Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 185 (2021) 1–8

Procedia Computer Science

www.elsevier.com/locate/procedia

Complex Adaptive Systems Conference Theme: Big Data, IoT, and AI for a Smarter Future
Malvern, Pennsylvania, June 16-18, 2021

Identification of Dominant Warm-Season Latent Heat Flux Patterns in the Lower Mississippi River Alluvial Valley

Andrew Mercer[a]* and Jamie Dyer[a]

[a]*Mississippi State University, 108 Hilbun Hall, Mississippi State, MS 39762, USA*

Abstract

Warm-season precipitation in the Lower Mississippi River Alluvial Valley (LMRAV) is heavily dominated by the rates of evapotranspiration and surface heat fluxes and is a primary water resource for agriculture. However, the stochastic nature of LMRAV warm-season thunderstorms makes precipitation forecasts challenging. The Weather Research and Forecasting Hydrologic (WRF-Hydro) model, coupled with the multi-parameter Noah land surface (Noah-MP) model, has improved estimates of important warm-season precipitation process. Given the widespread agriculture and dominance of crop and forested landscapes over the region, proper assessment of land use / land cover (LULC) is critical in predicting warm-season precipitation patterns. The objective of this study is to quantify simulated latent heat flux sensitivity (important for warm-season precipitation) to temporally updated LULC datasets. Both the model default and annually updated LULC conditions were used to initialize a 16-year WRF-Hydro simulation from which warm-season latent heat flux estimates were obtained. Annual root mean square difference was computed at each gridpoint. Cluster analysis preprocessed with kernel principal component analysis was used to identify spatial RMSD structures that quantified sensitivity to updated LULC conditions. Results showed the largest impacts occurred directly in the LMRAV and for points slightly east and revealed a meteorological link between these regions.

Keywords: Kernel Principal Component Analysis; Land-Atmosphere Interactions; Data Mining

* Corresponding author. Tel.: +1-662-325-0738; fax: +1-662-325-9423
E-mail address: a.mercer@msstate.edu

10.1016/j.procs.2021.05.001

1. Introduction

The Lower Mississippi River Alluvial Valley (LMRAV) is a robust agricultural region in the Southeastern United States (U.S.) that relies heavily upon abundant precipitation to ensure adequate availability of water. While much of this water is provided by cool-season synoptic-scale extratropical cyclone events [1], much of the region is still sensitive to fluctuations in warm-season precipitation when the growing season is active [2]. This is an important issue as precipitation predictability is much more limited during the Southeastern U.S. warm season [3-5] owing to limitations in rendering convective processes (which are the primary drivers of warm-season precipitation) in the dynamic models used for precipitation forecasting. Previous studies have specifically addressed limitations in this predictability [6,7] using dynamic atmospheric modeling approaches with limited success. The results in [7] actually revealed a negative skill to a support vector machine [8] model trained to predict the onset of warm-season rainfall at a lead time of 24 hours, suggesting that forecasting by simple climatology (i.e., a common baseline probability of precipitation for all points) yielded better forecasts than the dynamic model.

The work in [4] revealed that modification in the soil moisture characteristics yielded minimal improvements and small ensemble spread in the Weather and Research Forecasting (WRF – [9]) dynamic modeling framework. In their study, they specifically note that a better representation of the land surface-atmosphere interactions is essential if warm-season precipitation forecasts are to gain skill. However, there is little consensus as to the best approach to coupling the land-surface component of a numerical modeling system with the atmospheric component [10].

Currently, the National Oceanic and Atmospheric Administration (NOAA) National Water Model (NWM –[11]) is an operational framework based on the WRF hydrologic modeling system (WRF-Hydro) that utilizes precipitation and other atmospheric constituents to estimate surface hydrologic conditions in the United States. The WRF-Hydro system utilizes a blend of terrestrial and subsurface hydrologic models with the Noah multi-parameterization (Noah-MP [12]) land surface model and the WRF gridded pre-processing system. While limited success has been seen in approaching the warm-season rainfall prediction problem from the perspective of the WRF model [5,6], little work has investigated the potential improvements in quantifying critical land surface-atmosphere interactions when utilizing improved land use/land cover (LULC) input. One important challenge in this process is that land surface models such as the Noah-MP do not directly produce forecasts for precipitation; instead, they portray processes such as surface latent heat (LH) flux (essentially evaporative flux) and evapotranspiration that are associated with precipitation generation. The former of these is of particular importance for warm-season precipitation in the Southeast U.S. as it is the primary moisture source for convective rainfall events (i.e., thunderstorms).

The primary objective of this study is to assess the sensitivity of the Noah-MP land surface model's rendering of LH flux in the LMRAV to updates to the LULC input fields during the Southeast U.S. warm season. Composite analyses of LH flux differences between the Noah-MP simulations within WRF-Hydro using annually updated LULC and the model default configurations will be formulated using cluster analysis pre-processed by kernel principal component analysis (KPCA – [13,14]) to identify the most prevalent spatial patterns. The outcomes of this work will reveal spatial regions and land cover types that were most sensitive to changes in the Noah-MP initialization fields, which can help inform future forecast renderings of warm-season precipitation.

2. Datasets

Within the framework of WRF-Hydro [11], this study utilized the Noah-MP model [12,15] to estimate LH flux on a 1-km grid-spacing (499x499 grid) centered over the LMRAV (Figure 1). The Noah-MP model [12] is a land-surface modeling system that utilizes multiple configuration options for rendering land surface-atmosphere interactions. Its estimates of surface heat fluxes, including LH flux, are highly sensitive to the selection of these configuration options. This study used the configuration options from [15] as they have been shown to be appropriate for the LMRAV [16].

After the initial Noah-MP configuration was established, the study objectives required updating the LULC data when initializing the Noah-MP. Two separate Noah-MP input LULC dataset configurations were tested. First, the default configuration 20-class Moderate Resolution Imaging Spectrometer (MODIS) dataset [17] used in traditional WRF simulations was applied, which is based on the most prevalent LULC category at each grid point derived from MODIS satellite estimates between 2001-2010. These data are provided on a 30 arc second latitude/longitude spatial grid and were interpolated to a 1 km study grid.

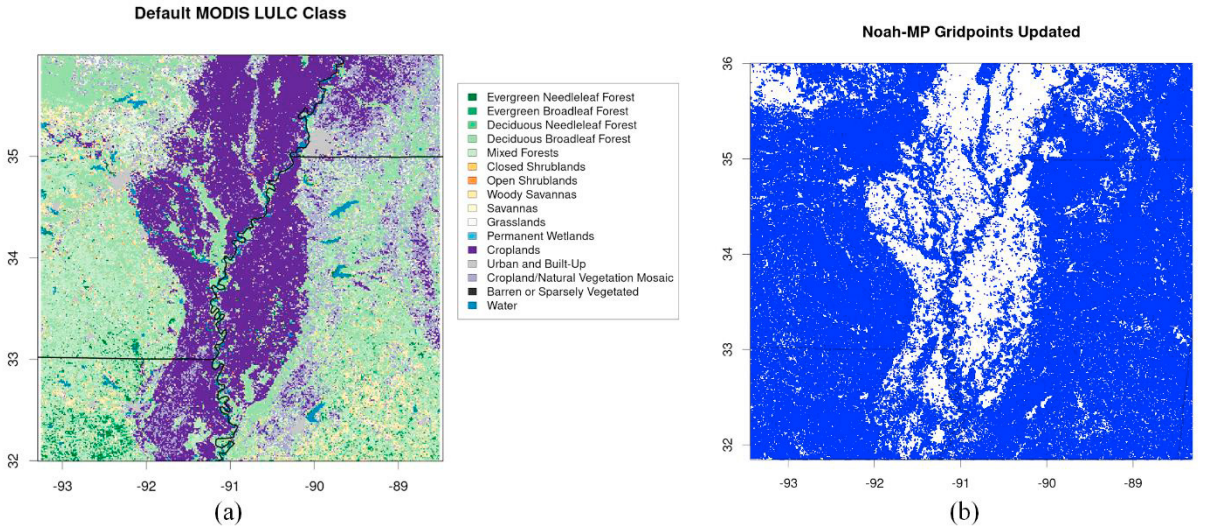

Fig. 1. (a) The default MODIS LULC category for each gridpoint used to initialize the default Noah-MP simulations. (b) Map of gridpoints that were altered by the annually updated MODIS LULC data (shaded).

For the second simulation, updated annual 20-class MODIS data (on the same spatial domain) were fed to the model for each simulation year, offering the Noah-MP model more representative LULC conditions for the second experiment. Both configurations were used to create daily Noah-MP LH flux simulations for the full warm season (June – August) for each of the study years (2003-2018). Note that the LH flux was computed within the Noah-MP model using the Penman-Monteith equation [18] such that changes to the LULC affect surface energy exchanges, surface radiation, and resistance to water vapor and sensible heat. Thus, significant modifications to the LH flux estimates were expected by updating the LULC fields. In total, 1472 daily gridded LH flux fields were obtained for the default Noah-MP simulations and the updated LULC Noah-MP simulations. Figure 1b shows the spatial regions that were modified by updating these configurations. The most common update was a shift towards MODIS class 8 (shrubland), as roughly 17% of the LULC updates were from class 4 (deciduous broadleaf forest) to class 8, roughly 14% were updates from class 5 (mixed forests) to class 8, and roughly 9% were changed from class 14 (cropland) to class 8. These updates certainly will impact LH flux results as shrubland has dramatically different heat flux characteristics than the other categories listed above.

3. Cluster Analysis Methodology

Once the Noah-MP simulation fields were obtained, the next step was to quantify the differences in the simulated LH flux fields between the two simulations. The root mean squared difference (RMSD) was computed between the default configured Noah-MP LH flux estimates and the simulated LH flux estimates from the updated LULC model configuration. The RMSD values were computed on the full warm season for each year so that a 16-year RMSD time series was obtained at each gridpoint, yielding a total of 16 RMSD maps. Spatial regions of LH flux that were consistently adjusted by the updated LULC fields exhibited elevated RMSD; therefore, spatial analysis was needed to examine sensitivity to updates in LULC by region.

These dominant spatial patterns were identified utilizing a traditional hierarchical cluster analysis [19] that employed Euclidean distance and Wards minimum variance method for cluster linkages. Previous work [20] has shown that improved clustering is possible if the data are preprocessed using a principal component analysis (PCA) [19] approach. Traditionally, PCA is formulated on a similarity matrix that describes linear covariability among the elements within the analysis, either by utilizing the covariance matrix or more commonly the Pearson correlation matrix. Recent studies [13,14,20] have shown that the kernel matrix **K** associated with support vector machines [8] can also serve as a similarity matrix (a technique known as kernel PCA – KPCA) but that this matrix has an advantage

over the covariance/correlation matrix since the kernel matrix **K** describes nonlinear variability among components. However, this does present an interpretation challenge since the relationship between **K** and the original data (here the LH flux data) is described by an unknown kernel map function φ. This means that direct physical interpretation of the PC loadings generated from **K** is not possible (unlike traditional PCA) such that the most common use of KPCA in atmospheric sciences is as a preprocessing step to cluster analysis. To assess the value of utilizing KPCA in this study, a cluster analysis that did not use the KPCA preprocessing step was compared against the results when using this KPCA preprocessing step. A linear PCA based on the correlation matrix was also tested to quantify the benefits, if any, the nonlinearity of KPCA offered to the analysis.

A metric to quantify the cluster quality was needed to determine what benefits the KPCA preprocessing step offered over the control methods. Two separate measures of cluster quality were employed. First, the silhouette coefficient S [21], a measure of cluster separation and cohesion, was used to quantify the cluster distinctness. S yields negative values for misclustered members and values near 1 for those that are completely distinct (i.e., singular points for each cluster). The mean of the silhouette coefficients for the 16 LH flux RMSD fields was computed for all cluster analysis configurations and was scaled by the percentage of correctly clustered members (those whose silhouette value was positive). The second measure of cluster success was a spatial comparison of the individual cluster members with their associated composite mean map, which is used for assessing underlying patterns.

The cluster analyses with varying preprocessing configurations required adjusting multiple tuning parameters to obtain the best setup. One important measure is the number of retained clusters. This value was carefully selected to ensure the yearly data were well distributed among the clusters (i.e., no cluster contained a vast majority of the yearly LH flux data). Additionally, both the KPCA and PCA preprocessing steps required selecting how many PCs to retain. Given only 16 years were considered, configurations retaining between 2 and 8 PCs were tested. Since KPCA does not explain variance linearly (unlike traditional PCA) but instead describes variance explained in nonlinear hyperspace, standard practice is to consider each number of retained PCs and instead optimize the S and maximize the correlation within the clusters. Finally, the kernel function in KPCA has various tuning parameters for different kernels, such as the spread parameter σ in the radial basis function kernel and the degree d in the polynomial kernel. In this study, values of σ from 2 to 50 with an interval of 2 were tested, and polynomials with degrees 1-4 were also considered (a total of 29 kernel configurations). All permutations of cluster count, PCs retained, and kernel configurations were considered for the KPCA (a total of 609 configurations). Results are provided below.

4. Cluster Analysis Results

The dendrogram from the control cluster analysis (without PCA preprocessing - Figure 2a) showed a clear 3 cluster grouping. The associated S (0.33) was among the highest of the configurations tested in the control phase and was associated with an average correlation among cluster members of 0.68. Notably, the 2009 cluster year was anomalous within its cluster (Figure 2a) and this had an impact on the average within-cluster correlation. Retaining 4 clusters resulted in keeping 2009 as its own single-member cluster and increased the average correlation metric among all clusters (0.77) while maintaining a similar S (0.32). Ultimately, since clusters with only one member are not desirable for composite analysis, 3 clusters were retained from the control group, resulting in cluster sizes of 3, 3, and 10. Interestingly, the linear PCA preprocessed cluster analysis yielded poorer results, as no PCA configuration provided an S higher than 0.2 despite modest average correlations (~0.7).

The KPCA results showed several interesting trends in terms of the sensitivity of the cluster analysis to the associated configuration parameters. First, radial basis function kernels generally had much poorer performance ($S <$ 0.2) than the polynomial kernel configurations tested. However, the polynomial kernels consistently produced identical clustering to the control group (Figure 2a). This result suggests that this clustering is likely the best possible configuration and that the KPCA preprocessing step offered little benefit to the analysis, a result attributed to the smaller sample size of LH flux years in the dataset and the dominance of the mean pattern mentioned above. Finally, the number of retained KPCs had no effect on the resulting clustering (Figure 2b) with the polynomial kernels, and the results between the linear kernel and the $d = 3$ (3rd degree) kernel were nearly identical. Ultimately, the optimal selected KPCA configuration was a linear kernel with 2 KPCs and 3 clusters, though this configuration produced a cluster configuration that was identical to the control (Figure 2a). Thus, for these data KPCA offered no real benefit to the cluster analysis, though regular PCA did produce worse results.

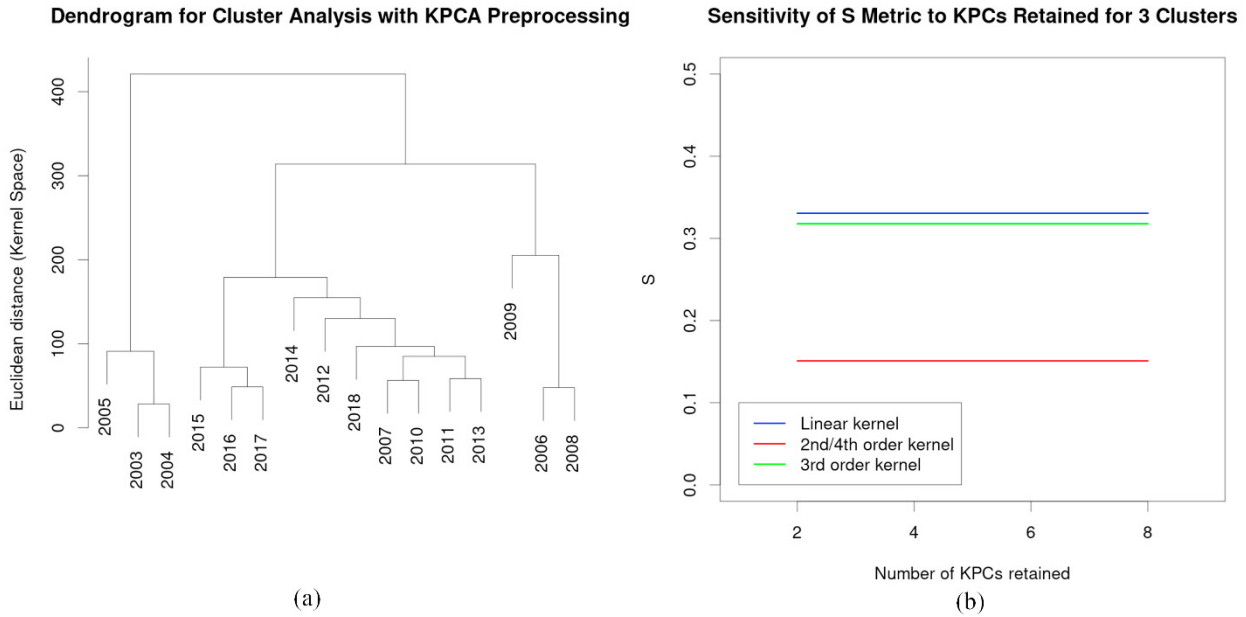

Fig. 2. (a) Dendrogram from the KPCA-preprocessed Ward's analysis with a linear polynomial kernel retaining 2 KPCs. (b) Sensitivity of the *S* metric to the selection of the number of KPCs for a 3-cluster KPCA preprocessed cluster analysis. Horizontal lines mean no changes to the clustering were observed when changing the number of retained KPCs. The linear kernel produced the highest overall *S* values and was selected.

5. Composite RMSD LH Flux Results

The mean LH flux RMSD field (Figure 3a) for the full study period (Figure 3a) showed that the smallest RMSD regions tended to be within the agricultural regions within the LMRAV, with the largest changes primarily located along boundaries between bodies of water and land, either with lakes or along the river itself. Additionally, large urban centers (e.g. Memphis, Tennessee, Little Rock, Arkansas, others) were revealed to have essentially no RMSD, which is logical given their general lack of LULC change over time and minimal LH flux with the urban canopy. The bias, estimated as the LH flux estimate from the default LULC configuration minus the updated LULC LH flux solution, was also computed (Figure 3b). A vast majority of the domain showed positive bias, meaning the updated LULC produced an overall decrease in LH flux estimates across the study region. This is supported by the global median difference in LH flux of 5.2 Wm^{-2}. The only notable exceptions were primarily along the eastern edge of the LMRAV in west-central Mississippi where estimates showed slight negative biases, suggesting elevated LH flux with the updated LULC information. These biases are important when assessing the cluster composite maps.

The cluster composite maps (Figure 4) showed dramatic differences in the patterns embedded within the 16-year RMSD LH flux fields. Cluster 1 (Figure 4a) primarily showed a near-average pattern across the entire domain, with slightly below-average conditions (less than 1 standard deviation) along the Arkansas-Mississippi border. This cluster had the highest member count (10 RMSD LH flux years), showing the prevalence of the mean pattern and the general lack of shifts from that mean for most cases. This result supports the notion that LULC updates resulted in predictable changes to the simulated LH flux fields. Cluster 2, which comprised the first three simulation years, showed dramatically elevated LH flux RMSD throughout the entire simulation domain. This result coupled with the negative positive bias suggested the first three simulation years had a dramatic reduction in LH flux with the updated LULC Noah-MP simulation. As all simulations utilized a five-year treadmill spin-up, it is highly unlikely this is a consequence of the model configuration but instead is indicative of a shift in land management practices as the study period progressed. Importantly, cluster 2 outlines the heavily agricultural areas of the LMRAV (purple regions in Fig. 2a) as most dramatically affected by the updated LULC, suggesting this cluster primarily focuses on heavy shifts in LH flux in the LMRAV itself. . The third cluster (Figure 4c) showed lower LH flux RMSD throughout the simulation

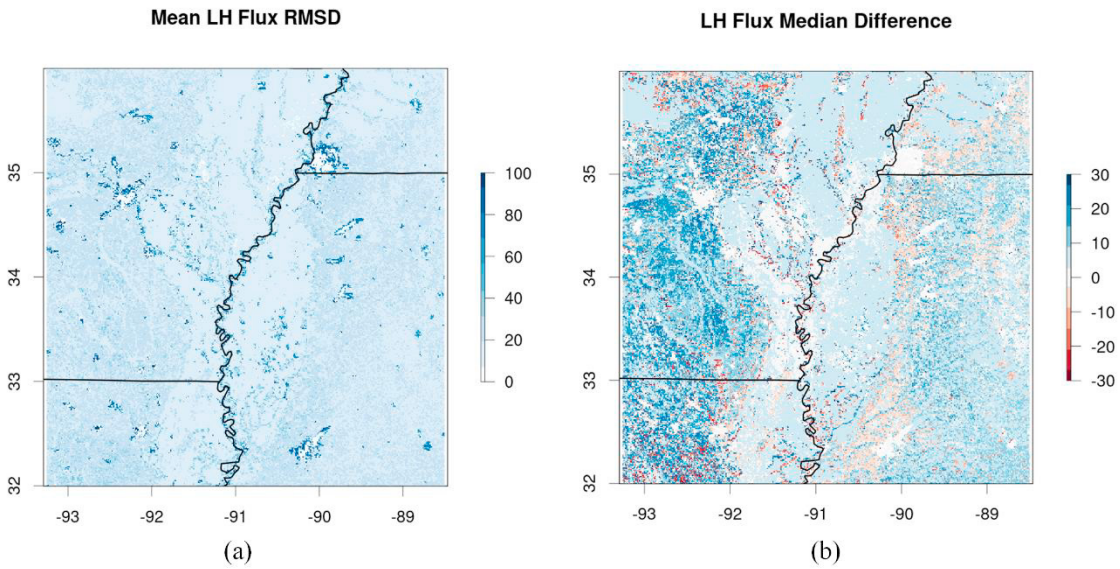

Fig. 3. (a) Mean LH flux RMSD (in Wm^{-2}) for all 16 warm seasons. (b) Median LH flux difference (Wm^{-2}) between the default LULC Noah-MP simulation and the updated LULC Noah-MP simulations for the full 16 warm seasons

region, except for the region just east of the LMRAV, meaning these simulation years were primarily associated with little to no LH flux modification by the updated LULC conditions. Notably, the results revealed a region of increased sensitivity slightly east of the cropland region in the LMRAV in northern Mississippi and southwestern Tennessee. This suggests that the small modifications to the LH flux in the LMRAV itself contributed to higher sensitivity in LH flux east of the region which supports the conclusion that updated LULC conditions within the Noah-MP can impact adjacent downstream regions. Clearly, a meteorological link exists between the LMRAV and points east, such that land management practices in the LMRAV agriculture could be affecting warm season precipitation downstream, which is an important result for agriculture in the study region.

In addition to isolating spatial structures within the data, the clusters revealed important LH flux sensitivities by LULC type. Though in cluster 1, no anomaly exceeded +/- 1 standard deviation, the results in clusters 2 and 3 revealed large areas with anomalously high or low sensitivity to the updated LULC information. In cluster 2, the highest percentage of points that experienced RMSD at least 1 standard deviation above the mean were cropland points that did not change type with the LULC update (~16.8%). This result is consistent with the previous conclusion regarding the dramatic sensitivity in the LMRAV itself with the cluster 2 member years. Other larger LULC changes that resulted in elevated RMSD for cluster 2 included shifts from class 4 (mixed dryland/irrigated cropland and pasture) to class 8 (shrubland), which comprised roughly 4.6% of the elevated RMSD gridpoints, and from class 4 to class 5 (cropland/grassland mosaic). These same regions showed a dramatic decrease in LH flux with the LULC updates, which were consistent with expectations as shrubland should produce greater sensible heat flux than cropland due to the higher vegetation density of agricultural areas. This result supports the notion that recent LULC modifications are having an important effect on LH flux in the LMRAV, which likely translates to modifications in the warm-season precipitation climatology as well. In cluster 3, most regions showed little to no modification as RMSD anomalies were highly negative. As expected, most LULC regions with highly negative anomalies (values < -1) showed consistent LULC types throughout the study period, with a large percentage (~17%) of points again being associated with cropland that was not modified over the study period. This result suggests that the cluster years portrayed here were consistently low with regards to LH flux. Interestingly, the member years in this cluster (2006, 2008, 2009) were anomalously low precipitation years in [2], and this lack of precipitation likely contributed to reduced LH flux as the LULC type had minimal impact (hence the lower RMSD anomaly values). Overall, some important sensitivities in LH flux, and thus warm-season precipitation, were observed, and future simulation years can better support the results seen in these simulations and solidify the relationships between LH flux and LULC.

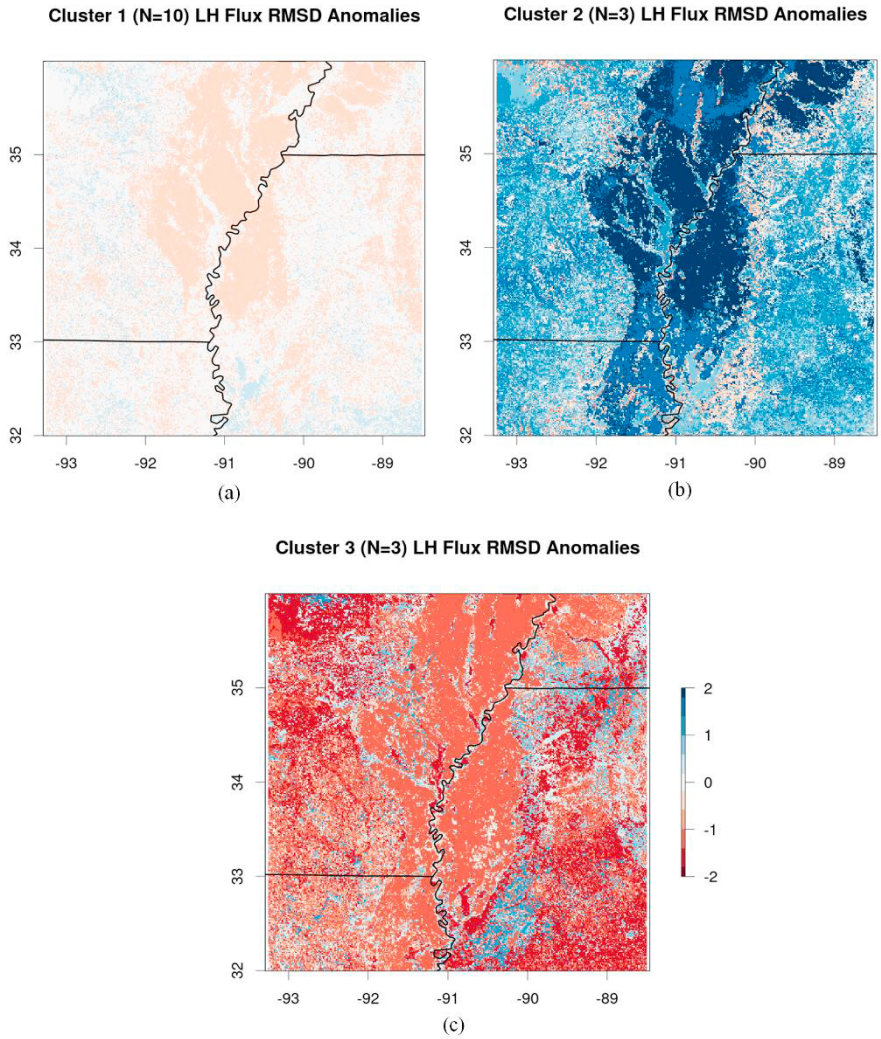

Fig. 4. (a) Composite mean LH flux RMSD for cluster 1 (panel a), cluster 2 (panel b), and cluster 3 (panel c). Shading are expressed in standard anomalies (numbers of standard deviations of RMSD above or below the mean).

6. Discussion and Conclusions

The results presented herein demonstrate the importance of updated LULC to LH flux estimates (used as a proxy for convection driving warm-season precipitation processes) from Noah-MP simulations. Most study years showed consistent impacts from updating the LULC fields (seen in the mean field, Figure 3a) as 10 of the 16 study years were included in the near-mean cluster 1 composite (Figure 4a). Other member years demonstrated large sensitivity to updated LULC, especially over the LMRAV region (Figure 4b), or showed essentially no sensitivity except east of the LMRAV region. The composite results demonstrated a clear inverse correlation between the LMRAV itself and regions just east of the LMRAV in both cluster 2 and 3's composites. This suggests an important linkage between LH flux over the LMRAV and points downstream and supports further investigation into the physical links between updated LULC in the LMRAV due to anthropogenic land use activities. This will require additional investigation and simulations to address these impacts more thoroughly.

There were some limitations to this study. The scope of the problem considered herein was massive but was still limited to utilizing a proxy variable (LH flux) for estimating warm-season precipitation. The KPCA preprocessing offered no difference in clustering, though it is expected that KPCA would offer important new insight with additional

study years. Future studies will incorporate additional study years, LULC modifications for the Noah-MP, and eventually formally coupling the atmospheric WRF model with the Noah-MP simulations to obtain true estimates of LH flux and precipitation with two-way surface-atmosphere feedbacks. Overall, this study demonstrated that sensitivities, while generally minor, did result from updated Noah-MP simulations and that these sensitivities were most demonstrative in proximity to the LMRAV and points east.

Acknowledgements

This work was supported by NOAA award #NA19OAR4590411. We wish to thank the reviewers for their careful review of our manuscript.

References

[1] Dyer, Jamie, and Andrew Mercer. (2013) "Assessment of spatial rainfall variability over the lower Mississippi River Alluvial Valley." *Journal of Hydrometeorology*, **14**: 1826-1843.

[2] Ouyang, Ying, Jiaen Zhang, Gary Feng, Yongshan Wan, and Theodor Leininger. (2020) "A century of precipitation trends in forecast lands of the Lower Mississippi River Alluvial Valley." *Scientific Reports,* **10**: 12802.

[3] Diem, Jeremy. (2006) "Synoptic-scale controls of summer precipitation in the Southeastern United States." *Journal of Climate*, **19**: 613-621.

[4] Dyer, Jamie. (2011) "Analysis of a warm-season surface-influenced mesoscale convective boundary in northwest Mississippi." *Journal of Hydrometeorology*, **12**: 1007-1023.

[5] Aligo, Eric, William Gallus, and Moti Segal. (2007) "Summer rainfall forecast spread in an ensemble initialized with different soil moisture analyses." *Weather and Forecasting,* **22**: 299-314.

[6] Ebert, Elizabeth. (2001) "Ability of a poor man's ensemble to predict the probability and distribution of precipitation." *Monthly Weather Review*, **129**: 2461-2480.

[7] Mercer, Andrew, Jamie Dyer, and Song Zhang. (2013) "Warm-season thermodynamically-driven rainfall prediction with support vector machines." *Procedia Computer Science*, **20**: 128-133.

[8] Cristianini, Nello, and John Shawe-Taylor. (2000) *An Introduction to Support Vector Machines and other Kernel-Based Learning Methods*. Cambridge, UK, 189.

[9] Skamarock, William, Joseph Klemp, Jimy Dudhia, David Gill, Zhiquan Liu, Judith Berner, Wei Wang, Jordan Powers, Michael Duda, Dale Barker, and Xiang-Yu Huang. (2019) A description of the Advanced Research WRF version 4. *NCAR Technical note NCAR/TN-556+STR*, 145.

[10] Santanello, Joseph, Patricia Lawston, Sujay Kumar, and Eli Dennis. (2019) "Understanding the impacts of soil moisture initial conditions on NWP in the context of land-atmosphere coupling." *Journal of Hydrometeorology,* **20**: 793-819.

[11] Gochis, David, Michael Barlage, Ryan Cabell, Matthew Casali, Aubrey Dugger, Katelyn FitzGerald, Molly McAllister, James McCreight, Arezoo Rafiee Nasab, Laura Read, Kevin Sampson, David Yates, and Yongxin Zhang. (2020) The WRF-Hydro modeling system technical description (version 5.1.1). *NCAR Technical Note,* 107.

[12] Niu, Guo-Yue, Zong-Liang Yang, Kenneth Mitchell, Fei Chen, Michael Ek, Michael Barlage, Anil Kumar, Kevin Manning, Dev Niyogi, Enrique Rosero, Mukul Tewari, and Youlong Xia. (2011) "The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements." *Journal of Geophysical Research*, **116**, D12109.

[13] Schölkopf, Bernhard, Alexander Smola, and Klaus-Robert Müller. (1998) "Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation,* **10**: 1299-1319.

[14] Mercer, Andrew, and Michael Richman. (2012) "Assessing atmospheric variability using kernel principal component analysis." *Procedia Computer Science,* **12:** 288-293.

[15] Yang, Zong-Liang, Guo-Yue Niu, Kenneth Mitchell, Fei Chen, Michael Ek, Michael Barlage, Laurent Longuevergne, Kevin Manning, Dev Niyogi, Mukul Tewari, and Youlong Xia. (2011) "The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins." *Journal of Geophysical Research*, **16**: D12110.

[16] Cai, Xitian, Zong-Liang Yang, Cédric David, Guo-Yue Niu, and Matthew Rodell. (2014) "Hydrologic evaluation of the Noah-MP land surface model for the Mississippi River Basin." *Journal of Geophysical Research Atmosphere*, **119**: 23-38.

[17] Broxton, Patrick, Xubin Zeng, Damien Sulla-Menashe, and Peter Troch. (2014) "A global land cover climatology using MODIS data." *Journal of Applied Meteorology and Climatology*, **53**, 1593-1605.

[18] Bonan, Gordon. (2008) *Ecological Climatology: Concepts and Applications, 2nd Ed.*, Cambridge, U.K., 550.

[19] Wilks, D. (2011) *Statistical Methods in the Atmospheric Sciences*. Academic Press, Burlington, MA, 704.

[20] Richman, M., and I. Adrianto. (2010) "Classification and regionalization through kernel principal component analysis." *Physics and Chemistry of the Earth*, **35**: 316-328.

[21] Rosseeuw, P. (1987) "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis." *Journal of Computational and Applied Mathematics*, **20**: 53-65.