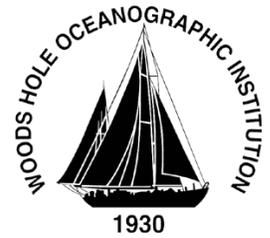


---

Woods Hole Oceanographic Institution  
**Biology/AOP&E Seminar**



Thursday, July 28, 2016  
Smith Conference Room – 12:00 Noon

---

## **Understanding Phytoplankton Communities Using Topic Models**

**Arnold Kalmbach**

**M.Sc. Student, McGill University**

Instruments such as the Imaging FlowCytobot provide us with novel, rich data about phytoplankton populations, and require new automated techniques to enable understanding beyond overall abundance and single taxa. During my guest student appointment, I have been investigating the use of topic models to discover communities of associated phytoplankton taxa. Topic models are a family of unsupervised probabilistic models that enable high-level interpretation of count data from many classes. In this project, the topic model is inferred from a timeseries of phytoplankton class counts recorded over 7 years at the Martha's Vineyard Coastal Observatory. It produces clusters representing the communities of phytoplankton taxa which are likely to be observed together, along with a description of each timestep by a mixture of these communities.

Using this plankton community model and a set of 18 environment variables, we can make much more accurate predictions of phytoplankton distributions compared to directly predicting the plankton populations. This suggests that while the relationship between particular taxa and environmental conditions may be too complex to understand directly, the relationship between communities and environmental conditions is much more straightforward, and easier to predict.

*Sponsored by WHOI Biology Department and Applied Ocean Physics and  
Engineering Department*