vegetation response across the landscape is helpful in constraining and validating hydrologic models that inform management decisions (Feddes et al., 2001; Bauerle and Bowden, 2011). In particular, studies have used vegetation response to provide insight into vegetation carbon and water use, as well as to provide an indicator of soil moisture content (Carlson et al., 1994; Gillies et al., 1997; Mallick et al., 2009). Analyzing the time-lagged correlations between remotely sensed vegetation greenness and hydrologic variables is a common method for parameterizing vegetation response (Hicke et al., 2003; Wang et al., 2004). In this context, vegetation sensitivity can be measured by the correlations between predictor variables, such as precipitation and soil moisture, and time-lagged spectral vegetation indices such as the Enhanced Vegetation Index (EVI). The vegetation response time is the lag at which the correlation is maximal.

The vegetation sensitivities measured through lagged correlations can be used as an indicator of soil moisture content (McNally et al., 2016). Wang et al. (2007) showed that remotely-sensed vegetation indices could be used to predict 42-74% of soil moisture variations in grassland and shrubland semi-arid regions, including in the CRB. Schnur et al. (2010) compared vegetation indices with in-situ soil moisture in the Southwestern U.S., and found that vegetation indices could account for 50-88% of the variation in observed soil moisture. These studies provide support for hydrologic models by constraining ecohydrological dynamics, and also serve to advance understanding of available datasets.

The lagged time at which the peak correlation occurs provides insight into the timing of hydrological processes that depend on vegetation conditions (Udelhoven et al., 2015). Zhang et al. (2011) found that vegetation responds to in-situ soil moisture anomalies after 8 to 56 days in the central and western United States. Cañón et al. (2011) explored Advanced Very High-Resolution Radiometer (AVHRR) Normalized Difference Vegetation Index (NDVI) response to precipitation in the CRB from 1986-2006, and presented ecoregion-averaged results. They found a 1-3-month NDVI response time to precipitation anomalies, with the more immediate 1-month responses falling in more arid ecoregions. Udelhoven et al. (2009) used spatially-distributed lag models and found that NDVI responded to precipitation in 1 month, with local deviations up to 3 months over semi-arid Spain. Colditz et al. (2017) found NDVI response times of 1 to 1.5 months with precipitation in Mexico. Yang et al. 1997 and Wang et al. (2003) found an NDVI response time to precipitation of 5-7 weeks in the Great Plains, U.S.

To date, satellite sensors alone are not capable of detecting water content below a depth of a few centimeters, which has been cited as a limitation to spatially-distributed correlation of vegetation response to hydrologic conditions (Bolten and Crow, 2012). The newly available Soil Merge (SMERGE) product holds promise to overcome this limitation by estimating RZSM at 40cm depth. SMERGE combines remotely sensed SSM data with model-based 40-cm depth RZSM from the North American Land Data Assimilation System (NLDAS, Tobin et al., 2019). SMERGE provides deeper soil moisture estimations than can be observed by satellite, though others have pointed out that the available distributed RZSM estimations may not be better correlated with vegetation than SSM estimations. Qiu et al. (2014 and 2016) and showed that SSM measurements give a similar amount of information as vertically integrated soil moisture for prediction of vegetation anomalies. They acknowledge that the integration method used is simple, but the results draw into question the assumption that estimated RZSM improves upon available surface measurements.

Given the utility of parameterizing vegetation response to hydrologic predictors, and the ongoing debate about RZSM versus SSM estimations, there is a need to compare the performance of different hydrologic variables in predicting vegetation response in different ecological settings. This is especially salient in arid and semi-arid regions experiencing increased drought as a result of climate change, where understanding and parameterizing vegetation response is critical to managing a scarce resource.

In this study we parameterize the sensitivity and lag time for vegetation response to precipitation, SSM, and new SMERGE RZSM estimations in the CRB using a spatially-distributed lagged correlation analysis. Our study period is 2001-2019, a period characterized by historic drought in the region. SMERGE RZSM has not previously been evaluated as a means to overcome

RZSM limitations in this context, nor has a study evaluated spatially-distributed vegetation response in the CRB over a long time series. We address the following specific research questions:

1. How does EVI response to hydrologic predictors vary over space and time in the CRB?
2. Which long-time series and spatially-distributed hydrologic datasets (precipitation, RZSM, or SSM) are better predictors of vegetation response in the CRB?

Our results provide a comparison of datasets for vegetation response studies in the CRB, as well as a spatially-distributed evaluation of vegetation sensitivity to hydrologic variables under drought conditions in a semi-arid region.

2 METHODS AND DATA

2.1 Data

The following subsections provide brief descriptions of each data source in the models for this study. All data are freely available to the public from the National Aeronautics and Space Administration (NASA) Earth Data Online portal (https://earthdata.nasa.gov/).

2.1.1 Vegetation: MODIS EVI, MOD13A1. MODIS EVI is a measure of vegetation greenness generated from atmospherically-corrected reflectance in the red, near, infrared, and blue wavebands. EVI corrects for some atmospheric conditions and solar incidence angle, and is more sensitive in areas with dense vegetation compared with NDVI (Huete et al., 2002). MODIS EVI has been validated extensively for vegetation dynamics studies. MOD13A1 is generated using a weighted temporal average of the MOD13A1 8-day repeat EVI product to reduce cloud cover and optimize accuracy for a 16-day and 0.05 ^{o} temporal repeat and spatial resolutions. A complete description of MODIS EVI product can be found in Didan et al. (2015).

2.1.2 Precipitation: IMERG. Integrated Multi-satellitE Retrievals for GPM (IMERG) is a unified satellite precipitation product that estimates global surface precipitation. IMERG V06 Final Run data acquired for this study are daily at 0.1° spatial resolution. We use the accumulated precipitation for each 16-day interval. Estimates from microwave and IR satellites are calibrated using the Global Precipitation Mission core satellite, and the result is merged with gauge data. IMERG data are validated through the Ground Validation Network. A complete description and validation of IMERG precipitation can be found in Huffman et al. (2015).

2.1.3 Root zone soil moisture (RZSM): SoilMERGE (SMERGE). SMERGE is a daily 40-cm depth soil moisture product for the continental United States that became available in 2018. It was developed by merging North American Land Assimilation System (NLDAS)-Noah 2 land surface model output with surface satellite retrievals from the European Space Agency (ESA) Climate Change Initiative (CCI). In accordance with NLDAS data, the spatial resolution is 0.125 °. Validation of SMERGE has demonstrated a significant improvement in estimating RZSM over using either product alone (Tobin et al., 2019). NLDAS-Noah 2 RZSM uses a single set of hydraulic parameters for 4 soil layers and calculates soil moisture with Richards Equation (Richards, 1931). ESA-CCI measurements are 0-7cm depth observations that merge inputs from multiple sensors (Dorigo et al., 2017). These two products are merged, with relative weights of each depending on the availability of data for CCI products. During 2001-2019, 97% of the Continental United States area had at least one monthly CCI measurement. A complete description and validation of SMERGE can be found in Tobin et al. (2019).

The optimal depth for RZSM measurements varies by vegetation type, but Wang et al. (2007) found major root zone depths of 5-50cm in the two semi-arid grass and shrub sites tested in the CRB. Other studies have shown much deeper rooting depth in desert biomes, with an average of 9.5 meters (Canadell et al., 1996).

2.2.4 Surface soil moisture (SSM): NLDAS-2 NOAH. NLDAS-2 produces land-surface model datasets for use in environmental modeling with 0.125 ^{o} spatial resolution. We selected the top soil layer from the NOAH model (0-10 cm) to represent SSM. Infiltration into the SSM layer is determined by a conceptual parameterization that considers the heterogeneity over the area of precipitation, as well as local potential for infiltration (Schaake et al., 2004). NLDAS-2 provides data from January 1979 to present at an hourly time step. The NASA's Soil Moisture Active Passive (SMAP) is currently collecting satellite SSM observations, but these data are only available since 2015. The NLDAS-2 NOAH SSM data have been shown to have a daily anomaly correlation of 0.59 with observed data in the southwest region. More information and complete validation of NLDAS SSM is available in Xia et al., 2014.

2.2 Study Area

The CRB is a 630,000 square kilometer basin in the semi-arid southwestern United States. Figure 1 shows the land cover as defined by the 2016 National Land Cover Database (NLCD, Homer et al., 2018). Across the CRB, most land with sufficient grass or shrubs is grazed by cattle, which has a significant impact on vegetation density and type. Vegetation density is further controlled by elevation, with higher elevations receiving more precipitation and cooler peak temperatures which support more forests. This is particularly clear in the northeastern quadrant, where scrub is dominant at low elevations, aspen, juniper-oak, and Douglas-fir forests dominant at middle elevations and coniferous forests dominant at high elevations (USEPA, 2013). For the purposes of this study, we have not included the 1.25% of the basin that falls in Mexico because this area lacks an EPA ecoregion definition.

*Insert Figure 1 here

2.2.1 Ecoregions. The CRB contains 12 EPA-defined Level-III ecoregions. Each ecoregion is distinguished by common geography and relatively uniform regional solar radiation and precipitation (USEPA, 2013). In this study, we refer to these ecoregions frequently as spatial references to discuss physical characteristics of our results. In some ecoregions, local variations in slope and aspect can result in significant heterogeneity in anticipated vegetation response. Figure 2 presents an overview of important characteristics of each ecoregion. More detailed definitions of Level III ecoregions can be found at https://www.epa.gov/eco-research/ecoregions.

*Insert Figure 2 here

2.2.2 Climate. Basin-wide climate and hydrologic averages from 2001-2019 are presented in Table 1. Temperature and evapotranspiration are from NLDAS-Noah 2 (Xia et al., 2012). Precipitation is derived from Integrated Multi-satellitE Retrievals for GPM (IMERG, Huffman et al., 2015). Most CRB ecoregions have distinct wet and dry seasons. The wet season in the southwestern Sonoran Basin and Range moves from winter to summer across a west-east gradient. The southeastern Chihuahuan Desert ecoregion has a summer wet season. The central Arizona/New Mexico Plateau, Colorado Plateau, Southern Rockies, and Wasatch and Uinta Mountains tend towards winter moisture regimes, with snow at higher elevations especially in the Southern Rockies. The Madrean Archipelago and Mojave Basin and range have mixed moisture regimes, receiving some precipitation in winter and summer (Cañón et al., 2011; USEPA, 2013).

*Insert Table 1 here

Figure 3A shows average annual precipitation across the CRB from IMERG data, and Figure 3B and C shows MODerate Resolution Imaging Spectroradiometer (MODIS) EVI annual mean and standard deviations, respectively. Vegetation in the CRB is water-limited, and clear patterns relate Panel A with Panel B. The Southern Rockies and Arizona/New Mexico Mountains ecoregions are distinguished by both the highest monthly rainfall and most dense forested vegetation. The Sonoran Basin and Range in the southwest has the lowest rainfall, and vegetation is dominated by Palo Verde-cactus shrub and Giant Saguaro cactus.

*Insert Figure 3 here

Figure 3B shows higher mean EVI values in forested regions in the Arizona/New Mexico Mountains and the Southern Rockies. EVI heterogeneity in the Southern Rockies ecoregion is associated with elevation banding, with less vegetation in the low and dry areas. Vegetation densities remain low throughout the drier portions of the basin in the Wyoming Basin, Arizona/New Mexico Plateaus, and the Mojave and Sonoran basins. The standard deviation in Figure 3C reflects seasonal flux intensities, and is largest in areas with deciduous forests and areas that receive seasonal moisture input from snowmelt. The Southern Rockies, Middle Rockies, and Wasatch and Uinta mountains demonstrate the strongest seasonal patterns. The conifer-dominated forests in the Arizona-New Mexico Mountains have high mean EVIs but low EVI standard deviation because greenness is stable year-round.

2.3 Pre-processing

Figure 4 provides a high-level overview of our data preprocessing workflow implemented in R coding software (R Core Team, 2017), from initial download through calculation of the standardized 16-day anomalies. Resampling of the rasters to match coordinates was accomplished using bilinear interpolation.

*Insert Figure 4 here

CRB ecosystems are subject to seasonal variability. Standardized monthly anomalies are used to avoid correlations that reflect seasonal periodicity rather than actual response to wet and dry conditions, as described in Udelhoven et al. (2009). Equation 1 was used to calculate the standardized 16-day anomalies (z-score time series), which has the seasonal trend removed:

$$z_{tj} = \frac{x_{tj} - \bar{x}_j}{s_j} \tag{1}$$

where x_{tj} is a variable value for day-of-year *j* and year *t*, and $\bar{x}_j$ is the mean value in day-of-year *j* for the period of study (2001-2019), and s_j is the standard deviation of $\bar{x}_j$. All datasets had seasonal fluctuations removed and anomalies calculated with this method.

2.4 Lagged correlation

A matrix of lagged z-score time series was generated up to 1 year, and Pearson's correlation coefficient was applied to each row to obtain correlation coefficients for each lag *l*, i.e., the cross-correlation function (CCF) *r(l)*, with *l* measured in 16-day intervals:

$$r(l) = \frac{\mathbb{E}[(x_{t-l} - \mu_x)(y_t - \mu_y)]}{\sigma_x \sigma_y}, \; l \in 1,2,\ldots \tag{2}$$

Where x and y are time series vectors of the Z-score anomalies, μ is the time average and σ is the average deviation from μ. Figure 5 shows how the lagged cross-correlation coefficient summarizes the alignment of two time series EVI and RZSM when EVI is shifted 0-2 months. On average, variations in *y* correspond to variations in *x* from *l* months prior. The lagged correlation is therefore a simple method for inferring sequential causation. If y is EVI, then we can infer the response strength of EVI to precipitation and soil moisture. $r(l)$ is a dimensionless value strictly between -1 and 1, and the proportion of variation in y explained by variation in x at lag l is calculated as $r(l)^2$.

*Insert Figure 5 here

We used lagged cross-correlation to study three vegetation responses. First, we found the lagged correlation between EVI and precipitation. Second, we found the lagged correlation of EVI and SSM, followed by the lagged correlation of EVI and RZSM. This was accomplished for all pixels across the CRB at 0.125 ° spatial resolution and 16-day timesteps.

To mask noise (i.e., sampling error) in visualizations of our results, the upper 97.5 percentile of the marginal null distribution per pixel was used as a simple threshold below which pixels are classified as having no effect when choosing the most prominent lag time. The threshold is given automatically by the autocorrelation function in *R*.

For reference, an overview of the analysis workflow described in this subsection is presented in Figure 6. The workflow starts with pre-processed data and ends with our final results.

*Insert Figure 6 here

3 RESULTS AND DISCUSSION

3.1 Overall EVI sensitivity and response times

Figure 7 shows the maximum lagged EVI correlations with each predictor variable. EVI was generally more sensitive to soil moisture than to precipitation, as expected. Overall basin average correlation (*r*) of EVI to precipitation was 0.20 (standard deviation of 0.08), EVI to SSM was 0.29 (standard deviation of 0.13), and EVI to RZSM was 0.27 (standard deviation of 0.14). Correlations increased along a Northeast-Southwest gradient for all hydrologic variables, with the highest correlations in the hottest and driest ecoregions. These areas with relatively higher correlations are primarily grass and desert regions with low mean annual EVI and high potential evapotranspiration (PET), where grass and new brush growth emerges directly following rain events. The immediate and strong vegetation response to both soil moisture and precipitation in the Mojave Basin and Range and western Arizona-New Mexico Plateau is indicative of desert-adapted vegetation with shallow root systems. This general pattern of higher correlations in the southwestern quadrant is in accordance with the findings of Cañón et al. (2011), who studied NDVI response to precipitation at the lumped ecoregion scale.

Forested areas tended to have lower EVI sensitivity to all of the hydrologic predictors. This pattern can be observed in the northern Colorado Plateau ecoregion, which supports a band of mixed deciduous and evergreen forest, in the coniferous forests in the northern Arizona/New Mexico Mountains, and throughout the Southern Rockies ecoregion. This low sensitivity can be described by several factors. First, the hydrologic predictors used are all relatively shallow compared with the rooting zone of trees, which may require moisture at depths of several meters (Canadell et al., 1996). Rooting depth can vary widely depending on local soil, climate and hydrology. Second, the high EVI standard deviation in the Southern Rockies ecoregion (shown in Figure 3C) suggests considerable seasonal changes in EVI. Though seasonal trends are removed in the z-score anomalies (see Equation 1), the low sensitivities in the deciduous areas indicates seasonal variation in anomaly correlations. This explains the relatively stronger sensitivities in the evergreen-dominated regions compared with the deciduous Southern Rockies.

Third, changes in the conditions of forests may not be well-captured by remotely-sensed vegetation indices including EVI, which provide more accurate measurements of primary production across grass and shrublands (Schloss et al., 2001). If the forest conditions are poorly represented by EVI, then these areas are likely more sensitive to precipitation and soil moisture than can be depicted with remotely sensed vegetation indices. Forested areas in the CRB have high annual precipitation relative to the rest of the basin, so may be more important to the overall CRB water balance than more arid regions.

The lower correlations in forested areas may also be caused by other disturbances, such as bark beetle outbreaks and fires. Drought and warming experienced in the CRB has led to an increase frequency and intensity in both of these events, which have caused sharp drops in EVI in forests throughout the CRB during our study period. In the steep and rugged Southern Rockies ecoregion, which primarily receives precipitation as snowfall, avalanches and blow-down events can also cause major EVI disturbances (Drummond, 2012).

Figure 8 illustrates the vegetation response time during a 365-day period (at 16-day intervals) across the CRB. Median EVI response to precipitation was 112 days, and response to both RZSM and SSM was 48 days. The delay in response for precipitation can be attributed to the snowmelt delay, where precipitation does not become available to vegetation until snowmelt occurs. In the southern CRB, where vegetation does not rely on snowmelt, the EVI response time to soil moisture was within the range found with in-situ measurements in Mexico by Zhang et al. (2011), who found responses between 0-56 days. A zero-day lag indicates that EVI is in-phase with the hydrologic predictor. In the southern CRB, lag times were not substantially longer for precipitation than soil moisture. Wang et al. (2007) found that vegetation in arid sites responded more quickly to soil moisture than non-arid sites by about 10 days, which is also supported by our findings.

*Insert Figure 7 here

*Insert Figure 8 here

Response times in the southern CRB are associated with the timing of the wet season. In all three panels of Figure 8, a gradient of response times can be observed across the southern CRB where the response time increases from 0 days to 48 days moving east to west. This follows same pattern as the summer-winter wet season gradient, such that summer wet season areas have a lag time of 0 to 16 days, and winter moisture regimes have a lag time of 32 to 48 days. The late summer rains in the southeastern CRB are heavy monsoon rains (Notaro et al., 2010), which trigger more immediate peak in vegetation response compared with the lighter winter rains, which likely are generating a more cumulative response.

In the Wyoming Basin and Colorado Plateau ecoregions, EVI showed a highly heterogeneous response time to RZSM, with patches of 0-16 day and up to 160 days response times (Figure 8 Panel C). When these areas are compared with Panels A and B, it is clear that SSM and precipitation responses reflect land cover pattens, but RZSM does not. This random noise in the RZSM response time corresponds to areas where correlations at all lags were small and thus had high sampling error variance. In such areas, the peak lag number is expected to be uniformly distributed across its tested values, as we would be taking the maximum value among noise.

Correlations shown in Figure 7 indicate how well vegetation represents the underlying hydrologic variables, and suggest that that in data-poor arid biomes EVI could be used to infer soil moisture and vice-versa. The vegetation response times to hydrologic variables shown in Figure 8 can be used to validate and constrain vegetation forecasts in the CRB depending on climate, but the weight of these validations should depend on the vegetation sensitivity shown in Figure 7. For example, the lag time of 0-32 days across the southern CRB can reasonably be used to parameterize vegetation response in hydrologic models. The noisy and longer lag times in the northern snowmelt-fed regions mean that validation with lagged EVI correlation offers less confident validations. Both correlation strength and lag plots should be considered jointly, with parameterization using lagged correlations given less weight in less sensitive areas.

3.2 Comparison of hydrologic variables for predicting EVI

Figure 9 evaluates the performance of the hydrologic variables SMERGE RZSM product by comparing EVI sensitivity to precipitation and SSM. Figure 9 Panel A shows that EVI is more sensitive to RZSM than precipitation over most of the basin, especially the more arid southwestern quadrant. In this region, much of the precipitation that falls returns to the atmosphere through evaporation before becoming available to vegetation.

*Insert Figure 9 here

It was surprising that SMERGE RZSM was not a better predictor of vegetation conditions compared with SSM over most of the CRB, even in areas with deep-rooted vegetation. This comparison is presented in Figure 9, Panel B. RZSM theoretically provides a more direct measure of water available to plants than precipitation or SSM, but our results suggest that SMERGE

RZSM does not better predict EVI. This could be because NLDAS simulations that underly SMERGE RZSM data do not account for lateral redistribution of snowmelt (Xia et al., 2012), and so the depth introduced in SMERGE RZSM does not improve upon the NLDAS SSM estimations. This deficiency may also be an indicator that the uniform hydraulic parameters used in the Noah NLDAS-2 model, which underly the SMERGE data, may not accurately describe subsurface conditions. If these parameters change significantly with depth, then the SMERGE RZSM estimates may introduce error to the estimations and detract from predictive power for the SSM alone.

This SMERGE RZSM limitation is especially apparent in the Southern Rockies ecoregion. Here, higher elevation areas contain vegetation that depends on snowmelt, and we find that RZSM is a particularly poor predictors of EVI (average peak r of 0.09), while SSM performs twice as well as measured by correlation in this ecoregion (average peak r of 0.18). We conclude that the SMERGE RZSM dataset does not provide added value for predicting vegetation conditions in the CRB compared with SSM. This finding supports the argument made by Qiu et al. (2014) that RZSM estimations may not enhance ecohydrological insights compared with SSM. The performance of RZSM estimations should be evaluated before RZSM estimations are applied in hydrological modeling to support management decisions.

3.3 Limitations and future directions

Our aim in plotting descriptive cross-correlations of pixels was to describe temporal dependence between SSM, RZSM, precipitation, and EVI with publicly available datasets. The primary limitation of this approach is the inability to rigorously incorporate hypothesis testing procedures and incorporate precise estimates of the sampling variability, false positive rates, and effect detection rates into our conclusions. Specifically, it is clear from our figures that cross-correlation estimates are themselves correlated over space and lag intervals. Future work aiming to conduct hypothesis testing should consider a multiple testing correction procedure that accounts for the shared information across pixels. For this reason, the threshold we have used to mask pixels should not be considered a complete null hypothesis significance test. Future work should take advantage of multi-level modeling may be used to mutually condition pixel-wise correlation estimates by proximity to one another, or to incorporate a regional class structure that reduces the dimensionality of the analysis. Modeling steps necessarily rely on a larger number of assumptions and design choices, whereas the current methods were chosen for simplicity and transparency.

Additional variables can also be tested in future work. In our analysis, SSM is a better overall predictor than 40cm RZSM estimations. However, the 40cm depth soil moisture data from SMERGE does not capture the entire root zone (Canadell et al., 1996), and this limitation may be apparent in the relatively low correlations identified in forested regions in the CRB. Deeper soil moisture estimations may provide higher sensitivity in certain areas.

4 CONCLUSION

This study provides an overview of the temporal and spatial relationships between vegetation and hydrologic datasets in the Colorado River Basin (CRB). We have compared 18 years of spatially-distributed soil moisture (SMERGE 0-40cm RZSM and NLDAS 0-10cm SSM) and precipitation (IMERG) data as predictors of MODIS EVI over the semi-arid CRB. We draw the following specific conclusions:

- NLDAS SSM is the best predictor of vegetation conditions of the three predictors tested over the CRB. This conclusion supports the findings of a previous study that showed that SSM measurements give a similar amount of information as RZSM estimations for prediction of vegetation anomalies (Qiu et al., 2014). Precipitation was the weakest predictor of vegetation conditions.

- The strongest vegetation sensitivities to hydrologic predictors were observed in hot and dry southwestern ecoregions dominated by shrub vegetation. Correlations approximately doubled in the southwestern CRB compared with the tree-dominated central and northern regions.
- All of the predictors had low correlations (less than 0.15) in areas dominated by deeper-rooted vegetation and snowmelt.
- Vegetation had a more immediate peak response to soil moisture compared with precipitation (median 48 days versus 112 days).
- In the arid southwestern CRB, peak vegetation response times were more immediate in areas with monsoon summer rains (0-16 days) than winter rains (32-48 days).

Vegetation plays a key role in mediating land-atmosphere interactions through transpiration. Therefore, these results can inform the use of these datasets in environmental management models in the CRB. The lag times reported demonstrate the timing and spatial patterns of vegetation-hydrology interactions, and correlations indicate how well the hydrologic data predicts vegetation. The correlation and lag time results should be considered jointly, as areas with high sensitivity will have a more accurate lag time.

References

Bauerle, W. L. and Bowden, J. D. (2011). Predicting transpiration response to climate change: insights on physiological and morphological interactions that modulate water exchange from leaves to canopies. *HortScience*, 46(2):163-166. http://dx.doi.org/10.21273/HORTSCI.46.2.163

Bolten, J. and Crow, W. (2012). Improved prediction of quasi-global vegetation conditions using remotely sensed surface soil moisture. *Geophysical Research Letters*, 39(19):1-5. http://dx.doi.org/10.1029/2012GL053470

Canadell, J., Jackson, R. B., Ehleringer, J. B., Mooney, H. A., Sala, O. E., & Schulze, E. D. (1996). Maximum rooting depth of vegetation types at the global scale. *Oecologia*, *108*(4), 583-595.

Cañón, J., Dominguez, F., and Valdes, J. B. (2011). Vegetation responses to precipitation and temperature: a spatiotemporal analysis of ecoregions in the Colorado River Basin. *International Journal of Remote Sensing*, 32(20):5665–5687. http://dx.doi.org/10.1080/01431161.2010.507259

Carlson, T.N., Gillies, R.R., Perry, E.M., (1994). A method to make use of thermal infrared temperature and NDVI measurements to infer surface soil water content and fractional vegetation cover. null 9, 161–173. https://doi.org/10.1080/02757259409532220

Colditz, R. R., Villanueva, V. L. A., Tecuapetla-G´omez, I., and G´omez Mendoza, L. (2017). Temporal relationships between daily precipitation and NDVI time series in Mexico. In *2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*, pp 1-4. http://dx.doi.org/10.1109/Multi-Temp.2017.8035202

Didan, K., Munoz, A. B., Solano, R., & Huete, A. (2015). MODIS vegetation index user's guide (MOD13 series). *University of Arizona: Vegetation Index and Phenology Lab*.

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., others, (2017). ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. Remote Sensing of Environment 203, 185–215.

Drummond, M.A. (2012). Status and Trends of Land Change in the Western United States. U.S. Geological Survey Professional Paper 1794-A.

Feddes, R. A., Hoff, H., Bruen, M., Dawson, T., De Rosnay, P., Dirmeyer, P., and Pitman, A. J. (2001). Modeling root water uptake in hydrological and climate models. *Bulletin of the American meteorological society*, 82(12):2797-2810. http://dx.doi.org/10.1175/1520-0477(2001)082%3C2797:MRWUIH%3E2.3.CO;2

Gillies, R.R., Kustas, W.P., Humes, K.S., (1997). A verification of the “triangle” method for obtaining surface soil water content and energy fluxes from remote measurements of the Normalized Difference Vegetation Index (NDVI) and surface e. null 18, 3145–3166. https://doi.org/10.1080/014311697217026

Hicke, J. A., Asner, G. P., Kasischke, E. S., French, N. H., Randerson, J. T., James Collatz, G., and Field, C. B. (2003). Postfire response of North American boreal forest net primary productivity analyzed with satellite observations. *Global Change Biology*, 9(8):1145-1157. http://dx.doi.org/10.1175/1520-0477(2001)082%3C2797:MRWUIH%3E2.3.CO;2

Homer, C. H., Fry, J. A., and Barnes, C. A. (2018). The national land cover database. *US Geological Survey Fact Sheet*, 3020(4):1–4.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment 83, 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Xie, P., and Yoo, S.-H., (2015). NASA global precipitation measurement (GPM) integrated multi-satellite retrievals for GPM (IMERG). *Algorithm theoretical basis document,* version 4:30.

Lukas, J., and Payton, E, (2020). Colorado River Basin Climate and Hydrology: State of the Science. Technical Report. https://doi.org/10.25810/3hcv-w477

Mahmood, R. and Hubbard, K. G. (2003). Simulating sensitivity of soil moisture and evapotranspiration under heterogeneous soils and land uses. *Journal of Hydrology*, 280(1-4):72-90. http://dx.doi.org/10.1016/S0022-1694(03)00183-5

McNally, A., Shukla, S., Arsenault, K. R., Wang, S., Peters-Lidard, C. D., & Verdin, J. P., (2016). Evaluating ESA CCI soil moisture in East Africa. International Journal of Applied Earth Observation and Geoinformation, 48, 96-109.

Mallick, K., Bhattacharya, B.K., Patel, N.K., (2009). Estimating volumetric surface moisture content for cropped soils using a soil wetness index based on surface temperature and NDVI. Agricultural and Forest Meteorology 149, 1327–1342. https://doi.org/10.1016/j.agrformet.2009.03.004

Notaro, M., Liu, Z., Gallimore, R.G., Williams, J.W., Gutzler, D.S., Collins, S., (2010). Complex seasonal cycle of ecohydrology in the Southwest United States. Journal of Geophysical Research: Biogeosciences 115. https://doi.org/10.1029/2010JG001382

Pulwarty, R. S., Jacobs, K. L., and Dole, R. M. (2005). The hardest working river: drought and critical water problems in the Colorado River Basin. *Drought and Water Crises. Science, Technology, and Management Issues, edited by Wilhite, DA, CRC Press, Boca Raton, FL, USA*, pp 249:285.

Qiu, J., Crow, W. T., Nearing, G. S., Mo, X., & Liu, S. (2014). The impact of vertical measurement depth on the information content of soil moisture times series data. *Geophysical Research Letters*, *41*(14), 4997-5004.

Qiu, J., W. T. Crow, and G. S. Nearing, (2016): The Impact of Vertical Measurement Depth on the Information Content of Soil Moisture for Latent Heat Flux Estimation. J. Hydrometeor., 17, 2419–2430, https://doi.org/10.1175/JHM-D-16-0044.1

R Core Team (2017). R: a language and environment for statistical computing. R foundation for statistical computing, Vienna. *www.R-project.org*.

Richards, L. A. (1931). Capillary conduction of liquids through porous mediums. *Physics*, 1(5): 318-333. http://dx.doi.org/10.1063/1.1745010

Schaake, J.C., Duan, Q., Koren, V., Mitchell, K.E., Houser, P.R., Wood, E.F., Robock, A., Lettenmaier, D.P., Lohmann, D., Cosgrove, B., Sheffield, J., Luo, L., Higgins, R.W., Pinker, R.T., Tarpley, J.D., (2004). An intercomparison of soil moisture fields in the North American Land Data Assimilation System (NLDAS). Journal of Geophysical Research: Atmospheres 109. https://doi.org/10.1029/2002JD003309

Schloss, A. L., Kicklighter, D. W., Kaduk, J., Wittenberg, U., and Intercomparison, T. P. O. T. P. N. M. (1999). Comparing global models of terrestrial net primary productivity (NPP): comparison of NPP to climate and the Normalized Difference Vegetation Index (NDVI). *Global Change Biology*, 5(S1): 25-34. http://dx.doi.org/10.1046/j.1365-2486.1999.00004.x

Schulze, E. D., Mooney, H. A., Sala, O. E., Jobbagy, E., Buchmann, N., Bauer, G., Ehleringer, J. R., (1996). Rooting depth, water availability, and vegetation cover along an aridity gradient in Patagonia. *Oecologia*, 108(3): 503-511. http://dx.doi.org/10.1007/BF00333727

Schnur, M.T., Xie, H., Wang, X., (2010). Estimating root zone soil moisture at distant sites using MODIS NDVI and EVI in a semi-arid region of southwestern USA. Ecological Informatics 5, 400–409. https://doi.org/10.1016/j.ecoinf.2010.05.001

Spruce, J.P., Gasser, G.E. and Hargrove, W.W., (2016). MODIS NDVI Data, Smoothed and Gap-filled, for the Conterminous US: 2000-2015. ORNL DAAC, Oak Ridge, Tennessee, USA. http://dx.doi.org/10.3334/ORNLDAAC/1299

Tobin, K. J., Crow, W. T., Dong, J., and Bennett, M. E., (2019). Validation of a New Root-Zone Soil Moisture Product: Soil MERGE. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(9):3351–3365. http://dx.doi.org/10.1109/JSTARS.2019.2930946

Udelhoven, T., Stellmes, M., del Barrio, G., and Hill, J., (2009). Assessment of rainfall and NDVI anomalies in Spain (1989–1999) using distributed lag models. *International Journal of Remote Sensing*, 30(8):1961– 1976. http://dx.doi.org/10.1080/01431160802546829

Udelhoven T., Stellmes M., Röder A., (2015). Assessing Rainfall-EVI Relationships in the Okavango Catchment Employing MODIS Time Series Data and Distributed Lag Models. In: *Remote Sensing Time Series. Remote Sensing and Digital Image Processing, vol 22.* Edited by Kuenzer C., Dech S., Wagner W. *Springer, Cham*. pp 225-245. http://dx.doi.org/10.1007/978-3-319-15967-6_11

USEPA (U.S. Environmental Protection Agency), (2013). Primary distinguishing characteristics of level III ecoregions of the continental United States. *Ecosystems Research*.

Van Wijk, M. T. (2011). Understanding plant rooting patterns in semi-arid systems: an integrated model analysis of climate, soil type and plant biomass. *Global Ecology and Biogeography*, 20(2): 331-342. http://dx.doi.org/10.1111/j.1466-8238.2010.00601.x

Wang, J., Rich, P.M., Price, K.P., (2003). Temporal responses of NDVI to precipitation and temperature in the central Great Plains, USA. International Journal of Remote Sensing, 24, 2345–2364. https://doi.org/10.1080/01431160210154812

Wang, J., Rich, P. M., Price, K. P., and Kettle, W. D., (2004). Relations between NDVI and tree productivity in the central Great Plains. *International Journal of Remote Sensing*, 25(16): 3127-3138. http://dx.doi.org/10.1080/0143116032000160499

Wang, X., Xie, H., Guan, H., Zhou, X., (2007). Different responses of MODIS-derived NDVI to root-zone soil moisture in semi-arid and humid regions. Journal of Hydrology 340, 12–24. https://doi.org/10.1016/j.jhydrol.2007.03.022

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, Helin; Meng, J., Livneh, B., Lettenmaier, D., Koren, V., Duan, Q., Mo, K., Fan, Y., and Mocko, D., (2012). Continental scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *Journal of Geophysical Research: Atmospheres*, 117(D03109):1-27. http://dx.doi.org/10.1029/2011JD016048

Xia, Y., Sheffield, J., Ek, M.B., Dong, J., Chaney, N., Wei, H., Meng, J., Wood, E.F., (2014). Evaluation of multi-model simulated soil moisture in NLDAS-2. Journal of Hydrology 512, 107–125. https://doi.org/10.1016/j.jhydrol.2014.02.027

Yang, W., Yang, L., Merchant, J.W., (1997). An assessment of AVHRR/NDVI-ecoclimatological relations in Nebraska, U.S.A. null 18, 2161–2180. https://doi.org/10.1080/014311697217819

Zhang, L., Ji, L., and Wylie, B. K. (2011). Response of spectral vegetation indices to soil moisture in grasslands and shrublands. *International journal of remote sensing*, 32(18):5267–5286. http://dx.doi.org/10.1080/01431161.2010.496471

Table 1. Climate and hydrologic averages for the CRB, 2001-2019.

Variable	*Mean monthly value*	*Standard deviation*
Temperature (degrees C)	11.6	8.3
Precipitation (mm)	26.9	15.3
Evapotranspiration (mm)	19.0	10.0

Figure 1. Land cover in the Colorado River Basin (CRB) within the U.S. from the National Land Cover Database (NLCD), 2016 (Homer et al., 2018). The 1.25% of the CRB that falls outside of the United States is excluded from this study.

Figure 2. Placement and description of ecoregions referred to in this study. Descriptions are summarized from the U.S. Environmental Protection Agency (USEPA, 2013). Images from https://www.epa.gov/eco-research/level-iii-and-iv-ecoregionscontinental-united-states.

Figure 3. A. Average annual rainfall (in millimeters) over the CRB from 2001-2019. B. Mean and C. standard deviation of standardized monthly Enhanced Vegetation Index (EVI) values, 2001-2019. Ecoregions are plotted for reference.

Figure 4. Data preprocessing workflow implemented in R open-source software (R Core Team, 2017). Pre-processed data is formatted as a matrix of standardized 16-day anomalies, with each row as a pixel (with X and Y properties). Columns are time points (2001-2019). CONUS= Continental United States.

Figure 5. Demonstrating the lagged correlation of anomalies for the time series of a single pixel. At Lag=0, the peaks are poorly aligned and the correlation between the time series is low. At Lag =1 the peaks are more closely aligned and the correlation is increased. At Lag = 2, the peaks are well-aligned and the highest correlation is recorded. In this example, the peak vegetation response to precipitation occurs at 2 months.

Figure 6. Analysis workflow implemented in R open-source software (R Core Team, 2017). SSM = Surface Soil Moisture. RZSM = Root Zone Soil Moisture. Noise filtering was accomplished by classifying pixels in the lowest 2.5 percentile of the marginal null distribution as having no effect.

Figure 7. Maximum correlations at the peak EVI response time across the CRB. Panel A shows the correlation for EVI and precipitation, and Panel B shows EVI and SSM, and Panel C shows EVI with RZSM. Higher correlations indicate higher sensitivity to and better prediction of EVI.

Figure 8: Peak response time in days for peak EVI response across the CRB. Panel A shows the peak response time for EVI to precipitation, Panel B shows the peak EVI response time to SSM and Panel C shows peak EVI response time to RZSM. Noisier areas correspond with greater sampling variance. A zero-day lag indicates that EVI is in-phase with the hydrologic predictor. *The upper 97.5 percentile of the marginal null distribution per pixel was used as a simple threshold below which pixels are classified as having no effect.

Figure 9. Comparison of EVI sensitivity to RZSM with precipitation (Panel A) and SSM (Panel B). Positive areas indicate that EVI was more sensitive to RZSM.