LDRD2

Abigail Snyder

#### Key staff

Abigail Snyder (postdoc), Robert Link, Ben Bond-Lamberty, Corinne Hartin, Ben Kravitz

#### Purpose and impact

Two important topics to researchers in earth sciences, climate change, and integrated assessment are the effects of extreme events and uncertainty in climate impacts. Further, understanding the uncertainty around extreme events also has direct relevance for policymakers, the U.S. energy grid, and other stakeholders. Nowhere is this more important than in the water cycle, for which the precise distribution of precipitation over space and time is a primary driver of extreme events like drought and flooding and a key contributor to the energy-water-land nexus through its impact on agricultural performance.

Studying these topics requires large ensembles of high-resolution future climate scenarios, which are generally too expensive to produce directly by running Earth System Models (ESMs). Climate model emulators attempt to solve this problem by approximating the output a climate model *would have* produced had it been run for a specified scenario. These emulators are cheap to run, but capture only the mean response of the climate variable being modeled and little to none of the variability that would be present in a real ESM output. Attempts to add variability to such emulations often lose important spatial and temporal correlations that fundamentally define ESM patterns (such as the El Niño–Southern Oscillation) (such as the El Niño–Southern Oscillation) and are critically important for understanding extreme events,events, or else cannot produce the number of realizations necessary for extreme event or uncertainty studies. In recently completed work we have solved many of these problems for the case of global *temperature*. Our current method generates random realizations of temperature global time series that retain the spatial and temporal variance and covariance structures of the input ESM data at a much lower computational cost compared to the ESMs being emulated.

Given the serious impacts of extreme precipitation events, we believe the additional work to extend our method to produce *joint realizations* of temperature and precipitation will be valuable to researchers in a variety of fields housed in EBSD. Specifically, extending the method to preserve the space and time correlation structure *between* temperature and precipitation that is present in ESM outputs will result in more realistic, and therefore more useful, realizations than would be achieved by arbitrarily pairing independent temperature and precipitation realizations. This extension of our statistical approach is perfectly suited for this LDRD call because it will enhance the impact of our established method.

Our technique provides a computationally feasible way to produce the large ensembles of high-resolution future climate scenarios needed for studying the effects of extreme events and uncertainty in climate impacts. Furthermore, the lessons learned from extending the method to produce joint realizations of temperature and precipitation will allow the procedure to be further extended to variables such as atmospheric aerosols in the future.

The papers and software produced by this project and anticipated follow on projects will help PNNL maintain its position as a leader in the fields of earth system modeling, uncertainty quantification, and integrated assessment modeling. The software produced will be directly useful to PNNL researchers interested in regional effects of future climate. Moreover, the results of this project will supplement ongoing PNNL research into applying machine learning techniques to climate model emulation and climate prediction.

#### Foundational work

We will adapt our method for generating spatially and temporally coherent realizations of global gridded temperatures. The procedure fits a mean response model to input data and then generates thousands of new residual fields via empirical orthognal vector decomposition, discrete Fourier Transforms, and random phase perturbations. Each residual field is added to the mean response to create new realizations of the temperature field. This ensures that the generated temperature fields have the same statistical properties as the ESM input used to train the emulator. At the same time, the realizations are statistically independent of one another and of the input data, allowing the generated data to express extremes that may not have been present in the training data. Additionally, because the statistical properties of the ESM are inherited by our generated realizations, the extremes present in our realizations could have been produced in enough (expensive) runs of the ESM.

#### Technical approach

Adapting this technique to produce joint realizations of temperature and precipitation global time series will require some additional work to account for the unique characteristics of the precipitation data, particularly its highly nonnormal distribution. Our method outputs normally distributed realizations of residuals, which would be inappropriate to add directly to mean precipitation responses. We propose to address this by transforming precipitation residuals to an intermediate representation that is normally distributed, applying our method, and inverting the transformation on the realizations before adding to the mean precipitation response. There will also be software development required to collate the separately provided temperature and precipitation data and to rearrange them into a form suitable for analysis. We will then develop and implement statistical tests suitable for testing the joint statistical properties of the temperature and precipitation fields and comparing them to the ESM input.

#### Products and timeline

This project will derive a computationally-efficient method to generate random joint time series of global precipitation and temperature at grid resolution. On average, across realizations, each generated time series will have the same variance and time, space, and inter-variable correlation structure as the ESM time series used to train the system. The method will be implemented as a package in the R programming language and released as open source software. Based on our experience developing the temperature field generation method, we expect to submit a manuscript on the extended technique in the summer of 2018.

#### No other funding

The current JGCRI SFA is funded to *use* the software and data products described above, but not to *create* them. It would be very difficult to produce work of our desired high quality on the SFA with the currently existing data methods and products. A successful outcome of this proposal could also lead to expanded funding to explore the developments described above.

#### Estimated cost

 l c c l c l c c l Staff & rate & hours & cost & & Staff & rate & hours & cost

ACS & 56.35 & 80 & 4808.00 & & RPL & 137.32 & 40 & 5492.80
BBL & 137.32 & 20 & 2746.40 & & CH & 119 & 20 & 2380.00
BK & 119 & 20 & 2380.00 & & & & &

Total:& 17507.20 &&&&&&

# References