Onomastic modelling for genetics

María José Ginzo Villamayor[1], *Rosa M. Crujeiras*[2]

[1]mariajose.ginzo@usc.es, [2]rosa.crujeiras@usc.es, Department of Statistics and Operations Research, University of Santiago de Compostela

The analysis of relations between surnames and genetic characteristics date back to the late 19th century, with the study of surnames to calculate the probability of first cousin marriages in Britain. Indeed, individuals who share location specific surnames are also likely to share a number of linguistic, genetic, historical and social characteristics as well as common ancestry. Usually, this analysis are based on isonymy measures. Isonymy refers to the possession of the same surname, being a premise in genetics that individuals with the same surname are more likely to share the same family lineage, so isonymy indicates biological relation. From an analogy with genetics, as it happens for alleles, drift of surnames is proportional to time, and then small values of isonymy suggests recent immigration or settlement. Isonymy can be also extended as a measure of population similarities between groups: isonymy between two regions. For instance, Cheshire et al. (2010) identified a strong relationship between surname regions and geographic locations in Great Britain. Other different measures of the isonymic distance between a pair of locations can be derived from Isonymy, for instance, the Lasker, Euclidean and Nei's distances. Euclidean and Nei's distances have been developed for purely genetic data, but they can be applied to the frequencies of surnames, such as done by Mikerezi *et al.* (2013).

This work will be focused on the introduction of new statistical methods for data processing and modeling in geolinguistics, specifically, on surnames in Galicia. The main objective is modeling spatial and spatio-temporal surname patterns in this region. The different research lines within the onomastic context have not taken into account the spatial and spatio-temporal dimension of the surnames evolution. By fixing administrative regions, for example, municipalities, spatial and spatio-temporal methods for count data can be applied in this setting. These methods will be useful for modeling evolution patterns for surnames. Hierarchical modeling, through BYM method (see Besag et al. (1991) and Rue and Held (2005) for a general reference in this type of models) will be used to meet this goal. In order to fit this type of models in practice, INLA (Integrated Nested Laplace Approximation) proposed by Rue et al. (2009), will be explored. The developed methods will be applied to th e data of surnames in Galicia, provided by the Galician Statistics Institute (census 2011).

Keywords: Hierarchical models; isonymy measures; surname patterns.