Estimating Crop Acreage over Regional Scale using Remote Sensing and Climate Data

Venkata Shashank Konduri[1,2], Jitendra Kumar[2], Forrest M. Hoffman[2], William W. Hargrove[3], Auroop R. Ganguly[1]

[1]Northeastern University, Boston, MA; [2]Oak Ridge National Laboratory, Oak Ridge, TN; [3]USDA Forest Service, Southern Research Station, Asheville, NC

OBJECTIVE

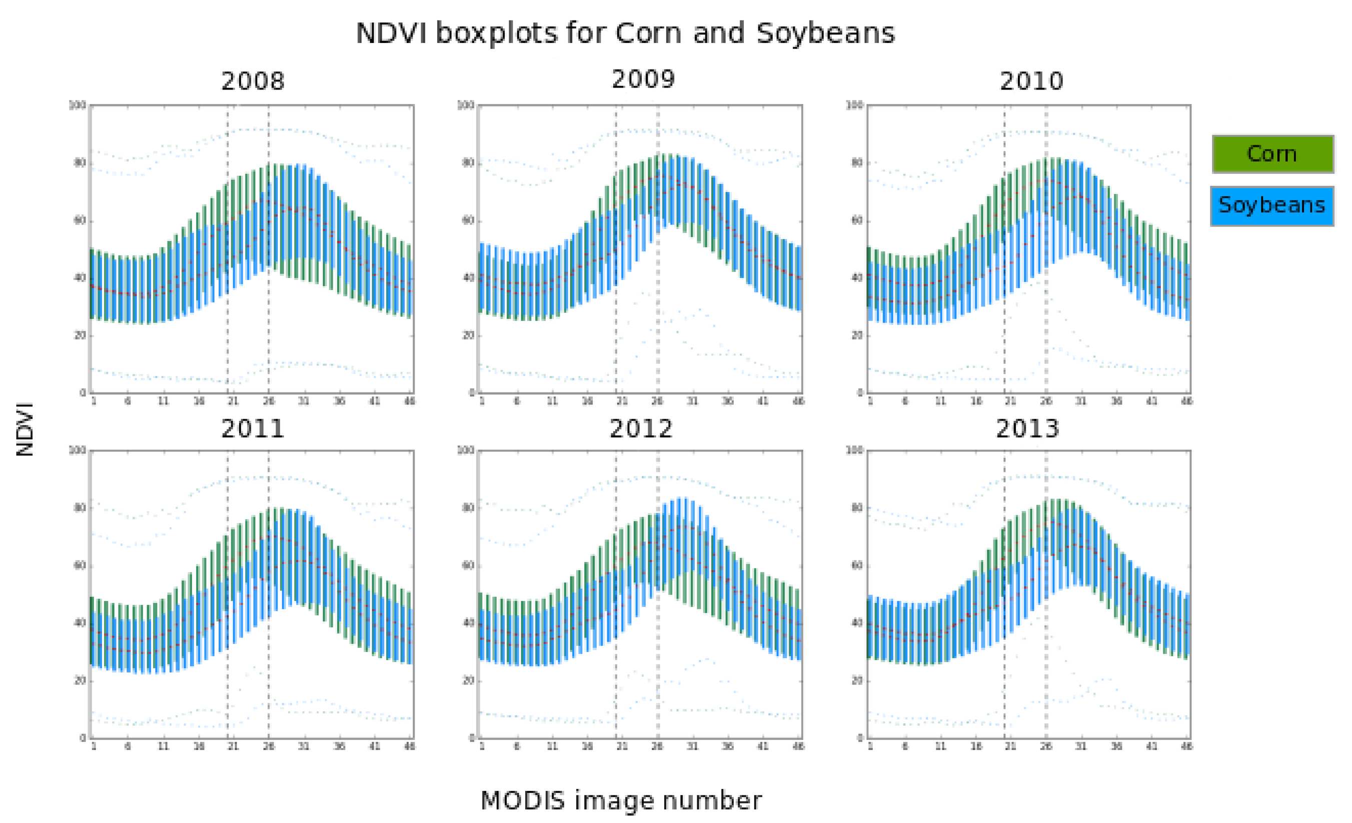
Estimate acreage under corn and soybeans using a neural network based regression approach for the major crop-producing states (Iowa, Illinois and Indiana) in the US midwest.

MOTIVATION

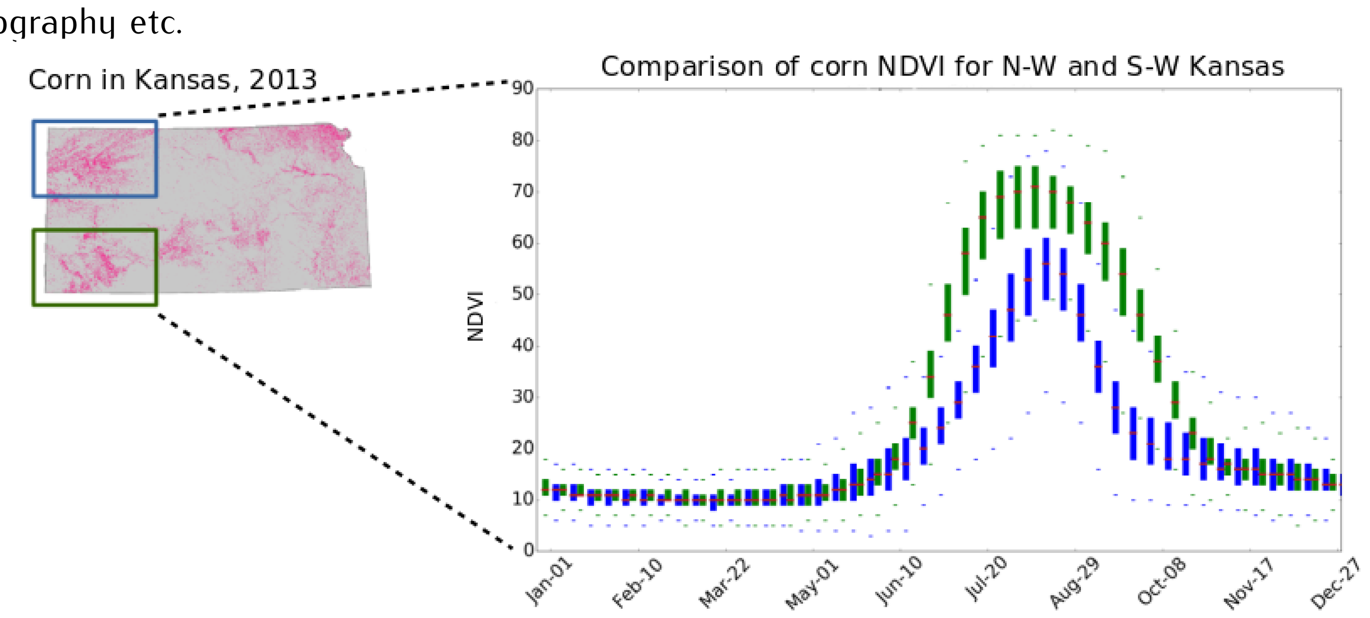
- Timely and accurate knowledge about the geospatial distribution of crops at regional to national scales is crucial for estimating crop water use, crop yield and for studying the impact of agriculture on the environment.
- The United States (US), a major producer of corn and soybeans in the world, currently lacks a spatially explicit crop data product available publicly during the growing season.
- An early-season crop-acreage product could potentially enable estimation of yield by the middle of the current growing season.

BACKGROUND

- Corn and soybeans have very similar NDVI (Normalized Difference Vegetation Index) signatures. Also, crop phenology varies every year mainly due to interannual variability in climate.



- Phenology for the same crop can vary across space due to differences in climate, soil properties, topography etc.

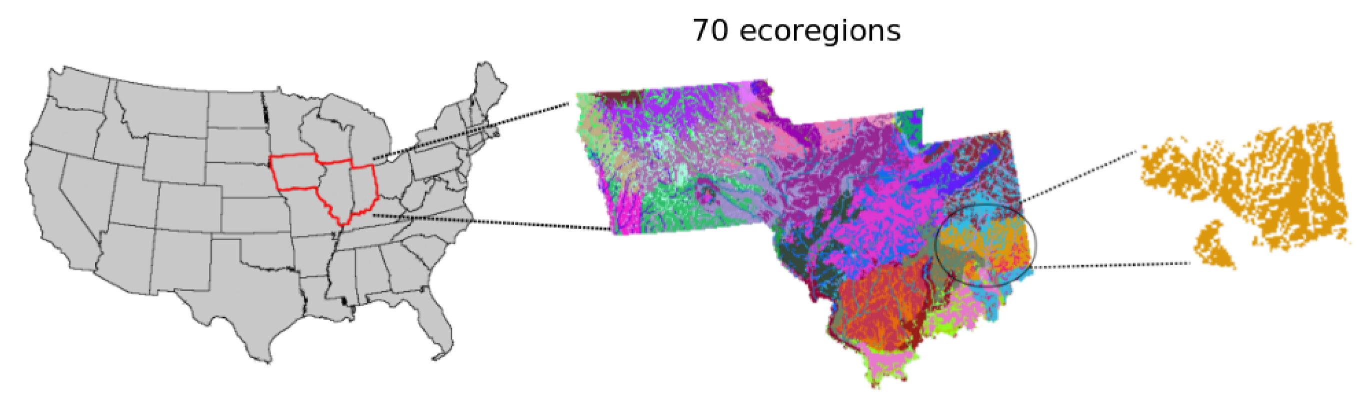


DATA

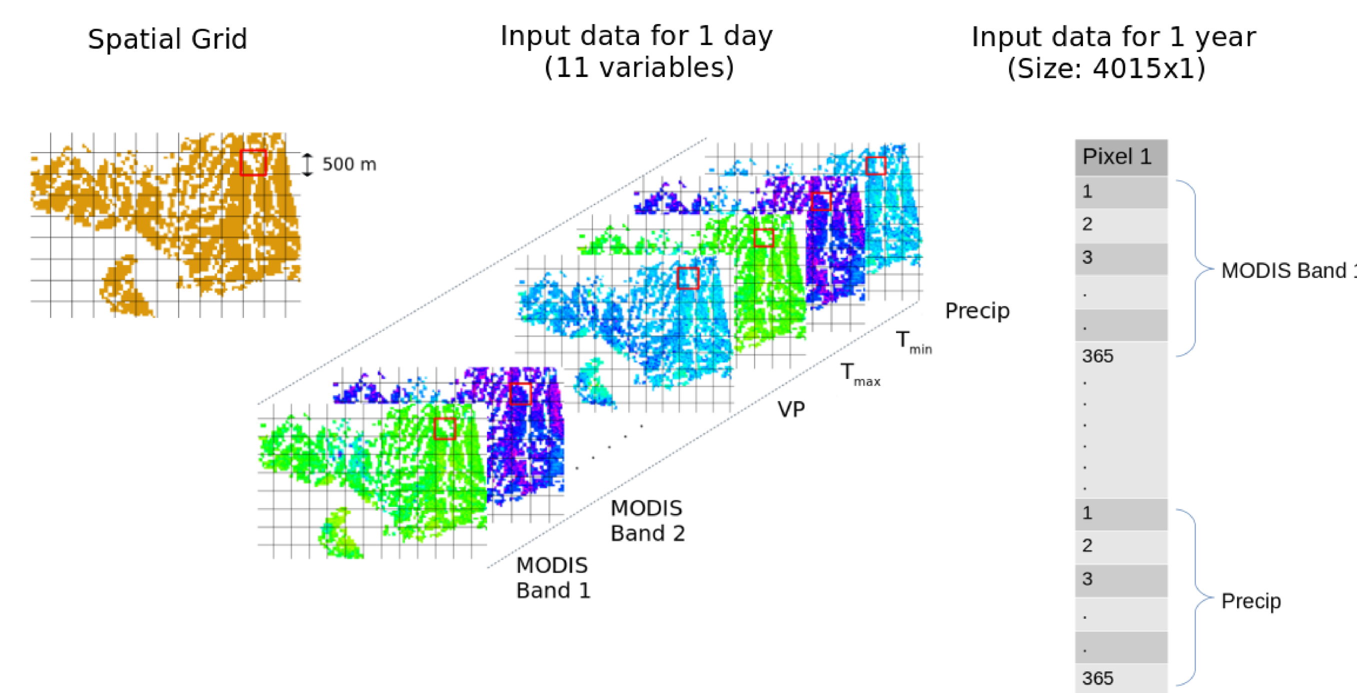
- **Remotely Sensed Surface Reflectance**: Data for bands 1 through 7 of MODIS MCD43A4 Version 6 NBAR data set available daily at 500 m resolution was collected for the years 2010–2011.
- **Climate**: Data for T_{max}, T_{min}, VP and $Precip$ available daily at 1 km $\times$ 1 km resolution was collected from the Daymet dataset for the years 2010–2011. The 1 km data was regridded to 500m to match the MODIS pixel resolution. Including climate data as input could help improve the model's robustness towards inter-annual variability in climate.
- **Ground truth**: Percent crop cover was derived from the USDA Cropland Data Layer [1] for the years 2010 and 2011.

METHODS

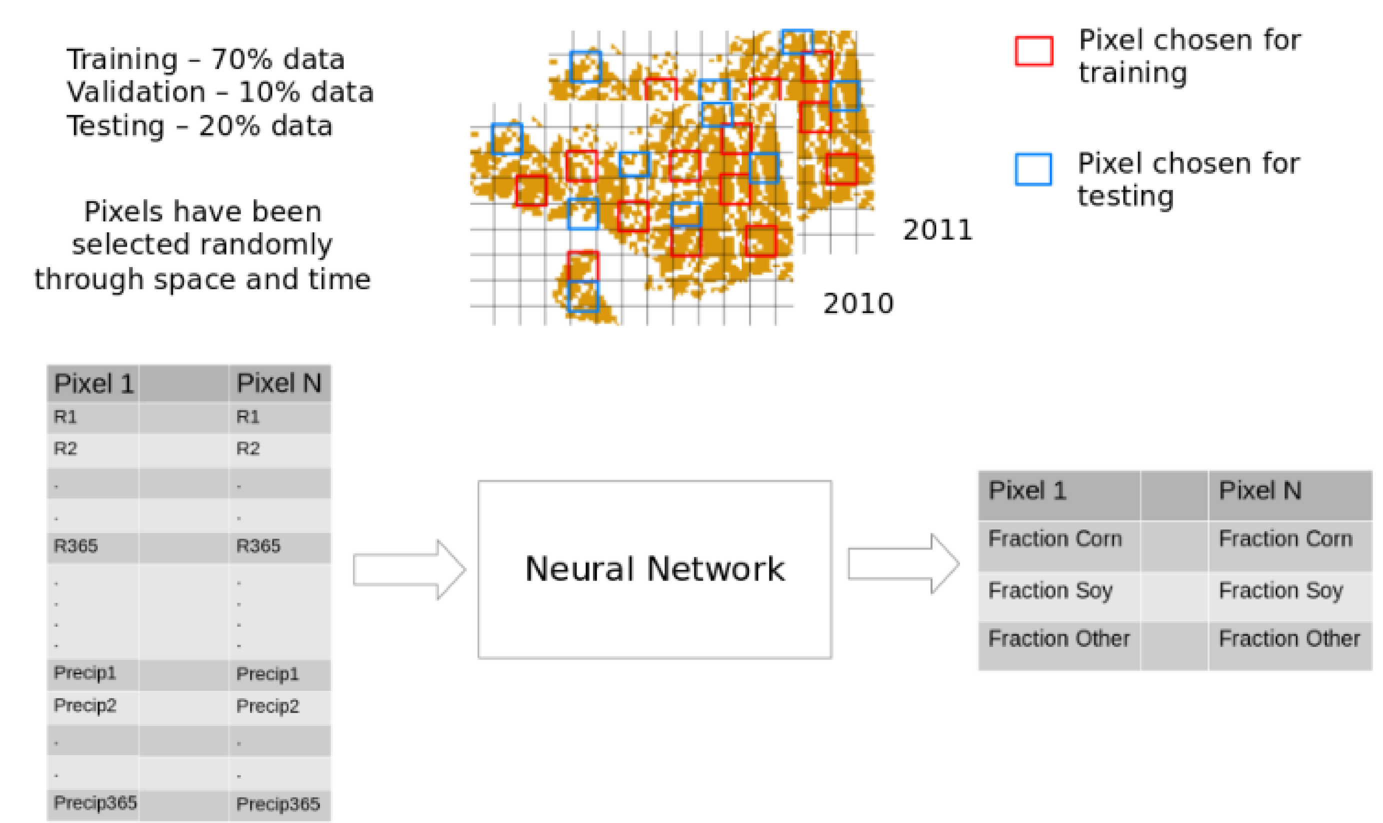
1. Spatial variability in phenology could be addressed by constraining the spatial extent of modeling to that of *ecoregions* - ecologically similar regions. These regions have been created by clustering bioclimatic, topographic and edaphic variables [2]. The study area was divided into 70 ecoregions.



2. Within every ecoregion, data for the 11 variables were extracted for every 500m pixel for each day of the year and saved as a column vector with 4015 (365 $\times$ 11) entries. The column vectors for different pixels were concatenated to form a dataframe.



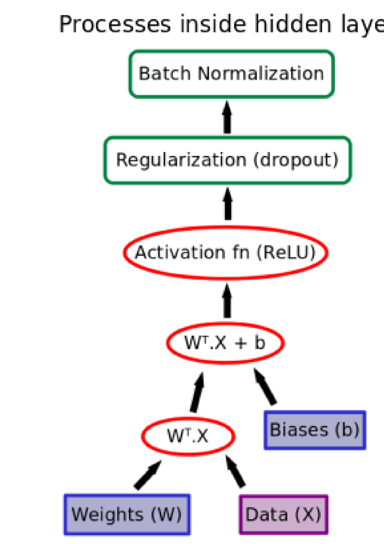
3. Fraction of 500 m pixel under corn and soybeans was estimated using a neural network based regression approach. The dataset was split into training, validation and testing by sampling pixels randomly through space and time.

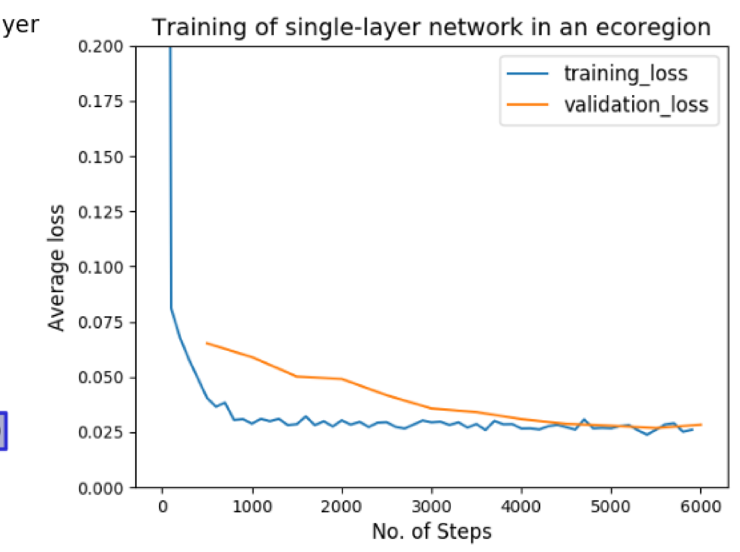


4. The best set of hyperparameters for the neural network were determined by performing a random search across a pre-defined search space. The set of hyperparameters selected for this study are as follows:

Number of layers = 1	Learn rate decay = 0.766	Optimizer = "Adam"
Number of units = 40	Dropout rate = 0.081	$\beta1$ = 0.9
Learn rate = 0.001	Batch size = 500	$\beta2$ = 0.999

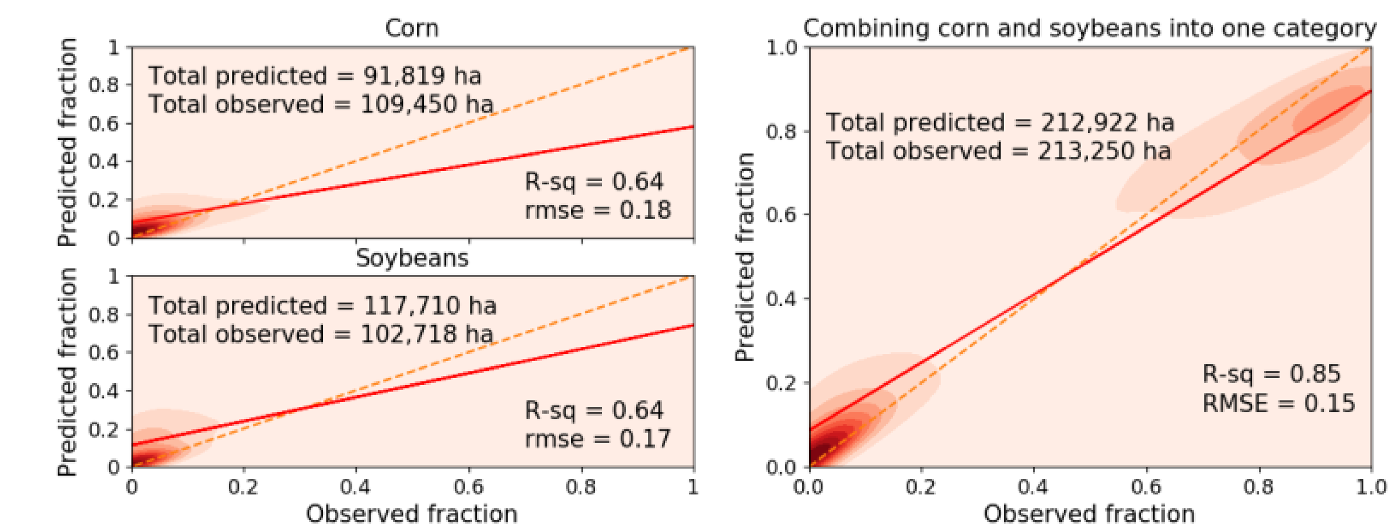
5. The neural network with the same hyperparameters was trained separately in all the ecoregions for 6000 steps. The modeling was performed in the *Tensorflow* environment on two 32 GB V100 GPUs of the NVIDIA-DGX station.



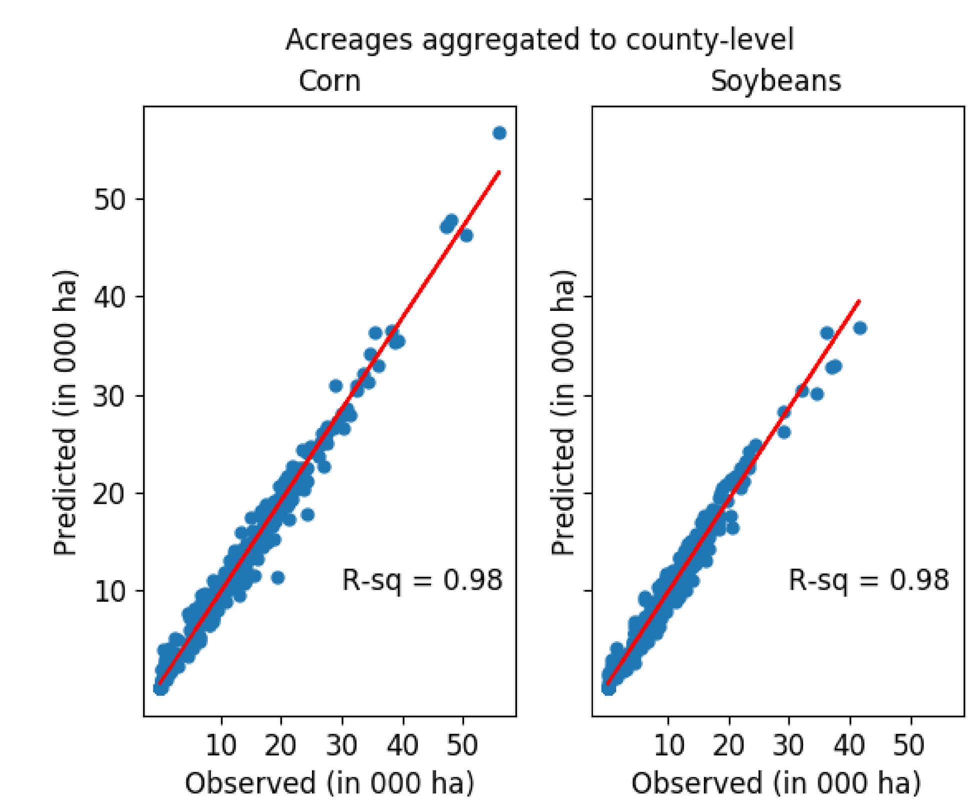


RESULTS

- Error metrics for the test data



- Aggregating corn and soy acreages to the county-scale



CONCLUSIONS

- Similarity in corn and soybeans phenology is the major reason behind poor error metrics at the 500m pixel level. Deeper network architectures could help in improving the error metrics even further.
- Corn and Soybean acreages agree very well with the actual acreages when aggregated to the county scale.

ACKNOWLEDGEMENTS

This research was sponsored by the U.S. Department of Agriculture Forest Service, Eastern Forest Environmental Threat Assessment Center (EFETAC) and the U.S. Department of Energy Biological and Environmental Research (BER) program. This research used resources of the Oak Ridge Leadership Computing Facility at Oak Ridge National Laboratory, which is managed by UT-Battelle, LLC, for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

REFERENCES

[1] Claire Boryan, Zhengwei Yang, Rick Mueller, and Mike Craig. Monitoring US agriculture: The US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. *Geocarto International*, 26(5):341–358, 2011.

[2] William W Hargrove and Forrest M Hoffman. Potential of multivariate quantitative methods for delineation and visualization of ecoregions. *Environmental management*, 34(1):S39–S60, 2004.