A comparison of statistical methods for epidemiological surveillance

There are several approaches to carry out surveillance studies in epidemiology and this project aims at comparing the most used ones in the context of chronic disease epidemiology (for infectious diseases things can be quite different). It is important that the method considered can somehow deal with multiple spatial locations (e.g. allow for spatial dependence)

We propose to focus on three methods:

CUSUM – I found a recent paper which develop CUSUM for spatial data

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4512475/>.

Following the reference of this paper you should be able to identify other relevant papers – I think Axel should be able to help here as well.

Scan statistics – worth looking at Satscan

<http://www.satscan.org/> which is a standalone software for surveillance using the scan statistics

For these two methods you might also find R packages. Here there is a review of software for surveillance

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2848213/>

BayesDetect (Space-time Bayesian model) – you have already the references for this.

For each method it would be important to identify: performance (based on false positive/negative etc.) as well as computational challenges and time to run

Data

We will use data that were simulated for 15 time points for 211 Clinical Commissioning Groups (CCGs) in England under a Poisson model with a BYM prior for the spatial component and a RW1 prior for the temporal component. A real dataset of asthma hospitalisation counts was used for the expected cases, adjusted for age and sex using direct standardisation.   
Fifteen unusual areas were selected in total according to the 10th, 25th, 50th, 75th and 90th percentiles of the median expected counts over time. At each percentile, 3 areas were selected, each corresponding to one of the three levels of the overall spatial risks, low (within the 10th-30th percentiles), medium (within the 45th-55th percentiles) and high (within the 70th-90th percentiles) with respect to the distribution of the estimated common spatial risks.

The following parameters were used:

* Marginal variance: 0.03
* Variance of unstructured spatial effects: 0.02
* Variance of temporal effects: 0.01

Given that the common temporal pattern is denoted by g(t), the unusual areas follow the scenario g\*(t) = g(t) + log(2) for t = 1, 10,11 and g\*(t) = g(t) - log(2) for t=4.

The following datasets are provided:

* simulated\_data.Rda: response data (matrix of 211x15)
* expected\_data.Rda: expected data used as offset in the Poisson model (vector of 211)
* CCG\_shapefile: the .shp data including England and Wales for mapping (NOTE: Wales has to be removed from the shapefile as the data consider only England)
* unusual\_areas.Rda: list of the selected unusual areas with their corresponding spatial risk (variables “SpRisk” and “RiskGroups”) and percentile (variable “Qgroups”).

Maps of the SMRs across time and a plot of the temporal effects (common VS unusual) are also provided.