Towards data-driven approaches for simulating rainfall in climate models

sarava@tamu.edu

J. Wang[1], J. Yang[1], M. Jun[1], R. Wong[1], C. Schumacher[2], and **R. Saravanan**[2]

[1]Department of Statistics and [2]Department of Atmospheric Sciences, Texas A&M University, College Station, TX

Poster A41R-2901

Motivation

- Climate models exhibit significant biases in their rainfall simulation (Fig. 1)
- Moist convection parameterizations that are physically motivated have failed to fully address this problem
- Why not try a purely data-driven (statistical or machine-learning) approach?

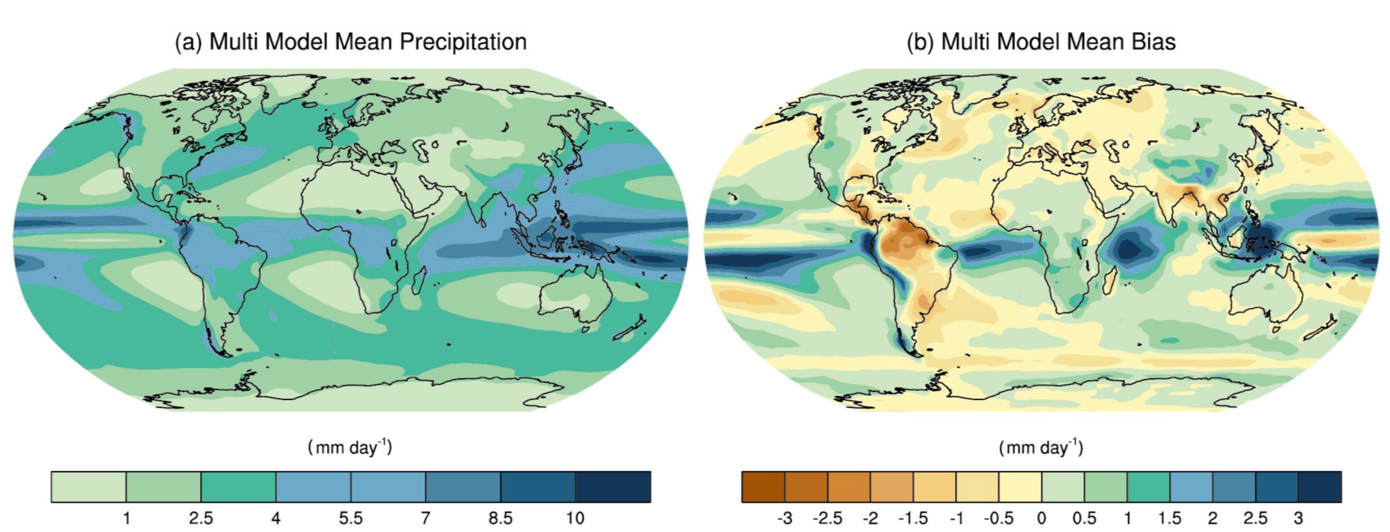
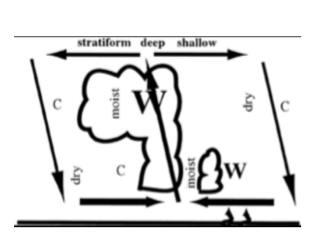


Figure 1. IPCC AR5 estimate of multi model mean precipitation and bias

Approach

Use available data to construct to a model predict rainfall from atmospheric state

- Atmospheric state data
  - MERRA 2 reanalysis
  - Interpolated to 0.5º, 6-hourly grid
- TRMM/GPM satellite radar rainfall data on same grid, divided into 3 rain types
  - Stratiform (STR)
  - Deep convective (DC)
  - Shallow convective (SC)



- Initial focus on East Pacific (EP) and West Pacific (WP) regions (Fig. 2)
  - Use one year of data to train model
  - Another year of data to validate model

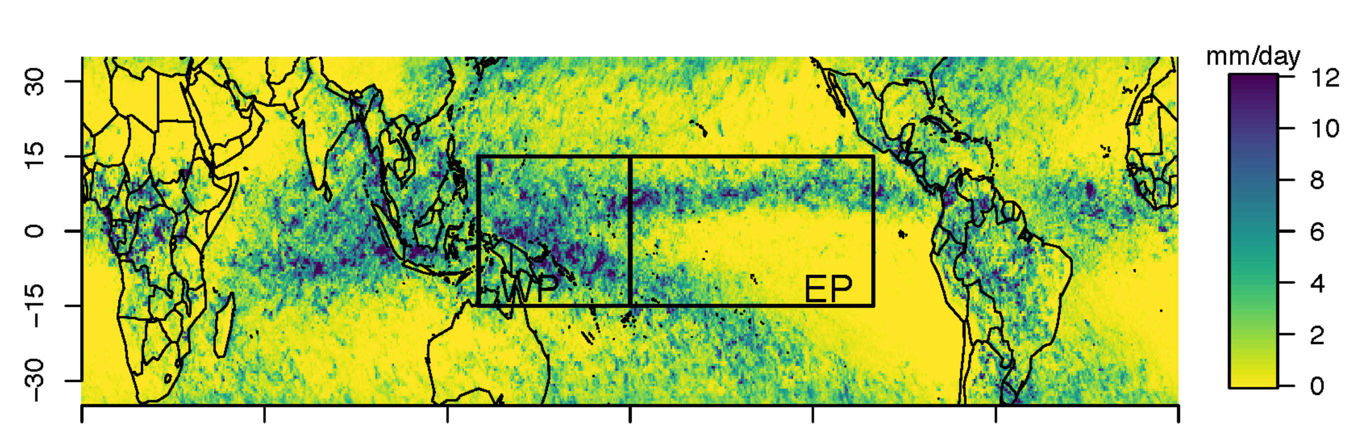


Figure 2. TRMM satellite annual average rainfall (2003)

Statistical modeling

First 3 EOFs of temperature, humidity, and other state variables used as predictors (Fig. 3). Generalized Linear Models (GLM) trained on 2003 data to predict occurrence and amount of rain in 2004. Average spatial pattern of rain is predicted well (Fig. 4).

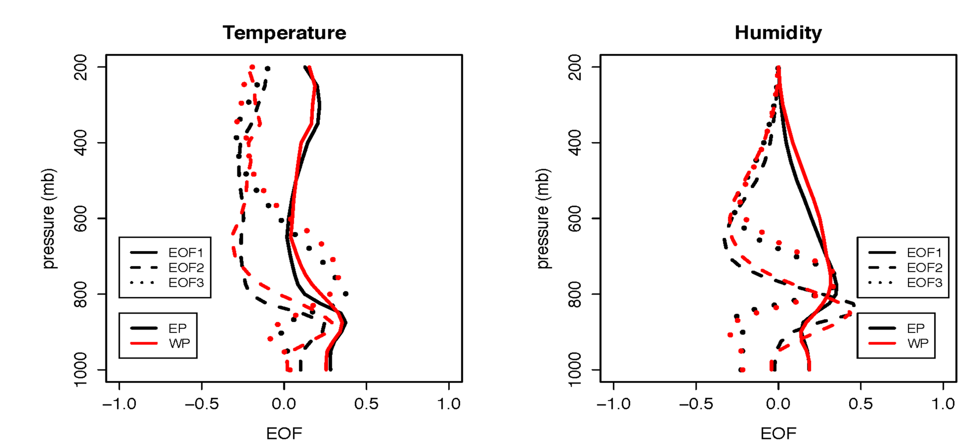


Figure 3. First 3 EOFs of T and q

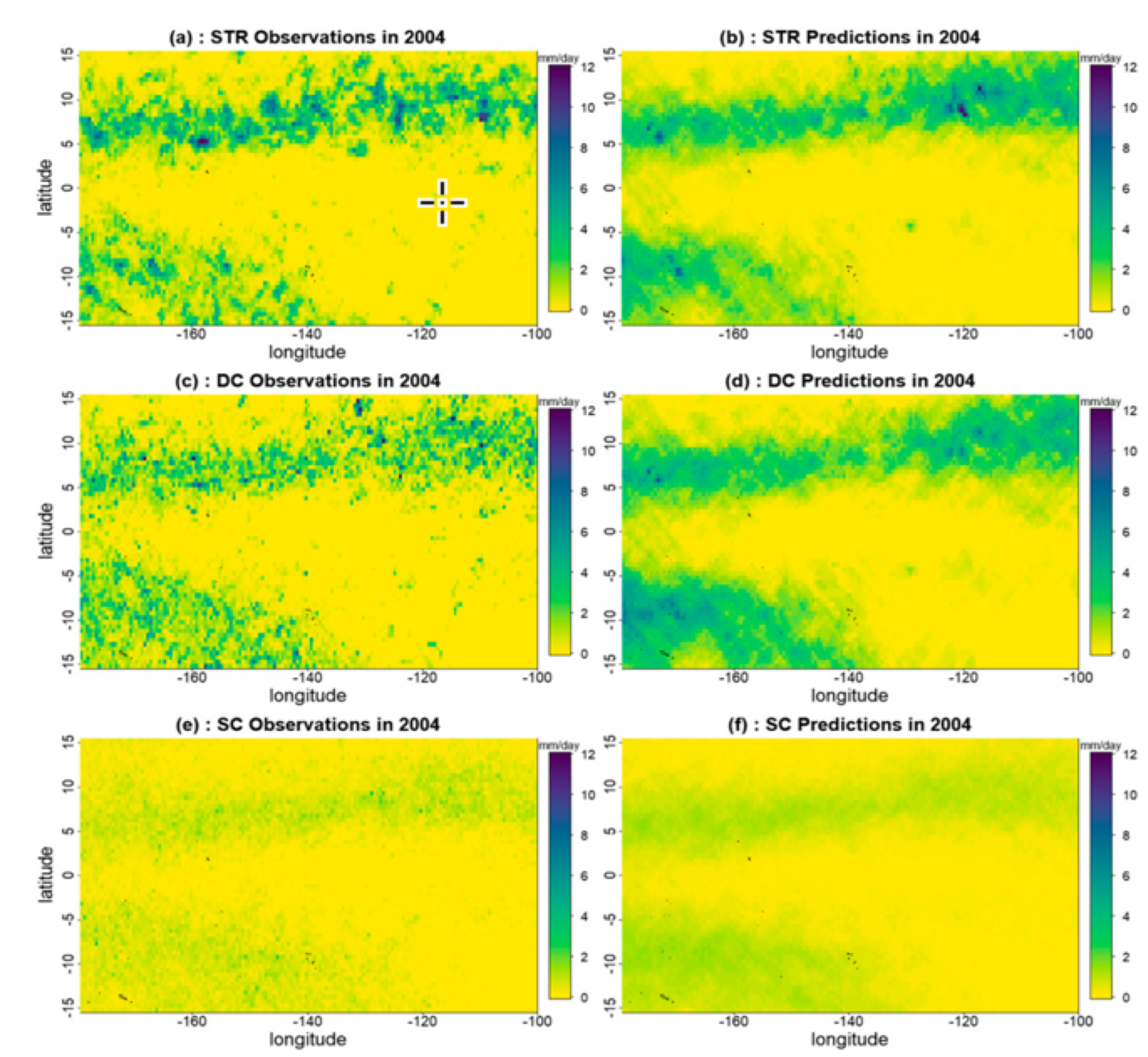


Figure 4. Observed (left) & predicted (right) average EP rainfall

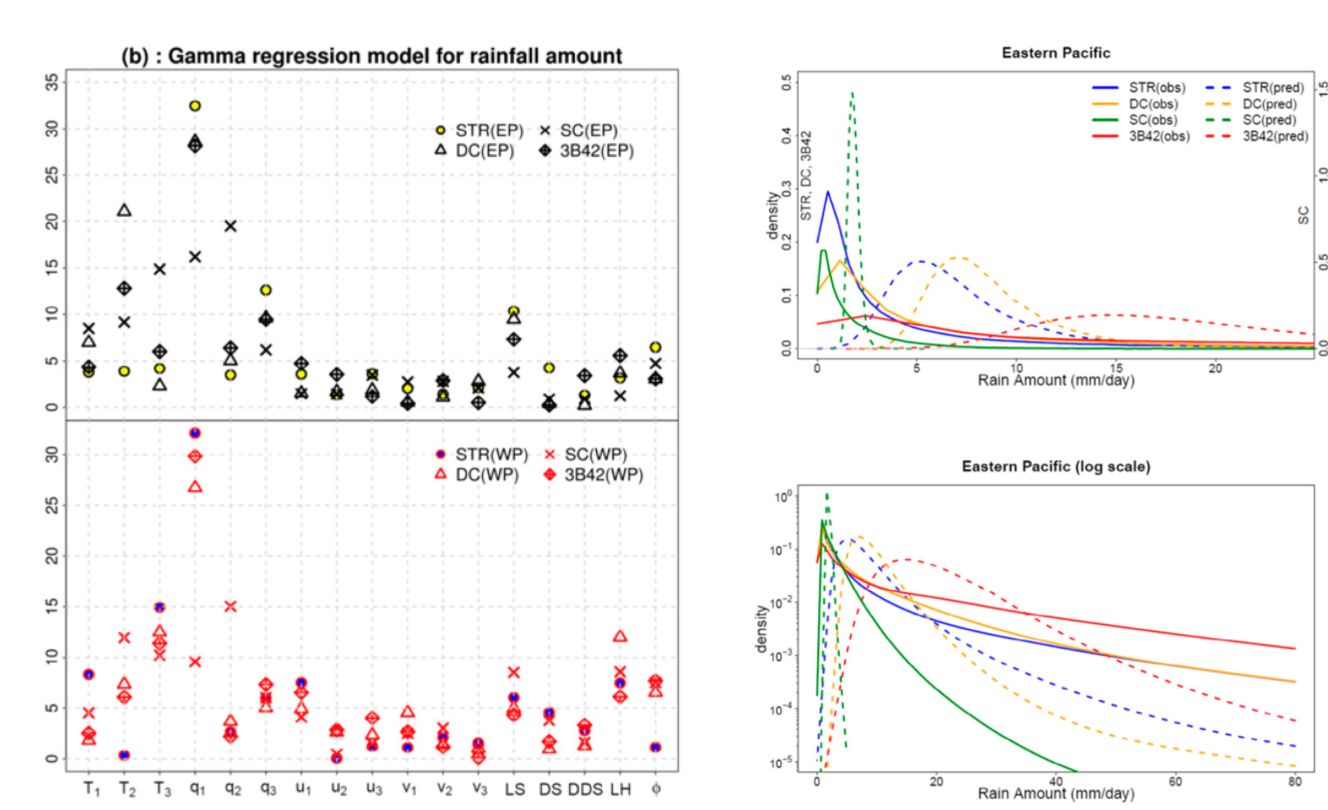


Figure 5. Predictive power of different variables (left). Observed and predicted EP rain rate distributions (right)

Machine Learning

Deep Feedforward Neural Networks (DFF-NN) were used to fit to 2017 GPM data and predict 2018 rain. The statistical GLM approach predicted mean spatial patterns well (Fig. 4), but not the rain rate (Fig. 5). DFF-NN appears to be able to predict both (Fig. 6)

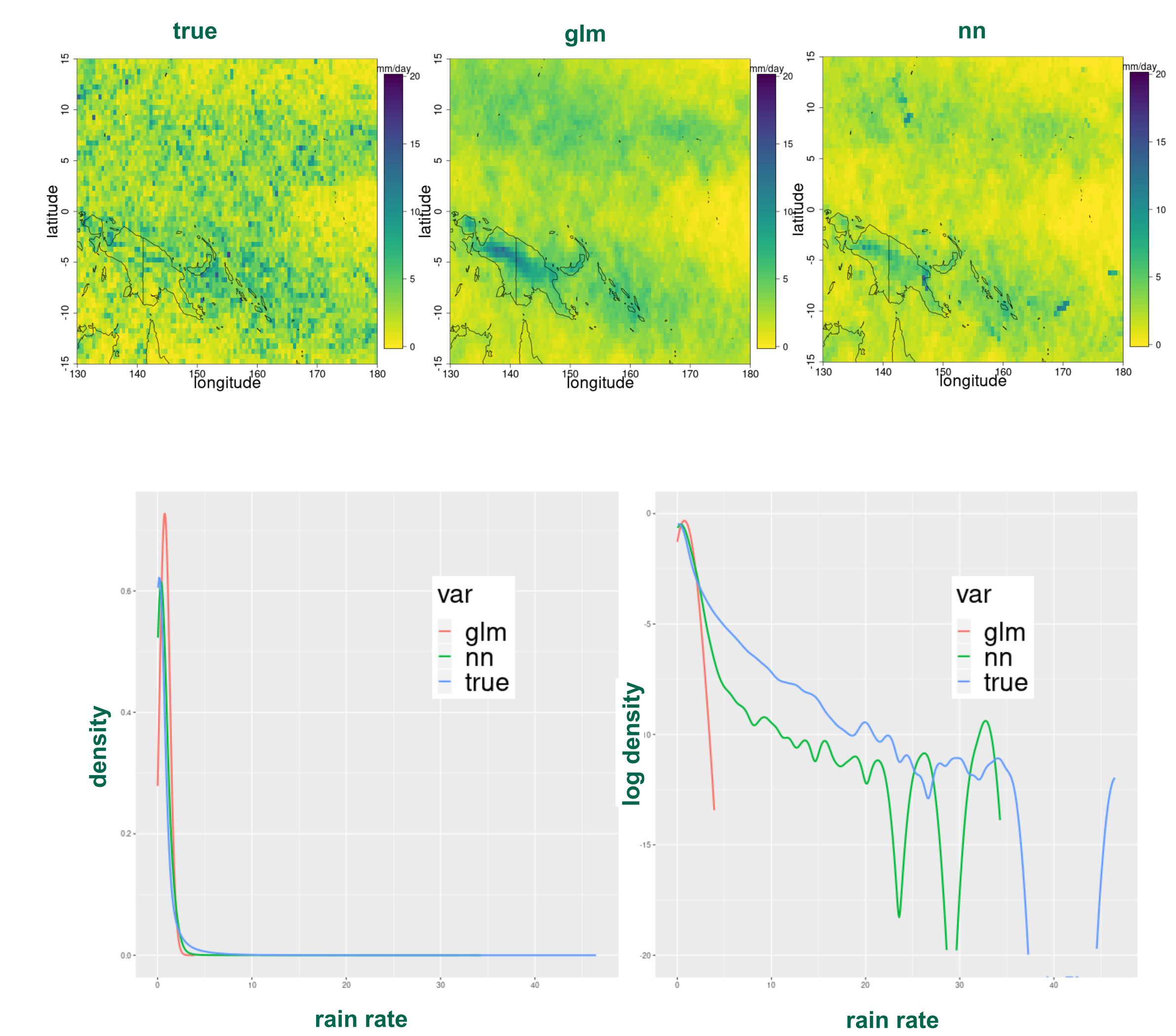


Figure 6. Observed, GLM, and DFF-NN prediction of average WP rainfall (top) and rain rates (bottom)

Conclusions

- Linear statistical models are able to predict the average spatial distribution of rainfall well (Fig. 4), but not so well the rain rate distribution (Fig. 5rt)
  - day-to-day rain occurrence is poorly predicted
- EOF1 of humidity and EOF2 of temperature provide the most predictive power (Fig. 3, Fig. 5lt)
- DFF Neural Networks appear to be able to predict the average spatial pattern and address some linear model deficiencies, e.g., rain rates (Fig. 6)

Acknowledgments

Texas A&M University Big Data Seed Grant and T3 Grant

References

Yang, J., Jun, M., Schumacher, C., Saravanan, R., 2019: Predictive statistical representations of observed and simulated rainfall using generalized linear models. *Journal of Climate*, v.32, 3409-3427pp doi:10.1175/JCLI-D-18-0527.1