

*Water Resources Research*

Supporting Information for

**Towards a robust, impact-based, predictive drought metric**

Sanaa Hobeichi<sup>1,2</sup>, Gab Abramowitz<sup>1,2</sup>, Jason P. Evans<sup>1,2</sup>, and Anna Ukkola<sup>1,2</sup>

<sup>1</sup>Climate Change Research Centre, UNSW Sydney, NSW 2052, Australia.

<sup>2</sup>ARC Centre of Excellence for Climate Extremes, UNSW Sydney, NSW 2052, Australia.

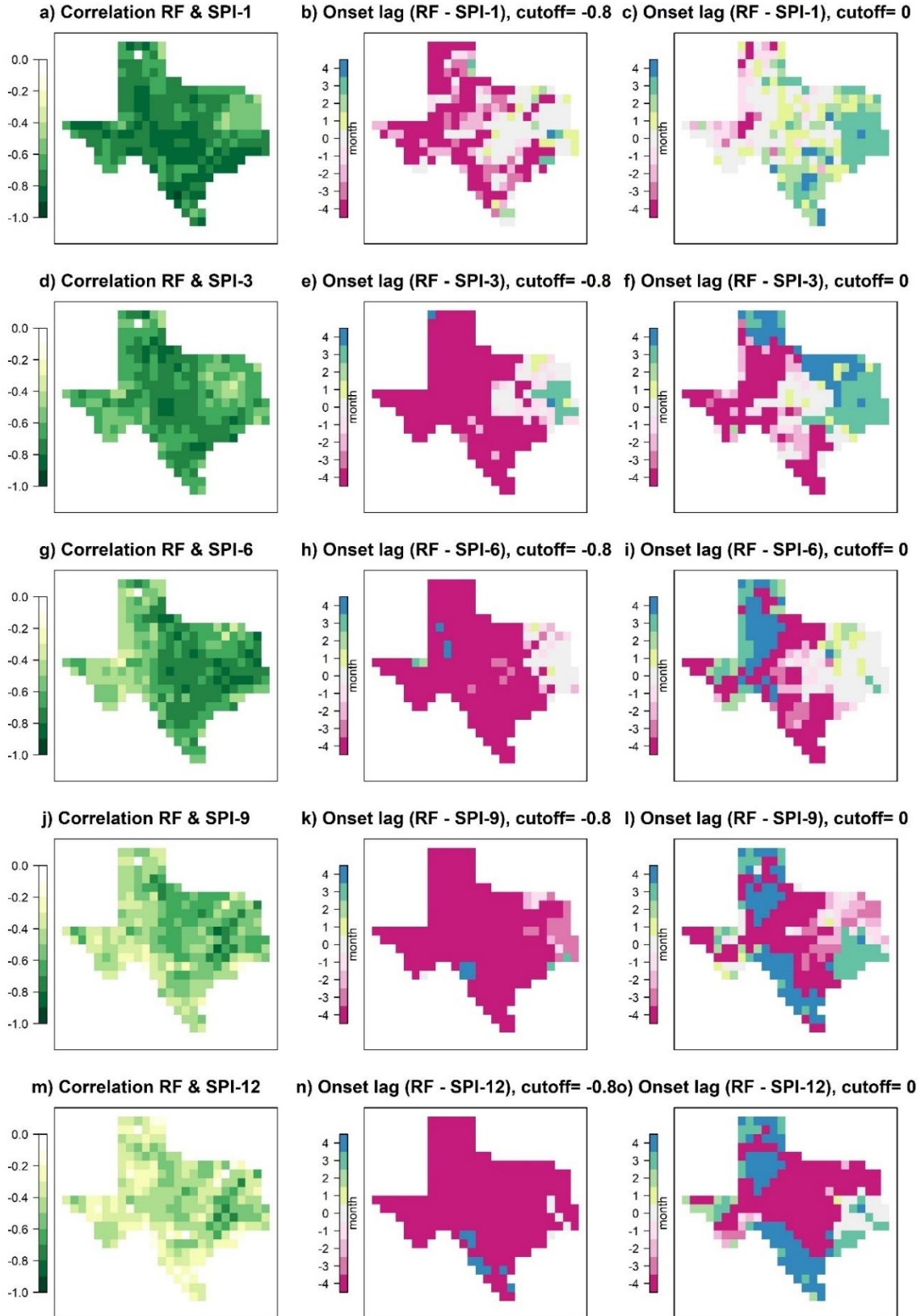
**Contents of this file**

Figure S1

Tables S2

**Additional Supporting Information (Files uploaded separately)**

Captions for Tables S1



**Figure S1.** Correlation between RF drought probabilities and SPI computed for 1, 3, 6, 9 and 12 month accumulation periods termed SPI-1 (a), SPI-3 (d), SPI-6 (g), SPI-9 (j) and SPI-12 (m) respectively; difference in drought onsets between RF and SPI, i.e.  $\text{OnsetRF} - \text{onsetSPI} - 0.8$ , where SPI has drought cutoff of -0.8, computed for each of the accumulation periods in b) e) h) k) and n); difference in drought onsets between RF and SPI, i.e.  $\text{onsetRF} - \text{onsetSPI}0$ , where SPI has drought cutoff of 0, computed for each of the accumulation periods in c) f) i) l) and o). Correlations and onsets are computed for the period spanning January 2010 – April 2012.

Large table uploaded in a separate document

**Table S1.** Database of drought and no drought events at 30 counties. The two labels '1' and '0' are used to indicate drought and no drought respectively. The database is spatiotemporally incomplete, and only months with data are included in the table. Drought and no drought information is extracted from the Drought Impacts Reporter, a national interactive drought impact database developed and maintained by the U.S. National Drought Mitigation Center (Wilhite et al. 2007) and the Texas Climate Monthly Reports produced by the Office of the State Climatologist at Texas A&M University that can be accessed at <https://climatexas.tamu.edu/products/texas-climate-bulletins/index.html>.

Machine learning classification algorithm.	Abbreviation	Reference
Random Forest	RF	(Breiman 2001)
Bagged Flexible Discriminant Analysis	BagFDA	(Friedman 1991)
Decision Tree	DT	(Swain and Hauska 1977)
Generalised Linear Models	GLM	(Nelder and Wedderburn 1972)
Lasso and Elastic-Net Regularised Generalised Linear Models	GLMnet	(Zou and Hastie 2005)
K-nearest Neighbors Algorithm	KNN	(Mitchell 1997)
Linear Discriminant Analysis	LDA	(Balakrishnama and Ganapathiraju 1998)
Support Vector Machine radial basis kernel	SVMRadial	(Scholkopf et al. 1997)
Support Vector Machine polynomial basis kernel	SVMPoly	(Scholkopf et al. 1997)

**Table S2.** List of the machine learning classification algorithms used in this study, and the abbreviation used in Figure 2. R software was used with Ranger package for Random Forest implementation, and the Caret package (Kuhn 2008) for implementation of the other machine learning algorithms. We refer the reader to the associated publications for details on each algorithm.