MSc Project Outline: Creating a predicted typology for cattle herds based on routine, pre-collected data using three sample areas of the UK

This project was developed using data from the Animal and Plant Health Agency. In Great Britain, it is a legal requirement for all cattle to be identified and registered on databases collated by the Animal and Plant Health Agency. Information is collected on both the individual animals and the herd that the animal is registered to, as part of the surveillance for bovine tuberculosis (Defra, 2016). However, recently, it has been discovered that that summary information collected on the herds may not be update (Animal and Plant Health Agency, unpublished data). The objective of this study was to identify herd types, based on the data provided by registered animals within a herd. A multiple correspondence analysis was used to identify the variables which provided the most contribution to the variance between herds, while a cluster analysis was used to determine homogenous groups of cattle.

This is the first work that aims to create a predicted typology for UK cattle herds, using routine, pre-collected data. In the UK, farms are generally classified based on the enterprise that generates the most income (Anderson et al., 2006). However, farm survey data can be used to characterise farm types. Published research for farm typification is limited (Gelasakis et al., 2012) but cattle clusters have been identified, based primarily on herd structure, management practices and land use, in areas such as Morocco (Srairi, and Lyoubi, 2003), the Dehesa (Escribano et al., 2014), Canada (Alemu et al., 2016) and Uganda (Muribu et al., 2007). Since herd type is a risk factor for bTb, the information gained from this study could potentially be used in order to help better target tuberculosis control.

Three variables were used in the analysis (age, sex and breed of animal). These variables were chosen as they are routinely collected and readily available on the databases. A preliminary literature review and consultation of expert opinion allowed creation of a list of what each breed of cattle is primarily used for in the UK, and of categories of age and sex in order to use the analysis. Risk factors for bovine tuberculosis are multiple, and include genetics, nutritional status, and age at the animal level (Broughan et al., 2016) and herd size, herd type and cattle movements at the herd level (Broughan et al., 2016). The epidemiology for bovine tuberculosis is complicated, with many confounders, for example dairy herds tend to be larger than beef herds (Broughan et al., 2016). As a result, it is important to correctly classify herd type. Potentially, correct classification of herd type could help in implementation of control programs.

The aim of the MCA was to explore the relationship between the three variables chosen for the analysis. It is a descriptive and exploratory technique (Panagiotakos and Pitsavos, 2004), designed to represent and model datasets as "clouds" of points in a multidimensional Euclidean space to uncover relationships within the data (Costa et al., 2013), and it results in a graphical map of interdependencies among variables (Hwang et al., 2006). Points that appear close to each other in the graph are thought to be closely related to each other, while points that appear further apart are less related. The MCA also calculates the mean variability in the data, and shows which categories had the largest contribution to the variance. Discrimination measures reflect the contribution of each variable to the overall variance in each dimension (Tascilar et al., 2014). Categories that explain more than the average variance are thought to be useful discrimination measures (Meesters, 2009). Sex had the largest contribution to the data, followed by percentage of primary purpose beef and dairy breed cattle.

.

The cluster analysis gave estimated cluster centres of the percentage of animals that were likely to be found in each grouping. The cluster analysis aims to maximise the difference between each grouping of animals. The clusters identified were a dairy type, dairy breed beef type, beef finisher, beef suckler, a more mixed beef production cluster (high percentage female animals, of beef breeds, but a more even age distribution) and a combination cluster (possibly of suckler and finisher animals).

In conclusion, this work provides an initial insight into herd types in Great Britain. Further work in this area would take account of cattle movement, and would find animals that entered/left the herd using the Cattle Tracing System and include them in the herd type. The work also assumes that the animal level information is more up to date than the summary information provided at the holding level, however since according to European Law, up-to-date records for each animal must be kept of its unique number, breed, sex, date of birth, movements and date of death (DEFRA, 2016) it is likely that information provided on actual animals registered in a herd is more up to date due to these regulations.

References

Alemu, A.W., Amino, B.D., Bittman, S., MacDonald, D. and Ominski, K.D., 2016, A typological characterisation of Canadian beef cattle farms based on a producer survey, *Candian Journal o Animal Science,* 96:187-202

Anderson, E., Verhoog, A.D., Elberson, B.S., Godeschalk, F.E. and Koole, B., 2006, A multidimensional farming system typology, *SEAMLESS Report,* 1:1-30

Animal and Plant Health Agency, unpublished data, Randomised Cluster Trials in Wales

Broughan, J.M., Judge, J., Ely, E., Delahay, R.J., Wilson, G., Clifton-Hadley, R.S., Goodchild, A.V., Bishop H., Parry J.E. and Downs, S.H., 2016, A review of risk factors for bovine tuberculosis infection in cattle in the UK and Ireland, *Epidemiology and Infection*, 1-28

Costa, P.S., Santos, N.C., Cunha, P., Cotter, J. and Sousa, N., 2013, The Use of Multiple Correspondence Analysis to Explore Associations between Categories of Qualitative Variables in Healthy Ageing, *Journal of Ageing Research,* 2013:1-12

DEFRA (Department for Environment, Food and Rural Affairs, 2016, Cattle identification, registration and movement, viewed 4th August 2016, < https://www.gov.uk/guidance/animal-identification-movement-and-tracing-regulations>

Escribaco, A.J., Gaspar, P., Mesias, F.J., Escribano, M., 2014, Cattle Production in Agroforestry System: An analysis on the role of intensification and dependence of subsidies, In *Book of Abstracts, 2014*

Gelasakis, A.I., Valergakis, G.E., Arsenos, G. and Banos, G., 2012, Description and typology of intensive Chios dairy sheep farms in Greece, *Journal of Dairy Science,* 95:3070-3079

Hwang, H., Dillon, W.R. and Takane, Y., 2006, An Extension of Multiple Correspondence Analysis for Identifying Heterogeneous Subgroups of Respondents, *Psychometrika,* 71(1):161-171

Meesters, J., 2009, The Meaning of Activities in the Dwelling and Residential Environment, IOS press

Muburi, S.L., Tenywa, J.S., Halberg, N., Romney, D., Nanyeenya, W., Baltenweek, I., and Staal, S., 2007, Categorisation of dairy production systems: a strategy for targeted development of the systems in Uganda, *Livestock Research for Rural Development,* 19(7):

Panagio, D.B. and Pitsavos, C., 2004, Interpretation of Epidemiological Data Using Multiple Correspondence Analysis and Log-linear Models, *Journal of Data Science,* 2:75-86

Sraïri, M.T., Kiade, N., Lyoubi, R., Messad, S. and Faye, B., 2009, A comparison of dairy cattle systems in an irrigated perimeter and in a suburban region: case study from Morocco. *Tropical Animal Health and Production*, *41*(5):835-843.

Tascilar, K., Melikogo, M., Ugurlu, S., Sut, N., Caglar, E., and Yazici, H., 2014, Vascular involvement in Behçet's syndrome: a retrospective analysis of associations and the time course, *Rheumatology*, *53*(11):2018-2022