# Population size has a weak effect on phoneme inventory size

Muhammad Rehan (rehanmuh@g.ucla.edu)
University of California, Los Angeles

## Background

- A negative correlation alleged between population size and complexity (cf. Bromham 2025).
- Investigation of phonological complexity centered on phoneme inventory sizes.
- Does population size effect phoneme inventory size?
  - Yes: Hay and Bauer (2007), Atkinson (2011, 2012), Brentari et al. (2021), and Hou and Vos (2022)
  - No: Moran et al. (2012), Donohue and Nichols (2011) and Jäger (2025)

## Data

- 1,532 languages from PHOIBLE
- Population from LinguaMeta
- Trees from Glottolog

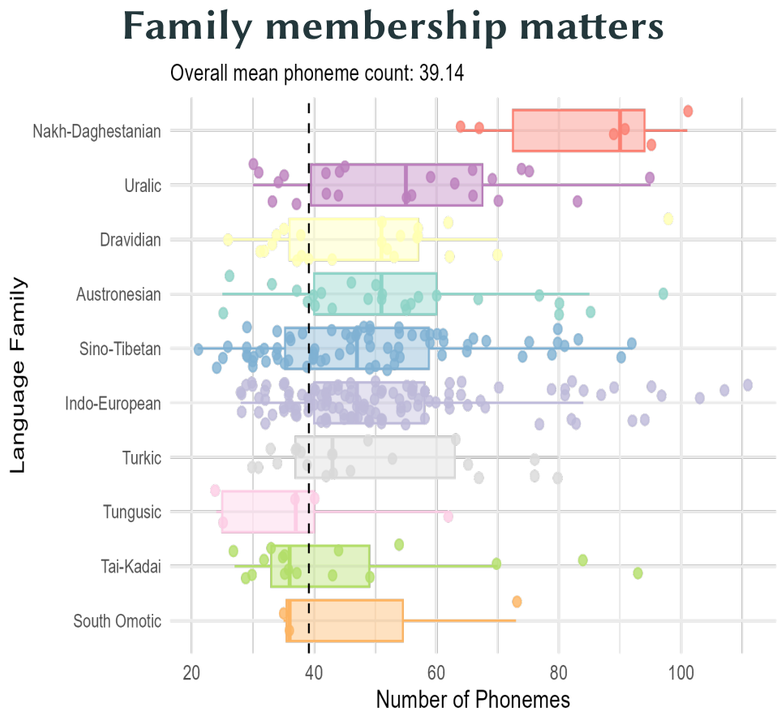


**Figure 1:** Phylogenetic dependencies in the distribution of phoneme inventory sizes

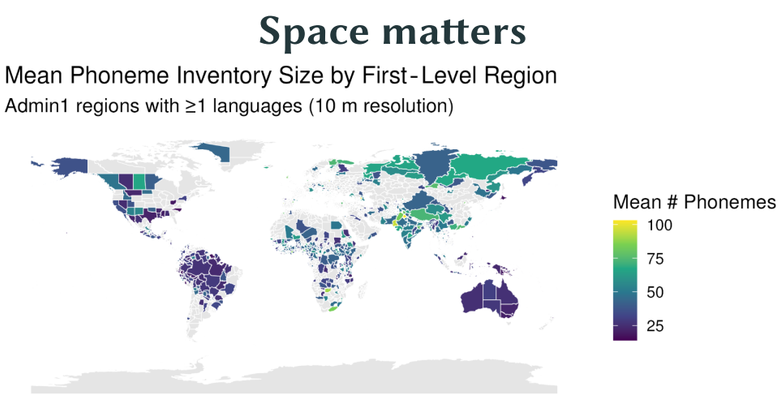


**Figure 2:** Spatial dependencies in the distribution of phoneme inventory sizes (Moran's $I$ = 0.185, $\mathbf{p} \approx 0$)

## Methods

- Bayesian hierarchical regression with weakly informative priors:

  $$y_i \sim t_3(\mu_i,\, \sigma)$$

- *Phylogenetic modeling*
  - Family-level intercepts:
    $\mu_i = \beta_0 + \beta\ \log(\text{pop}_i) + \alpha_{f[i,j]}$
  - + Phylogenetic regression:
    $\beta_0 + \beta\ \log(\text{pop}_i) + \alpha_{f[i,j]} + \gamma_{p[i,j]}$
    - Phylogenetic random intercepts constrained by a covariance matrix reflecting the tree structure
- *Spatial modeling*
  - + Intrinsic conditional auto-regressive (ICAR): $\mu_i = \beta_0 + \beta\ \log(\text{pop}_i) + \alpha_{f[i,j]} + \gamma_{p[i,j]} + \phi_i$
    - Each $\phi_i$ conditioned on neighbor's $\phi_j$
  - / Besag York Mollié 2: $\beta_0 + \beta\ \log(\text{pop}_i) + \alpha_{f[i,j]} + \gamma_{p[i,j]} + u_i$
    - Structured and unstructured spatial effects

## Results




**Figure 3:** Weak effect of population with genetic and spatial varying effects

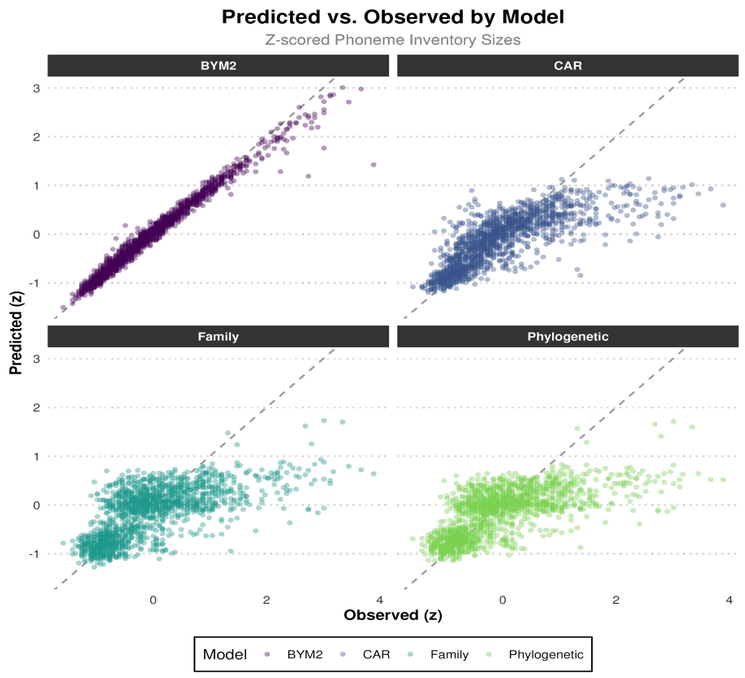


**Figure 4:** Spatial models have better posterior predictive performance

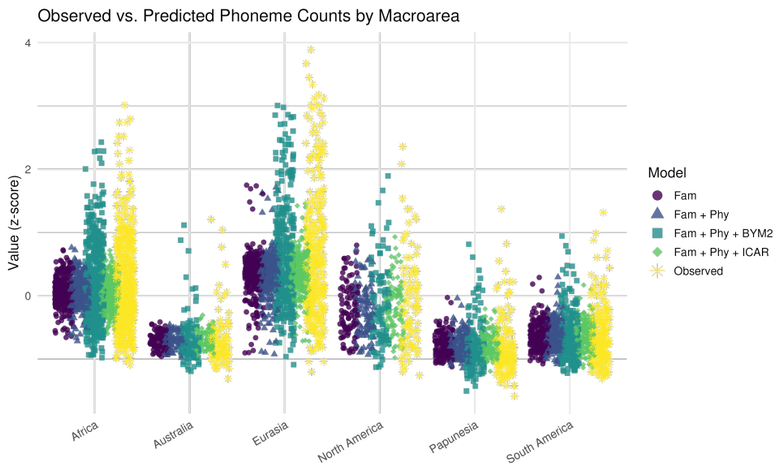


**Figure 5:** BYM2 model captures the spatial effects best

## Discussion & Conclusions

- *Weak effect* (0.11 S.D. increase in log(pop) per S.D. increase in phoneme inventory size) consistent with Jäger (2025) and Moran et al. (2012)
- Why the weak effect?
  - Measurement error & L2 speakers
  - Phoneme inventory sizes a crude measure of phonological complexity
  - Explosion in population sizes and colonial decimation of populations
  - Multicolinearity in predictor variables
- Using conditional auto-regressive models for spatial modeling:
  - Sparse covariance matrix makes inference quicker
  - Interpretable structured and unstructured spatial dependencies in the BYM2 model
  - No assumption of stationarity
  - Viable option for spatial modeling beside spatial GPs

## Acknowledgments & References

- Many thanks to David Goldstein and Steven Moran for their help.
- Fuller presentation with references available at the QR code.