Erik Erhardt

Research Statement

Assistant Professor of Statistics
Department of Mathematics and Statistics
University of New Mexico

A theme of my research is the development of statistical models motivated by real-world problems to reveal scientifically meaningful structure in multivariate data by decomposing a mixture into its constituent sources. Method development is based on parsimonious models which can extend to uncover more complicated structure in the data. This strategy allows a researcher to either address elemental, pre-specified scientific questions or reveal more complicated structure given enough evidence. Two areas of application are in estimating animal diets in ecological foodwebs and in using functional imaging of the human brain to estimate how connections between brain regions differ between patient populations. Implications include better evidence-based decisions for environmental management and improved understanding of mental differences and treatment.

Models for ecological foodwebs

What animals have eaten can often be determined by matching the ratios of organic atomic isotopes between the animal's tissues and potential food sources. While analysis of an animal's stomach contents provide evidence of their last meal, and scat provides evidence for what wasn't digested, stable isotope ratios indicate which potential sources were digested and assimilated into the animal's body tissues over weeks (e.g., blood) and years (e.g., bone). For this reason, isotopic methods are invaluable for investigating the ecology of individuals and populations, and for making evidence-based environmental management decisions.

A class of models well-suited for diet inference is the mixing model, where an estimated probability vector indicates the proportional contribution of each potential food source to the diet of the consumer animal. When there are many sources, there are many possible diet proportion combinations that can explain the consumer's tissues; this so-called "underconstrained" situation has been addressed in two ways. First, our multivariate Bayesian model can combine information from isotope data with scientists' knowledge of animal behavior to estimate the likely range of consumer animal diets (**Erhardt** and Bedrick 2013). The parsimonious version consists of three independent submodels, one for each source of data, which are combined by the mixing model equation. We illustrate extensions to include covariates such as time, sex, and age, since an animal's dietary needs change during life events, such as during migration or reproduction. Second, we developed an R software package (sisus) to provide a quick representative set of diet vectors from an estimate of the most likely solution to the Bayesian model (**Erhardt**, Wolf, Ben-David, and Bedrick 2014).

When there are few sources, there may be only one or no possible diet proportion combinations of sources that can explain the consumer's tissues; this so-called "overconstrained" situation can be addressed with the Bayesian model above, or by an alternative (nonBayesian) model. Previous methods applied only to the situation when the diet proportion vector could be uniquely solved for in terms of one or two isotope ratios. We developed a model that applies to the "overconstrained" situation for an arbitrary number of isotope ratios, and generalized to allow diet constraints (such as "source 1 greater than source 2") and modeling over time (**Erhardt** and Bedrick 2014). This modelling alternative provides a closed-form solution (that can be computed by hand) to the diet vector. These models have been widely applied to many situations.

Plans for future research. Since the first depictions of a foodweb in 1880, an illustration of the directed graph of who eats whom in an ecosystem, people have attempted to estimate its structure. An important extension of the Bayesian model above is to estimate the edges of the foodweb directed graph. A paper in revision illustrates this model with a marine foodweb spanning from algae to dolphins, additionally illustrating the extensive use of supplemental diet and isotopic information from the literature and model assessment by comparing model predictions with observed data (**Erhardt**, Wilson, Nelson, and Chanton 2015). The development of software will facilitate the use of our model by

Dept of Mathematics and Statistics, Room SMLC 312, MSC 01 1115, 1 University of New Mexico
Albuquerque, New Mexico, 87131-0001 – USA
☎ *+1 (505) 750-4424* • ✉ *erike@stat.unm.edu*

ecologists.

There are many other extensions to pursue. A model selection strategy for mixing models is needed to understand whether an animal's diet is associated with covariates such as location, time of year, sex, etc. Current strategies fit a model for each condition, then compare the results; a single model-based strategy is preferred (**Erhardt** and Bond 2016). Furthermore, extensions will estimate when an animal has changed it's diet, and to understand the migration patterns of animals.

Collaborators are primarily animal ecologists from North America and England.

Erhardt, **EB** and EJ Bedrick (2013). "A Bayesian framework for stable isotope mixing models". *Environmental and Ecological Statistics* 20 (3). pdf, pp. 377–397. ISSN: 1352-8505. DOI: 10.1007/s10651-012-0224-1.

Erhardt, **EB**, BO Wolf, M Ben-David, and EJ Bedrick (May 2014). "Stable Isotope Sourcing using Sampling". *Open Journal of Ecology* 4 (6). pdf, pp. 289–298. DOI: 10.4236/oje.2014.46027.

Erhardt, **EB** and EJ Bedrick (2014). "Inference for stable isotope mixing models: a study of the diet of dunlin". *Journal of the Royal Statistical Society: Series C* 63 (4). pdf, pp. 579–593. DOI: 10.1111/rssc.12047.

Erhardt, **EB**, RM Wilson, J Nelson, and JP Chanton (2015). "An extended Bayesian stable isotope mixing model for trophic level inference". pdf, In Revision.

Erhardt, **EB** and Alexander Bond (2016). "Stable isotope sourcing with Bayesian covariate model selection".

Models for brain imaging

The large-scale neural activity of the human brain can be understood as a mixture of connected source networks (components) that vary over time. A general question we answer is "which regions of the brain work together, and how do these regions and their neuronal activity differ between patient populations?" Spatial independent component analysis (ICA) applied to functional magnetic resonance imaging (fMRI) data is a decomposition strategy applied with great success to identify functionally connected networks by estimating spatially independent patterns from their linearly mixed fMRI signals. Each component has two parts: (1) a set of spatial regions of the brain whose neuronal activity is represented by (2) a shared time course. Adding the estimated components together approximates the full activity of the brain.

Since 1998, ICA has been applied to fMRI data for individual subjects, followed by several models for performing multi-subject ICA. The general strategy for multi-subject ICA is typically to combine all subjects to estimate group components, then estimate subject-specific components. Multi-subject ICA aleviates two problems of analyzing subjects individually by having a higher signal-to-noise ratio when subjects are combined and removing the difficult component-matching problem between subjects.

My contributions in brain imaging follow the pattern of identifying an important community need and attempting to provide a general solution that is easy to implement. My first contribution to this area was to provide a general model framework for which each of several multi-subject ICA approaches was a special case (**Erhardt**, Rachakonda, et al. 2011). From the general framework we both identified contradictory assumptions in one popular method and identified a superior strategy that had not yet been used (now implemented in the popular GIFT matlab toolbox),

Applying our improved model to the largest multi-subject ICA at that time ($n = 603$), we recommend and illustrated a multivariate analysis strategy for evaluating the association of covariates on a range of response variables (Allen, **Erhardt**, Damaraju, et al. 2011). The method is implemented in our MANCOVAN matlab toolbox and in regular use.

Most method proposals include both a simulation as a proof-of-concept exercise and an application to real data. These custom one-shot simulations often lack realism or complexity, or are tuned to the strengths of the proposed method. To address the need for quick and realistic fMRI simulations, we developed a multi-subject fMRI simulation matlab toolbox, SimTB, to make it easy for researchers to assess their proposed methodology on a "ground truth" (**Erhardt**, Allen, Wei, Eichele, and Calhoun 2012; Allen, **Erhardt**, Wei, Eichele, and Calhoun 2011a). Our simulation model provided a middle ground to what already existed, strategies that either required real data as an input or others that simulated the brain's bloodflow process and can take hours to generate data. SimTB simulates block and event-related task data, as well as resting, and can mix conditions. In a companion paper, we used SimTB to better understand and document the capabilities and limitations of spatial multi-subject ICA approaches under conditions of spatial, temporal, and amplitude variability between subjects (Allen, **Erhardt**, Wei, Eichele, and Calhoun 2012).

We were one of the first groups to propose a strategy for understanding dynamic functional connectivity (FC), how the correlation of activity between different regions of the brain changes over short time scales, up to a few times per

minute (Allen, Damaraju, et al. 2012). This method was applied to understanding resting-state dynamics from a large sample ($n = 405$) of young adults.

Plans for future research. Continuing work in dynamic FC, in a well-scored NIH grant proposal (under revision) with collaborators from Rice University, MD Anderson, and the University of Bergen, Norway, we are developing a Bayesian dynamic graphical model to understand time-varying connectivity changes.

Erhardt, **EB**, S Rachakonda, et al. (2011). "Comparison of multi-subject ICA methods for analysis of fMRI data". *Human Brain Mapping* 32 (12). pdf, pp. 2075–2095. DOI: 10.1002/hbm.21170.

Allen, EA, **EB Erhardt**, E Damaraju, et al. (2011). "A baseline for the multivariate comparison of resting state networks". *Frontiers in Systems Neuroscience* 5. pdf, pp. 1–23. DOI: 10.3389/fnsys.2011.00002.

Erhardt, **EB**, EA Allen, Y Wei, T Eichele, and VD Calhoun (2012). "SimTB, a simulation toolbox for fMRI data under a model of spatiotemporal separability". *NeuroImage* 59 (4). pdf, pp. 4160–4167. DOI: 10.1016/j.neuroimage.2011.11.088.

Allen, EA, **EB Erhardt**, Y Wei, T Eichele, and VD Calhoun (2011a). *simtb: simulation toolbox for fMRI*. Software for Allen, **Erhardt**, Wei, Eichele, and Calhoun 2011b on "Capturing inter-subject variability". pdf. Mind Research Network. Albuquerque NM 87111. URL: http://mialab.mrn.org/software/simtb/index.html.

Allen, EA, **EB Erhardt**, Y Wei, T Eichele, and VD Calhoun (2012). "Capturing inter-subject variability with group independent component analysis of fMRI data: a simulation study". *NeuroImage* 59 (4). pdf, pp. 4141–4159. DOI: 10.1016/j.neuroimage.2011.10.010.

Allen, EA, E Damaraju, et al. (2012). "Tracking whole-brain connectivity dynamics in the resting-state". *Cerebral Cortex* 24 (3). pdf supp suppvid1 suppvid2 suppvid3, pp. 663–676. DOI: 10.1093/cercor/bhs352.

Allen, EA, **EB Erhardt**, Y Wei, T Eichele, and VD Calhoun (2011b). "Capturing inter-subject variability with group independent component analysis of fMRI data: a simulation study". In: *Organization for Human Brain Mapping*. pdf.

Visualizing scientific data

Effective visual communication remains an active interest. Because of the quality the visualizations in my papers with Elena Allen, we were invited to respond to a critique on brain functional network connectivity estimation, interpretation, assessment, and visualization (**Erhardt**, Allen, Damaraju, and Calhoun 2011). This led to a followup survey of visualizations in the neurosciences where we found that the field could improve our visual communication standards. We provided suggestions for "required" components of a plot and easy-to-implement "improved visualizations", and provided one novel idea with MATLAB code to "show more, hide less" in functional displays over structural MRI images (Allen, **Erhardt**, and Calhoun 2012). Dr. Allen and I were then invited to coauthor a book chapter titled "Visualizing Scientific Data" to appear in the 2016 in Handbook of Psychophysiology, 4th Edition by Cambridge University Press (Allen and **Erhardt** 2016).

Erhardt, **EB**, EA Allen, E Damaraju, and VD Calhoun (2011). "On network derivation, classification, and visualization: a response to Habeck and Moeller". *Brain connectivity* 1 (2). pdf, pp. 105–110. DOI: 10.1089/brain.2011.0022.

Allen, EA, **EB Erhardt**, and VD Calhoun (2012). "Data visualization in the neurosciences: overcoming the curse of dimensionality". *Neuron* 74. pdf, pp. 603–608. DOI: 10.1016/j.neuron.2012.05.001.

Allen, EA and **EB Erhardt** (2016). "Handbook of psychophysiology". In: ed. by John T Cacioppo, Louis G Tassinary, and Gary Berntson. 4th ed. pdf. Cambridge University Press. Chap. Visualizing Scientific Data, In Press.

Dept of Mathematics and Statistics, Room SMLC 312, MSC 01 1115, 1 University of New Mexico
Albuquerque, New Mexico, 87131-0001 – USA
☎ *+1 (505) 750-4424* • ✉ *erike@stat.unm.edu*