UCLA Department of Statistics

Spatial Models for Bird Origin Assignment Using Genetic and Isotopic Data

Colin Rundel

July 31, 2011

Project Background

Ongoing research at the Center for Tropical Research (CTR) at UCLA seeks to identify patterns of continental scale migratory connections

- Current methods are too coarse for most applications
- Large amounts of data are available (>150,000 feather samples from >500 species)
    - Genetic data - microsatellites, mitochondrial haplotypes, SNPs (soon)
    - Isotopic data - δ^2H

Species of interest

Hermit Thrush
Catharus guttatus

Photo by John Ingram

138 Individuals
14 Locations
6 Loci
9-27 Alleles

Wilson's Warbler
Wilsonia pusilla

© Glenn Bartley

163 Individuals
8 Locations
9 Loci
15-31 Alleles

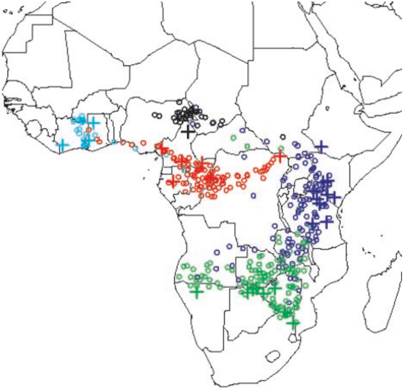
Model Framework

Assuming that the genetic(G) and isotopic(I) models are conditionally independent, then for a sample S and location k:

$$\begin{aligned}P(k|S,G,I) \propto\ & P(S|k,G,I)\ \pi(k)\\ &= P(S_G,S_I|k,G,I)\ \pi(k)\\ &= P(S_G|k,G,I)\ P(S_I|k,G,I)\ \pi(k)\\ &= P(S_G|k,G)\ P(S_I|k,I)\ \pi(k)\end{aligned}$$

Previous work

Wasser, et. al [2004] developed an approach for assigning genetic samples from illegal ivory shipments to geographic locations.



Model Basics

Using a multinomial error structure

$$p(y_{l\cdot k}|f_{l\cdot k}) = \frac{s_{lk}!}{\prod_i y_{lik}!} \prod_i (f_{lik})^{y_{lik}}$$

where:

- f_{lik} is the allele frequency of allele i from locus l at location k.
- y_{lik} is the count of allele i from locus l at location k.
- $s_{lk} = \sum_i y_{lik}$ is the total count of alleles from locus l at location k.

Modeling allele frequency

Allele frequencies are modeled using normalized values:

$$f_{lik} = \frac{\exp(\theta_{lik})}{\sum_j \exp(\theta_{ljk})}$$

where θ_{li} is a gaussian process:

$$\underset{[r\times 1]}{\boldsymbol{\theta}_{li}} \sim \text{MVN}(\underset{[r\times 1]}{\mathbf{M}_{li}}, \underset{[r\times r]}{\boldsymbol{\Sigma}})$$

Model Parameters

Mean

$$\underset{[r\times 1]}{\mathbf{M}_{li}} = \xi_l \ \eta_{li} \ \mathbf{1}_{[r,1]}$$

$$\xi_l \sim \text{Unif}(-\infty, \infty)$$

$$\eta_{li} \sim \text{N}(0, \beta_l)$$
$$\beta_l \sim \text{Unif}(0, 10^6)$$

Variance

$$\{\boldsymbol{\Sigma}\}_{k_1,k_2} = \sigma(d_{k_1,k_2}|\boldsymbol{\alpha})$$

Assumes process is stationary and isotropic

Covariance Functions

Powered Exponential Covariance:

$$\sigma(d|\boldsymbol{\alpha}) = \alpha_0 \exp\left[-\left(\frac{d}{\alpha_1}\right)^{\alpha_2}\right] + \alpha_3 \; I_{d=0}$$

Matérn Covariance:

$$\sigma(d|\boldsymbol{\alpha}) = \alpha_0 \frac{1}{\Gamma(\alpha_2)\; 2^{(\alpha_2-1)}} \left(\frac{d}{\alpha_1}\right)^{\alpha_2} K_{\alpha_2}\left(\frac{d}{\alpha_1}\right) + \alpha_3 \; I_{d=0}$$

with the following priors on $\boldsymbol{\alpha}$:

$$\alpha_0, \log(\alpha_1), \log(\alpha_2), \alpha_3 \sim \text{Unif}$$

Model Fitting via MCMC

All parameters are updated by random walk Metropolis Hasting with normal jump proposals.

However, we first reparameterize as follows:

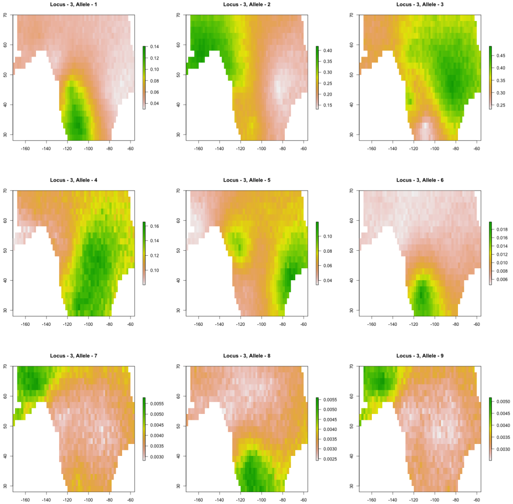
$$\underset{[r\times 1]}{\mathbf{V}_{li}} \sim \text{MVN}(0, \underset{[r\times r]}{\mathbf{\Sigma}})$$

$$\underset{[r\times 1]}{\mathbf{V}_{li}} = \underset{[r\times r]}{\text{Chol}(\mathbf{\Sigma})} \cdot \underset{[r\times 1]}{\mathbf{X}_{li}}$$

where

$$\{\mathbf{X}_{li}\}_k \sim \text{N}(0,1)$$

Allele Frequency - Hermit Thrush - Locus 3



Probability of a Sample

For a sample S with alleles i_l and j_l at locus l and location k

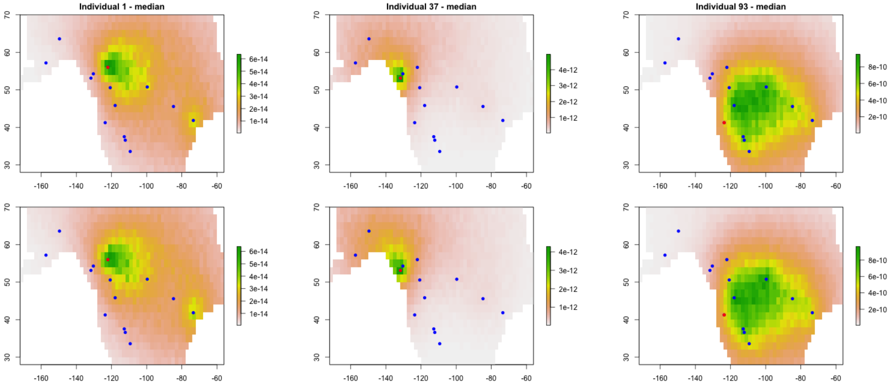
$$p(S|f,k) = \prod_l p_l(i_l, j_l|f,k)$$

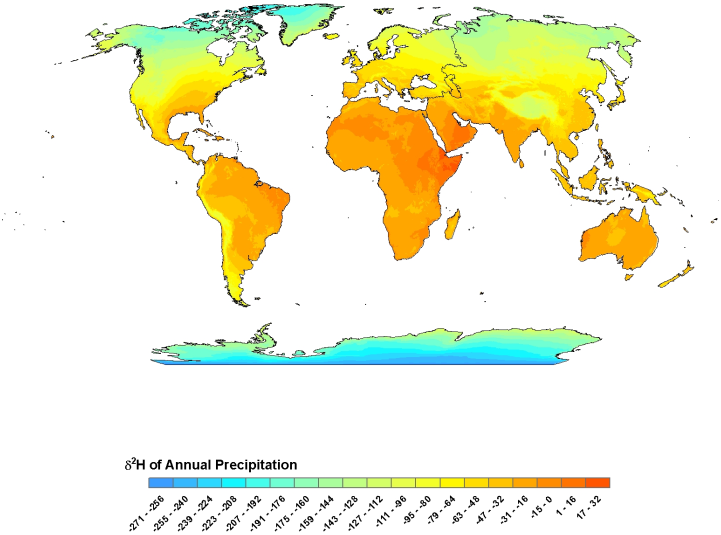
$$p_l(i_l, j_l|f,k) = \begin{cases} \gamma \; p_l(i_l|f,k) + (1-\gamma)\; p_l(i_l|f,k)^2 & \text{if } i_l = j_l \\ (1-\gamma)\; p(i_l|f,k)\; p(j_l|f,k) & \text{if } i_l \neq j_l \end{cases}$$

$$p_l(i_l|f,k) = (1-\delta) f_{lik} + \delta/m_l$$

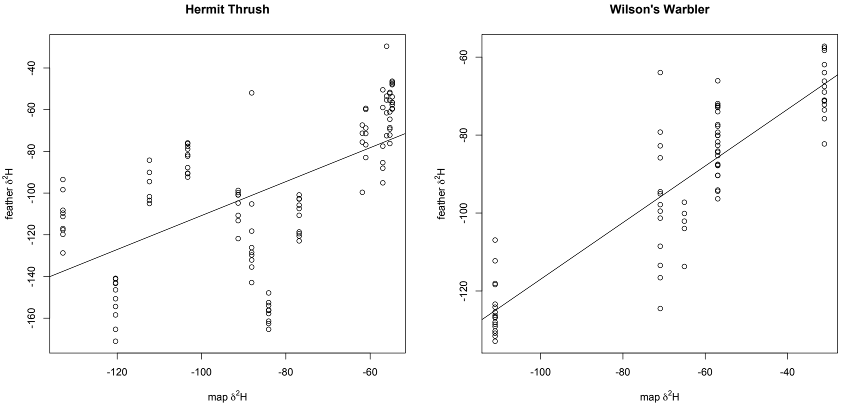
where δ is the probability only one of the alleles amplified and γ is the probability of a genotyping error.

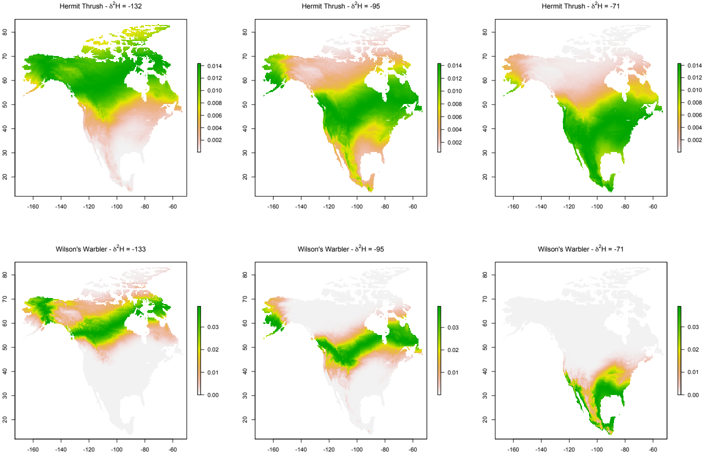
Spatial Posteriors - Hermit Thrush



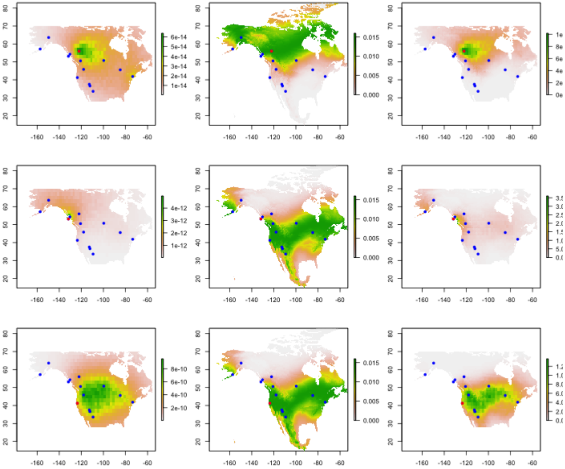

δ^2H of Annual Precipitation
-271 - -256
-255 - -240
-239 - -224
-223 - -208
-207 - -192
-191 - -176
-175 - -160
-159 - -144
-143 - -128
-127 - -112
-111 - -96
-95 - -80
-79 - -64
-63 - -48
-47 - -32
-31 - -16
-15 - 0
1 - 16
17 - 32

Mapping isotope values

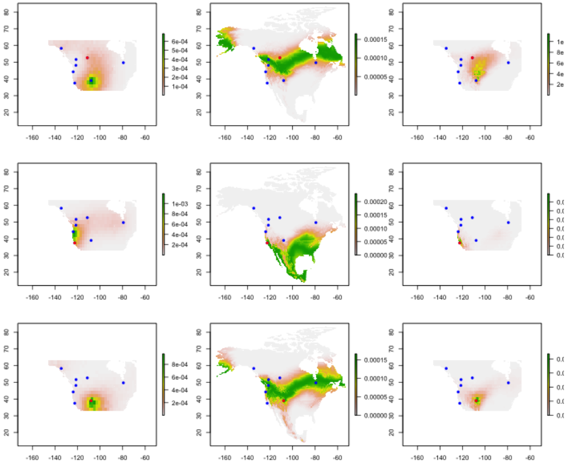



Hermit Thrush - δ^2H = -132
Hermit Thrush - δ^2H = -95
Hermit Thrush - δ^2H = -71
Wilson's Warbler - δ^2H = -133
Wilson's Warbler - δ^2H = -95
Wilson's Warbler - δ^2H = -71
80
70
60
50
40
30
20
-160
-140
-120
-100
-80
-60
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.03
0.02
0.01
0.00

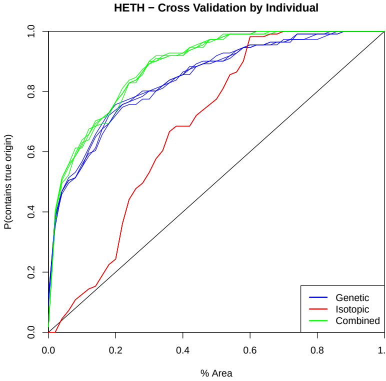
Combined - Hermit Thrush

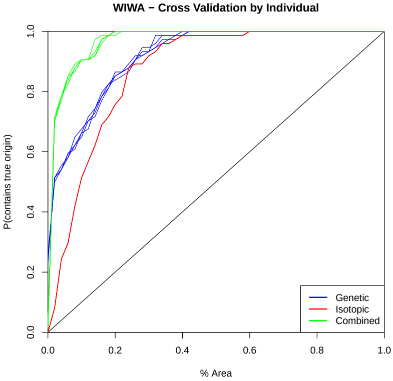


Combined - Wilson's Warbler

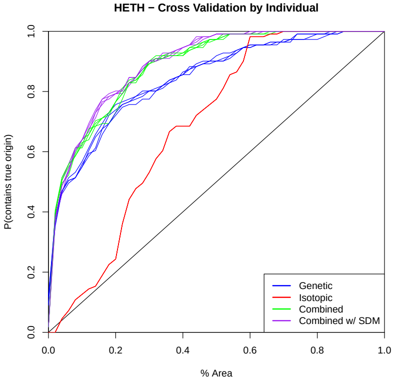


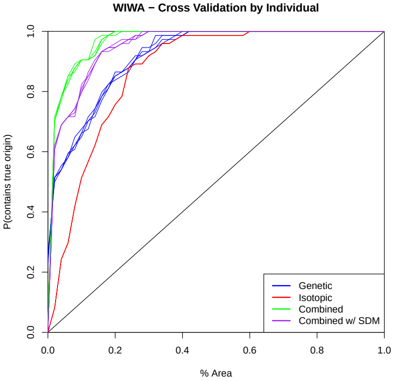
Classifier Results





Classifier Results + SDM





Conclusion

- Simple unified framework for combining Genetic and Isotopic models
- Combined results dramatically outperform either model alone
- Future Work
  - Fully bayesian isoscape model
  - Refine SDM priors
- R packages (available soon):
  - Genetic - Rscat
  - Isotopic - isoscape

Acknowledgements

Genetic Methods:

- John Novembre, UCLA
- Matthew Stephens, U of Chicago

Isotopic Methods:

- Michael Wunder, UC Denver
- Andrew Schuh, Colorado State

Feather Analysis:

- Tom Smith, UCLA
- Kristen Ruegg, UCSC
- Allison Alvarado, UCLA
- Ryan Harrigan, UCLA