On the correlation between phoneme inventory size and elevation: the case of Dagestan

George Moroz

Linguistic Convergence Laboratory (HSE University)

September 2, 2021

presentation is available here: tinyurl.com/ydv9w364

Overview

Connection of language diversity to physical environment

Prediction of phonological inventory size

East Caucasian languages' data

On the correlation between phoneme inventory size and elevation

Discussion

Connection of language diversity to physical environment

There are a lot of work that connects language diversity to physical environment:

- famous and controversial [Everett 2013] that shows connection between altitude and distribution of ejectives and even propose a hypothesis of the reason of such a correlation
- in [Bentz et al. 2018] shows that phylogenetic signals correlates with different environmental factors (including altitude)
- a lot of works by Johanna Nichols and summary in [Urban 2020]

Overview

Connection of language diversity to physical environment

Prediction of phonological inventory size

East Caucasian languages' data

On the correlation between phoneme inventory size and elevation

Discussion

Prediction of phonological inventory size

There are several works [Hay and Bauer 2007; Atkinson 2011] that report connection of phoneme inventory sizes and language population sizes. There are several replies on those works [Maddieson et al. 2011; Donohue and Nichols 2011; Wichmann et al. 2011; Moran et al. 2012].

Overview

Connection of language diversity to physical environment

Prediction of phonological inventory size

East Caucasian languages' data

On the correlation between phoneme inventory size and elevation

Discussion

About East Caucasian phonology

Most East Caucasian (Nakh-Dagestanian) languages are located in the Republic of Dagestan, Chechen Republic and Republic of Ingushetia.

- phonology
  - [Kibrik and Kodzasov 1990]
  - The indigenous languages of the Caucasus: [Job and Smeets 1994], [Smeets 1994]
  - Yaziki mira: [Alekseev et al. 2001] and predecessor projects
  - [Grawunder 2017]
  - The Oxford Handbook of Languages of the Caucasus: [Beguš 2021], [Boris 2021a], [Boris 2021b]
  - Handbook of Caucasian Languages: [Koryakov and Maisak Fort]
  - and phonological parts in more general descriptions of East Caucasian languages
- historical-comparative phonetics
  - S. M. Gasanova, G. Kh. Ibragimov, P. T. Magomedova, U. A. Meylanova, B. B. Talibov, S. M. Khaydakov, A. E. Alekseev, Ju. D. Desheriev, B. K. Gigineishvili, D. S. Imnaishvili, J. Nichols, S. L.

Languages

- Nakh
  - Chechen
  - Ingush
  - Tsova-Tush (Batsbi)
- Tsezic (Didoic)
  - Tsez
  - Hinuq
  - Bezhta
  - Hunzib
  - Khwarshi-Inkhoqwari
- Avar–Andic
  - Avar
  - Andi
  - Botlikh
  - Godoberi
  - Chamalal
  - Karata
  - Tindi
  - Bagvalal
  - Akhvakh
- Dargwa
- Lak
- Lezgic
  - Archi
  - Tabasaran
  - Aghul
  - Lezgian
  - Udi
  - Budukh
  - Kryz
  - Rutul
  - Tsakhur
- Khinalug

Languages not present in [Kibrik and Kodzasov 1990]

- Nakh
  - Chechen
  - Ingush
  - Tsova-Tush (Batsbi)
- Tsezic (Didoic)
  - Tsez
  - Hinuq
  - Bezhta
  - Hunzib
  - Khwarshi-Inkhoqwari
- Avar–Andic
  - Avar
  - Andi
  - Botlikh
  - Godoberi
  - Karata
  - Chamalal
  - Tindi
  - Bagvalal
  - Akhvakh
- Dargwa
- Lak
- Lezgic
  - Archi
  - Tabasaran
  - Aghul
  - Lezgian
  - Udi
  - Budukh
  - Kryz
  - Rutul
  - Tsakhur
- Khinalug

Languages not present in [Kibrik and Kodzasov 1990]

- Andic, Bagvalal [Kibrik et al. 2001]
- Andic, Botlikh [Alexeyev and Verhees Fort]
- Andic, Godoberi [Moroz et al. Fort]
- Andic, Karata [Magomedbekova 1971]
- Dargwa, Mehweb [Moroz 2019]
- Nakh, Chechen [Nichols 1994a]
- Nakh, Ingush [Nichols 1994b]
- Nakh, Tsova-Tush [Holisky and Gagua 1994]

Phonological peculiarities

- East Caucasian consonant systems are more or less typical
  - except uvular and laryngeal consonants
  - except ejective consonants
  - except laterals in Avar-Andic and Tsezic branches
  - except labialization
  - except gemination
- East Caucasian vowel systems are typical

Phonological data

1905 × 9 table with the following columns:

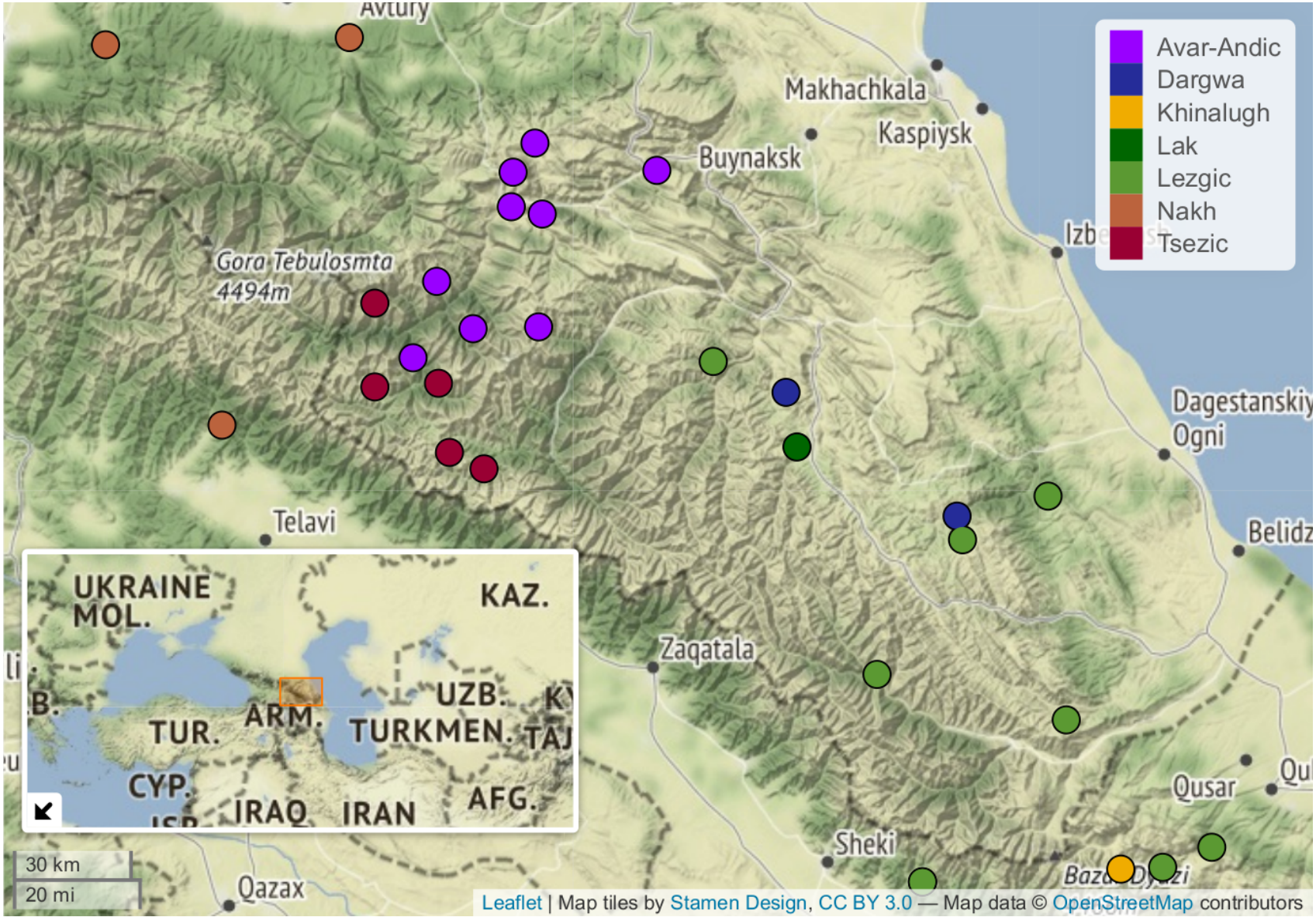
- `id`
- `branch`
- `language`
- `segment IPA`
- `segment sorce`
- `source`
- `sound type`
- `comments`
- `glottocode`

Phonological data

1905 × 9 table with the following columns:

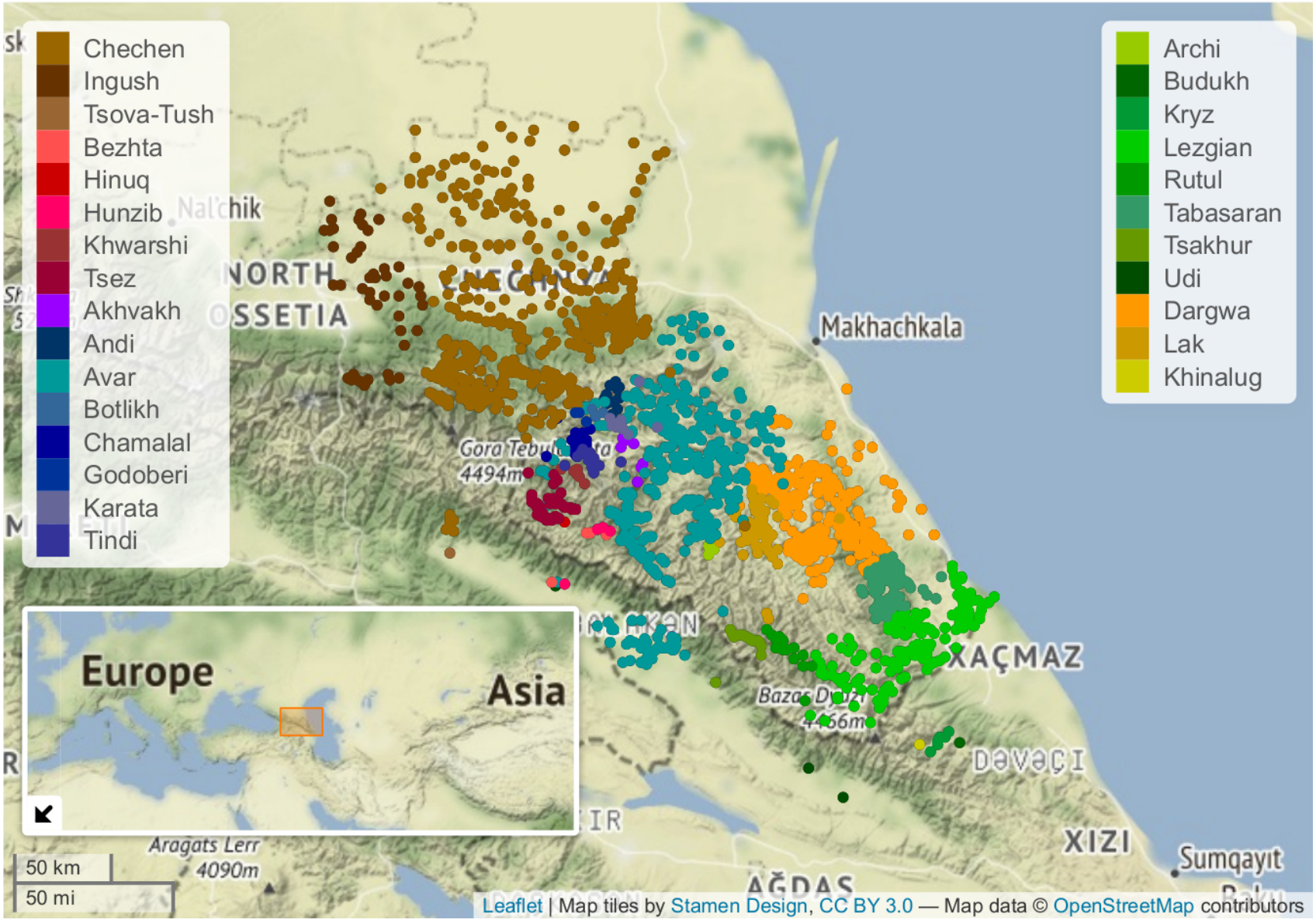
- `id` – 1407
- `branch` – Lezgic (possible values are Avar-Andic, Dargwa, Khinalugh, Lak, Lezgic, Nakh, Tsezic)
- `language` – Tsakhur
- `segment IPA` – tʃː
- `segment sorce` – c̆
- `source` – [Kibrik, Kodzasov 1990: 343-344]
- `sound type` – consonant (possible values are consonant, vowel, diphthong)
- `comments` – rare sound
- `glottocode` – tsak1249

Geographic data [Moroz and Verhees 2020]



created with `lingtypology` [Moroz 2017]

Geographic Data [Moroz and Verhees 2020]



created with `lingtypology` [Moroz 2017]

Elevation Data

Elevation for this data was retrieved using the GLOBE digital elevation model from the National Centers for Environmental Information (after [Urban and Moran 2021]).

Overview

Connection of language diversity to physical environment

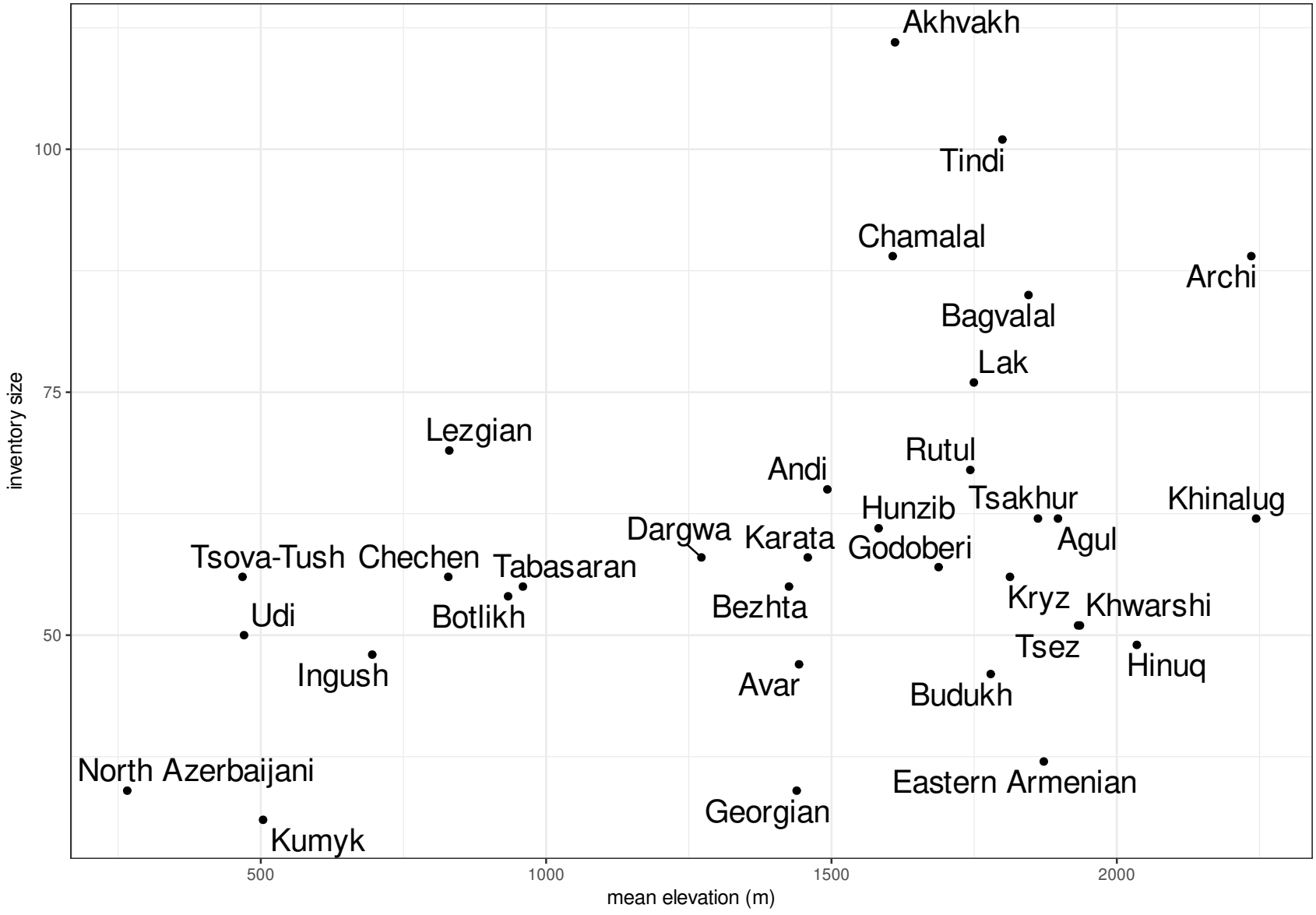
Prediction of phonological inventory size

East Caucasian languages' data

On the correlation between phoneme inventory size and elevation

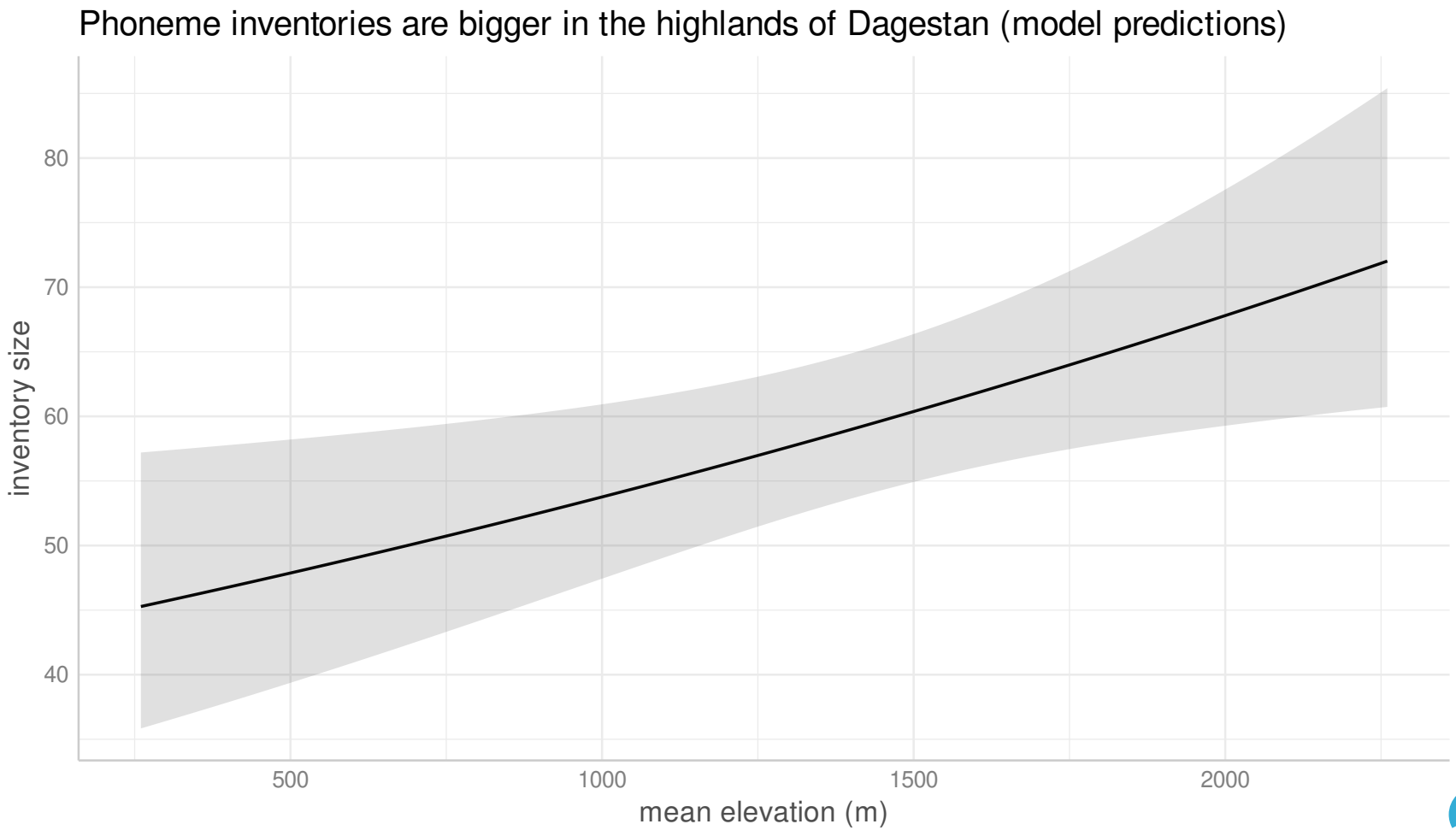
Discussion

East Caucasian data



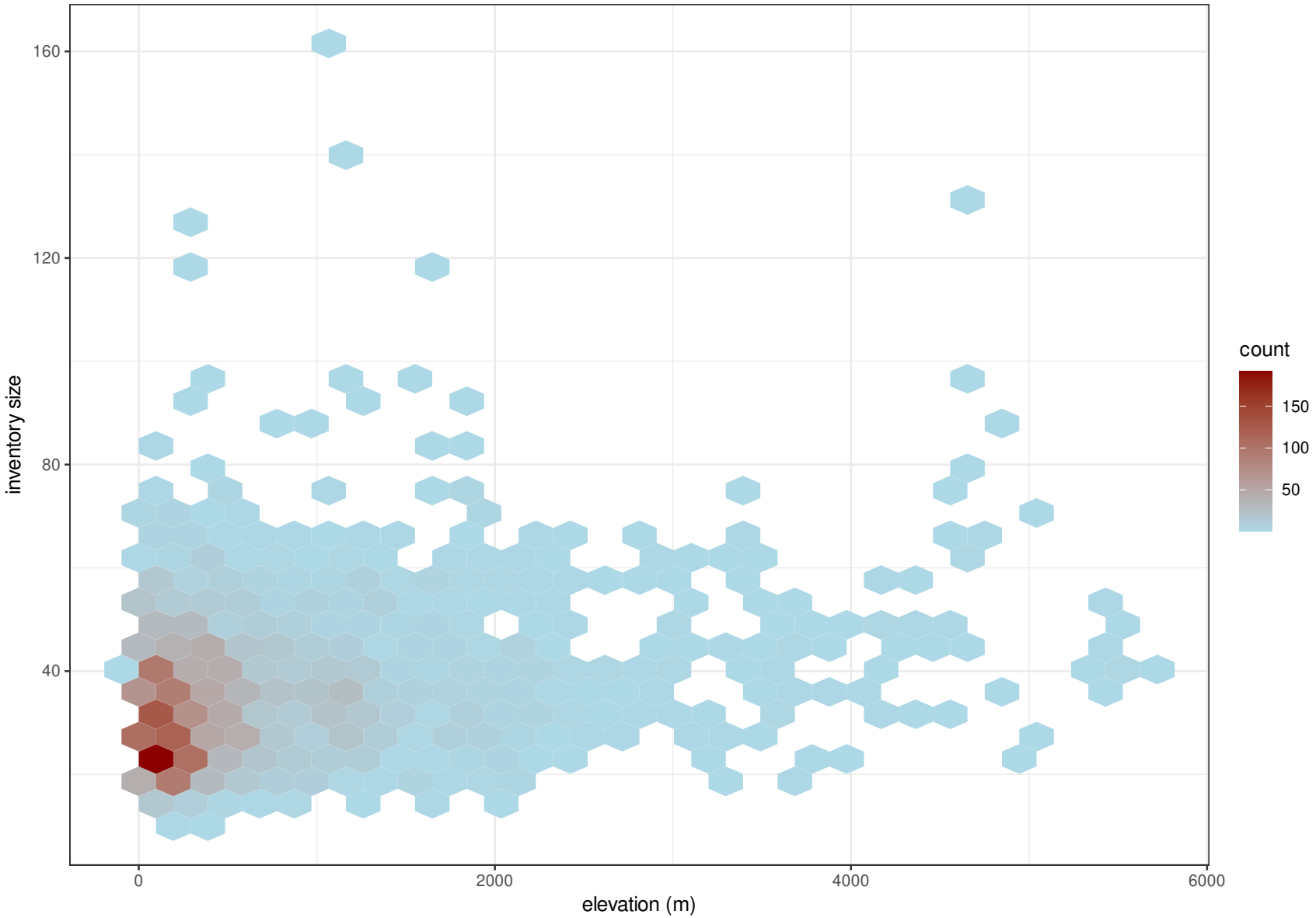
Negative binomial regression model (East Caucasian languages)

$$\text{number of phonemes} = \exp\left[3.7525 + 0.0002 \times \text{elevation}\right]$$



based on the sample of East Caucasian languages

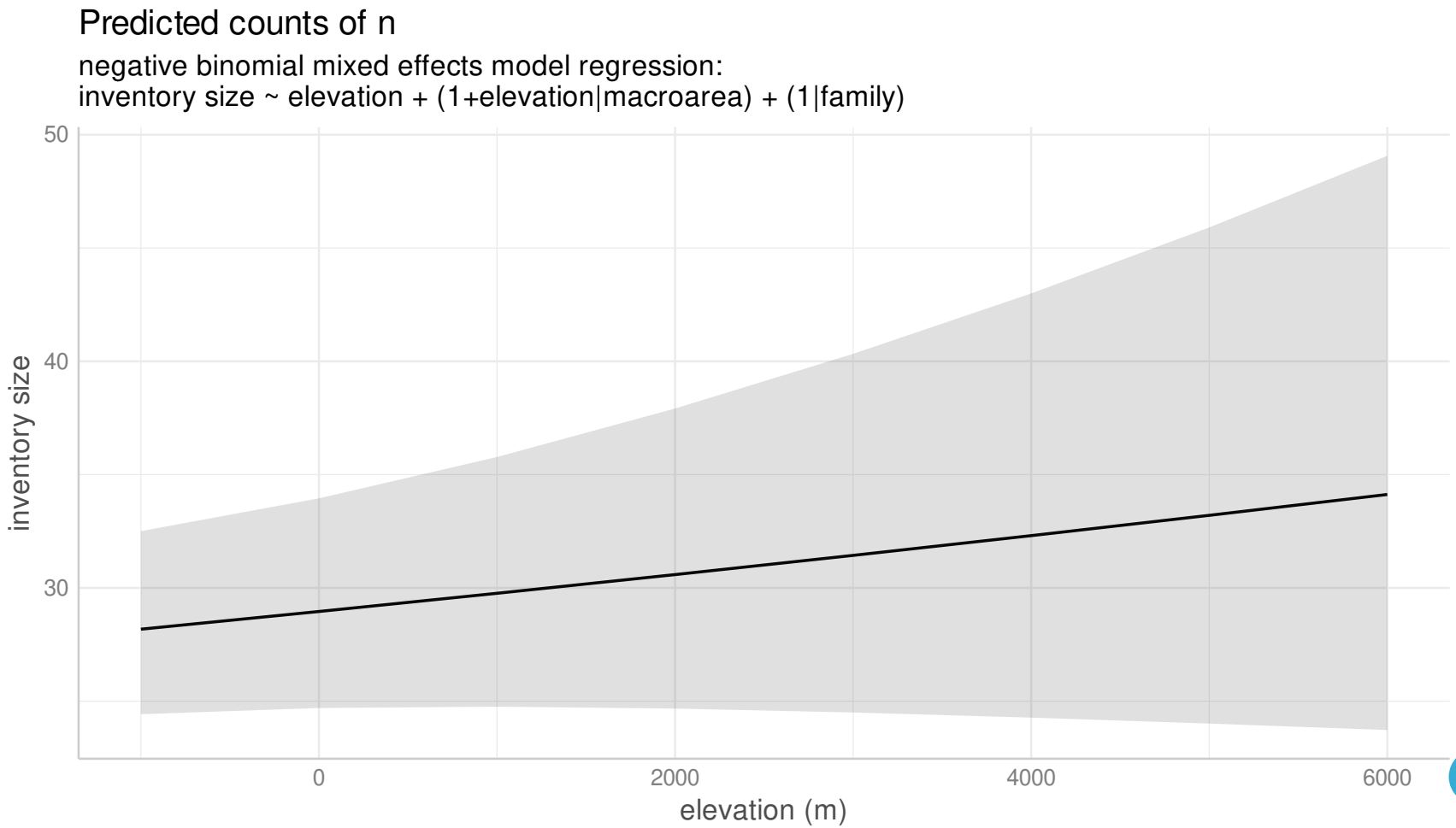
PHOIBLE data [Moran et al. 2014]



Negative binomial regression model (PHOIBLE)

elevation is not statistically significant

$$\text{number of phonemes} = \exp\left[3.3659 + 0.0273 \times \text{elevation}\right]$$



Overview

Connection of language diversity to physical environment

Prediction of phonological inventory size

East Caucasian languages' data

On the correlation between phoneme inventory size and elevation

Discussion

False? Results:

As a result, I discovered a local effect in Dagestan that should be compared with other mountain areas (see e. g. affricate-rich languages in the Eastern Himalayan and in the Hindukush linguistic areas [Nikolaev and Grossman 2018]). If they show similar results, it would confirm Johanna Nichols' [2013] hypothesis that mountain areas provide isolation and "isolation favors complexity".

False? Results:

- However mountains can be different: see the Altai argument in [Urban 2020];
- It make sense to think more about other obstacles (e. g. rivers and travel costs, see [Allassonnière-Tang, Her this conference], [Koile et al. forthcoming]);
- Ignoring social, cultural and historical changes in language comunities leads to oversimplification: see about multilingualism patterns and their collapse in Dagestan [Dobrushina et al. 2019] and Urban's note during this conference;
- Pure altitude values are useless: we can't take this result and try to predict something in e. g. South America;
- I also wonder, what is the difference between Mountain area patterns and just Isolation patterns. Is it possible to treat Mountains as a complicated isolation situation?

To future researcher

- I personally do not believe that the results that I obtained are any different from spurious correlations (like the amount of pirates and Earth's temperature).

To future researcher

- I personally do not believe that the results that I obtained are any different from spurious correlations (like the amount of pirates and Earth's temperature).
- My model marks elevation as a statistically significant predictor of number of segments in a language, but it is not the same about pure vowel nor consonant inventory size.

To future researcher

- I personally do not believe that the results that I obtained are any different from spurious correlations (like the amount of pirates and Earth's temperature).
- My model marks elevation as a statistically significant predictor of number of segments in a language, but it is not the same about pure vowel nor consonant inventory size.
- Despite the fact that Johanna Nichols has proposed general model for language dynamics in mountain areas, I think we need more prediction oriented theory of Mountain linguistics.

To future researcher

- I personally do not believe that the results that I obtained are any different from spurious correlations (like the amount of pirates and Earth's temperature).
- My model marks elevation as a statistically significant predictor of number of segments in a language, but it is not the same about pure vowel nor consonant inventory size.
- Despite the fact that Johanna Nichols has proposed general model for language dynamics in mountain areas, I think we need more prediction oriented theory of Mountain linguistics.
- … and research preregistration!

Thank you for your attention!

References

Alekseev, M. E., Starostin, S. A., Klimov, G. A., and Testelets, Y. G. (2001). *Yazyki mira. Kavkazskiye yaziki.* Academia.

Alexeyev, M. and Verhees, S. (Fort). Botlikh. In Koryakov, Y. and Maisak, T., editors, *Handbook of Caucasian Languages*, volume 2. De Gruyter Mouton, Berlin/New York.

Atkinson, Q. D. (2011). Phonemic diversity supports a serial founder effect model of language expansion from africa. *Science*, 332(6027):346–349.

Beguš, G. (2021). Segmental phonetics and phonology. In Polinsky, M., editor, *The Oxford Handbook of Languages of the Caucasus*, pages 1–40. Language Science Press.

References

Bentz, C., Dediu, D., Verkerk, A., and Jäger, G. (2018). The evolution of language families is shaped by the environment beyond neutral drift. *Nature Human Behaviour*, 2(11):816–821.

Boris, L. (2021a). Tone and intonation in languages of the caucasus. In Polinsky, M., editor, *The Oxford Handbook of Languages of the Caucasus*, pages 1–30. Oxford University Press.

Boris, L. (2021b). Word stress in the languages of the caucasus. In Polinsky, M., editor, *The Oxford Handbook of Languages of the Caucasus*, pages 1–30. Language Science Press.

Dobrushina, N., Kozhukhar, A., and Moroz, G. (2019). Gendered multilingualism in highland daghestan: story of a loss. *Journal of Multilingual and Multicultural Development*, 40(2):115–132.

References

Donohue, M. and Nichols, J. (2011). Does phoneme inventory size correlate with population size? *Linguistic Typology*, 15(2):161–170.

Everett, C. (2013). Evidence for direct geographic influences on linguistic sounds: The case of ejectives. *PloS one*, 8(6):e65275.

Grawunder, S. (2017). The caucasus. In Hickey, R., editor, *The Oxford Handbook of Languages of the Caucasus*, pages 356–395. Cambridge University Press.

Hay, J. and Bauer, L. (2007). Phoneme inventory size and population size. *Language*, 83(2):388–400.

Holisky, D. A. and Gagua, R. (1994). Tsova-Tush (Batsbi). *The indigenous languages of the Caucasus*, 4(part 2):147–212.

References

Job, D. M. and Smeets, R., editors (1994). *The indigenous languages of the Caucasus*, volume 3. Caravan Books.

Kibrik, A. E. and Kodzasov, S. V. (1990). *Sopostavitelnoye izucheniye dagestanskix yazykov: Imya. Fonetika*, volume 2. Moskovskij Gosudarstvennyj Universitet.

Kibrik, A. E., Tatevosov, S. G., Lyutikova, E. A., and Kazenin, K. I. (2001). *Bagvalinskiy yazik: Grammatika. Teksti. Slovari.* Naslediye, Moscow.

Koryakov, Y. and Maisak, T., editors (Fort). *Handbook of Caucasian Languages*, volume 2. De Gruyter Mouton, Berlin/New York.

Maddieson, I., Bhattacharya, T., Smith, D. E., and Croft, W. (2011). Geographical distribution of phonological complexity.

References

Magomedbekova, Z. M. (1971). *Karatinskiy yazyk: grammaticheskiy analiz, teksty, slovar'*. Metsniereba, Tbilisi.

Moran, S., McCloy, D., and Wright, R. (2012). Revisiting population size vs. phoneme inventory size. *Language*, pages 877–893.

Moran, S., McCloy, D., and Wright, R., editors (2014). *PHOIBLE Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig.

Moroz, G. (2017). *lingtypology: easy mapping for Linguistic Typology*.

Moroz, G. (2019). Phonology of Mehweb. In Daniel, M., Dobrushina, N., and Ganenkov, D., editors, *The Mehweb language: Essays on phonology, morphology and syntax*, pages 17–39. Language Science Press.

References

Moroz, G., Naccarato, C., and Verhees, S. (Fort). Godoberi. In Koryakov, Y. and Maisak, T., editors, *Handbook of Caucasian Languages*, volume 2. De Gruyter Mouton, Berlin/New York.

Moroz, G. and Verhees, S. (2020). East Caucasian villages dataset.

Nichols, J. (1994a). Chechen. *The indigenous languages of the Caucasus*, 4(part 2):3–77.

Nichols, J. (1994b). Ingush. *The indigenous languages of the Caucasus*, 4(part 2):81–145.

Nichols, J. (2013). The vertical archipelago: Adding the third dimension to linguistic geography. *Space in language and linguistics: Geographical, interactional, and cognitive perspectives*, pages 38–60.

References

Nikolaev, D. and Grossman, E. (2018). Areal sound change and the distributional typology of affricate richness in eurasia. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"*, 42(3):562–599.

Smeets, R., editor (1994). *The indigenous languages of the Caucasus*, volume 4. Caravan Books.

Urban, M. (2020). Mountain linguistics. *Language and Linguistics Compass*, 14(9):e12393.

Urban, M. and Moran, S. (2021). Altitude and the distributional typology of language structure: Ejectives and beyond. *Plos one*, 16(2):e0245522.

Wichmann, S., Rama, T., and Holman, E. W. (2011). Phonological diversity, word length, and population sizes across languages: The asjp evidence. *Linguistic Typology*, 15(2):177–198.