David Leifer

09/20/2016

Notes

Reading 1

Adaptive Spatio-Temporal Exploratory Models: Hemisphere-wide species distributions from massively crowdsourced eBird data

**Abstract**

-Multiple scales across range of spatial and temporal

-STEM- fixed scale

-AdaSTEM- Adaptive Spatio-Temporal Exploratory Models

-crowdsource eBird data- estimates distribution of Barn Swallows

**Introduction**

-phenomena should be studied at a variety of scales (Levin, 1992)

-exploratory models using tree data structures

-“Understanding distributional patterns in fine detail across broad extents is a key concern for biodiversity studies” lmao

-“During the breeding season, the location of good foraging habitat and nest sites determine bird occurrence at the same location at much smaller scales (Fortin and Dale 2005).”

-crowdsourced data follows human activity patterns because it allows people to choose broad observation points

-Executed on *Lonestar* cluster on XSEDE: Extreme Science and Engineering Discovery Envi

**STEM: Spatio-Temporal Exploratory Models**

-stixel = spatiotemporal pixel

-λ is fixed [lamb-da]

-mixture model creating a dense mixture of local regression models with overlapping *support*

-takes advantage that stixels overlap- taking average of all base models where stixel includes location

-combines bias reducing properties of local models (decision trees) with variance-reducing properties of randomized ensembles (bagging).

**The Mixture Model**

-STE[A]M is like a spatiotemporal wrapper for any base model

-linear models fit via least squares for synthetic experiments

-logistic Generalized Additive Models (GAM) for binary classification

---honestly still not sure what this means---

**Ensemble Theory­**

-size of Dm impacts the range of spatiotemporal correlation within/between base[ic] models

-results establish guidelines on how to construct Dm to improve predictive performance

**Adaptive Multiscale Modeling with AdaSTEM**

-λ controls size of stixels and min range of spatial correlation φˆ s

-similar to spatiotemporal BVC decomposition- tree data structures: decreasing function of density ρ at *s*

-based on tree-data structures

-data density = ρ [rho]

-locality = *s*

**Empirical Analysis**

-The base model for the m-th stixel, Dm is: zi = fm(x, y) = βm + βx,mxi + βy,myi + Ei

-RMSE increases sharply at around a stixel sample size of about 50 [38 minimizes GAM estimate of RMSE]

-“range of RMSE values for case B are comparable to those in case A. This demonstrates the ability of the ensemble averaging to control the increased variation as AdaSTEM adapts to *high* density observations.”

-AdaSTE[a]M outperforms STEM when multiscale structure exists

**eBird**

-broad-scale Barn Swallow citizen science collection checklist protocol: time, location, effort, count

-model accounts for effort

-coverage computed from 3000 location sampled from Stratified Random Design (SRD)

**Predictive Performance Comparison**

-test data was subsampled-computed independently because of seasonal variance

-AdaSTEAM > STEM in all months

-greatest performance difference is during **breeding** **season**

**Fall Migration Estimates**

- Notably, Barn Swallows are known to be absence from the Sierra Nevada, the southwestern portion of California and Arizona, and Florida

**Discussion**

-AdaSTEM demon-strated adaptability to multiscale signal when data is **dense** enough

-**density**-based adaptation = good