

# Creating a Baseline to Evaluate Correlations Between Language and Environment

Catherine Arnett & Maho Takahashi (ccarnett@ucsd.edu, mtakahas@ucsd.edu)



UC San Diego

← Scan me for access to our data and code!

## Background

- Correlations between certain environmental features and linguistic features have been proposed through a series of statistical analyses
  - Ejective consonants ~ elevation (Everett, 2013)
  - Complex tone systems ~ humidity (Everett et al. 2015)
  - Consonant-vowel ratio ~ humidity (Everett, 2017)
- To further investigate these relationships we:
  - Replicated the findings** with a unified statistical method
  - Created a baseline** to compare the replicated correlations to

## Replication

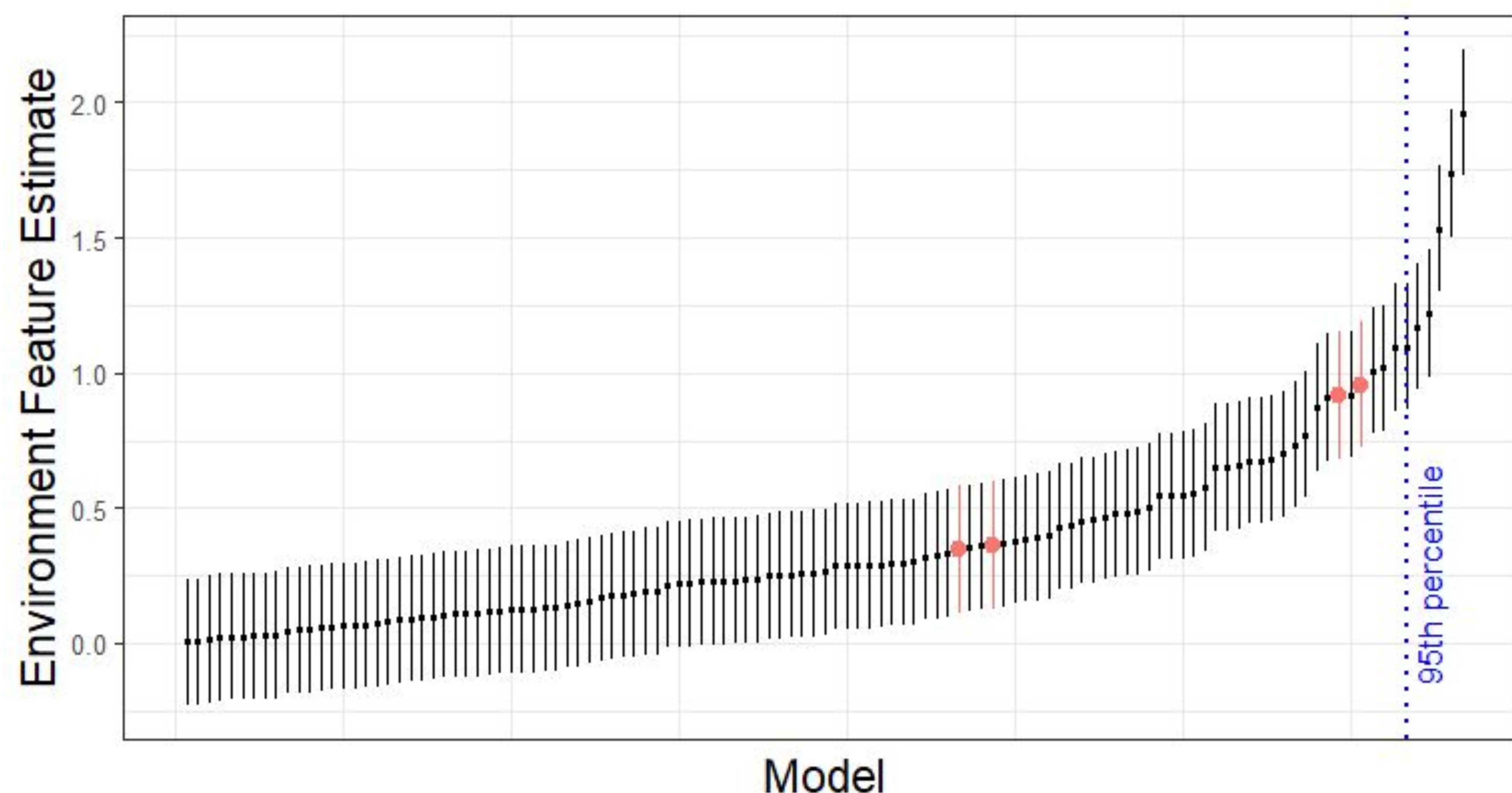
- Method:** The same data source was used but analyzed uniformly with a generalized linear mixed-effects model (Bates et al., 2014) with random effects of language family and region
- Results:**
  - Ejective consonants ~ elevation -> **not replicated** ( $\beta = 0.346$ , SE = 0.325,  $p = 0.286$ )
  - Complex tone systems ~ humidity -> **replicated** ( $\beta = 176.22$ , SE = 58.44,  $p < 0.01$ )
  - Consonant-vowel ratio ~ humidity -> **replicated** ( $\beta = 0.206$ , SE = -0.955,  $p < 0.01$ )

## Creating a Baseline

- Create a real distribution of effect sizes, rather than estimating a distribution (Trott & Bergen 2020)
- Methods**
  - wide range of variables that include data from at least 400 languages were extracted from WALS and AUTOTYP
  - same environmental data and statistical model (glmer) was used as in the replication
  - create a distribution of estimates (with SE), NHST for significance based on this distribution
- Results:**
  - many of the new relationships were correlated and statistically significant
  - the effect size of the tone ~ humidity model was not more extreme than 95% of the models in the distribution we created

## Language Feature ~ Environment Feature Models

All models



Everett replications are plotted in red

## Conclusions

- Out of the three correlations between environment and language features that were previously proposed, **we failed to replicate one**: ejective consonant ~ elevation
- We also revealed a series of such correlations where the former is highly unlikely to affect the latter (e.g., numeral classifier systems)
- In a real distribution of effect sizes of correlations between language features and environment features, the replication results are **not significant**.
- These results are consistent with the agent-based model, which will take into account other factors, e.g. social and cultural factors or genetic differences

## References

- Bates, D. et al. (2014) Fitting linear mixed-effects models using lme4. **Everett, C.** (2013). Evidence for direct geographic influences on linguistic sounds. *PloS One*. **Everett, C.** (2017). Languages in drier climates use fewer vowels. *Frontiers*. **Everett, C.** et al. (2015) Climate, vocal folds, and tonal languages. *PNAS* 112. **Roberts, S. G.** (2018) Robust, causal, and incremental approaches to investigating linguistic adaptation. *Frontiers* 9. **Trott, S., & Bergen, B.** (2020). Why do human languages have homophones?. *Cognition*, 205, 104449. **Urban, M. & Moran, S.** (2021) Altitude and the distributional typology of language structure. *PloS One* 16.