

## Research summary

Academic supervision for the postdoctoral project will be provided by Dr. Patrick Brown at the Centre for Global Health Research (CGHR) and the Department of Statistical Sciences (DSS) at the University of Toronto. Dr. Brown has worked extensively in the areas of spatio-temporal modelling (Brown et al. 2000; Brown et al. 2001), statistical computing in spatial statistics (Brown 2015; Brown 2016), and kernel smoothing methods for spatially aggregated data (Nguyen, Brown, and Stafford 2012; Lee et al. 2017). The DSS houses a strong spatio-temporal research group, which in addition to Dr. Brown includes Prof. Jamie Stafford, Dr. Daniel Simpson, and a number of graduate students, research associates, and postdocs. The collaborative research projects will be led by Prof. Prabhat Jha from CGHR and the University of Toronto's Dalla Lana School of Public Health.

The methodological project will concern using data from health surveys (i.e. Jha 2014) with the Generalized Linear Geostatistical Model (or GLGM, see P. J. Diggle and Ribeiro 2006). Writing  $Y_{ijk}$  as the outcome for the  $k$ th individual in sampling unit  $i$  (located  $s_i$ ) at time  $t_j$ , a spatio-temporal GLGM might be written as:

$$\begin{aligned} Y_{ijk} &\sim \text{Bernoulli}(\rho_{ijk}) \\ g(\rho_{ijk}) &= X_{ijk}\beta + U(s_i) + V(t_j) + W(s_i, t_j) \\ V(t) &\sim \text{time series model} \\ U(s) &\sim \text{Gaussian random field} \\ W(s, t) &\sim \text{Spatio-temporal process} \end{aligned}$$

In the context of the Indian mortality data from the Million Death Study, there are roughly 8000 unique locations  $s_i$  and 15 years of data (2001 to 2016 and counting). The temporal component  $V(t)$  reflects the observed drop in mortality from most causes since 2001. Inclusion of the spatial process  $U(s)$  stems from knowledge that many health outcomes exhibit substantial regional variation - for example, deaths due to stroke are more common in the north-east. The spatio-interaction  $W(s, t)$  is the key to the research question, namely identifying which areas (if any) have experienced exceptionally slow or rapid changes in mortality in relation to the remainder of the country.

Issues to be addressed in the project include the following:

- The high dimensionality of the random effects requires a reduced rank or sparse matrix approximation (or both, see Lindgren, Rue, and Lindstrom 2011) of the covariance functions.
- Penalized Complexity prior distributions (Simpson et al. 2017) should reflect the belief that  $V$  is known to be important,  $U$  somewhat less so, and  $W$  is often negligible or absent.
- The model might require age, period, and cohort effects, although the latter is not present above.
- The locations  $s_i$  are determined by a sampling design and sampling weights could be relevant.

The direction from which these problems will be tackled will depend on the strengths and interests of the successful candidate, and will likely involve one of MCMC, INLA, local likelihood, or ADMB/TMB. To discuss the project further, please contact Patrick Brown at [patrick.brown@utoronto.ca](mailto:patrick.brown@utoronto.ca).

## References

Brown, Patrick E. 2015. "Model-Based Geostatistics the Easy Way." *Journal of Statistical Software* 63. <http://www.jstatsoft.org/v63/i12>.

———. 2016. "Maps, Coordinate Reference Systems and Visualising Geographic Data with mapmisc." *R-Journal* 8 (1): 64–91. <https://journal.r-project.org/archive/2016-1/brown.pdf>.

Brown, Patrick E, P J Diggle, Martin E Lord, and Peter C Young. 2001. "Space-Time Calibration of Radar Rainfall Data." *Journal of the Royal Statistical Society, Series C Applied Statistics* 50 (2): 221–41. doi:10.1111/1467-9876.00230.

Brown, Patrick E, Kjetil F Karesen, Gareth O Roberts, and Stefano Tonellato. 2000. "Blur-Generated Non-Separable Space-Time Models." *Journal of the Royal Statistical Society, Series B Statistical Methodology* 62 (4): 847–60. doi:10.1111/1467-9868.00269.

Diggle, Peter J, and Paulo Justiniano Ribeiro. 2006. *Model-Based Geostatistics*. New York: Springer-Verlag.

Jha, Prabhat. 2014. "Reliable Direct Measurement of Causes of Death in Low- and Middle-Income Countries." *BMC Medicine* 12 (1): 19. doi:10.1186/1741-7015-12-19.

Lee, Jonathan S W, Paul Nguyen, Patrick E Brown, Jamie Stafford, and Nathalie Saint-Jacques. 2017. "A Local-EM Algorithm for Spatio-Temporal Disease Mapping with Aggregated Data." *Spatial Statistics* 21: 75–95. doi:10.1016/j.spasta.2017.05.001.

Lindgren, Finn, Håvard Rue, and Johan Lindstrom. 2011. "An Explicit Link Between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach." *Journal of the Royal Statistical Society B* 73 (4). Blackwell Publishing Ltd: 423–98. doi:10.1111/j.1467-9868.2011.00777.x.

Nguyen, Paul, Patrick E Brown, and Jamie Stafford. 2012. "Mapping Cancer Risk in Southwestern Ontario with Changing Census Boundaries." *Biometrics* 68. doi:10.1111/j.1541-0420.2012.01792.x.

Simpson, Daniel, Haavard Rue, Andrea Riebler, Thiago G. Martins, and Sigrunn H. Sorbye. 2017. "Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors." *Statistical Science* 32 (1). The Institute of Mathematical Statistics: 1–28. doi:10.1214/16-STS576.