



## **Bayesian ANOVA for the calculation of climate anomalies.**

Martin Tingley

National Center for Atmospheric Research and Harvard University, United States (tingley@fas.harvard.edu)

For the purposes of studying climate, space-time variables such as temperature are generally analyzed only after the mean over some reference time interval has been removed from each time series. Given the underlying assumption that the climate variable is changing through time, a common interval must be used to calculate the mean of each time series. In general, the time series feature missing data, resulting in a reference interval shorter than the length of the data set. As a result, the variance across the estimated anomaly time series (i.e. spatial variance) is reduced within the reference interval, and inflated elsewhere.

I present an alternative approach to calculating climate anomalies, based on maximizing the length of the reference period and accounting for the increased uncertainty that results from the missing values. The original data set is decomposed using a two-factor ANOVA model, with the factors being location and year; estimates of the anomalies then result from removing the location effects from each time series. Within a Bayesian inference framework, the missing values in the data set are treated as additional parameters that must be estimated, while the posterior distributions of the year and location effects account for the uncertainty introduced by the missing observations.

This Bayesian ANOVA scheme is used to re-express an annually averaged version of the Climate Research Unit's gridded temperature anomaly product as anomalies with respect to means calculated over the entire 1850-2009 interval spanned by the data set. The choice of reference interval has a demonstrable effect on the second spatial moment time series of this temperature data set, but the first moment time series is in this case largely unaffected by the change in reference interval.