Modelling the Afghanistan conflict from the Wikileaks data

Guido Sanguinetti

Institute for Adaptive and Neural Computation
School of Informatics
University of Edinburgh

St Andrews, Dec 2015

Outline of the talk

Serendipity

- Serendipity: the occurence and development of events by chance in a happy or beneficial way

Serendipity

- Serendipity: the occurence and development of events by chance in a happy or beneficial way
- Most of the work in this talk done by Dr Andrew Zammit-Mangion
- My friend Visakan Kadirkamanathan becomes HoD → his student Andrew looks for help in other corners
- My friend Mike Dewar does a visualisation of Wikileaks (for the fun of it) and sends me the link at about the same time when Andrew was working on spatio-temporal point processes

Serendipity

- Serendipity: the occurence and development of events by chance in a happy or beneficial way
- Most of the work in this talk done by Dr Andrew Zammit-Mangion
- My friend Visakan Kadirkamanathan becomes HoD → his student Andrew looks for help in other corners
- My friend Mike Dewar does a visualisation of Wikileaks (for the fun of it) and sends me the link at about the same time when Andrew was working on spatio-temporal point processes

This work was not funded by anyone and was purely curiosity driven

Serendipity

- Serendipity: the occurence and development of events by chance in a happy or beneficial way
- Most of the work in this talk done by Dr Andrew Zammit-Mangion
- My friend Visakan Kadirkamanathan becomes HoD → his student Andrew looks for help in other corners
- My friend Mike Dewar does a visualisation of Wikileaks (for the fun of it) and sends me the link at about the same time when Andrew was working on spatio-temporal point processes

This work was not funded by anyone and was purely curiosity driven

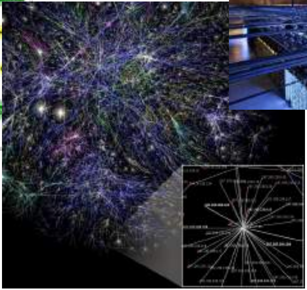
Next time the guy next door asks if you want a coffee, just go

How it was until recently (last decade)

- Data was scarce and hard to collect (e.g. intelligence, surveys of attitudes, official statistics)
- Data was often qualitative without time resolution
- Quantitative modelling was only attempted on highly controlled scenarios in the Social Sciences (e.g. the famous monks data set)
- Qualitative prediction based on intuition from limited data sources was the norm

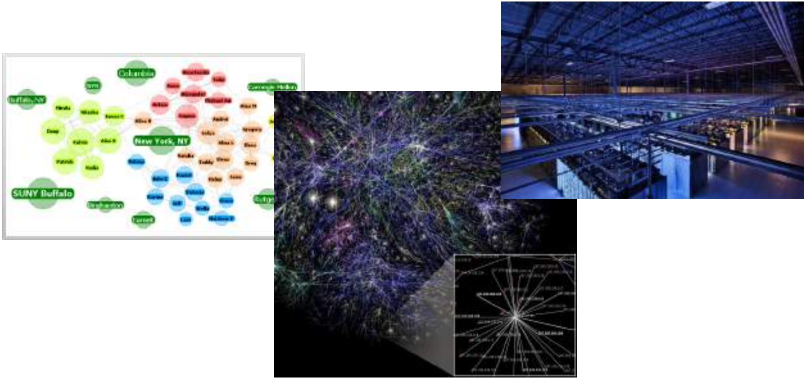
The data revolution

What's all this stuff?



The data revolution

What's all this stuff?



We think that by 2007 there may be 100.000 www domains (pitch to investors by NetCarta founder in 1992)

The data revolution and conflict modelling

What are these and what do they have in common?

Quantitative conflict modelling

- Descriptive statistical modelling: e.g. clustering of events (identification of hotspots), spatial interpolation with non-parametric statistical methods (O'Loughlin et al, 2010)
- Model-free signal processing approaches: e.g. estimate the impulse response (transfer function) of the Israeli/ Palestinian conflict by analysing temporal patterns of response to attacks (Haushofer et al, 2010)
- Postulate simple models that can predict general patterns that can be qualitatively compared with data (Johnson et al, 2011)
- Attempts at capturing some spatial/ temporal correlations using auto-regressive models (Weidmann and Ward, 2010)

Quantitative conflict modelling

- Descriptive statistical modelling: e.g. clustering of events (identification of hotspots), spatial interpolation with non-parametric statistical methods (O'Loughlin et al, 2010)
- Model-free signal processing approaches: e.g. estimate the impulse response (transfer function) of the Israeli/ Palestinian conflict by analysing temporal patterns of response to attacks (Haushofer et al, 2010)
- Postulate simple models that can predict general patterns that can be qualitatively compared with data (Johnson et al, 2011)
- Attempts at capturing some spatial/ temporal correlations using auto-regressive models (Weidmann and Ward, 2010)
- Our approach will be to use a flexible spatio-temporal dynamical system and Bayesian methods for prediction

The Bayesian view of the world

- We are interested in explaining data Y (spatio-temporal location of events) in terms of an unobserved stochastic process X
- The dynamics of the stochastic process are governed by (unknown) parameters Θ
- This is a hierarchical model

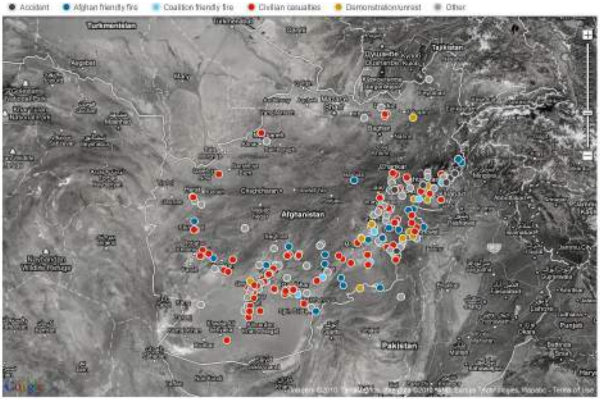
$$p(Y) = \int d\Theta dX p(Y|X)p(X|\Theta)p(\Theta) \tag{1}$$

- We are interested in computing the **joint posterior distribution**

$$p(X,\Theta|Y) = \frac{p(Y|X)p(X|\Theta)p(\Theta)}{p(Y)} \tag{2}$$

- This is generally impossible computationally, approximations required

Conflict data as random sets



Poisson point processes

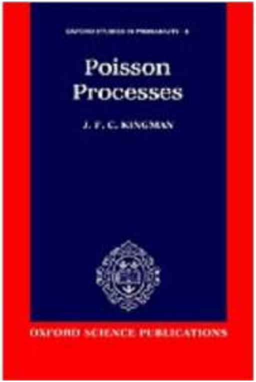
- Poisson Point Process (PPP): random process on a measurable space A whose samples are random points of A
- We assume tha an intensity function $\lambda(\boldsymbol{s})$ exists, such that

$$p(N|a \subset A) = Poiss\left(\int_a \lambda(\boldsymbol{s})d\boldsymbol{s}\right)$$

- Corollary: number of events in disjoint sets are independent random variables
- Given a set of points $\mathcal{P} = \{\boldsymbol{s}_i\} \subset A$, we can compute a *likelihood*

$$p(\mathcal{P}|\lambda(\boldsymbol{s})) = \exp\left\{\sum_i \log\lambda(\boldsymbol{s}_i) - \int_A \lambda(\boldsymbol{s})\mathrm{d}\boldsymbol{s}\right\} \tag{3}$$

One of the most enjoyable books ever

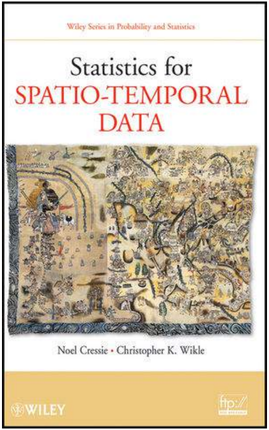


Published in 1993; Kingman had been Vice-Chancellor of Bristol University since 1985...

Cox processes

- Independence property of PPPs is too strong
- Cox processes remove it by assuming the intensity is also a random variable
- We will work in the log-Gaussian framework (LGC= log-Gaussian Cox), i.e.
$$\log \lambda \sim \mathrm{GP}\left(m, k\right)$$
- The dynamics will enter the game through a (spatio-temporal) stochastic process which determines this GP

An excellent recent book (2011)



Contains an excellent discussion of modelling and statistics, as well as essential reference for spatio-temporal stats.

Stochastic Partial Differential Equations (SPDEs)

- The most general spatio-temporal dynamical system?
- Let $z \in \mathbb{L}^2(\Omega)$ be a function (field) on a domain $\Omega \in \mathbb{R}^2$
- Let $\mathcal{A} \colon \mathbb{L}^2(\Omega) \to \mathbb{L}^2(\Omega)$ be a differential operator on functions over Ω
- A Stochastic Partial Differential Equation (SPDE) is defined as

$$dz(\boldsymbol{s}) = \mathcal{A}[z](\boldsymbol{s})dt + dW(\boldsymbol{s}, t)$$

  where $dW(s, t)$ is temporally white (spatially coloured) Wiener noise
- Nasty (or nice according to taste) infinite dimensional object

Time discretisation and integro-difference equations

- In some cases, the mean behaviour of the SPDE (a PDE) can be solved exactly using integral operators (green functions)
- The (discrete time) evolution then becomes

$$z(\mathbf{s})_{t+1} = \int d\zeta K(\mathbf{s}, \zeta) z(\zeta)_t + \epsilon_t$$

  where K is the integral kernel of the SPDE differential operator and the statistics of the noise term ϵ_t are determined by the original noise and dynamics
- In many cases, it is easier to start from this form, the so called *stochastic integro-difference equation* (SIDE)

Finite dimensional reduction

- The SIDE is equivalent to an SPDE and hence still infinite dimensional
- We can reduce it to finite dimensions by expanding the field $z(\boldsymbol{s})$ and kernel K on a fixed set of basis functions $\phi_1(\boldsymbol{s}), \dots, \phi_N(\boldsymbol{s})$
- Denoting by Π_ϕ the projection operator onto the subspace spanned by the basis functions, let A be the matrix representation of the projected operator $\Pi_\phi K \Pi_\phi^T$ in the ϕ basis, and let $\Pi_\phi z(\boldsymbol{s}) = \sum_i x_i \phi_i(\boldsymbol{s})$
- Rewriting the Poisson process likelihood as a function $f(\mathbf{x})$ of the coefficients of this expansion, we have the following latent linear-Gaussian evolution

$$p(\mathcal{P}_t|\mathbf{x}_t) = f(x_t) \qquad \mathbf{x}_{t+1} = A\mathbf{x}_t + \epsilon_t \tag{4}$$

  with ϵ_t the projection of the (Gaussian) noise terms.
- Inference can therefore be performed by Kalman smoothing.

How to choose basis functions?

- A simple choice would be to use a spectral representation; however, these are not necessarily localised
- Number of basis functions is essentially limited by computational power (Kalman smoothing scales cubically)
- Compact support radial basis functions are convenient, but how to choose width?
- The basic idea is to choose the width of the basis function to be *smaller* than the characteristic lengthscale with which the intensity field changes
- This is done by using (spatial) Fourier transforms (using the fact that the FT of a radial basis function is again a radial basis function) and appealing to Shannon's theorem

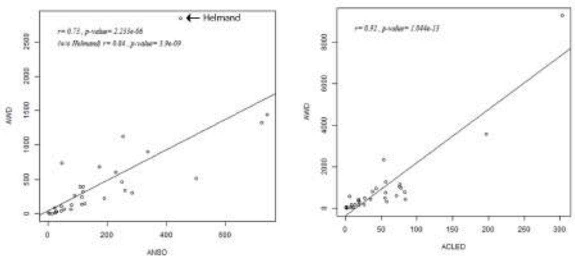
Variational inference

- Parameters also random variables: exact inference intractable
- We use a factorised *variational* approximation to the joint posterior, based on minimising the *Kullback-Leibler* (KL) divergence within the class of factorised distributions
- Due to the non-Gaussian observation model, we use a Laplace approximation within the state estimation part
- Offers good computational speed at a fraction of the costs of sampling methods
- Surprisingly accurate distributionally (see Zammit-Mangion et al 2012, Yuan et al 2012)

Wikileaks Afghan war diary

- In July 2010, Wikileaks released tens of thousands of logs of "activities" in Afghanistan between 2004 and 2009
- Activities range from stop and search to pitched battles
- A log consists of a time stamp (day) and GPS location, plus annotation (not used here)
- Alternatively hailed as a resource for historians, a bastion for free speech or damned as the blackest treason, endangering life etc
- We wanted to know whether it is more than noise

Preliminary analysis - data validation



Wikileaks logs per province vs ANSO (Afghanistan NGO Safety Organisation) (left) and ACLED (Armed Conflict Location and Event Dataset) (right)

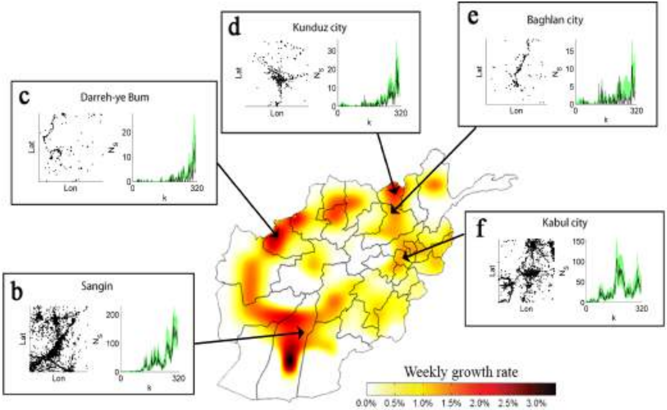
Model choice

- Qualitative patterns reminiscent of ecological data, motivating use of dynamical framework (even if underlying mechanism unknown)
- Relative increments normally distributed (K-S test at p=0.05)
- Pairwise autocorrelations very similar to cross-correlations → negligible spatial interactions
- This suggest a *geometric Brownian motion* prior for the intensity Λ

$$\mathrm{d}\Lambda(s,t) = \left[R(s) + \frac{1}{2}\sigma^2\right]\Lambda(s,t)\mathrm{d}t + \sigma\Lambda(s,t)\mathrm{d}W(s,t)$$

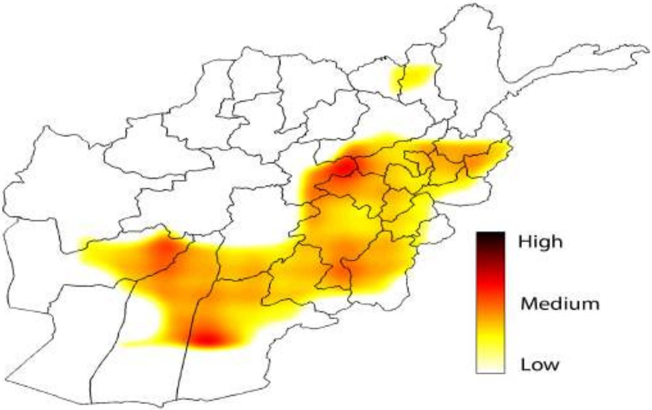
where $R(s)$ is a spatially varying growth coefficient and $W(s,t)$ is spatially coloured and temporally white noise.

Results- Growth coefficient



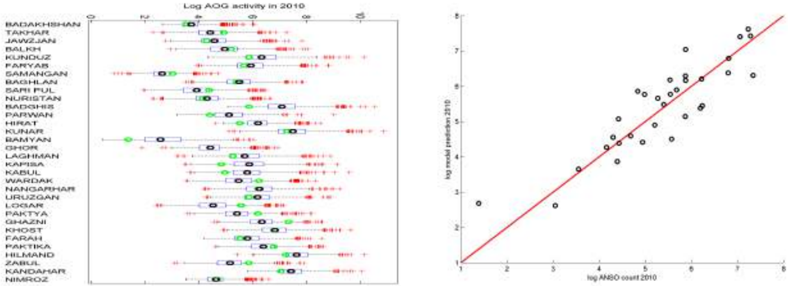
Spatially dependent growth coefficient and posterior fits of selected provinces. Growth hotspots may occur in areas with relatively low counts.

Results- Volatility



Spatially dependent volatility (noise variance).

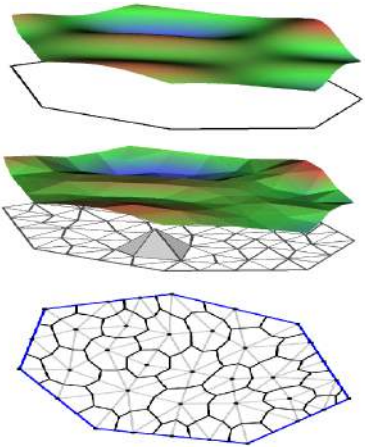
More results



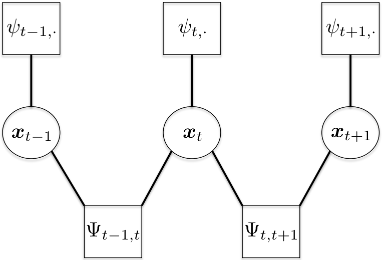
Correspondence between model predictions and AOG events in 2010 as recorded by ANSO. AOGs were estimated as a constant fraction of events on 2009 data; model predictions were generated as 10K Monte Carlo runs.

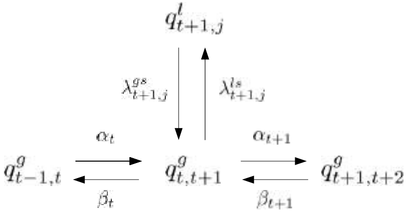
Finite elements and SPDEs

- Finite element projection leads to sparse matrices
- Can use sparsity for efficient matrix inversions
- Problem: sparsity is lost in dynamical systems (intuition: the exponential of a sparse matrix is not sparse)



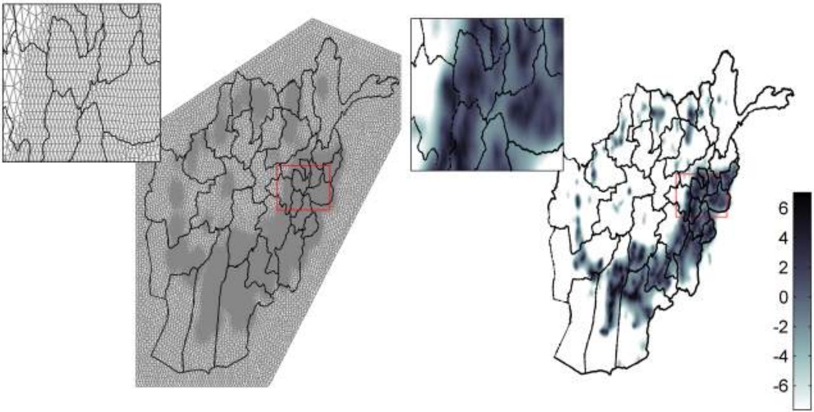
Approximating messages





- Use message passing algorithm to approximate marginals (Expectation-Propagation)
- Enforce approximate sparsity of the messages

Scaling up inference



- Spatial resolution increased 100-fold
- Parameter inference can reveal micro-scale phenomena

Conclusions

- We can (and should) do more than visualisation with spatio-temporal data
- Large scale modelling of heterogeneous spatio-temporal data likely to become more important with social networking etc.
- Dynamic modelling provides a flexible but interpretable tool for spatio-temporal data modelling
- Model-based approach returns far richer statistics that can be used for downstream analyses
- Variational approximation enables large-scale modelling

References

- A. Zammit-Mangion, G. S. and V. Kadirkamanathan, IEEE Trans. Sign. Proc, 73(2):1Đ37, (2012)
- A. Zammit-Mangion, M.A. Dewar, V. Kadirkamanathan and G. S., Point Process Modelling of the Afghan War Diary, PNAS 109(31) (2012)
- A. Zammit-Mangion, M. Dewar, V. Kadirkamanathan, A. Flesken and G.S., Modeling Conflict Dynamics with Spatio-temporal Data, Springer Briefs, 2013
- B. Cseke, A. Zammit-Mangion, T. Heskes and G. Sanguinetti, Sparse approximate inference in spatio-temporal point process models, arXiv:1305.4152, to appear in J. Am. Stat. Ass.