

XXXI KKZMBM

Kościelisko, 15th–19th September 2026

KRONECKER-STRUCTURE COVARIANCE ESTIMATION IN THE PHYLOGENETIC SMALL-SAMPLE-HIGH-DIMENSION CASE

Johan Alenlöv¹, Krzysztof Bartoszek¹, Jesualdo Fuentes-González², Wang Liao¹, Monika Mokrzycka³, Joao Muñoz-Durán⁴, Marcin Okrój⁵, Jolanta Pielaszkiewicz¹

¹ The Division of Statistics and Machine Learning, Department of Computer and Information Science, Linköping University, 581 83, Linköping, Sweden,

² Biomolecular Sciences Institute, Florida International University, Miami, FL 33199, USA,

³ Institute of Plant Genetics, Polish Academy of Sciences, Poznań, Poland

⁴ Department of Biology, National University of Colombia, CR. 30 # 45-03, Bogotá D.C., Colombia

⁵ Intercollegiate Faculty of Biotechnology, University of Gdańsk and Medical University of Gdańsk, Dębinki 1 street, 80-211 Gdańsk, Poland

ABSTRACT

Phylogenetic comparative methods (PCMs) are concerned with modeling the evolution of traits on a between-species level. This implies that they need to take into account the phylogeny describing the evolutionary history of the species. Mathematically speaking, the considered models are Branching Markov Processes (BMPs), i.e., a Markov process evolves as a “normal one” until it reaches a branching event and there it splits into independent copies with starting points equaling the value of the ancestral process at the split. For Gaussian models of evolution, this phylogeny translates into a non-diagonal between observations covariance matrix. If the considered traits are “typical” morphological ones, like body weight, length, width, e.t.c., then direct analytical or numerical optimization of the likelihood function can be employed. However, if the traits are gene expression or coordinate-based morphological characters, then we can very easily have more unknowns (corresponding to different trait dimensions), than observations (species).

For the small-sample-high-dimension case Statistics offer alternative techniques, e.g., based on regularization. We explore here one such approach in the phylogenetic setting. We assume that the multivariate traits evolve as a Brownian motion and then the between-species-between-traits variance-covariance matrix has a Kronecker structure. In such a setting newly developed regularized structured approximations based on improved entropy loss [3] can provide more stable and comparable covariance representations.

By simulations we explore these techniques in the phylogenetic setting finding that their performance is influenced by sample size, the strength of regularization, and the degree of separation among candidate structures [5]. We illustrate the approach with a lung cancer gene expression dataset [1] and phylogenetically structured [2] canid morphological measurements [4]. For the lung cancer clone-expression data, both global gene-set and the GO-module analyses favour a diagonal diffusion matrix, indicating that gene expression variation, after taking into account between clone dependencies is domination by gene-specific variation.

ACKNOWLEDGMENTS

KB is supported by the Swedish Research Council (Vetenskapsrådet) project grant nr. 2025–03823 and a travel grant from Stiftelsen Lars Hiertas Minne (FO2025–0432).

REFERENCES

- [1] A. Felberg, M. Bieńkowski, T. Stokowy, K. Myszczyński, Z. Polakiewicz, K. Kitowska, R. Sądej, F. Mohlin, A. Kuźniowska, D. Kowalska, G. Stasiłóć, I. Jongerius, R. Spaapen, M. Mesa-Guzman, L. M. Montuenga, A. M. Blom, R. Pio, and M. Okrój: *Elevated expression of complement factor I in lung cancer cells associates with shorter survival—potentially via non-canonical mechanism*, Translational Research **269** (2024), 1–13.
- [2] J. Fuentes-González and J. Muñon-Duran: *Comportamiento social y preferencias alimentarias en la evolución de los cánidos actuales*, Revista Mexicana de Biodiversidad **88** (2017), 192–206.
- [3] M. Mokrzycka, J. Pielaszkiewicz, J. Alenlöv, and A. Tillander: *Covariance structures approximation in doubly multivariate models under high-dimensionality via improved Kullback-Leibler divergence*, under review.
- [4] J. Muñon-Duran and J. Fuentes-González: *Evolución de la socialidad, estrategias alimentarias y anatomía craneo-dental en la subfamilia Caninae*, Acta Biológica Colombiana **17** (2012), 173–200.
- [5] W. Liao: *Kronecker-structure covariance estimation in the small-sample-high-dimension case with phylogenetic applications*, 2026, Master thesis in Statistics, Division of Statistics and Machine Learning, Department of Computer Science, Linköping University.