Prequential Approaches for the Assessment of Stochastic Spatio-temporal Models for Infectious Disease Data

Malesios C[1]., Demiris N[2]. and Koutroumanidis Th[1].
[1]Department of Agricultural Development, Democritus University of Thrace, 193 Pantazidou str., Orestiada, Greece
[2]Department of Statistics, Athens University of Economics and Business, 76, Patission Str., Athens

Introduction

➢ Despite their theoretical advantages, prequential methods (*Dawid, 1984*) for model selection and comparison have not been extensively explored.

➢ Only requirement: models under consideration are able to generate predictions for the next observation.

➢ We investigate the advantages of one-step-ahead predictions for assessing model fit in a Bayesian framework.

➢ The methodology is applied to (sequential in nature) sheep pox and foot and mouth disease epidemic data.

Prequential methodology

➢ Examines each model's prediction for the next data point x_{n+1} after fitting the model to data $x_1, x_2,...,x_n$.

➢ Criterion for model selection: distances, known as scoring rules, measuring the discrepancy between the forecasts and the data.

➢ The best model is the one with the smallest average score among the entertained models.

➢ We utilize four scoring rules for the evaluations:

Single-valued predictions:

$$AES = \left| x_t - \mu_{p_t} \right|$$

$$SES = \left(x_t - \mu_{p_t} \right)^2$$

Probabilistic predictions:

$$\log S = S\left(x_t, F_t \right) = -\log q(x_t)$$

$$RPS = \sum_{K=0}^{\infty} \left(P\left(X \leq k \right) - 1\left(x \leq k \right) \right)^2$$

Spatio-temporal models

Let y_i be the number of occurrences of infected farms at time t_i. Poisson, NB and ZIP/ZINB models are special cases of:

$$y_i \sim g\left(y_i \mid \theta_i, p_i \right)$$

$$g\left(y_i \mid \theta_i, p_i \right) = p_i I_{\{y_i=0\}} + \left(1 - p_i \right) f\left(y_i \mid \theta_i \right)$$

$$\theta_i = h\left(\mu_i \right) = \exp\left(\mu_i \right)$$

$$d\lambda_t = \phi\left(\lambda_t - \mu_t \right) dt + dB_t$$

$$\mu_t = \mathbf{X}_{(i)} \cdot \mathbf{\Theta}_{\beta} + \theta_\tau \cdot y_{(i-1)} + K\left(d_i, \mathbf{\Theta}_K \right)$$

$$\mathbf{\Theta} = \left(\mathbf{\Theta}_{\beta}, \theta_\tau, \mathbf{\Theta}_K \right)^t$$

with model parameters including environmental factors, transmission kernel functions and the mean reversion rate of the O-U process .

Implementation details

➢ Inference and sampling from the predictive distributions was performed in WinBUGS.

➢ Prequential analysis was done in MatBUGS, a tool developed for combining WinBUGS and Matlab.

Results

FMD data:

| Model	$\overline{SES}$	$\overline{AES}$	$\overline{\log S}$	$\overline{RPS}$	$\overline{D}$	DIC
Poisson	0.824	0.603	0.933	0.824	142	167.6
Negative binomial	0.914	0.633	0.956	0.914	146.2	174.3
ZIP$_h$	**0.798**	**0.451**	**0.491**	**0.700**	113.4	---
ZINB	0.924	0.597	1.109	0.851	116.4	---
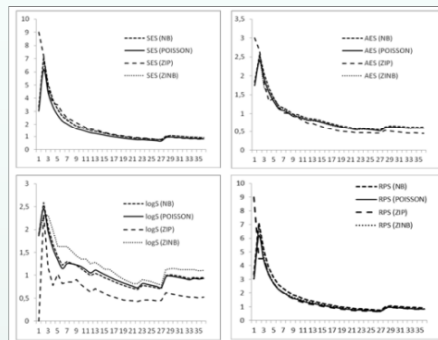
Series of one-step-ahead average prediction error estimates per mean prediction score for each data point

The analysis did not confirmed theoretical claims, such as high sensitivity of RPS to distance (*Czado et al., 2009*), or sensitivity of logS to extreme cases (*Gneiting and Raftery, 2007*)

Sheep pox data:

Model	$\overline{SES}$	$\overline{AES}$	$\overline{\log S}$	$\overline{RPS}$	$\overline{D}$	DIC
Poisson	7.549	**1.242**	5.115	7.569	288.9	324.7
NB	6.746	1.422	8.636	10.338	276.48	329.6
ZIP$_h$	**5.008**	1.295	**0.568**	**5.972**	228	---
ZINB	5.818	1.275	0.591	6.049	277	---
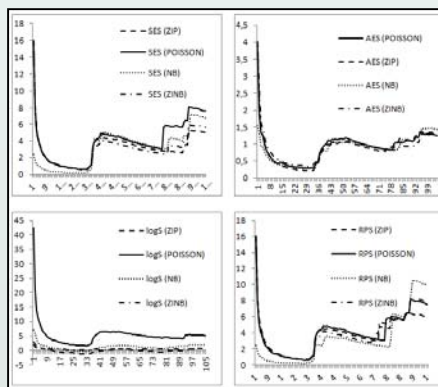
Discussion

➢ Assessment based upon the scoring rules indicates the superiority of the ZIP distributional assumption.

➢ Main difference with the deviance-based criteria: scoring rules tend to select simpler models.

➢ A principled advantage of the prequential approach is that it respects the time ordering of the data.

➢ Allows for the temporal assessment of each model's performance.

➢ We prefer the probabilistic one-step-ahead predictions based on logS and RPS since they properly account for uncertainty.

➢ Efficient (sequential) calculation of predictions represents an important area for future research.