Towards an Early Drought Warning System

Advisor/Project Partner: Dr. Trevor Keenan   Discovery Group: Jason Zou
Data Science Discovery: December 2019

Long-Term Goal

Flash droughts are characterized by sudden and/or short periods of low rainfall and high atmospheric evaporative demand, which can be reasonably forecasted using knowledge of seasonal/sub-seasonal vegetation cycles.

Through predicting environmental attributes such as photosynthesis, evapotranspiration, and respiration, we seek to develop a **machine-learning based early warning system for flash droughts** at a local and global scale.

Current Goal

Environmental variables such as precipitation can influence environmental products over a longer time-scale.

Long Short-Term Memory networks (LSTMs) may be more effective than simple ANNs at modeling these influences.

Model

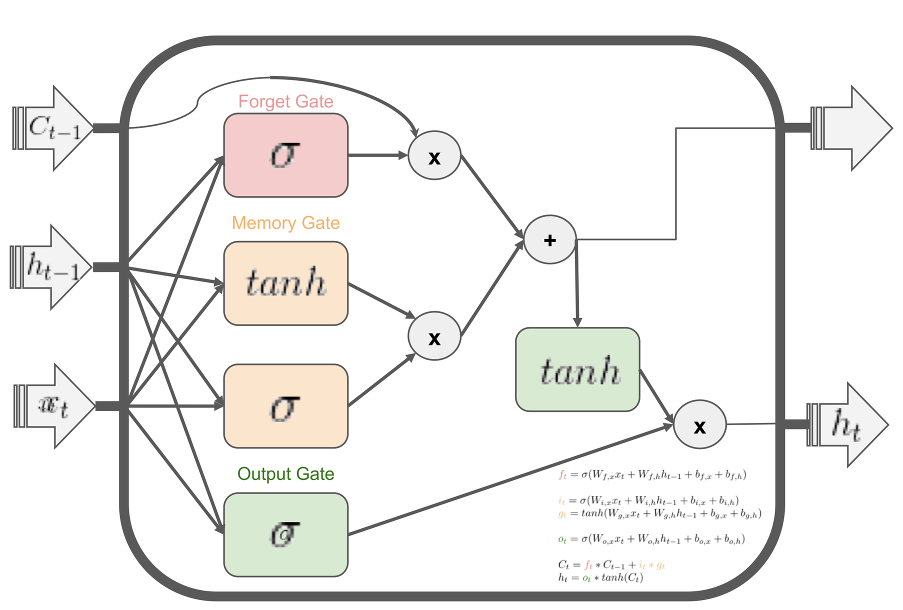


Figure 1. Simplified LSTM cell. Arithmetic details shown within.

Results

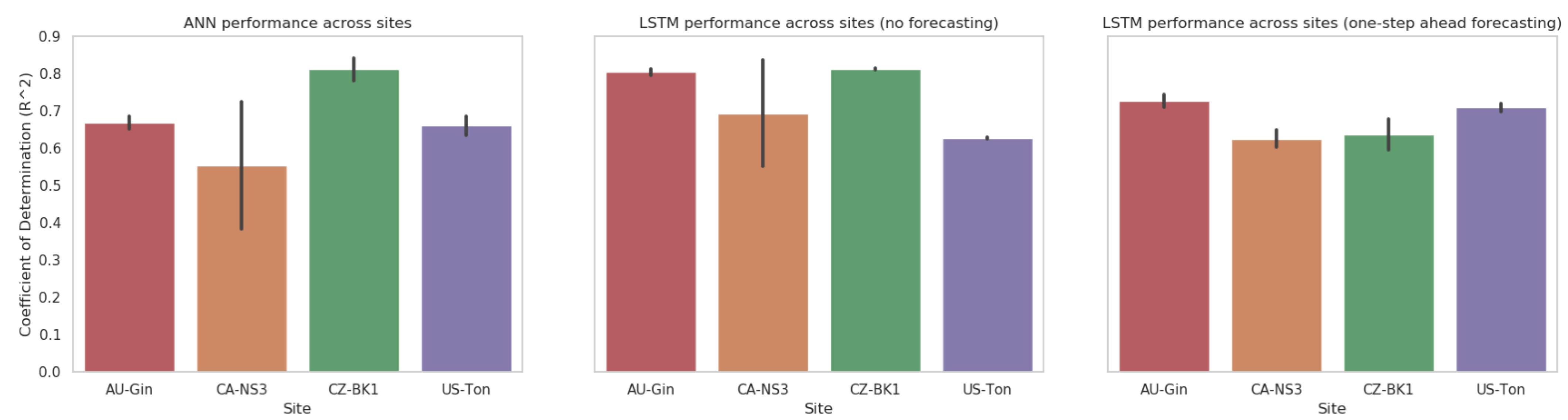


Figure 2. Average r^2 values per site for each model. Error bars represent the standard deviation of the value across 100 runs. (Left) ANN baseline; (Middle) LSTM without forecasting; (Right) LSTM with one-step ahead forecasting

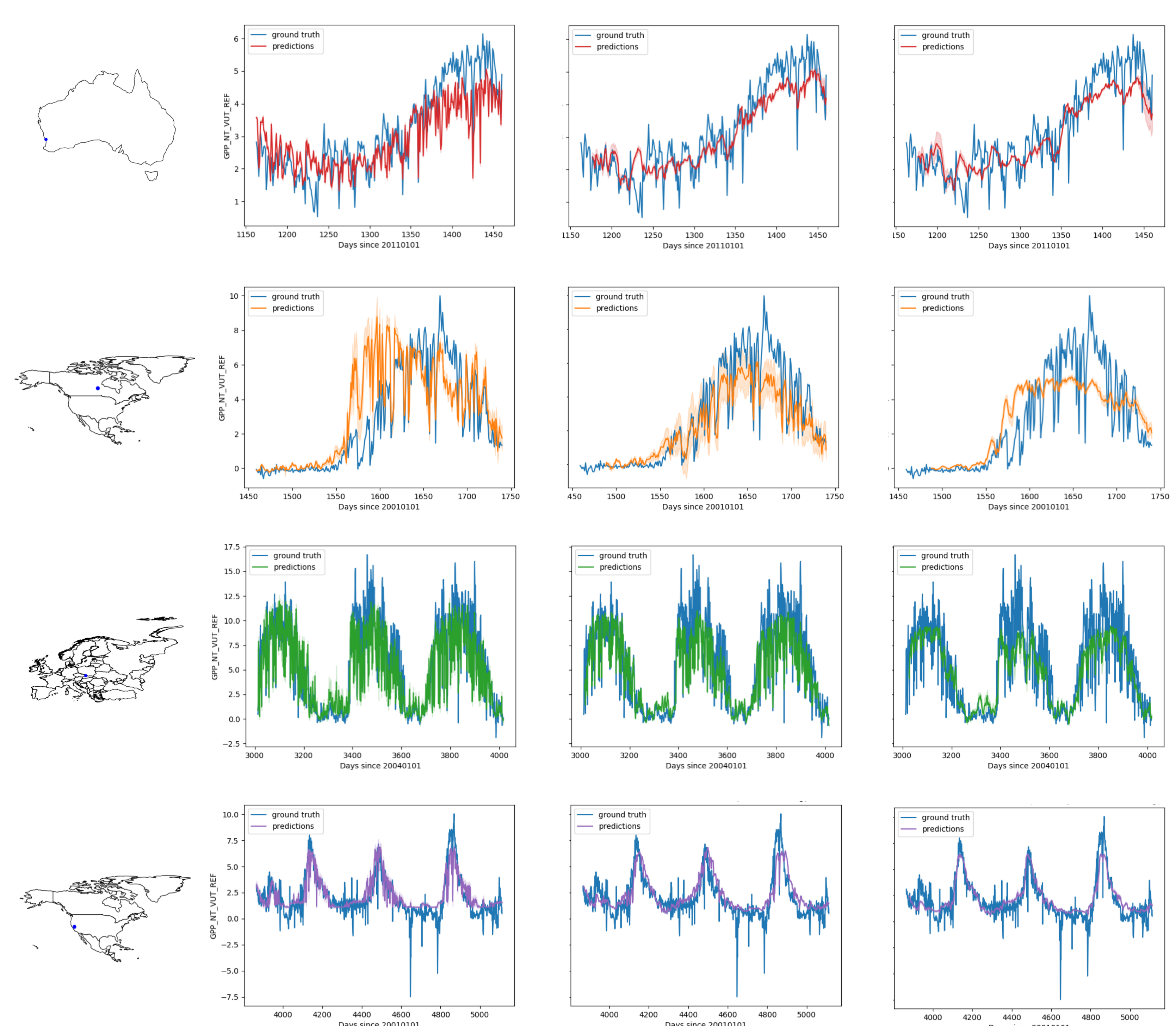
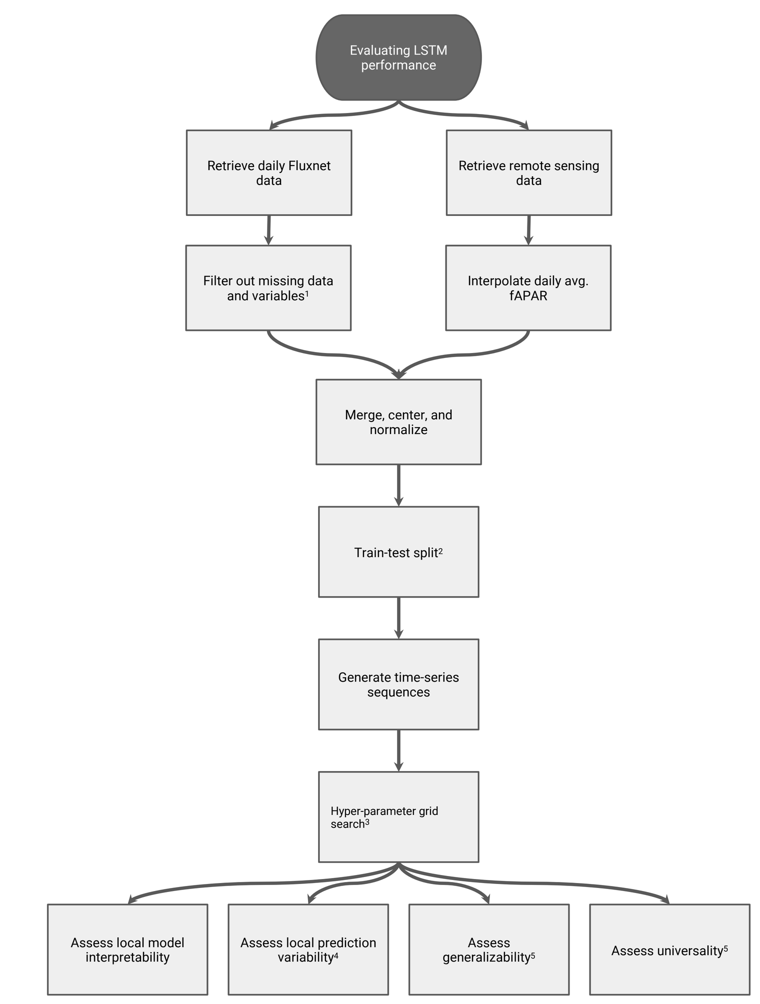


Figure 3. Test set predictions for each model-site pair. (Left-Right) ANN, LSTM without forecasting, LSTM with one-step ahead forecasting; (Top-Bottom) AU-Gin, CA-NS3, CZ-BK1, US-Ton. Note: US-Ton had no SWC data available during training.

Methods



[1] Target variables: TA, SW_IN, P, WS, VPD, CO2, SWC. Prioritized normal gap-filled data and used MDS gap-filled data as back-up
[2] 75-25 split; LSTM used non-shuffled splits due to to preservation of chronology
[3] ANN: 5-fold CV; LSTM: 2-fold CV. Hidden dimension was kept the same as ANN
[4] Trained best model 100 times and calculated the standard deviation for each predicted point
[5] Tested across sites that share the same vegetational class

Figure 4. Flowchart for evaluating model performance across sites.

Discussion

LSTMs generally performance better than ANNs and generate less noisy predictions. Difficulty in capturing abnormalities remains.

Forecasting remains a difficult task, although still performs better than ANN in some cases.

Optimal sequence length/"working memory" often varied for each site, likely due to dataset size..

Lack of overfitting suggests need for more extensive model optimization in future work.

Interpretability of variable-prediction relationships and model generalizability beyond local site remain unclear.

References:

Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., Reichstein, M., Arain, M. A., Cescatti, A., Kiely, G., Merbold, L., Serrano-Ortiz, P., Sickert, S., Wolf, S., and Papale, D.: Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms, Biogeosciences, 13, 4291–4313, https://doi.org/10.5194/bg-13-4291-2016, 2016.