The ‘Out-of-Africa’ hypothesis

Peter McCullagh

Department of Statistics
University of Chicago

South Africa, November 2013
Outline

Atkinson paper
  Phonemic inventories and linguistic diffusion hypothesis
  WALS data
  Fitting Atkinson’s cline model

Spatial and variance-component models

Controversies: Feb/Mar 2012
The boundaries and names shown and the designations used on this map do not imply official endorsement or acceptance by the United Nations.

Final boundary between the Republic of the Sudan and the Republic of South Sudan has not yet been determined.
Quentin Atkinson paper: Science 2011

Linguistic diffusion hypothesis:
Single point of origin (where?)
diversity process accumulates like genetic mutation
language spreads by migration/diffusion

At migration/fragmentation:
the ancestral population maintains linguistic inventory
the migrating group loses inventory

Implications of the diversity/inventory hypothesis:
older languages have greater vowel inventory
older languages have greater consonant inventory
older languages have greater tone inventory
older languages have greater phoneme inventory

Greater phoneme inventory ↔ closer to geographic source
Less phoneme inventory ↔ more distant from source
⇒ Phoneme inventory gradient
Atkinson’s WALS data

From *World Atlas of Language Structures*

1. Lname Abkhaz, Aikan??, B?t??, . . . , Zuni
2. Fam Arawakan, Indo-European, Sino-Tibetan,...
3. Lat as a decimal number: −12.67
4. Long e.g. −60.67 (meaning 60° 40’W)
5. Nvd Normalized vowel diversity: WALS feature No. 2
6. Ncd Normalized consonant diversity: WALS feature No. 1
8. Tnpd Total normalized phoneme diversity
9. Popn Estimated speaker population: integer
   1–873 014 298
10. Dbo Distance in km. from Atkinson’s best-fit origin

504 languages; 109 families Niger-Congo 62; Austronesian (42); Indo-European (42);
   Afro-Asiatic (??)
Atkinson’s WALS data

From *World Atlas of Language Structures*

1. **Lname** Abkhaz, Aikan??, B?t?©, . . ., Zuni
2. **Fam** Arawakan, Indo-European, Sino-Tibetan,...
3. **Lat** as a decimal number: −12.67
4. **Long** e.g. −60.67 (meaning 60° 40’W)
5. **Nvd** Normalized vowel diversity: WALS feature No. 2
6. **Ncd** Normalized consonant diversity: WALS feature No. 1
7. **Ntd** Normalized tone diversity: WALS feature No. 13
8. **Tnpd** Total normalized phoneme diversity
9. **Popn** Estimated speaker population: integer
   1–873 014 298
10. **Dbo** Distance in km. from Atkinson’s best-fit origin

504 languages; 109 families Niger-Congo 62; Austronesian (42); Indo-European (42);
Afro-Asiatic (??)
Atkinson’s WALS data

From *World Atlas of Language Structures*

1. **Lname** Abkhaz, Aikan??, B?©t?©, . . ., Zuni
2. **Fam** Arawakan, Indo-European, Sino-Tibetan,…
3. **Lat** as a decimal number: −12.67
4. **Long** e.g. −60.67 (meaning 60° 40’W)
5. **Nvd** Normalized vowel diversity: WALS feature No. 2
6. **Ncd** Normalized consonant diversity: WALS feature No. 1
7. **Ntd** Normalized tone diversity: WALS feature No. 13
8. **Tnpd** Total normalized phoneme diversity
9. **Popn** Estimated speaker population: integer
   1–873 014 298
10. **Dbo** Distance in km. from Atkinson’s best-fit origin

504 languages; 109 families  
Niger-Congo 62; Austronesian (42); Indo-European (42);  
Afro-Asiatic (??)
Atkinson’s WALS data
From *World Atlas of Language Structures*

1. **Lname** Abkhaz, Aikan??, B?t?©, . . ., Zuni
2. **Fam** Arawakan, Indo-European, Sino-Tibetan,...
3. **Lat** as a decimal number: −12.67
4. **Long** e.g. −60.67 (meaning 60° 40’W)
5. **Nvd** Normalized vowel diversity: WALS feature No. 2
6. **Ncd** Normalized consonant diversity: WALS feature No. 1
7. **Ntd** Normalized tone diversity: WALS feature No. 13
8. **Tnpd** Total normalized phoneme diversity
9. **Popn** Estimated speaker population: integer
   1–873 014 298
10. **Dbo** Distance in km. from Atkinson’s best-fit origin

504 languages; 109 families  Niger-Congo 62; Austronesian (42); Indo-European (42); Afro-Asiatic (??)
Atkinson’s WALS data

From World Atlas of Language Structures

1. Lname Abkhaz, Aikan??, B?©t?©, . . ., Zuni
2. Fam Arawakan, Indo-European, Sino-Tibetan,...
3. Lat as a decimal number: −12.67
4. Long e.g. −60.67 (meaning 60° 40’W)
5. Nvd Normalized vowel diversity: WALS feature No. 2
6. Ncd Normalized consonant diversity: WALS feature No. 1
8. Tnpd Total normalized phoneme diversity
9. Popn Estimated speaker population: integer
   1–873 014 298
10. Dbo Distance in km. from Atkinson’s best-fit origin

504 languages; 109 families Niger-Congo 62; Austronesian (42); Indo-European (42);
Afro-Asiatic (??)
Atkinson’s WALS data

From *World Atlas of Language Structures*

1. **Lname** Abkhaz, Aikan??, B?t?©, . . ., Zuni
2. **Fam** Arawakan, Indo-European, Sino-Tibetan, . . .
3. **Lat** as a decimal number: $-12.67$
4. **Long** e.g. $-60.67$ (meaning $60° 40′ W$)
5. **Nvd** Normalized vowel diversity: WALS feature No. 2
6. **Ncd** Normalized consonant diversity: WALS feature No. 1
7. **Ntd** Normalized tone diversity: WALS feature No. 13
8. **Tnpd** Total normalized phoneme diversity
9. **Popn** Estimated speaker population: integer
   1–873 014 298
10. **Dbo** Distance in km. from Atkinson’s best-fit origin

504 languages; 109 families  
Niger-Congo 62; Austronesian (42); Indo-European (42);
Afro-Asiatic (??)
Atkinson’s WALS data

From *World Atlas of Language Structures*

1. **Lname** Abkhaz, Aikan??, B?©t?©, . . ., Zuni
2. **Fam** Arawakan, Indo-European, Sino-Tibetan,...
3. **Lat** as a decimal number: −12.67
4. **Long** e.g. −60.67 (meaning 60° 40’W)
5. **Nvd** Normalized vowel diversity: WALS feature No. 2
6. **Ncd** Normalized consonant diversity: WALS feature No. 1
7. **Ntd** Normalized tone diversity: WALS feature No. 13
8. **Tnpd** Total normalized phoneme diversity
9. **Popn** Estimated speaker population: integer
   1–873 014 298
10. **Dbo** Distance in km. from Atkinson’s best-fit origin

504 languages; 109 families  Niger-Congo 62; Austronesian (42); Indo-European (42);
Afro-Asiatic (??)
Atkinson’s WALS data

From *World Atlas of Language Structures*

1. **Lname** Abkhaz, Aikan??, B?t??, . . ., Zuni
2. **Fam** Arawakan, Indo-European, Sino-Tibetan,...
3. **Lat** as a decimal number: \(-12.67\)
4. **Long** e.g. \(-60.67\) (meaning 60° 40’W)
5. **Nvd** Normalized vowel diversity: WALS feature No. 2
6. **Ncd** Normalized consonant diversity: WALS feature No. 1
7. **Ntd** Normalized tone diversity: WALS feature No. 13
8. **Tnpd** Total normalized phoneme diversity
9. **Popn** Estimated speaker population: integer
   
   1–873 014 298
10. **Dbo** Distance in km. from Atkinson’s best-fit origin

504 languages; 109 families  
Niger-Congo 62; Austronesian (42); Indo-European (42);  
Afro-Asiatic (??)
Atkinson’s WALS data

From *World Atlas of Language Structures*

1. **Lname** Abkhaz, Aikan??, B?t?, . . ., Zuni
2. **Fam** Arawakan, Indo-European, Sino-Tibetan, . . .
3. **Lat** as a decimal number: \(-12.67\)
4. **Long** e.g. \(-60.67\) (meaning 60° 40’W)
5. **Nvd** Normalized vowel diversity: WALS feature No. 2
6. **Ncd** Normalized consonant diversity: WALS feature No. 1
7. **Ntd** Normalized tone diversity: WALS feature No. 13
8. **Tnpd** Total normalized phoneme diversity
9. **Popn** Estimated speaker population: integer 1–873 014 298
10. **Dbo** Distance in km. from Atkinson’s best-fit origin

504 languages; 109 families Niger-Congo 62; Austronesian (42); Indo-European (42); Afro-Asiatic (??)
Atkinson’s WALS data

From World Atlas of Language Structures

1. Lname Abkhaz, Aikan??, B?©t?©, . . ., Zuni
2. Fam Arawakan, Indo-European, Sino-Tibetan, ...
3. Lat as a decimal number: −12.67
4. Long e.g. −60.67 (meaning 60° 40’W)
5. Nvd Normalized vowel diversity: WALS feature No. 2
6. Ncd Normalized consonant diversity: WALS feature No. 1
8. Tnpd Total normalized phoneme diversity
9. Popn Estimated speaker population: integer
   1–873 014 298
10. Dbo Distance in km. from Atkinson’s best-fit origin

504 languages; 109 families Niger-Congo 62; Austronesian (42); Indo-European (42);
    Afro-Asiatic (??)
Worldwide language distribution
Restrictions on migration routes: choke points
A sample of Atkinson’s WALS data

<table>
<thead>
<tr>
<th>Lname</th>
<th>WALS</th>
<th>Fam</th>
<th>Lat</th>
<th>Long</th>
<th>Tnpd</th>
<th>ISO</th>
<th>Popn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abkhaz</td>
<td>abk</td>
<td>Nw Caucasian</td>
<td>43.08</td>
<td>41</td>
<td>-1.186</td>
<td>abk</td>
<td>105952</td>
</tr>
<tr>
<td>Acoma</td>
<td>aco</td>
<td>Keresan</td>
<td>34.92</td>
<td>-107.58</td>
<td>0.220</td>
<td>kjq</td>
<td>3391</td>
</tr>
<tr>
<td>Andoke</td>
<td>adk</td>
<td>Andoke</td>
<td>-0.67</td>
<td>-72</td>
<td>0.127</td>
<td>ano</td>
<td>619</td>
</tr>
<tr>
<td>Aleut E</td>
<td>aea</td>
<td>Eskimo-Aleut</td>
<td>54.75</td>
<td>-164</td>
<td>-0.348</td>
<td>ale</td>
<td>490</td>
</tr>
<tr>
<td>Arabic (Egyptian)</td>
<td>aeg</td>
<td>Afro-Asiatic</td>
<td>30</td>
<td>31</td>
<td>-0.377</td>
<td>arz</td>
<td>46321000</td>
</tr>
<tr>
<td>Aghem</td>
<td>agh</td>
<td>Niger-Congo</td>
<td>6.67</td>
<td>10</td>
<td>0.965</td>
<td>agq</td>
<td>26727</td>
</tr>
<tr>
<td>Ahtna</td>
<td>aht</td>
<td>Na-Dene</td>
<td>62</td>
<td>-145</td>
<td>-0.936</td>
<td>aht</td>
<td>80</td>
</tr>
<tr>
<td>Aikan??</td>
<td>aik</td>
<td>Arawakan</td>
<td>-12.67</td>
<td>-60.67</td>
<td>0.061</td>
<td>tba</td>
<td>90</td>
</tr>
<tr>
<td>Ainu</td>
<td>ain</td>
<td>Ainu</td>
<td>43</td>
<td>143</td>
<td>0.061</td>
<td>ain</td>
<td>15</td>
</tr>
<tr>
<td>Akan</td>
<td>akn</td>
<td>Niger-Congo</td>
<td>6.5</td>
<td>-1.25</td>
<td>0.127</td>
<td>aka</td>
<td>8300000</td>
</tr>
<tr>
<td>Akawaio</td>
<td>akw</td>
<td>Cariban</td>
<td>6</td>
<td>-59.5</td>
<td>-0.311</td>
<td>ake</td>
<td>5000</td>
</tr>
<tr>
<td>Alamblak</td>
<td>ala</td>
<td>Sepik</td>
<td>-4.67</td>
<td>143.33</td>
<td>0.247</td>
<td>alh</td>
<td>1527</td>
</tr>
<tr>
<td>Albanian</td>
<td>alb</td>
<td>Indo-European</td>
<td>41</td>
<td>20</td>
<td>0.247</td>
<td>aln, als</td>
<td>5823075</td>
</tr>
<tr>
<td>Alladian</td>
<td>ald</td>
<td>Niger-Congo</td>
<td>5.17</td>
<td>-4.33</td>
<td>0.127</td>
<td>ald</td>
<td>23000</td>
</tr>
<tr>
<td>Alawa</td>
<td>alw</td>
<td>Australian</td>
<td>-15.17</td>
<td>134.25</td>
<td>-0.348</td>
<td>alh</td>
<td>17</td>
</tr>
<tr>
<td>Amele</td>
<td>ame</td>
<td>Trans-New Guinea</td>
<td>-5.25</td>
<td>145.58</td>
<td>0.181</td>
<td>aey</td>
<td>5300</td>
</tr>
<tr>
<td>Amharic</td>
<td>amh</td>
<td>Afro-Asiatic</td>
<td>10</td>
<td>38</td>
<td>0.526</td>
<td>amh</td>
<td>17417913</td>
</tr>
<tr>
<td>Amo</td>
<td>amo</td>
<td>Niger-Congo</td>
<td>10.33</td>
<td>8.67</td>
<td>0.686</td>
<td>amo</td>
<td>12263</td>
</tr>
<tr>
<td>Arrernte</td>
<td>amp</td>
<td>Australian</td>
<td>-24</td>
<td>136</td>
<td>-0.348</td>
<td>aer</td>
<td>2175</td>
</tr>
<tr>
<td>Amuesha</td>
<td>amu</td>
<td>Arawakan</td>
<td>-10.5</td>
<td>-75.42</td>
<td>-0.348</td>
<td>ame</td>
<td>9831</td>
</tr>
<tr>
<td>Amuzgo</td>
<td>amz</td>
<td>Oto-Manguean</td>
<td>16.83</td>
<td>-98</td>
<td>0.845</td>
<td>amu</td>
<td>23000</td>
</tr>
<tr>
<td>Araona</td>
<td>ana</td>
<td>Tacanan</td>
<td>-12.33</td>
<td>-67.75</td>
<td>-0.069</td>
<td>aro</td>
<td>81</td>
</tr>
<tr>
<td>Angas</td>
<td>anc</td>
<td>Afro-Asiatic</td>
<td>9.5</td>
<td>9.5</td>
<td>0.566</td>
<td>anc</td>
<td>400000</td>
</tr>
<tr>
<td>//Ani</td>
<td>ani</td>
<td>Khoisan</td>
<td>-18.92</td>
<td>21.92</td>
<td>-0.218</td>
<td>hnh</td>
<td>1000</td>
</tr>
<tr>
<td>Angaatiha</td>
<td>ant</td>
<td>Trans-New Guinea</td>
<td>-7.22</td>
<td>146.25</td>
<td>0.339</td>
<td>agm</td>
<td>2100</td>
</tr>
<tr>
<td>Yulu</td>
<td>yul</td>
<td>Nilo-Saharan</td>
<td>8.5</td>
<td>25.25</td>
<td>1.682</td>
<td>yul</td>
<td>7000</td>
</tr>
<tr>
<td>Yurok</td>
<td>yur</td>
<td>Algic</td>
<td>41.33</td>
<td>-124</td>
<td>-0.098</td>
<td>yur</td>
<td>12</td>
</tr>
<tr>
<td>Yupik</td>
<td>yus</td>
<td>Eskimo-Aleut</td>
<td>65</td>
<td>-173</td>
<td>-0.348</td>
<td>ess</td>
<td>1350</td>
</tr>
<tr>
<td>Zande</td>
<td>zan</td>
<td>Niger-Congo</td>
<td>4</td>
<td>26</td>
<td>0.685</td>
<td>zne</td>
<td>1142000</td>
</tr>
<tr>
<td>Zoque</td>
<td>zqc</td>
<td>Mixe-Zoque</td>
<td>17</td>
<td>-93.25</td>
<td>-0.657</td>
<td>zoc</td>
<td>10000</td>
</tr>
<tr>
<td>Zulu</td>
<td>zul</td>
<td>Niger-Congo</td>
<td>-30</td>
<td>30</td>
<td>0.339</td>
<td>zul</td>
<td>9563422</td>
</tr>
<tr>
<td>Zuni</td>
<td>zun</td>
<td>Zuni</td>
<td>35.08</td>
<td>-108.83</td>
<td>-0.377</td>
<td>zun</td>
<td>9651</td>
</tr>
</tbody>
</table>
Statistical modelling: Atkinson style

Response: Total normalized phoneme inventory $Y_i$

Geographic coords: $x_i = (x_1, x_2)_i$ of language $i$

No of speakers: $s_i$

Language origin coordinates: $\tau = (\tau_0, \tau_1)$

$$E(Y_i) = \beta_0 + \beta_1 \|x_i - \tau\| + \gamma \log(s_i)$$

$$\text{cov}(Y_i, Y_j) = \sigma^2 \delta_{ij}$$

Parameters: $\beta_0, \beta_1, \gamma, \tau, \sigma^2$

distance $\|x_i - \tau\|$ versus log distance

RSS as a function of $\tau$ or $F$-ratio

$$F(\tau) = \frac{(RSS(\tau) - RSS(\hat{\tau}))/2}{RSS(\hat{\tau})/(n - 5)}$$
Phonemic inventory versus distance from $\vec{r}$
Phonemic inventory versus distance (by region)
Test for consistency of pattern across continents:

Null: $E(Y) = \text{dist} + \log(\text{pop})$

Alt: $E(Y) = \text{Reg} + \text{Reg.dist} + \log(\text{pop})$

$$F = \frac{\text{Reduction in RSS}/(\text{Reduction in df})}{\text{Resid SS}/\text{Resid df}}$$

F-ratio = 8.6 on (10, 491) df. (should be < 2)

No clear evidence of a consistent cline
RSS ($F$-ratio) as a function of origin $\tau$
RSS ($F$-ratio) as a function of origin $\tau$
Atkinson’s posterior density for $\tau$
Confidence regions, HPD regions

Fisher’s $F$-ratio and the excess variance criterion

$$F(\tau) = \frac{(RSS(\tau) - RSS(\hat{\tau})) / 2}{RSS(\hat{\tau}) / (n - 5)}$$

Distributed as $F_{2,n-5}$ approx; 95% quantile $F < 3.01$

BIC $\equiv$ LR:

$$n \log RSS(\tau) - n \log(RSS(\hat{\tau})) = n \log(1 + 2F / (n - 5)) \simeq 2F$$

distributed as $2F_{2,n-5} \sim \chi^2_2$ approx

95% frequentist region $LR(\tau) < 6$
Atkinson’s region as illustrated: $\text{BIC} < 4$
whence 4 BIC units?
coverage prob: 86%
Further modelling considerations

\[ E(Y_i) = \beta_0 + \beta_1 \|x_i - \tau\| + \gamma \log(s_i) \]
\[ \text{cov}(Y_i, Y_j) = \sigma^2 \delta_{i,j} \]

- Diversity implies \( \beta_1 \leq 0 \): \( F \)-ratio does not distinguish
  In fact: \( \hat{\beta}_1 < 0 \) for Africa, Europe, Asia
  \( \hat{\beta}_1 > 0 \) for America, Oceania (incl Malagasy)

- Covariates versus relationships
  covariate \( i \mapsto s_i \) (number of speakers)
  \( (i, j) \mapsto R_1(i, j) \) same linguistic family
  \( (i, j) \mapsto \|x_i - x_j\| \) geographic or linguistic distance

- Independence
  Possible correlation within linguistic families
  Possible correlation due to geographic proximity

- Covariates: \( E(Y_i) = \beta_0 + \beta_1 x_i + \cdots \)
  Relationships: \( \text{cov}(Y_i, Y_j) = \sigma^2_0 \delta_{i,j} + \sigma^2_1 R_1(i, j) + \cdots \)
Further modelling considerations

\[ E(Y_i) = \beta_0 + \beta_1 \|x_i - \tau\| + \gamma \log(s_i) \]
\[ \text{cov}(Y_i, Y_j) = \sigma^2 \delta_{i,j} \]

- Diversity implies \( \beta_1 \leq 0 \): \( F \)-ratio does not distinguish
  In fact: \( \hat{\beta}_1 < 0 \) for Africa, Europe, Asia
  \( \hat{\beta}_1 > 0 \) for America, Oceania (incl Malagasy)

- Covariates versus relationships
  covariate \( i \mapsto s_i \) (number of speakers)
  \( (i, j) \mapsto R_1(i, j) \) same linguistic family
  \( (i, j) \mapsto \|x_i - x_j\| \) geographic or linguistic distance

- Independence
  Possible correlation within linguistic families
  Possible correlation due to geographic proximity

- Covariates:
  \( E(Y_i) = \beta_0 + \beta_1 x_i + \cdots \)

Relationships:
\[ \text{cov}(Y_i, Y_j) = \sigma^2_0 \delta_{i,j} + \sigma^2_1 R_1(i, j) + \cdots \]
Further modelling considerations

\[ E(Y_i) = \beta_0 + \beta_1 \| x_i - \tau \| + \gamma \log(s_i) \]
\[ \text{cov}(Y_i, Y_j) = \sigma^2 \delta_{i,j} \]

- Diversity implies \( \beta_1 \leq 0 \): F-ratio does not distinguish
  In fact: \( \hat{\beta}_1 < 0 \) for Africa, Europe, Asia
  \( \hat{\beta}_1 > 0 \) for America, Oceania (incl Malagasy)

- Covariates versus relationships
  Covariate \( i \mapsto s_i \) (number of speakers)
  \( (i, j) \mapsto R_1(i, j) \) same linguistic family
  \( (i, j) \mapsto \| x_i - x_j \| \) geographic or linguistic distance

- Independence
  Possible correlation within linguistic families
  Possible correlation due to geographic proximity

- Covariates: \( E(Y_i) = \beta_0 + \beta_1 x_i + \cdots \)
  Relationships: \( \text{cov}(Y_i, Y_j) = \sigma_0^2 \delta_{i,j} + \sigma_1^2 R_1(i, j) + \cdots \)
Further modelling considerations

\[
E(Y_i) = \beta_0 + \beta_1 \|x_i - \tau\| + \gamma \log(s_i)
\]
\[
\text{cov}(Y_i, Y_j) = \sigma^2 \delta_{i,j}
\]

- Diversity implies \(\beta_1 \leq 0\): \(F\)-ratio does not distinguish
  - In fact: \(\hat{\beta}_1 < 0\) for Africa, Europe, Asia
  - \(\hat{\beta}_1 > 0\) for America, Oceania (incl Malagasy)

- Covariates versus relationships
  - covariate \(i \mapsto s_i\) (number of speakers)
  - \((i, j) \mapsto R_1(i, j)\) same linguistic family
  - \((i, j) \mapsto \|x_i - x_j\|\) geographic or linguistic distance

- Independence
  - Possible correlation within linguistic families
  - Possible correlation due to geographic proximity

- Covariates: \(E(Y_i) = \beta_0 + \beta_1 x_i + \cdots\)
  - Relationships: \(\text{cov}(Y_i, Y_j) = \sigma_0^2 \delta_{i,j} + \sigma_1^2 R_1(i, j) + \cdots\)
A linear Gaussian variance-components model

\[ E(Y_i) = \beta_0 + \beta_1 \| x_i - \tau \| + \gamma \log(s_i) \]
\[ \text{cov}(Y_i, Y_j) = \sigma^2 \delta_{i,j} + \sigma_1^2 F_{ij} + \sigma_2^2 \exp(-d_{ij}/\rho) \]

\[ \tau = (9.5, -1.25); \quad \rho = 1000 \text{ km fixed} \]

Model comparison via likelihood-ratio statistics

<table>
<thead>
<tr>
<th>Covariance model</th>
<th>( \delta )</th>
<th>( F )</th>
<th>( V )</th>
<th>( F + V )</th>
<th>( + F \cdot V )</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR * 2</td>
<td>0.00</td>
<td>88.80</td>
<td>127.17</td>
<td>143.69</td>
<td>147.70</td>
</tr>
<tr>
<td>( \hat{\beta}_1 \times 10^4 )</td>
<td>-0.391</td>
<td>-0.313</td>
<td>-0.351</td>
<td>-0.331</td>
<td>-0.334</td>
</tr>
<tr>
<td>SE</td>
<td>0.036</td>
<td>0.057</td>
<td>0.077</td>
<td>0.079</td>
<td>0.074</td>
</tr>
</tbody>
</table>

Conclusions:
- Evidence for both family and geographic correlatons
- Dbo coefficient \( \hat{\beta}_1 \) not much affected
- Ordinary least-squares SE is misleading
Confidence regions for source (exponential model)

95%: all of Africa
99%: Afr, Eur, Levant
Confidence regions for source (Bessel0 model)

95%: Afr, Eur, Levant
99%: World
Summary

- Distinction between covariates and relationships
  - Rel₁: family as a block factor $F_{ij}$
  - Rel₂: geographic separation $D_{ij}$
  - Spatial covariance models: Matérn’s Bessel family
  - Effect of correlation on OLS coefficients
  - Effect of correlation on SE of OLS coefficients
  - Selection of languages for inclusion in analysis
  - Choice of total phoneme inventory as response
  - Alternative forms of data using specific phonemes
Summary

- Distinction between covariates and relationships
- Rel$_1$: family as a block factor $F_{ij}$
- Rel$_2$: geographic separation $D_{ij}$
- Spatial covariance models: Matérn’s Bessel family
- Effect of correlation on OLS coefficients
- Effect of correlation on SE of OLS coefficients
- Selection of languages for inclusion in analysis
- Choice of total phoneme inventory as response
- Alternative forms of data using specific phonemes
Summary

- Distinction between covariates and relationships
- $\text{Rel}_1$: family as a block factor $F_{ij}$
- $\text{Rel}_2$: geographic separation $D_{ij}$
- Spatial covariance models: Matérn’s Bessel family
- Effect of correlation on OLS coefficients
- Effect of correlation on SE of OLS coefficients
- Selection of languages for inclusion in analysis
- Choice of total phoneme inventory as response
- Alternative forms of data using specific phonemes
Summary

- Distinction between covariates and relationships
- $\text{Rel}_1$: family as a block factor $F_{ij}$
- $\text{Rel}_2$: geographic separation $D_{ij}$
- Spatial covariance models: Matérn’s Bessel family
- Effect of correlation on OLS coefficients
- Effect of correlation on SE of OLS coefficients
- Selection of languages for inclusion in analysis
- Choice of total phoneme inventory as response
- Alternative forms of data using specific phonemes
Summary

- Distinction between covariates and relationships
- Rel₁: family as a block factor $F_{ij}$
- Rel₂: geographic separation $D_{ij}$
- Spatial covariance models: Matérn’s Bessel family
- Effect of correlation on OLS coefficients
  - Effect of correlation on SE of OLS coefficients
- Selection of languages for inclusion in analysis
- Choice of total phoneme inventory as response
- Alternative forms of data using specific phonemes
Summary

- Distinction between covariates and relationships
- \( \text{Rel}_1 \): family as a block factor \( F_{ij} \)
- \( \text{Rel}_2 \): geographic separation \( D_{ij} \)
- Spatial covariance models: Matérn’s Bessel family
- Effect of correlation on OLS coefficients
- Effect of correlation on SE of OLS coefficients
- Selection of languages for inclusion in analysis
- Choice of total phoneme inventory as response
- Alternative forms of data using specific phonemes
Summary

- Distinction between covariates and relationships
- Rel₁: family as a block factor $F_{ij}$
- Rel₂: geographic separation $D_{ij}$
- Spatial covariance models: Matérn’s Bessel family
- Effect of correlation on OLS coefficients
- Effect of correlation on SE of OLS coefficients
- Selection of languages for inclusion in analysis
- Choice of total phoneme inventory as response
- Alternative forms of data using specific phonemes
Summary

- Distinction between covariates and relationships
- Rel₁: family as a block factor $F_{ij}$
- Rel₂: geographic separation $D_{ij}$
- Spatial covariance models: Matérn’s Bessel family
- Effect of correlation on OLS coefficients
- Effect of correlation on SE of OLS coefficients
- Selection of languages for inclusion in analysis
- Choice of total phoneme inventory as response
- Alternative forms of data using specific phonemes
Summary

- Distinction between covariates and relationships
- Rel\(_1\): family as a block factor \(F_{ij}\)
- Rel\(_2\): geographic separation \(D_{ij}\)
- Spatial covariance models: Matérn’s Bessel family
- Effect of correlation on OLS coefficients
- Effect of correlation on SE of OLS coefficients
- Selection of languages for inclusion in analysis
- Choice of total phoneme inventory as response
- Alternative forms of data using specific phonemes
Published commentary: *Science* Feb/Mar 2012

Four commentaries: Munich, Shanghai, Stanford, Rochester/MIT

Summary of criticisms

- WALIS data: vowels and tones exaggerated: (M, Sh, St, R)
  (1 tone ≡ 2.2 vowels ≡ 5.7 consonants)
- Inconsistency of pattern: cline in Africa but not elsewhere (St)
- Inconsistency of pattern for v/c/t: (Sh, St)
- BIC-4 allowance: (M)
- Neighbour influence (St)
- Global geography (R)
Published commentary: Science Feb/Mar 2012

Four commentaries: Munich, Shanghai, Stanford, Rochester/MIT

Summary of criticisms

- WALSD data: vowels and tones exaggerated: (M, Sh, St, R)
  (1 tone $\equiv$ 2.2 vowels $\equiv$ 5.7 consonants)

- Inconsistency of pattern: cline in Africa but not elsewhere (St)

- Inconsistency of pattern for v/c/t: (Sh, St)

- BIC-4 allowance: (M)

- Neighbour influence (St)

- Global geography (R)
Four commentaries: Munich, Shanghai, Stanford, Rochester/MIT

Summary of criticisms

- WALS data: vowels and tones exaggerated: (M, Sh, St, R)
  (1 tone ≡ 2.2 vowels ≡ 5.7 consonants)
- Inconsistency of pattern: cline in Africa but not elsewhere (St)
- Inconsistency of pattern for v/c/t: (Sh, St)
- BIC-4 allowance: (M)
- Neighbour influence (St)
- Global geography (R)
Published commentary: *Science* Feb/Mar 2012

Four commentaries: Munich, Shanghai, Stanford, Rochester/MIT

Summary of criticisms

- **WALS data: vowels and tones exaggerated**: (M, Sh, St, R)
  (1 tone $\equiv$ 2.2 vowels $\equiv$ 5.7 consonants)

- **Inconsistency of pattern**: cline in Africa but not elsewhere (St)

- **Inconsistency of pattern for v/c/t**: (Sh, St)

- **BIC-4 allowance**: (M)

- **Neighbour influence**: (St)

- **Global geography**: (R)
Published commentary: *Science* Feb/Mar 2012

Four commentaries: Munich, Shanghai, Stanford, Rochester/MIT

Summary of criticisms

- WALSS data: vowels and tones exaggerated: (M, Sh, St, R)
  (1 tone ≡ 2.2 vowels ≡ 5.7 consonants)
- Inconsistency of pattern: cline in Africa but not elsewhere (St)
- Inconsistency of pattern for v/c/t: (Sh, St)
- BIC-4 allowance: (M)
- Neighbour influence (St)
- Global geography (R)
Four commentaries: Munich, Shanghai, Stanford, Rochester/MIT

Summary of criticisms

- WALSD data: vowels and tones exaggerated: (M, Sh, St, R)
  (1 tone ≡ 2.2 vowels ≡ 5.7 consonants)
- Inconsistency of pattern: cline in Africa but not elsewhere (St)
- Inconsistency of pattern for v/c/t: (Sh, St)
- BIC-4 allowance: (M)
- Neighbour influence (St)
- Global geography (R)
http://www.en.uni-muenchen.de/news/newsarchiv/2012/2012_
http://www.stat.uchicago.edu/~pmcc/prelims/2011/