

Background

- ▶ Emerging and re-emerging pathogens usually lead to small scale epidemics
- ▶ Analysis of historical outbreaks informs control of future outbreaks
- ▶ Risk assessment is crucial to inform policy on potential new incursions

Aim

- ▶ Infer historic outbreak characteristics from just observed cases (and small)
- ▶ Predict future evolution of ongoing outbreaks from its early phase
- ▶ Use data from small historic and ongoing outbreaks to select between models

Modelling framework

- ▶ An individual location i , eg a farm, makes an infectious contact with a susceptible individual j at rate β_{ij} assumed to be

$$\beta_{ij} = \beta_0 h_{ij},$$

β_0 being the contact rate. β_{ij} is known as a spatial kernel transmission function and 4 forms widely used in the disease modelling literature are considered:

- K_1 : $h_{ij} = \exp\{-\tau d_{ij}\}$
- K_2 : $h_{ij} = \left(1 + \left(\frac{d_{ij}}{d}\right)^\tau\right)^{-1}$
- K_3 : $h_{ij} = \left(1 + \frac{d_{ij}}{d}\right)^{-1}$
- K_4 : $h_{ij} = 1 - \exp\left(-\left(\frac{d_{ij}}{d}\right)^{-\tau}\right)$

- ▶ Once infected the time to detection is assumed to follow a left-truncated gamma distribution:

$$R_i - I_i \sim \mathcal{TG}(\alpha, \gamma, c),$$

- ▶ The likelihood of observing the detections R_i , $i = 1, \dots, n_R$, is

$$L = \prod_{i=1, i \neq v}^{n_I} \left(\sum_{j \in \mathcal{Y}_i} \beta_{ji} \right) \times \exp\{-S\} \gamma^{\alpha n_R} \exp\left\{-\gamma \sum_{i=1}^{n_R} (R_i - I_i)\right\} \prod_{i=1}^{n_R} \frac{(R_i - I_i)^{\alpha-1}}{\Gamma(\alpha, \gamma c)},$$

where $S = \sum_{i=1}^{n_I} \sum_{j=1}^N \beta_{ij} (\min(R_i, I_j) - \min(I_i, I_j))$

Inference

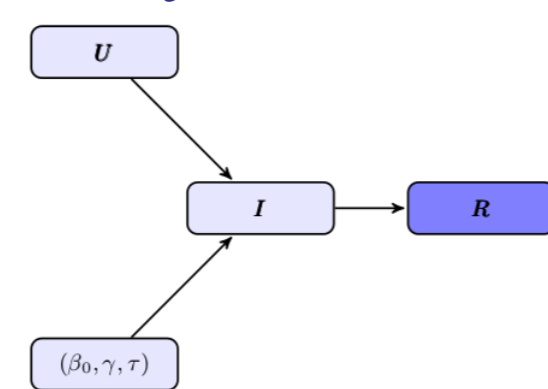
Bayesian inference

- The distribution of the model parameters and latent variables given the data is

$$\pi(\theta, \mathbf{I} | \mathbf{R}) \propto L(\mathbf{R}, \mathbf{I}; \theta) \pi(\theta),$$

- Metropolis-Hastings within Gibbs algorithm
- Non-centered parameterisation for computational efficiency

Parameters and latent variables updated simultaneously and helps reduce autocorrelation

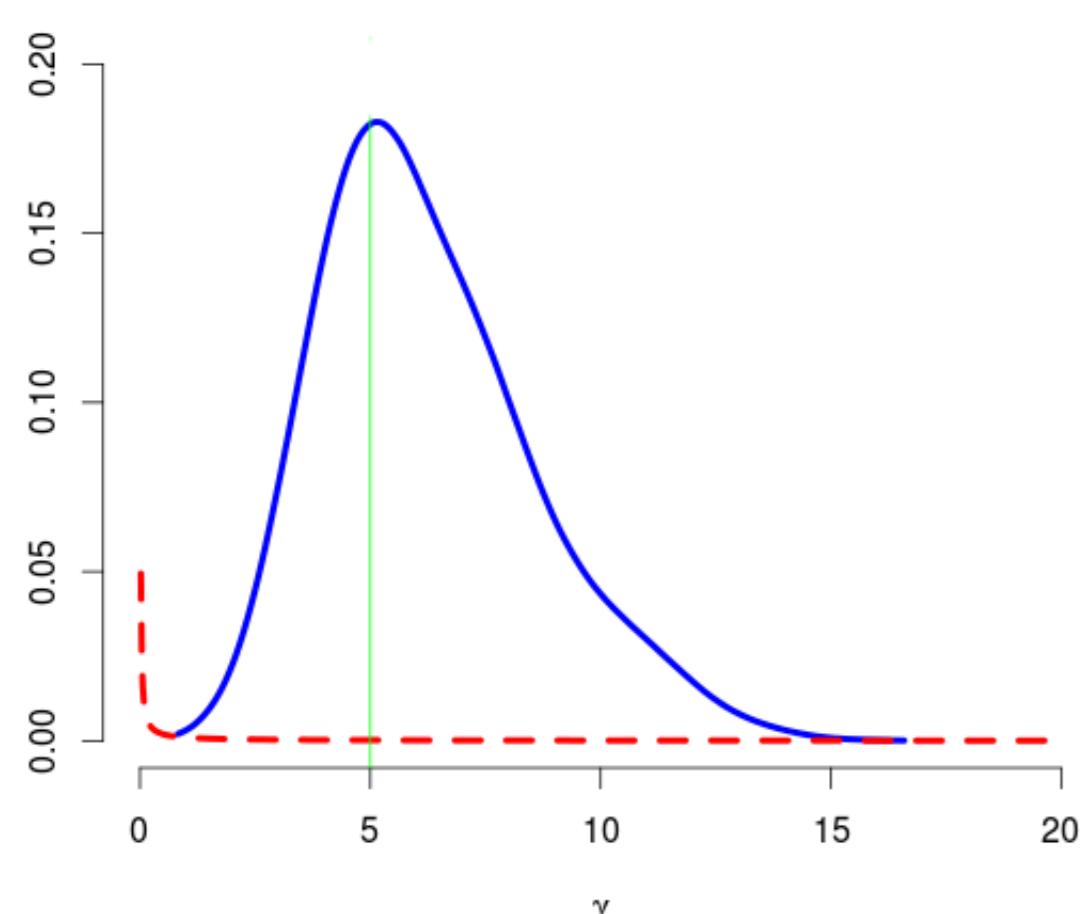


Model assessment and selection

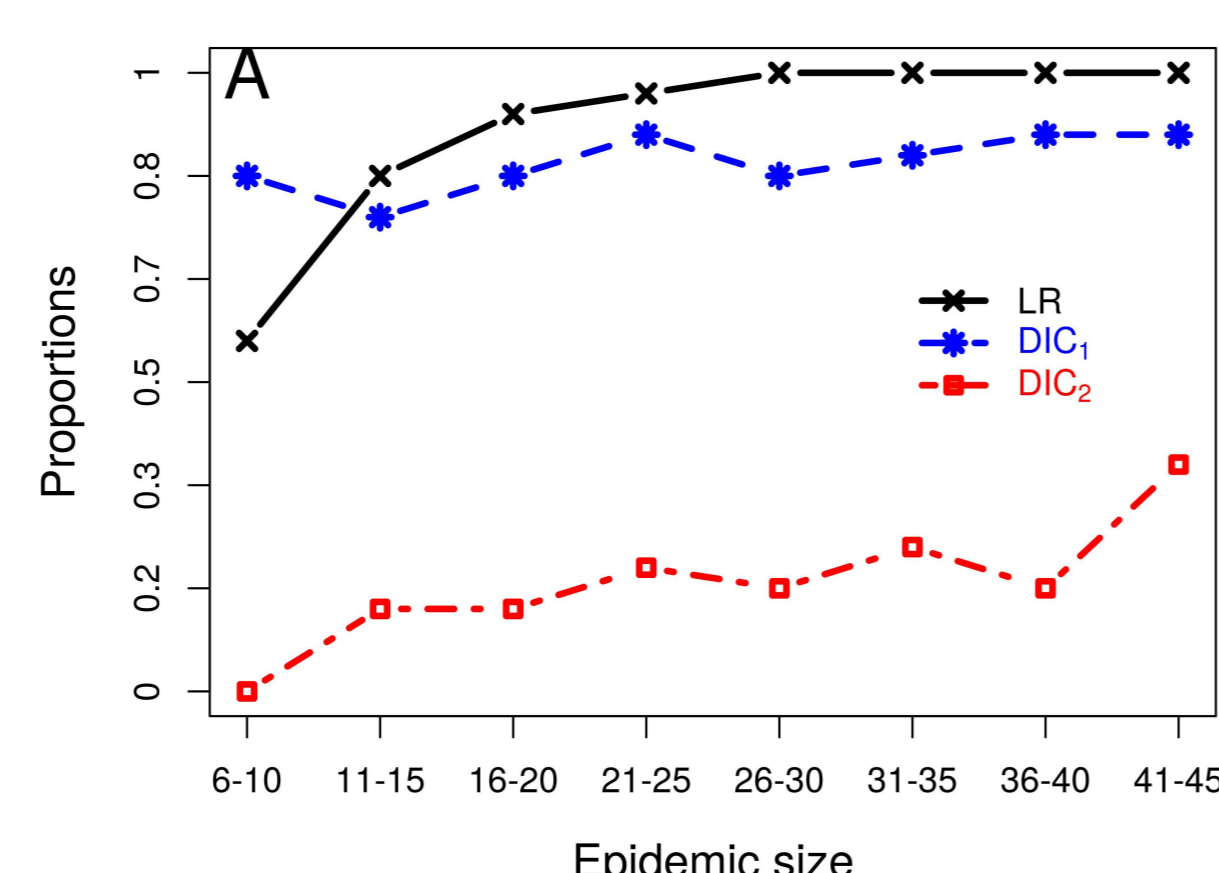
- DIC = "goodness of fit" + "complexity"
- Problems: non invariance to reparameterisation, lack of consistency, weak theoretical justification with multiple definitions in the case of latent variables – 2 used
- Bayesian Latent residuals (LR): inferred from data, iid uniform r.v. if fitted model consistent with data generation process
- LR constructed to test different components of model

Assess performance for historic outbreaks: simulated data

- * Coverage properties: true parameters are contained $\approx 95\%$ in the CI
- * Uncertainty of the estimates reduces as the epidemic size increases
- * Evaluate the amount of data size needed for model selection
- * Increasing epidemic size increases the accuracy of identifying the correct model with the LR



Posterior distribution γ with true values $\alpha = 5$, $\beta_0 = 0.35$, $N = 201$ premises and $n = 43$ infected. The posterior is in blue, the prior in red and the true value in green



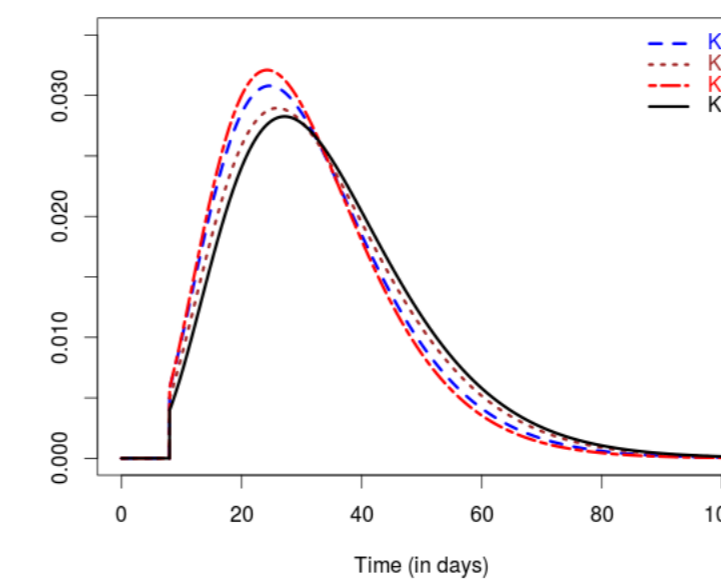
LR and DIC comparisons: proportion of simulated data sets where correct model is selected as function of outbreak size

Classical Swine Fever in East Anglia Norfolk in 2000

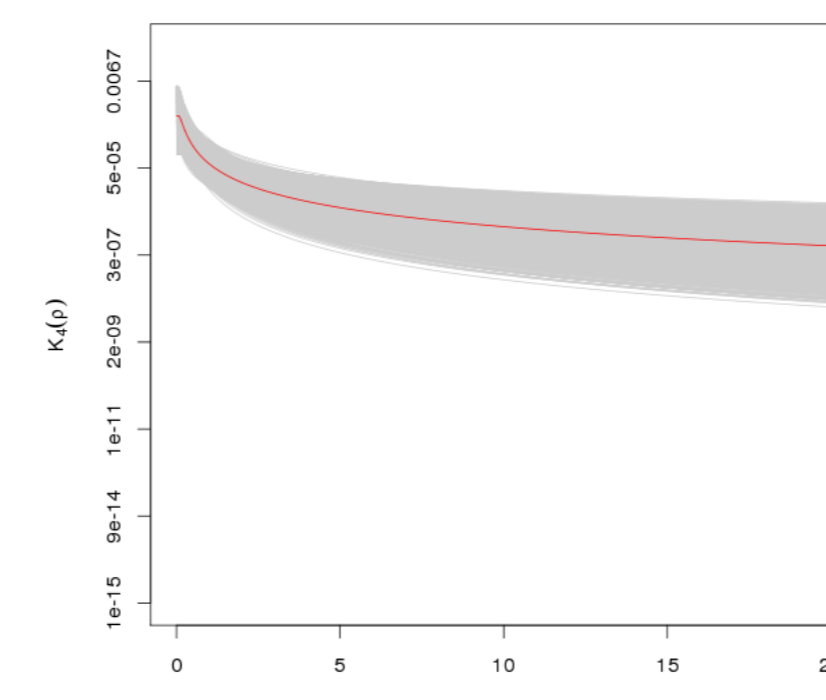
Data

- * $N = 1703$ farms with exact location or coordinates
- * Times and location of 16 detected cases

Inference and predictions



mean infectious period



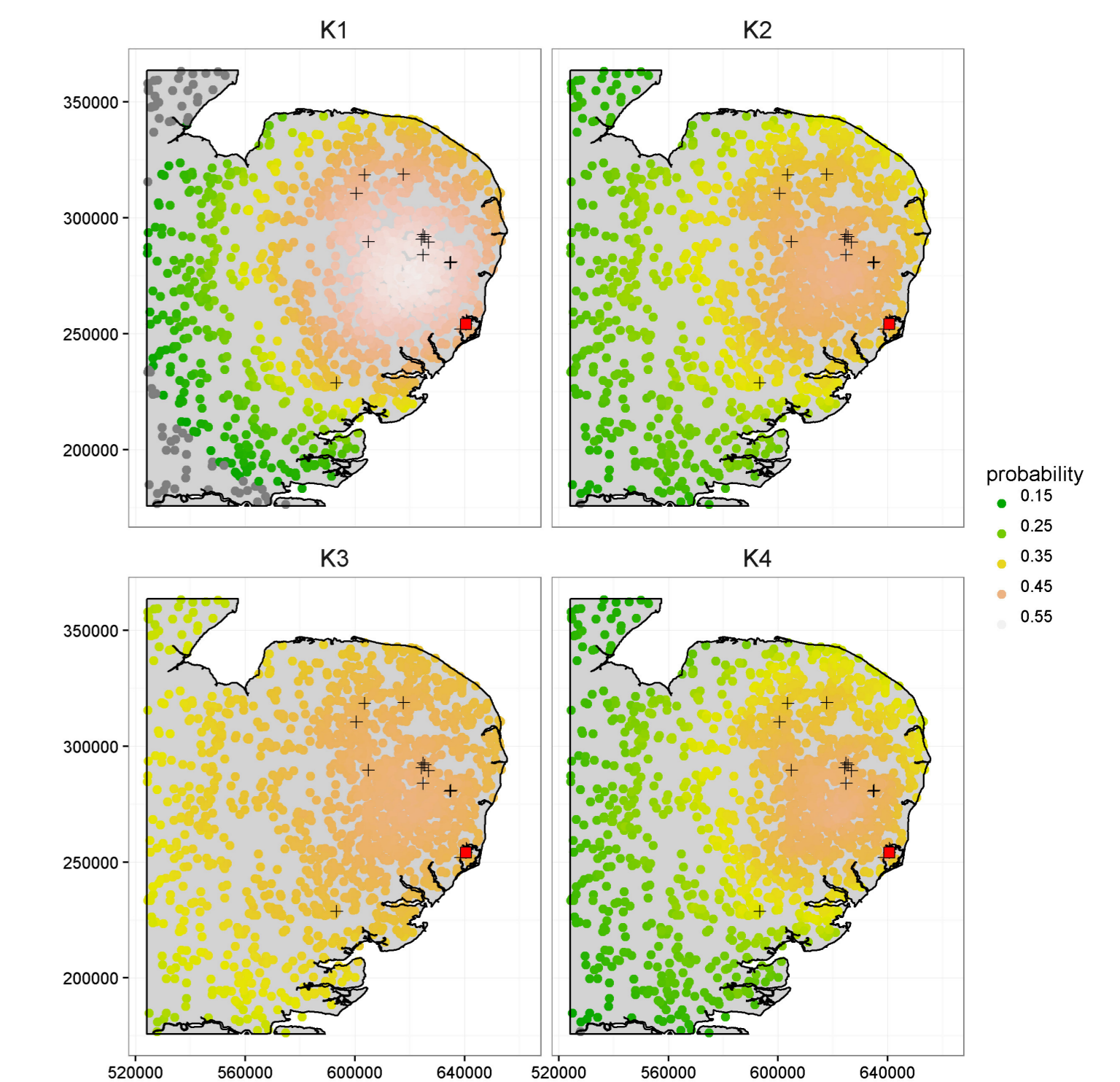
K_4 with 95% CI

Model selection

- DIC and proportion of p-values less than 5% ($\Pr(p < 5\%)$) for the LR under the 4 kernels

	DIC ₁	DIC ₂	$\Pr(p < 5\%)$
K_1	429.301	156.682	27.78%
K_2	317.229	157.021	10.67%
K_3	352.671	156.055	32.83%
K_4	411.096	157.703	19.47%

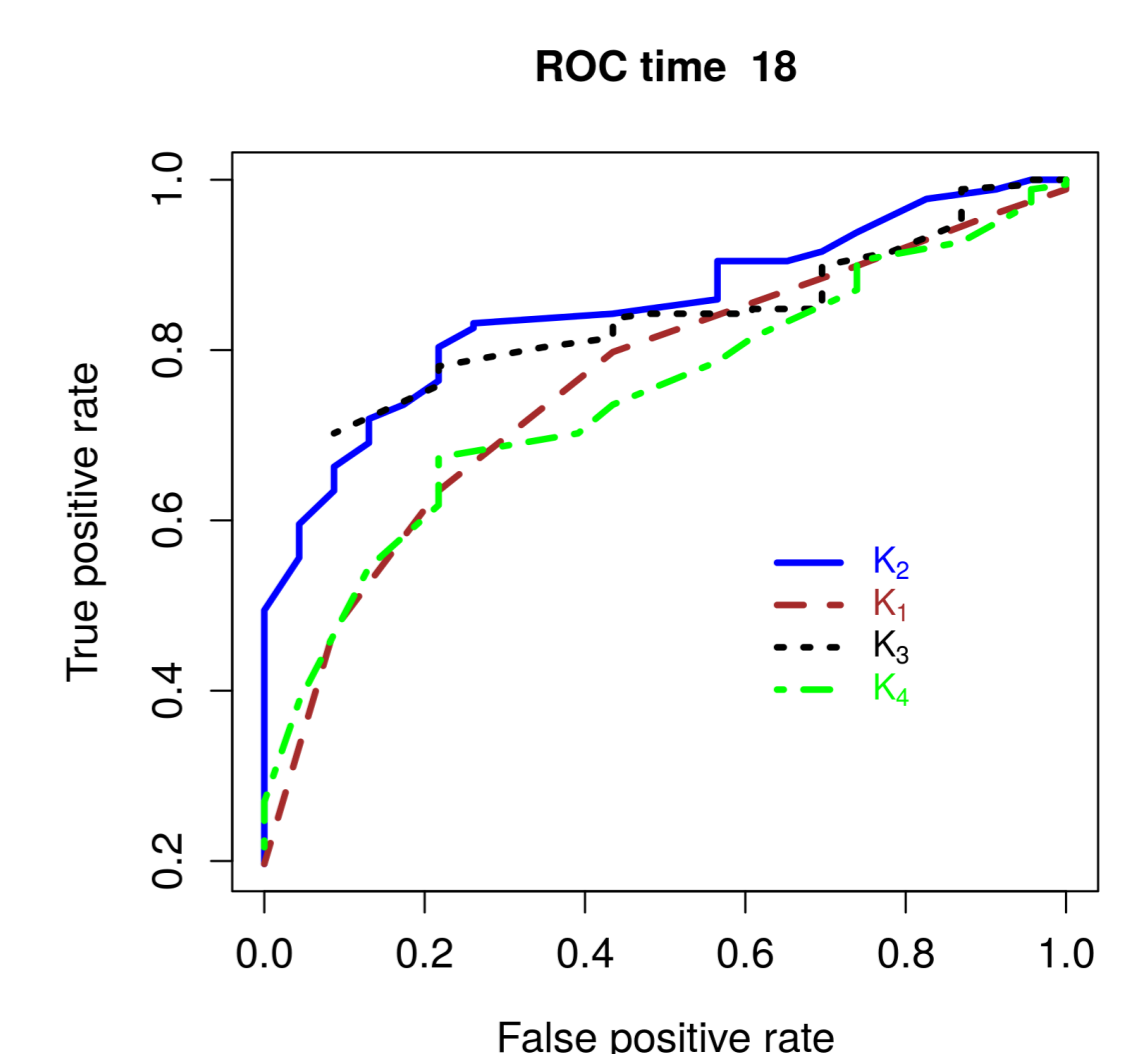
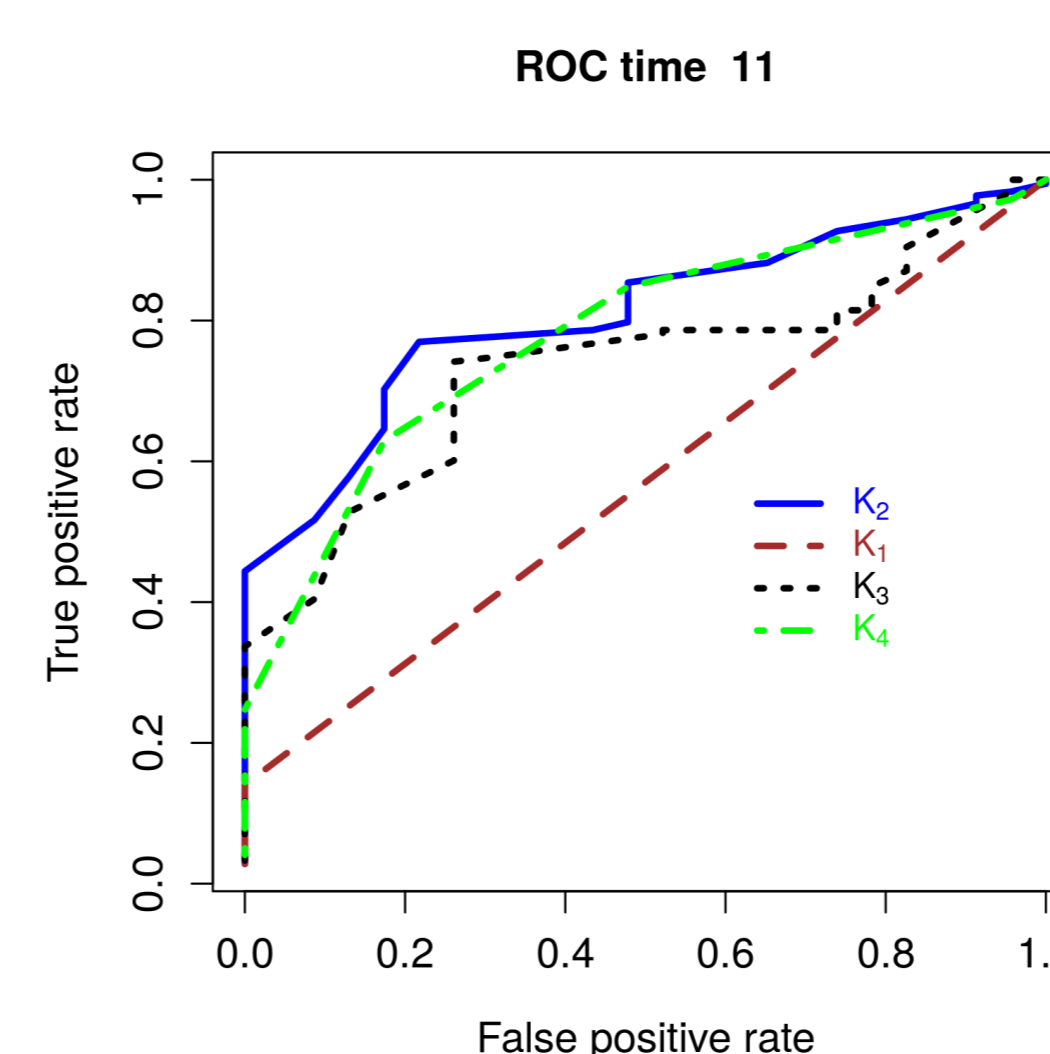
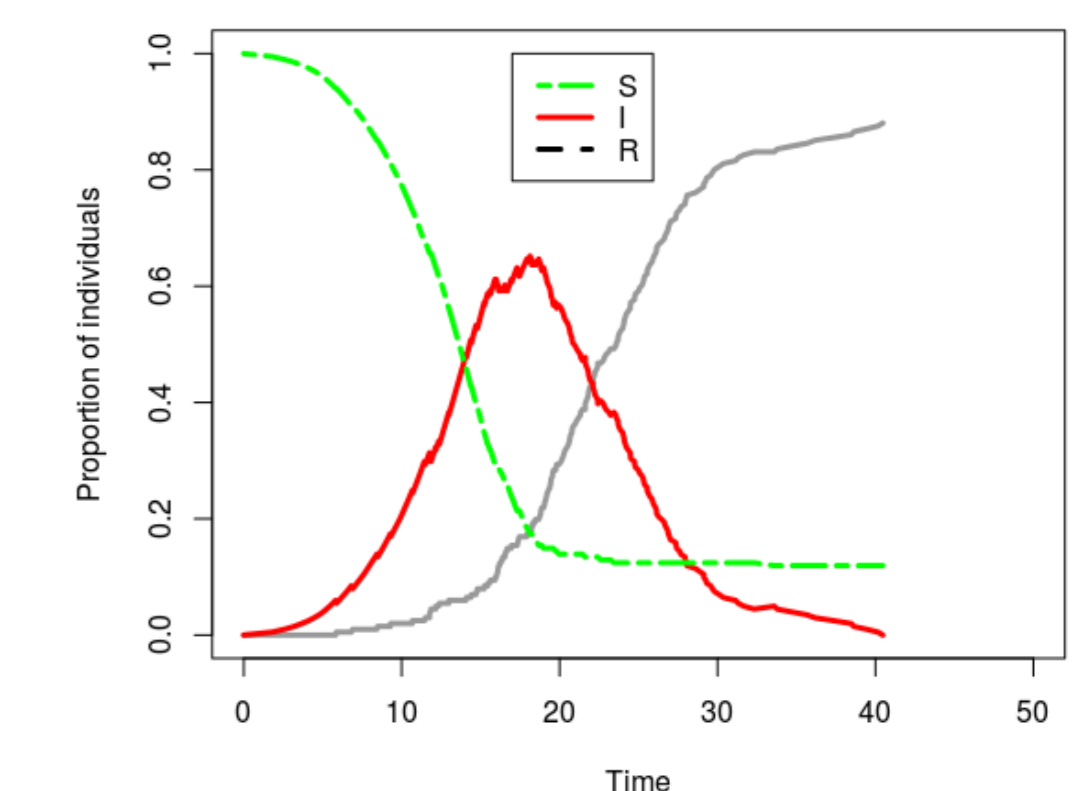
LR assessment correlates better with predicted risk



Predicted risk maps under the 4 kernels

Inference during early phase of outbreak

- ▶ Predict future evolution of outbreak under various kernels
- ▶ Assess using ROC curve
- ▶ More observations imply better prediction



LR identify the correct kernel (K_2) and all kernels with similar prediction

Summary

- ▶ Infer outbreak characteristics from small and ongoing outbreaks e.g. 16 CSF cases
- ▶ Novel model selection tools based on LR allow selection of models e.g. kernels
- ▶ LR approach leads to more reliable risk assessment

Reference

K Gamado, G Marion & T Porphyre. Data-Driven Risk Assessment from Small Scale Epidemics: Estimation and Model Choice for Spatio-Temporal Data with Application to a Classical Swine Fever Outbreak. *Frontiers in Veterinary Science* (2017) 4:16

Acknowledgment

