Identification of Flash Drought Events using Machine Learning Techniques

Stuart G. Edris[1,5], Jeffrey B. Basara[1,2], Jordan I. Christian[1], Amy McGovern[1,3], Jason C. Furtado[1], Xiangming Xiao[4]

[1]School of Meteorology, University of Oklahoma, Norman, Oklahoma
[2]School of Civil Engineering and Environmental Science, University of Oklahoma, Norman, Oklahoma
[3]School of Computer Science, University of Oklahoma, Norman, Oklahoma
[4]Department of Microbiology and Plant Biology, University of Oklahoma, Norman, Oklahoma

Introduction

Flash droughts (FDs) are droughts that develop over rapid time scales ($\sim$ 1 month) and have high impact on vegetation and agriculture. This study seeks to investigate the capability of supervised learning (SL) techniques to identify FD. However, since there are multiple methods for identifying FD, each method will be treated as a set of label data for the SL methods to identify. The particular SL techniques investigated in this study are methods that have been previously found to be effective for identifying drought, namely random forests (RFs) and artificial neural networks (ANNs). Finally, this study uses indices that have been recently developed to represent rapidly changing conditions associated with FD.

Data and Methods

Data was collected from NARR 32km $\times$ 32km data, with temporal data averaged to the pentad timescale. Data was spatially split into 8 different regions.

Training Data:

- SESR (ET/PET)
- SEDI ($ET - PET$)
- EDDI (PET)
- SPEI ($P - PET$)
- SAPEI ($P - PET + memory$)
- SMI (0 to 40cm SM)
- SODI ($[P + L + R_0] - [PET + SMD]$)
- FDII (0 to 40cm SM)

Each variable (except FDII) is also accompanied by the 30-day mean change in that variable.

Label Data: FD was identified using the following methods:

- Christian et al. 2019 (SESR)
- Nogeura et al. 2020 (SPEI)
- Liu et al. 2020 (SM)
- Pendergrass et al. 2020 (EDDI)
- Otkin et al. 2021 (FDII)

SL Algorithms: The following SL have been implemented:

- Random Forests
  - Some preliminary runs show 100 tree forests produce sufficient results for all methods
- Traditional Artificial Neural Networks
  - Some preliminary runs show 1 layer model with 25 nodes works best for the Christian and Nogeura methods, and a 2 layer model with 15 nodes each work best for the rest

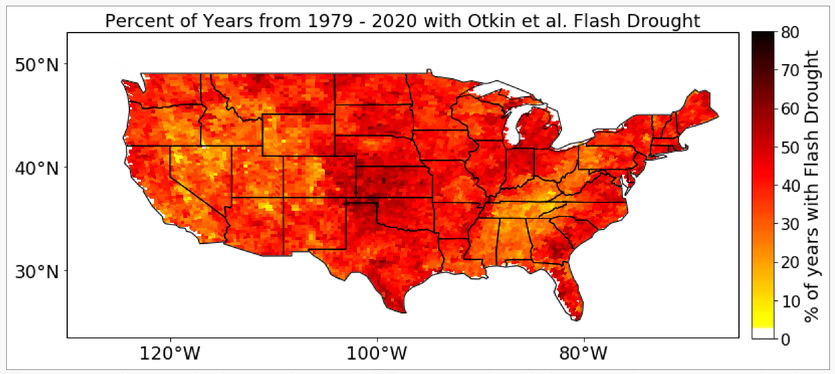

Figure 1. Original climatology of FD using the Otkin et al. 2021 method (i.e., climatology of FDII).

Results

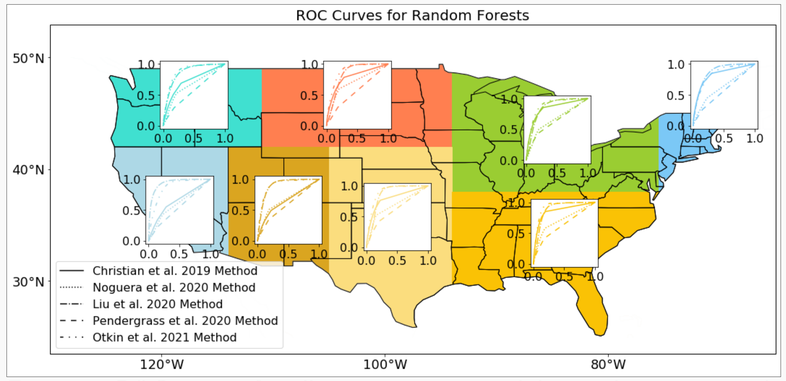

Figure 2. ROC curves for all regions using a validation dataset on each RF model (one for each FD identification method per region).

	SESR	ΔSESR	SEDI	ΔSEDI	SPEI	ΔSPEI	SAPEI	ΔSAPEI	EDDI	ΔEDDI	SMI	ΔSMI	SODI	ΔSODI	FDII
Christian et al.	–	–	11.1	10.5	7.2	7.4	8.5	8.7	4.4	7.0	9.3	9.7	7.3	7.0	2.0
Noguera et al.	5.5	5.1	11.3	8.8	–	–	8.5	5.9	23.2	12.6	4.9	4.6	4.2	4.1	0.6
Liu et al.	8.8	6.9	6.9	6.2	6.3	5.2	7.5	6.3	4.0	5.6	10.5	7.1	6.8	6.5	4.4
Pendergrass et al.	8.6	8.5	8.4	8.5	8.3	8.7	8.9	9.1	–	–	3.4	9.2	8.5	8.5	1.1
Otkin et al.	11.2	6.6	8.4	6.2	7.2	5.9	7.8	6.8	3.7	5.8	10.2	7.1	7.2	6.6	–

Table 1. The average (over all regions) contribution (percentage) of each index to the RF models for each method.

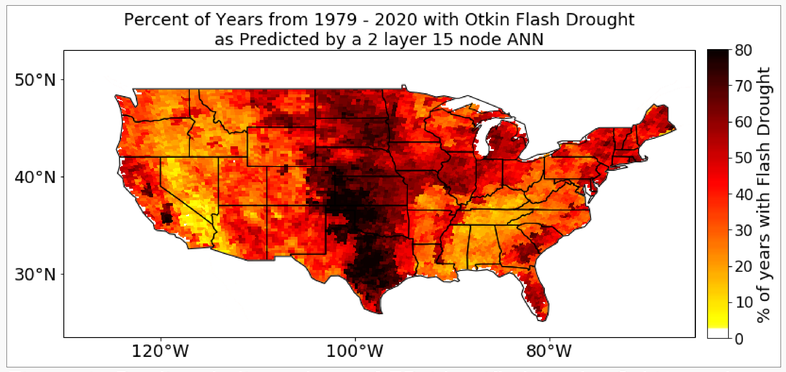

Figure 4. Predicted climatology of FD identified by the Otkin et al. 2021 method by 2 layer, 15 node each, ANNs.

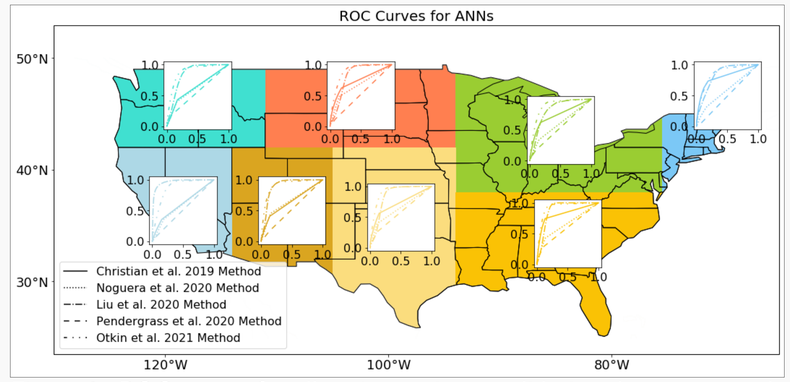

Figure 3. ROC curves for all regions using a validation dataset on each ANN model (one for each FD identification method per region).

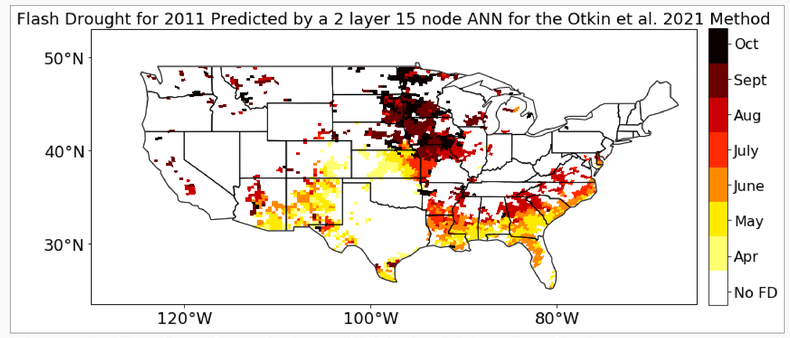

Figure 5. Predicted evolution of FD for 2011. Prediction was performed by 2 layer, 15 node ANNs trained on the whole NARR dataset.

Future Work

- Add more SL models to the analysis (e.g., support vector machines)
- Perform analysis with additional datasets (e.g., NLDAS, ERA5) for robustness and to create a global analysis
- Try identifying drought and rapid intensification seperately (FD is the intesection of the two)

Summary and Conclusions

- Both SL techniques had a hard time identifing FD with methods using just PET and P, while they were more able of identifying FD based on SM (methods using ET varied from region to region)
- The SL techniques were able to recreate the climatology of FD and capture the spatial hotspots, but had a harder time with smaller time scales (e.g., year-to-year events)

Acknowledgements

This work was funded by the NASA Water Resources Program grant 80NSSC19K1266, the National Science Foundation grants OIA-1920946 and OIA-1946093, and the USDA Southern Great Plains Climate Hub.

[5]Contact: sgedris@ou.edu