Consumer Data Research Centre

An ESRC Data Investment

Forecasting network faults with Bayesian spatiotemporal statistical models

Thomas Statham[1], Gavin Dong[1], Robert Ogden[2] and Andy Simpson[2]

[1]University of Liverpool, [2]National Broadband Provider

Project Background

Fast and reliable broadband services are increasingly important as the United Kingdom transitions to an information society. As broadband plays an increasingly critical role in society, the supporting network infrastructure must keep pace to ensure reliable services. Moreover, providing reliable services is important for retaining customers in a competitive market. Therefore, major broadband service providers must accurately forecast network faults to determine the number of engineer and truck call outs necessary to fix any service disruptions. This is important for operational efficiency and network maintenance costs. As such, the aim of this study is to apply different forecasting methods to compare which method gives the highest forecast accuracy of network faults for a broadband service provider.

Data and Methods

We extended the conventional time-series approaches used in telecommunications literature by incorporating space. We applied four different model specifications: a stationary time-series process, a non-stationary time-series process, a spatiotemporal model, and a space-time interaction model. For customer confidentiality, we aggregated all faults at the postcode level to the Middle Layer Super Output Area, modelled through a Besag-York-Mollié prior to aggregation. We applied these complex models using Integrated Nested Laplace Approximation for fast Bayesian inference. We first applied this methodology to all faults, and then to the top 5 most common fault types. We also looked at potential socio-demographic factors from open-source datasets and their impacts on network faults. These included fixed-line broadband usage from Ofcom to indicate broadband demand, the Internet User Classification from CDRC, the level of user income and education, and the proportion of elderly population. The study area was approximately 4215km^2 in North West England.

Key Findings

The results support the value of incorporating space into conventional time-series approaches. Although the model fit of the most complex space-time interaction model was the best, the spatiotemporal model had the highest forecast accuracy or difference between observed and forecasted network faults. Incorporating space allows the forecaster to identify how the spatial distribution of faults changes over time at a much finer spatial scale than conventional time-series models (Figure 1). For example, there was a higher probability of forecasted faults exceeding 15 faults in North Liverpool, Saint Helens and Wigan. Due to the anomalous observed faults in January 2018, the model significantly overestimated network faults for February 2018. By applying the preferred spatiotemporal model to the top 5 network fault types, we found that the anomalous results were associated with an increase in Customer Equipment related faults and this was higher for the Merseyside area.

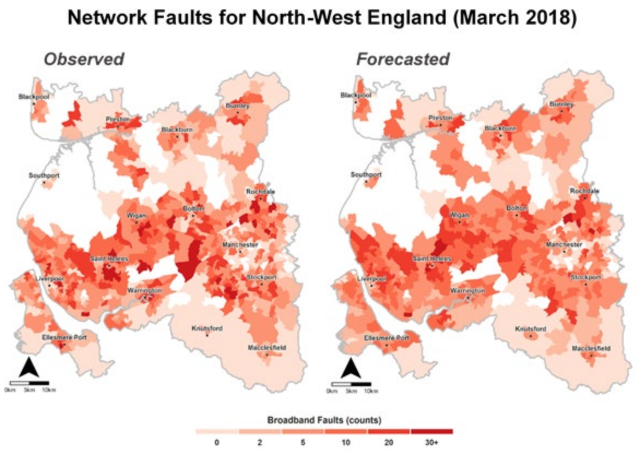

Figure 1. Observed and forecasted broadband faults (March 2018)

The results of this study also support the digital divide in the literature. Interestingly, network faults are lowest in rural and inner-city areas. Inner cities are associated with the e-Professionals and Youthful Urban Fringe IUC groups. These represent a young socio-economic group, who are typically students and young professionals are actively engaged with the internet. Conversely, the Passive and Uncommitted and E-Withdrawn IUC groups had the highest positive effect on network faults. These represent low income, ethnically diverse groups. However, it's unclear whether it's the characteristics of these groups that influence faults, or the characteristics of the areas they live in.

Value of the Research

Applying this methodology gives fast estimates of forecasted network faults that would not be possible with Markov Chain Monte Carlo approaches. Incorporating space gives a higher forecast accuracy at a much finer scale than conventional time-series approaches. Moreover, this approach has applied this methodology at a much larger scale than conventional small area studies in the literature. This methodology could be extended to other applications, such as forecasting broadband demand or be applied to different fields.