

A Spatio-Temporal Forecasting Approach for Health Indicators

Peter Congdon

Queen Mary University of London

Abstract: Progress towards government health targets for health areas may be assessed by short term extrapolation of recent trends. Often the observed longitudinal series for a set of health areas is relatively short and a parsimonious model is needed that is adapted to varying observed trajectories between areas. A forecasting model should also include spatial dependence between areas both in representing stable cross-sectional differences and in terms of changing incidence. A fully Bayesian spatio-temporal forecasting model is developed incorporating flexible but parsimonious time dependence while allowing spatial dependencies. An application involves conception rates to women aged under 18 in the 32 boroughs of London.

Key words: Autoregression, forecasting, fractional polynomial, spatial effects, teenage conceptions.

1. Introduction

Government health improvement targets to reduce adverse events or health inequity increasingly involve forecasts to assess how far much progress towards the target has been achieved. When the targets are set for health areas, this requires a forecast for a panel data situation, with observations formed by a collection of incidence rates by areas $i = 1, \dots, N$ and times $t = 1, \dots, T$, and with forecasts required for times $t = T + 1, \dots, T + R$. This paper sets out a spatio-temporal forecasting model suited to such analysis that allows both for enduring spatial patterning in the health outcome and for spatial clustering of growth or decline in incidence.

Enduring health contrasts reflect features such as the impact of social structure and health behaviours on health outcomes. Such contrasts tend to be relatively stable in established urban areas, only subject to slow change (e.g. due possibly to selective migration over a number of decades), though more pronounced change may occur in fast growing urban areas (e.g. if there is substantial building of new homes). By contrast, trends towards improvement (or deterioration) in health outcomes may well not be positively correlated with existing differentials, and may reflect a wide range of factors including success or otherwise of health

sector interventions, though may still show spatial patterning. This is true for the application considered here which involves short term forecasts of conception rates to women under 18 in the 32 London boroughs, using observations over 1992-2003; the percent change in this rate between 1992-95 and 2000-2003 has a correlation of -0.14 with the average rate over the entire 12 year period.

Considerable work has been done in developing spatio-temporal models for descriptive analysis (e.g. Waller et al, 1997) but the temporal mechanisms assumed are not necessarily suited to short to medium term extrapolation (between one to ten years ahead). Growth curve models in biostatistics and psychometrics also assume relatively simple time effects (e.g. linear growth) and are not necessarily suited to extrapolation. Bernardinelli et al (1995) present a model for spatial disease outcomes through time that allows for spatially correlated linear growth, but this change pattern may be a simplification in many applications.

Some have argued for an integration of growth curve principles and autoregressive models, improving forecasting potential, since autoregressive dependence on lagged observations provides a basis for short term forecasting that reflects the recent history of the series (Bollen and Curran, 2004). There is also scope for modelling nonlinear evolution within a growth path context, e.g. by introducing fractional polynomial effects in time (Royston and Altman, 1994). Fractional polynomials allow a flexible polynomial shape without necessarily requiring high order polynomial terms in time. The model developed here combines autoregressive and growth curve principles within a spatio-temporal forecasting framework; the model allows for spatial patterning both in the health outcome and in changing incidence of the outcome, with area change profiles modelled using fractional polynomials.

The application involves a relatively short panel series of conception rates to women aged under 18 in 32 London boroughs. Event totals and denominators (women aged 15-17) are available back to 1992 and the series ends in 2003. Here the last year's observations are held out to provide a cross-validation benchmark for models based on the 11 years 1992-2002. Teenage conceptions in London in 2003 total 6500 in relation to a risk population (over all 32 areas combined) of 127 thousands. The borough rates in 2003 vary from 22.4 per 1000 to 105 per 1000 (see Figure 1) with the highest rates in socially deprived areas in east and south east inner London (cf. Diamond et al, 1999), while lower rates occur in generally more affluent suburban boroughs. Specification of the model variants is based on fully Bayesian principles, involving prior densities on parameters and their updating via the likelihood to produce estimated posterior densities. Estimation involves repeated sampling via Markov Chain Monte Carlo procedures, and uses the WINBUGS package¹

¹See Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. (2003) WinBUGS User Manual.

Figure 1 Actual Rates in 2003

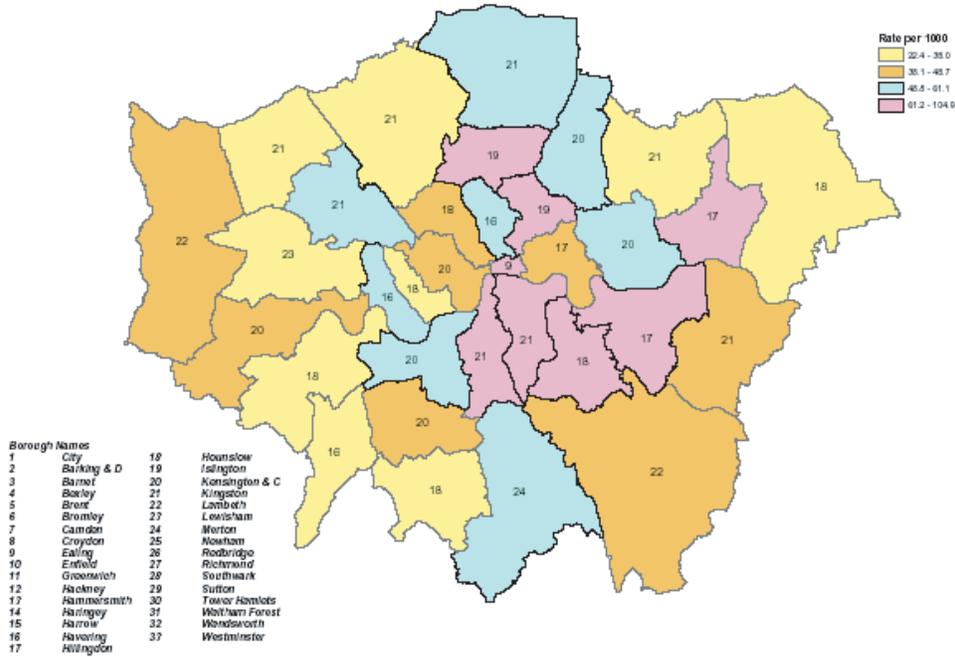


Figure 1: Actual rates in 2003

2. Modelling Framework

For event totals Y_{it} and populations at risk N_{it} , binomial or Poisson sampling is frequently assumed in conjunction with a logit and log links respectively. Thus $Y_{it} \sim \text{Bin}(N_{it}, r_{it})$ where r_{it} are the event probabilities, or $Y_{it} \sim \text{Po}(N_{it}\mu_{it})$ where μ_{it} are relative risks. The framework suggested here allows for time specific spatial morbidity effects s_{it} that are expected to change relatively slowly. However, forecasting is facilitated by lagged effects of previous outcomes and by including possibly nonlinear growth effects. Here binomial sampling is assumed, and lagged effects are based on moment estimates of rates $p_{it} = Y_{it}/N_{it}$.

For modelling spatial effects s_{it} , one possible prior structure is provided by the intrinsic conditionally autoregressive (ICAR) Normal prior of Besag et al (1991),

Version 1.4 (<http://www.mrc-bsu.cam.ac.uk/bugs>)

with weights w_{ijt} relating areas i and j in year t ,

$$s_{it} | s_{jt, j \neq i}, \tau_t \sim N \left(\frac{\sum_{j \neq i} w_{ijt} s_{jt}}{\sum_{j \neq i} w_{ijt}}, \frac{\tau_t}{\sum_{j \neq i} w_{ijt}} \right)$$

where τ_t is the conditional variance of the spatial effects. Several other forms of spatial prior have been suggested, including priors that mix heterogeneity with spatial correlation (Leroux et al, 1999). The s_{it} can be envisaged as residuals with spatial structure so that s_{it} and s_{jt} have positive spatial dependence. Such spatial effects are often used in cross-sectional disease mapping applications (Pascutto et al, 2000) to represent spatially correlated and unmeasured influences on disease risk. Here they can be taken to represent time varying spatially correlated influences on the risk of teenage conception. It is not necessary to assume Normality regarding the s_{it} and other options are Student t and double exponential densities (Lawson and Clark, 2002), especially if outliers are thought likely to distort an otherwise relatively smooth spatial interdependence. For identification, the restriction $\sum_i s_{it} = 0$ within each period is needed; this is achieved by centering the effects at each MCMC iteration. The weights w_{ijt} define the form of spatial contiguity or connectiveness² and are typically assumed to be non-stochastic. They make take account of geographic or even social distances between areas, but are commonly defined in terms of adjacency (i.e. $w_{ijt} = 1$ if areas i and j are adjacent and zero otherwise).

The regression model is here confined to a changing intercept α_t to model the overall change in incidence, though other known predictors for areas and/or times might be relevant to explaining conception trends (e.g. youth employment prospects). Then a model for r_{it} including lags at $1, \dots, L$ in p_{it} has the form

$$\text{logit}(r_{it}) = \alpha_t + s_{it} + \beta_1 p_{i,t-1} + \dots + \beta_L p_{i,t-L}$$

For t exceeding $T+L$, the lagged $p_{i,t-L}$ are replaced by lagged $r_{i,t-L}$. Stationarity is not imposed a priori on the autoregressive (AR) coefficients. It is likely that autoregressive effects differ by area, and a generalization assumes random effects AR coefficients, either with no spatial structure or spatially structured, so

$$\text{logit}(r_{it}) = \alpha_t + s_{it} + \beta_{i1} p_{i,t-1} + \dots + \beta_{iL} p_{i,t-L}$$

where the β_{il} for lag l could have ICAR form separately, though a multivariate version of the CAR model (an MCAR prior) can also be used to model the β_{il} jointly (Gamerman et al, 2003).

²Lesage, J. (1999). Spatial Econometrics. Web manuscript. Department of Economics, University of Toledo, see <http://www.spatial-econometrics.com>

Following Bollen and Curran (2004), albeit concerned with non-spatial applications, a more inclusive approach to modelling longitudinal change includes growth curve as well as autoregressive influences. While several forms of time series model have been proposed for nonlinear time effects, such as functional autoregressive models (Huang and Shen, 2004), these are difficult to extend to the case where the nonlinear effects differ by area. By contrast, the fractional polynomial structure allows flexible modelling of nonlinear effects and also permits the coefficients to vary by area. A fractional polynomial (FP) relation between r_{it} and time t can be added to the above model via

$$\text{logit}(r_{it}) = \alpha_t + s_{it} + \beta_{i1}p_{i,t-1} + \dots + \beta_{iL}p_{i,t-L} + \phi_M(t; \xi, q) \quad (2.1)$$

where $q = (q_1, q_2, \dots, q_M)$, ($q_1 < q_2 < \dots < q_M$), is a vector of powers in t , including possibly $q_j = 0$ leading to $\log(t)$. The summation

$$\phi_M(t; \xi, q) = \sum_{j=1}^M \xi_j H_j(t)$$

involves polynomial terms $H_j(t)$, $j = 1, \dots, M$ defined as $H_j(t) = t^{q_j}$ if $q_j \neq q_{j-1}$ but $H_j(t) = H_{j-1}(t) \log(t)$ if $q_j = q_{j-1}$. To reflect the spatial panel data application, the coefficients ξ_{ij} can be made area specific, with spatially structured (possibly multivariate) random effects priors being one option for $(\xi_{i1}, \xi_{i2}, \dots, \xi_{iM})$.

In practice models with large L and/or M may not lead to pronounced improvements in forecasting accuracy, and one may adopt an incremental modelling strategy starting with a simple model. For example, a baseline model oriented to forecasting might be defined by $\{L = 1, M = 1, q = (1)\}$ and varying area lag and FP effects, namely

$$\text{logit}(r_{it}) = \alpha_t + s_{it} + \beta_i p_{i,t-1} + \xi_i t \quad (2.2)$$

and one may generalize this model by adding (area specific) lags and/or further fractional polynomial terms. The model in (2.2) would also tend to accurately describe observed cross-sectional patterns (within the estimation period) due to the s_{it} parameters. However, reduced parameterization of the descriptive side of the model may be possible without adversely affecting short-term forecasts. This is illustrated by the case study.

3. Case Study

The above framework is now applied to the spatial panel for London boroughs over 1992-2003 (source: Department of Health for England), with the first ten years of data providing the estimation period (see the web site of Journal of Data

Science for the data, including area neighbour structure). For teen conception totals Y_{it} and female populations at risk N_{it} , binomial sampling is assumed in conjunction with a logit link. The goal is parsimonious modelling that produces a good forecast to (known) values in the cross validation year, 2003. Comparisons of model fit within the estimation period use the deviance information criterion (DIC) of Spiegelhalter et al (2002), where the DIC is the posterior average of the deviance plus a measure of complexity d_e . This measure can also be regarded as an estimate of the effective dimension of the model. It is estimated as the difference between the average deviance and the deviance $\text{Dev}(\bar{\theta})$ at the posterior mean of the full parameter set θ . To assess out of sample fit to known rates in the period $T + 1$ (i.e. the year 2003, so $R = 1$), predictive error sum of squares

$$E_1 = \sum_i \frac{(\bar{r}_{i,T+1} - p_{i,T+1})^2}{p_{i,T+1}} \quad (3.1)$$

and average absolute deviation measures

$$E_2 = \frac{1}{N} \sum_i \frac{|\bar{r}_{i,T+1} - p_{i,T+1}|}{p_{i,T+1}} \quad (3.2)$$

are used, where $\bar{r}_{i,T+1}$ denotes the posterior mean forecast. Then $100E_2$ expresses the average percent absolute relative error in forecasts to 2003.

Five models are considered, with the rationale in model development being to introduce enhancements to improve forecast performance while avoiding excess parameterization. A baseline model (model 1) resembles descriptive models used in disease mapping studies; this will tend to produce a good within sample fit but not necessarily a good forecast fit. Model 1 allows for spatially correlated effects s_{it} specific to time and also for changing overall incidence. The spatial effects follow the ICAR model of Besag et al (1991), with weights w_{ijt} . This model allows for changes in the degree of spatial correlation in a health outcome over time (Waller et al, 1997); specifically, the changing variance τ_t of the spatial effects is modelled using a first order random walk in $\log(\tau_t)$. Changing overall incidence is modelled via a changing intercept α_t , also assumed to follow a first order random walk prior, with identifiability achieved by centering the s_{it} . The weights w_{ijt} are assumed constant over time and to equal 1 when boroughs are adjacent and zero otherwise. Hence model 1 is

$$\begin{aligned} Y_{it} &\sim \text{Bin}(N_{it}, r_{it}) \\ \text{logit}(r_{it}) &= \alpha_t + s_{it} \\ s_{it} | s_{jt}, j \in L_i, \tau_t &\sim N(\bar{S}_{it}, \tau_t/k_i) \\ \varphi_t &= 1/\tau_t = \exp(\kappa_t) \end{aligned}$$

$$\begin{aligned} \kappa_t &\sim N(\kappa_{t-1}, \phi_\kappa) & t = 2, \dots, T + R \\ \alpha_t &\sim N(\alpha_{t-1}, \phi_\alpha) & t = 2, \dots, T + R \end{aligned}$$

where L_i denotes the neighbourhood of borough i (the set of k_i boroughs adjacent to it) and \bar{S}_{it} is the average of the spatial effects in that neighbourhood in year t . Diffuse Normal $N(0, 1000)$ priors (i.e. variance of 1000) are assumed on κ_1 and α_1 , while $1/\phi_\kappa$ and $1/\phi_\alpha$ follow $Ga(1, 1)$ gamma priors.

A two chain run of 5000 iterations with disparate initial values shows convergence after 1000 iterations according to Gelman-Rubin criteria (Gelman et al, 1995). There is some evidence that overall incidence is increasing early in the observation span (the early and mid 1990s), but that this is tailing off by the 2000s. Hence the forecast (posterior mean) for α_t for 2003 is for very similar to the mean for 2002. By contrast, the precisions of the spatial effects are increasing and the mean of φ_{T+1} exceeds the mean of φ_T .

This model is relatively heavily parameterized ($d_e = 488$ compared to $NT = 352$ observations) while not facilitating forecasts tailored to the growth path in each area. The fit within the estimation period (1992-2002) is good as judged by criteria 3.1 and 3.2 evaluated for 2002 (Table 1). However, the forecasts for 2003 show lack of fit, with average relative absolute error of 41% (100 times E_2). Hence a second model (model 2) introduces area specific autoregression on the preceding years rate $r_{i,t-1}$ and also allows a borough specific linear growth effect; this is the case $M = L = 1$ with $q = (1)$, as mentioned above.

Table 1: Insample and forecast model fit

	Within Estimation Sample			Forecast			
	DIC	ED ^a	DPMP ^b	SREES ^c	ARAD2 ^d	SRES ^e	ARAD3 ^f
Model 1	3412.8	488.1	2436.6	0.0007	0.017	0.323	0.407
Model 2	2961.0	78.9	2803.2	0.0060	0.053	0.0280	0.094
Model 3	2915.7	96.5	2722.8	0.0045	0.048	0.0254	0.095
Model 4	2948.6	91.1	2766.4	0.0051	0.048	0.0264	0.093
Model 5	2933.6	101.9	2729.8	0.0033	0.040	0.0284	0.100

^aED = Effective Dimension d_e (Complexity), ^bDPMP= Deviance at posterior mean of parameters, ^cSREES= Sum of Relative Errors Squared (2002), ^dARAD2 = Average of Relative Absolute Deviations (2002), ^eSRES = Sum of Relative Errors Squared (2003), ^fARAD3 = Average of Relative Absolute Deviations (2003).

To reduce parameterization the time specific spatial effect is replaced by a spatial factor model (Wang and Wall, 2003; Congdon, 2002), so that s_{it} is replaced by $\lambda_t s_i$ where $\lambda_t = \exp(\eta_t)$ are positive coefficients(loadings), with $\lambda_1 = 1$, and η_t

following an RW1 prior to allow for pooling strength over time. The terminology ‘spatial factor’ describes the use of a shared error with coefficients linking the error through times. The autoregressive effect and the linear growth effect are both assumed spatially correlated. The resulting model 2 has the form

$$\begin{aligned} \text{logit}(r_{it}) &= \alpha_t + \exp(\eta_t)s_i + \beta_i p_{i,t-1} + \xi_i t & t = 2, \dots, T + R \\ \text{logit}(r_{i1}) &= \alpha_1 + s_i + \xi_i \\ s_i | s_j, j \in L_i, \tau &\sim N(\bar{S}_i, \tau/k_i) \\ \xi_i | \xi_j, j \in L_i, \omega_\xi &\sim N(\bar{\xi}_i, \omega_\xi/k_i) \\ \beta_i | \beta_j, j \in L_i, \omega_\rho &\sim N(\bar{\beta}_i, \omega_\beta/k_i) \\ \eta_t &\sim N(\eta_{t-1}, \phi_\eta) \\ \alpha_t &\sim N(\alpha_{t-1}, \phi_\alpha) \end{aligned}$$

with $\eta_1 = 0$, $\alpha_1 \sim N(0, 1000)$, $1/\tau \sim Ga(1, 1)$, $1/\omega_\beta \sim Ga(1, 1)$, $1/\omega_\xi \sim Ga(1, 1)$, $1/\phi_\eta \sim Ga(1, 1)$ and $1/\phi_\alpha \sim Ga(1, 1)$. For t exceeding $T + L$, the lagged $p_{i,t-L}$ are replaced by lagged $r_{i,t-L}$.

A two chain run of 5000 iterations (convergent from 1000) shows that this model leads to a some loss of fit (to 2002 data) within the estimation period. However, the loss is slight in view of the much reduced parameterization as compared to model 1. The DIC actually falls due to the reduced parameter count (Table 1) though the deviance at the posterior parameter mean is higher. Despite the reduced parameter count, the forecast accuracy for 2003 (when forecasts can be compared to known conception rates) improves considerably compared to model 1. The average percent relative forecast error falls to 9%.

Model development could be in extending the fractional polynomial, the AR components or both. To assess the gain through developing the fractional polynomial component of the model, the fractional polynomial components is expanded to $\phi_3(t; \xi_i, q)$, where $q = (-1, 0, 1)$ and $\xi_i = (\xi_{i1}, \xi_{i2}, \xi_{i3})$. Hence model 3 is

$$\begin{aligned} \text{logit}(r_{it}) &= \alpha_t + \exp(\eta_t)s_i + \\ &\quad \beta_i p_{i,t-1} + \xi_{i1}/t + \xi_{i2} \log(t) + \xi_{i3}t & t = 2, \dots, T + R \\ \text{logit}(r_{i1}) &= \alpha_1 + s_i + \xi_{i1} + \xi_{i3} \\ \eta_t &\sim N(\eta_{t-1}, \phi_\eta) \\ \alpha_t &\sim N(\alpha_{t-1}, \phi_\alpha) \\ s_i | s_j, j \in L_i, \tau &\sim N(\bar{S}_i, \tau/k_i) \\ \beta_i | \beta_j, j \in L_i, \omega_\rho &\sim N(\bar{\beta}_i, \omega_\beta/k_i) \end{aligned}$$

with $\eta_1 = 0$, $\alpha_1 \sim N(0, 1000)$, $1/\tau \sim Ga(1, 1)$, $1/\omega_\beta \sim Ga(1, 1)$, $1/\phi_\eta \sim Ga(1, 1)$ and $1/\phi_\alpha \sim Ga(1, 1)$. The increased parameterization may be contained by pooling strength under a multivariate version of the CAR Normal prior (Gamerman

et al, 2003) for $\xi_i=(\xi_{i1}, \xi_{i2}, \xi_{i3})$. Thus the pairwise difference prior with precision matrix Ψ_ξ is

$$P(\xi_i | \Psi_\xi) \propto |\Psi_\xi|^{n/2} \exp\left\{-\sum_{i,j} w_{ij}(\xi_i - \xi_j)' \Psi_\xi (\xi_i - \xi_j)\right\}$$

As above a contiguity assumption is made for w_{ij} , and a Wishart prior with identity scale matrix and 3 degrees of freedom is adopted on Ψ_ξ ; so $\Psi_\xi \sim W(3, I)$.

This model produces a gain in fit within the estimation period at the expense of an increase in complexity d_e from 78.9 to 96.5. There is a gain in forecasting performance according to criterion (3.1). Figure 2 shows the forecast rates for 2003 under model 3, and Table 2 shows the posterior summary of the forecasts.

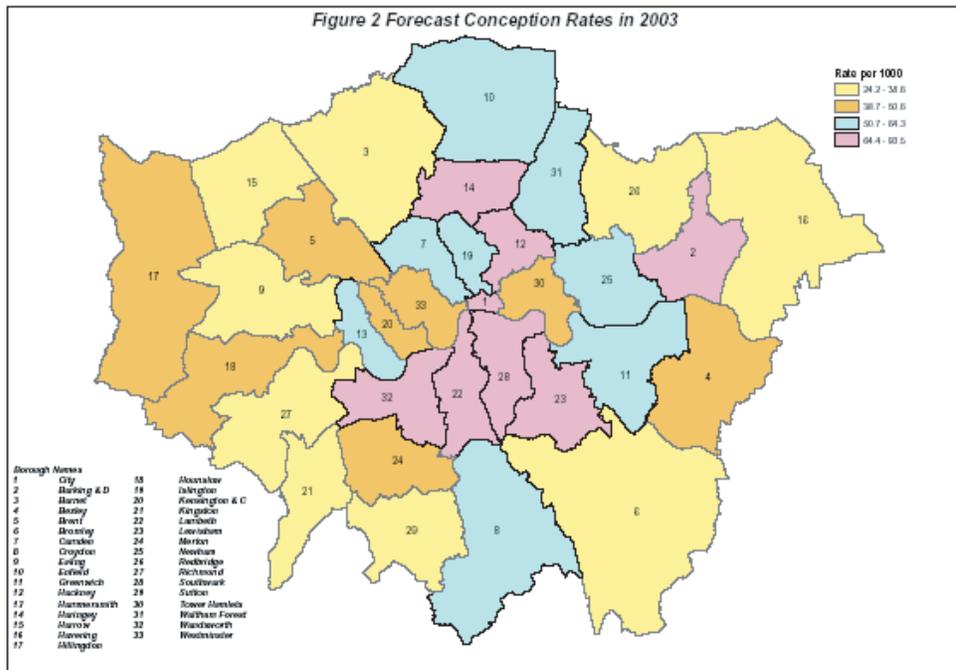


Figure 2: Forecast conception rates in 2003

Table 2: Actual 2003 rates and model 3 prediction summary

Borough	Actual	FPM ^a	2.5%	97.5%
Barking	71.9	68.0	61.7	74.2
Barnet	30.7	31.6	28.2	35.3
Bexley	41.2	42.8	38.4	47.5
Brent	56.2	50.6	46.2	55.4
Bromley	38.6	35.4	31.7	39.4
Camden	48.7	51.0	45.5	56.7
Croydon	52.6	58.2	53.4	62.9
Ealing	38.0	37.6	34.1	41.3
Enfield	57.5	53.7	48.8	59.7
Greenwich	69.1	64.3	58.6	70.5
Hackney & City of London (combined)	69.3	77.6	71.5	84.3
Hammersmith	51.7	63.3	56.6	70.6
Haringey	71.0	77.9	71.7	84.7
Harrow	26.9	26.6	23.3	30.2
Havering	35.9	38.6	34.0	43.4
Hillingdon	48.4	43.7	39.5	48.4
Hounslow	45.8	42.6	38.3	47.4
Islington	61.1	64.1	57.0	71.1
Kens-Chelsea	27.6	45.9	39.9	52.4
Kingston	30.8	30.1	25.8	34.8
Lambeth	104.9	93.5	85.8	102.2
Lewisham	74.0	70.8	64.0	77.1
Merton	39.7	43.7	38.7	49.1
Newham	55.2	56.4	51.8	61.2
Redbridge	32.2	32.7	29.0	36.9
Richmond	22.4	24.2	20.6	28.1
Southwark	86.8	80.7	74.0	87.6
Sutton	32.1	35.8	30.6	41.2
Tower Hamlets	42.9	49.1	44.2	54.0
Waltham Forest	55.4	59.5	54.3	65.3
Wandsworth	58.8	65.8	59.6	71.9
Westminster	41.5	46.8	41.2	53.0

^aFPM= forecast posterior mean

Observation level forecasting performance in this and other models is dominated by overestimation of the conception rate in the central London borough of

Kensington and Chelsea in 2003; this has an actual rate of 27.6 per 1000 in 2003 compared to a posterior mean forecast of 46 per 1000 under model 3. The number of conceptions here fell sharply to 57 in 2003, compared to totals of 80,80 and 81 in the years 2000, 2001 and 2002 respectively. Excluding this borough reduces criterion (3.2) 9.5% to 7.7%. However, some discrepancies are related to shifts in the denominator population estimates. For example, the denominator population in Westminster (women aged 15-17) shifted upwards by 12% between 2002 and 2003 due to an upward revision of populations, reflecting concerns over 2001 census enumeration errors³. The official rate per 1000 (Department of Health, England) for 2003, with the revised population denominator, is 41.5 compared to a posterior mean forecast of 46.8 per 1000.

Model 4 considers instead considers an extension of the lag effects in model 2, namely adding spatially varying AR2 coefficients. A bivariate CAR form is assumed for the two autoregressive coefficients. Thus model 4 is defined by

$$\begin{aligned} \text{logit}(r_{it}) &= \alpha_t + \exp(\eta_t)s_i + \beta_{i1}p_{i,t-1} + \beta_{i2}p_{i,t-2} + \xi_i t \quad t = 3, \dots, T + R \\ \text{logit}(r_{i2}) &= \alpha_2 + \exp(\eta_2)s_i + \beta_{i1}p_{i1} + 2\xi_i \\ \text{logit}(r_{i1}) &= \alpha_1 + s_i + \xi_i \\ s_i | s_j, j \in L_i, \tau &\sim N(\bar{S}_i, \tau/k_i) \\ \xi_i | \xi_j, j \in L_i, \omega_\xi &\sim N(\bar{\xi}_i, \omega_\xi/k_i) \\ P(\beta_i | \Psi_\beta) &\propto |\Psi_\beta|^{n/2} \exp\left\{-\sum_{i,j} w_{ij}(\beta_i - \beta_j)' \Psi_\beta(\beta_i - \beta_j)\right\} \\ \eta_t &\sim N(\eta_{t-1}, \phi_\eta) \\ \alpha_t &\sim N(\alpha_{t-1}, \phi_\alpha) \end{aligned}$$

with $\Psi_\beta \sim W(2, I)$, $\eta_1 = 0$, $\alpha_1 \sim N(0, 1000)$, $1/\tau \sim Ga(1, 1)$, $1/\omega_\xi \sim Ga(1, 1)$, $1/\phi_\eta \sim Ga(1, 1)$ and $1/\phi_\alpha \sim Ga(1, 1)$.

This model produces a slight loss in fit to the estimation sample as compared to model 3, and also no gain in forecast accuracy as compared to model 3. Which model among models 1 to 4 is preferred depends on the relative weights attached to insample fit and out-of-sample forecast accuracy; relative forecasting performance also depends on which of (3.1) or (3.2) is used. Since model 3 has the best in-sample fit (among models 2 to 4), and best forecasting fit under (3.1), it seems to show the best overall performance. If one were primarily concerned with description then model 1 is preferred.

Model 3 has a simple fractional polynomial form. However, extending the fractional polynomial did not improve forecast or in-sample performance. Thus

³See Office of National Statistics (2004). News Release: Results of 2001 Census based Local Authority Population Studies, July 2004 (www.statistics.gov.uk/about/methodology_by_theme/population.asp).

model 5 replicates model 3 except in expanding the FP component to $\phi_5(t; \xi_i, q)$, where $q = (-2, 1, 0, 1, 2)$ and $\xi_i = (\xi_{i1}, \xi_{i2}, \xi_{i3}, \xi_{i4}, \xi_{i5})$ following an MCAR form. The extra complexity of this model leads to a worse DIC and some deterioration in forecast accuracy too (Table 1).

4. Conclusion and Possible Methodological Extensions

Many spatio-temporal models for disease and mortality outcomes have focussed on description and not on extrapolation to the future. As model 1 in the case study application illustrated, a descriptive model providing a close fit to an existing dataset may not produce a good forecast. By contrast, models allowing for short term extrapolation of growth trends and/or autoregression on recent rates may improve on descriptive models in terms of out-of-sample accuracy.

For health and social indicators relatively short observed series over sets of constant boundary areas are often common. Short panel series for area units may be due to boundary changes, changes in data collection or indicator definition, or intermittent collection (as in the case health status data from Censuses). Hence there may be only a small gain in adopting highly complex models with high order lags or complex polynomial forms. However, the case study demonstrated that improved fit did result from adopting a spatially varying fractional polynomial of a relatively simple form.

The above analysis has been for a single outcome, but there may be gains in pooling strength over two or more outcomes. For example, area conception rates for women aged under 16, aged 17-18 and aged 19-20 tend to be positively correlated. For outcomes k assumed binomial with rates r_{ikt} , one might envisage models that develop on (2.1), such as

$$\text{logit}(r_{ikt}) = \alpha_{tk} + \lambda_{tk}s_{ik} + \beta_{ik1}p_{i,k,t-1} + \dots + \beta_{ikL}p_{i,k,t-L} + \phi_{M_k}(t; \xi_{ik}, q_k)$$

where the s_{ik} model stable morbidity differences, the λ_{tk} are positive time and output specific coefficient loadings, the β_{ikl} model area specific autoregressive own lags in outcome k , and the fractional polynomial allows area and outcome specific coefficients and also possibly q_k vectors different by outcome. Pooling over outcomes in the autoregressive coefficients might be based on a common factor approach, for example expressing β_{ikl} as $\kappa_k\beta_{il}$. More complex models might involve cross-lags (of r_{ikt} on $p_{im,t-l}$ for $m \neq k$).

References

- Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C. and Ghislandi, M. (1995). Bayesian analysis of space-time variations in disease risk. *Statistics in Medicine* **11**, 983-1007.

- Besag, J., York, J. and Mollié, A. (1991). Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistics and Mathematics* **43**, 1-59.
- Bollen, K. and Curran, P. (2004). Autoregressive latent trajectory (ALT) models: A synthesis of two traditions. *Sociological Methods and Research* **32**, 336-383.
- Congdon, P. (2002). A life table approach to small area health need profiling. *Statistical Modelling* **2**, 1-26.
- Diamond, I., Clements, S., Stone, N. and Ingham, R. (1999). Spatial variation in teenage conceptions in south and west England. *Journal of the Royal Statistical Society, Series A* **162**, 273-289.
- Gamerman, D., Moreira, A. and Rue, H. (2003). Space-varying regression models: Specifications and simulation. *Computational Statistics and Data Analysis* **42**, 513-533.
- Gelman, A., Carlin, J., Stern, H. and Rubin, D. (1995). *Bayesian Data Analysis*. Chapman and Hall.
- Huang, J. and Shen, H. (2004). Functional coefficient regression models for non-linear time series: A polynomial spline approach. *Scandinavian Journal of Statistics* **31**, 515-534.
- Lawson, A. and Clark, A. (2002). Spatial mixture relative risk models applied to disease mapping. *Statistics in Medicine* **21**, 359-370.
- Leroux, B., Lei, X. and Breslow, N. (1999). Estimation of disease rates in small areas: A new mixed model for spatial dependence. In *Statistical Models in Epidemiology, the Environment and Clinical Trials* (Edited by Halloran, M. and Berry, D.), 135-178. Springer-Verlag.
- Pascutto, C., Wakefield, J., Best, N., Richardson, S., Bernardinelli, L., Staines, A. and Elliott, P. (2000). Statistical issues in the analysis of disease mapping data. *Statistics in Medicine* **19**, 2493-2519.
- Royston, P. and Altman, D. (1994). Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling. *Applied Statistics* **43**, 429-467.
- Spiegelhalter, D., Best, N., Carlin, B. and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* **64**, 583-639.
- Waller, L., Carlin, B., Xia, H. and Gelfand, A. (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association* **92**, 607-617..
- Wang, F. and Wall, M. (2003) Generalized common spatial factor model. *Biostatistics* **4**, 569-582.

Received April 19, 2005; accepted June 15, 2005.

Peter Congdon
Department of Geography
Queen Mary University of London
Mile End Rd, London E1 4NS, UK
p.congdon@qmul.ac.uk