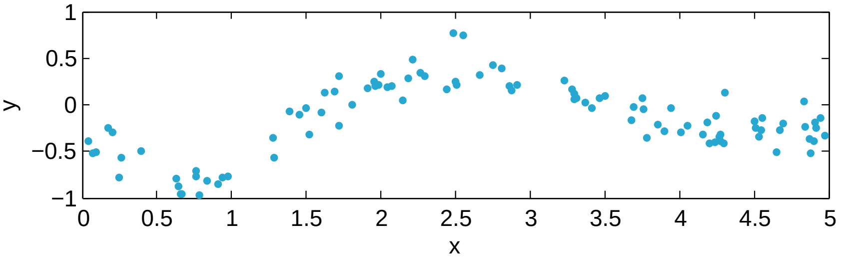
Quantile Emulation for Non-Gaussian Stochastic Models

R. Barillec, A. Boukouvalas, D. Cornford

UQ2012, Raleigh, NC, April 2-5, 2012

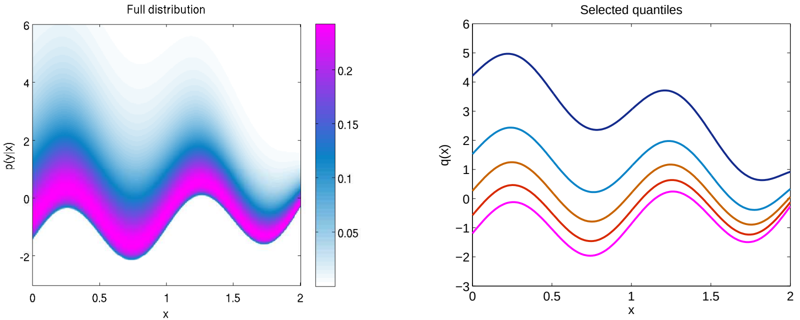
Outline

Stochastic emulation



- We consider a stochastic simulator: $f(x) \sim p(y|x)$
- The simulator is run at a set of inputs $\mathbf{x} = (x_1, \ldots, x_N)$ and produces a set of scalar outputs $\mathbf{y} = (y_1, \ldots, y_N)$ where y_i is a realisation from $p(y|x_i)$
- Due to the variability of the simulator output, it is likely a large number of runs will be required for typical applications (e.g. calibration, ...)
  ➡ build an emulator

Emulating distributions



- $p(y|x)$ is usually unknown (but we can sample from it)
- We assume that $p(y|x)$ varies smoothly in x
- Ideally, we would like to emulate the full $p(y|x)$
  ➡ tricky, unless we make an assumption about the shape of $p(y|x)$
- Decision problems often involve a summary of the distribution only (moments, threshold probability, quantiles)
- Quantiles provide a simple, yet flexible way to characterise a distribution

Quantile regression - existing approaches

2 main trends:

- Model-based approach
  - Estimate the full $p(y|x)$ and invert the CDF to retrieve the quantiles
  - Several possible options: Dirichlet process[1], indicator cokriging[2], Gaussian process[3]
  - Allows for a full Bayesian treatement
  - Often requires expensive Monte-Carlo parameter estimation

[1] Taddy and Kottas (2010)
[2] Pardoiguzquiza and Dowd (2005)
[3] Quadrianto et al. (2009)

Quantile regression - existing approaches

2 main trends:

- Direct quantile estimation
  - Based on minimisation of an appropriate loss function to find model parameters
  - Various options: linear model[4], splines[5]...
  - Provides estimates that are not Bayesian (although sometimes claim to be...)

    ➡ Our method falls in this category

---

[4] Yu and Moyeed (2001)

[5] Koenker (2005)

Direct quantile regression

- Generally, the risk (or expected loss) function at a given input x is given by:

$$R(t) = \int L(y - t)\, p(y|x)\, dy$$

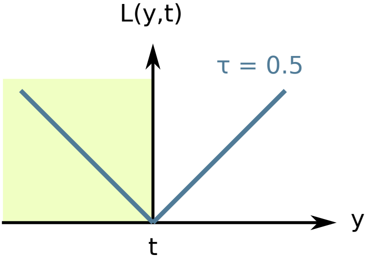
where $L(.)$ is a chosen loss function

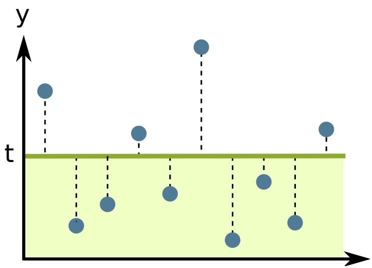
- Different choices of loss function yield different “optimal” estimates:

Loss function	$L(y-t)$	**Optimal estimate**
Square loss	$(y-t)^2$	$E[p(y\|x)]$
Absolute loss	$\|y-t\|$	median of $p(y\|x)$

- The quantile can be defined as the optimal solution to a risk minimisation problem for a specific choice of loss function

Tilted absolute loss function

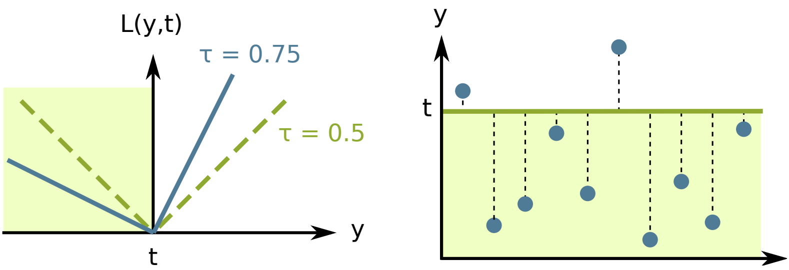




- The result on the median can be extended to any quantile τ by using a "tilted" absolute loss function:

$$L_\tau(y-t) = \begin{cases} \tau \times (y-t) & \text{if } y \geq t \\ -(1-\tau) \times (y-t) & \text{if } y < t \end{cases}$$

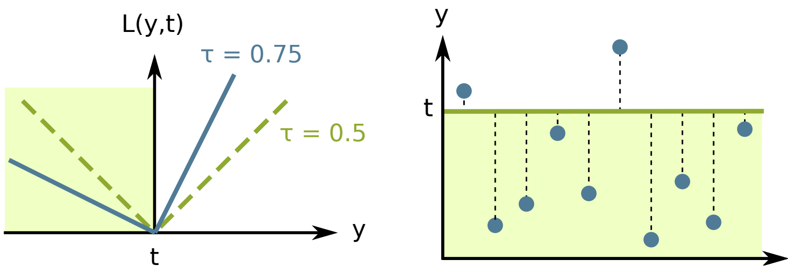
Tilted absolute loss function



- The result on the median can be extended to any quantile τ by using a "tilted" absolute loss function:

$$L_\tau(y-t) = \begin{cases} \tau \times (y-t) & \text{if } y \geq t \\ -(1-\tau) \times (y-t) & \text{if } y < t \end{cases}$$

Tilted absolute loss function



Property

Minimising the expected loss

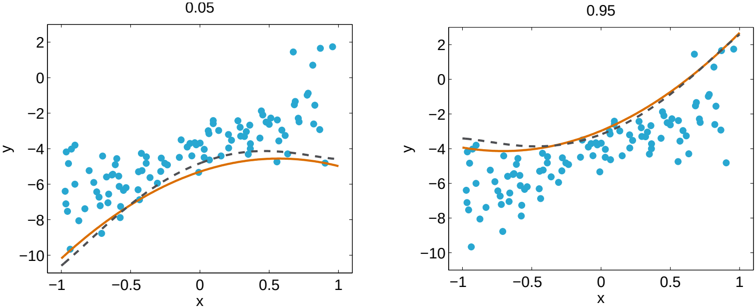
$$\int L_\tau(y-t)p(y|x)dy$$

yields the τ-th quantile of $p(y|x)$.

Direct quantile regression

- Direct quantile regression typically involves:
  1. Choosing a class of models t (e.g. polynomial) depending on some parameters θ
  2. Minimising the empirical expected loss to find the optimal $\hat{\theta}$

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{N} L_\tau(y_i - t_i)$$



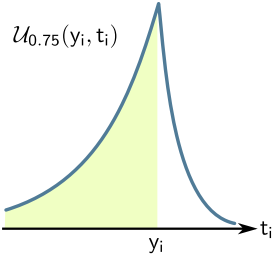
Aston University

"Bayesian" quantile regression

- A mathematically equivalent formulation consists in using the utility function:

$$\mathcal{U}_\tau(\mathbf{y},\mathbf{t}) = \prod_{i=1}^{N} \frac{1}{Z_i} \exp\left[-L_\tau(y_i - t_i)\right]$$

which is a product of Asymmetric Laplace distributions



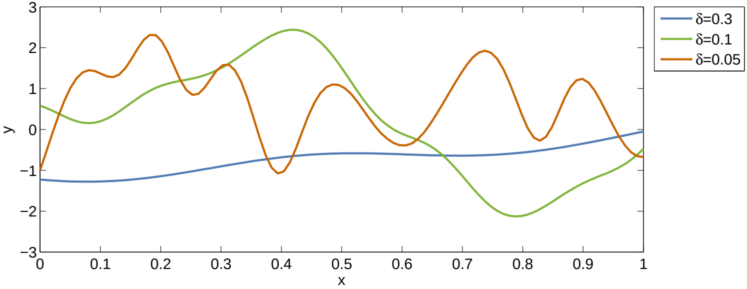
- $\mathcal{U}_\tau$ is sometimes referred to as a "likelihood" leading to a maximum likelihood formulation of the problem

Gaussian Process quantile regression

- We put a Gaussian process prior on the quantile regression functions

$$p(\mathbf{t}) = GP\left[0, c(\mathbf{x}, \mathbf{x}')\right]$$

- $c(\mathbf{x}, \mathbf{x}')$ is a covariance function (i.e. kernel) with parameters $\theta = (\sigma, \delta)$



Gaussian Process quantile regression

- We train the model by maximising the empirical utility directly

$$\operatorname*{argmax}_{\theta} \int_{\mathbf{t}} \mathcal{U}_{\tau}(\mathbf{y}, \mathbf{t}) p(\mathbf{t}) \mathrm{d}\mathbf{t} \tag{1}$$

where

$$\mathcal{U}_{\tau}(\mathbf{y}, \mathbf{t}) = \frac{1}{Z} \exp \left[- \sum_{i=1}^{N} L_{\tau}(y_i - t_i) \right] \tag{2}$$

- Unfortunately, integral (1) is not analytically tractable
- However, because of the independence assumption of the utility, we can employ the Expectation Propagation (EP) algorithm[6] to approximate the product in the integral

[6] Minka (2001)

Expectation Propagation (EP)

- What we want (but can't have):

$$Z = \int p(\mathbf{t}) \prod_i \pi(y_i | t_i) \, d\mathbf{t}$$

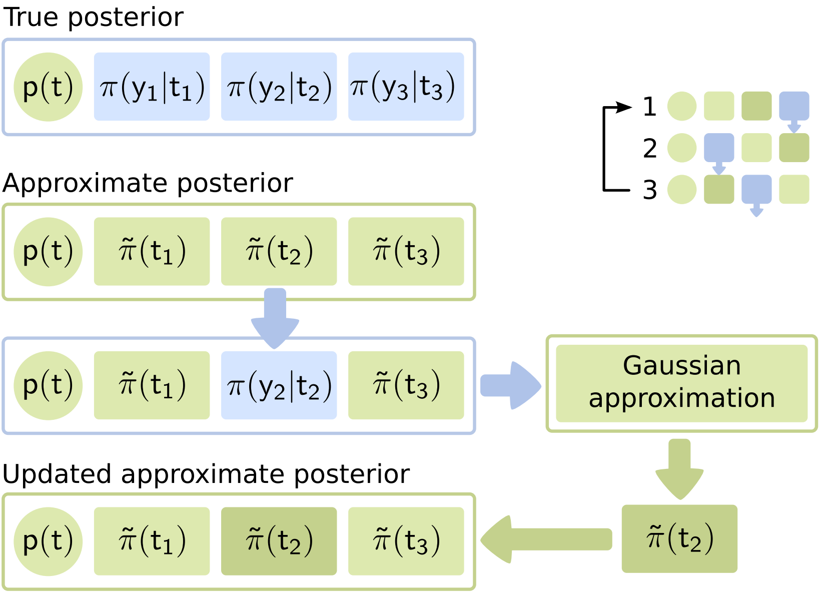
- EP approximates a posterior distribution of the form:

$$p(\mathbf{t}|\boldsymbol{y}) = \frac{1}{Z} p(\mathbf{t}) \prod_i \pi(y_i | t_i)$$

with one of the form:

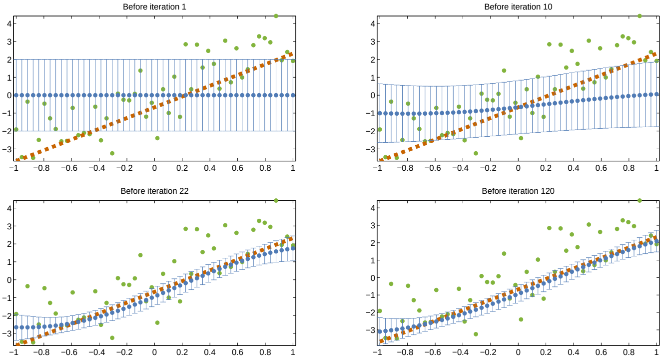
$$u(\mathbf{t}|\boldsymbol{y}) = \frac{1}{Z_{EP}} p(\mathbf{t}) \prod_i \tilde{\pi}(t_i)$$

where the sites $\tilde{\pi}(t_i) = \tilde{Z}_i \mathcal{N}(\tilde{\mu}_i, \tilde{\sigma}_i^2)$ are unnormalised Gaussian functions.


True posterior
$p(t)$ $\pi(y_1|t_1)$ $\pi(y_2|t_2)$ $\pi(y_3|t_3)$
Approximate posterior
$p(t)$ $\tilde{\pi}(t_1)$ $\tilde{\pi}(t_2)$ $\tilde{\pi}(t_3)$
$p(t)$ $\tilde{\pi}(t_1)$ $\pi(y_2|t_2)$ $\tilde{\pi}(t_3)$
Gaussian approximation
$\tilde{\pi}(t_2)$
Updated approximate posterior
$p(t)$ $\tilde{\pi}(t_1)$ $\tilde{\pi}(t_2)$ $\tilde{\pi}(t_3)$
1
2
3

Parameter inference

- Repeat until convergence:
  1. Run the EP algorithm to convergence
     ➡ Get a tractable approximation to the expected utility
  2. Maximise the expected utility with respect to θ



Prediction

- The predictive expectation at a new point x for the quantile is:

$$\begin{aligned} \mathrm{E}[t|\mathbf{x},x] &= c(\mathbf{x},x)^T c(\mathbf{x},\mathbf{x})^{-1}\mu \\ &= c(\mathbf{x},x)^T \left[c(\mathbf{x},\mathbf{x}) + \tilde{\Sigma} \right]^{-1} \tilde{\mu} \end{aligned}$$

  where $\mathbf{x}$ is the training data
- We note here that the prediction is on the latent variables for the quantile τ and not for the noisy observations y
- The predictive “variance” for the quantile would be:

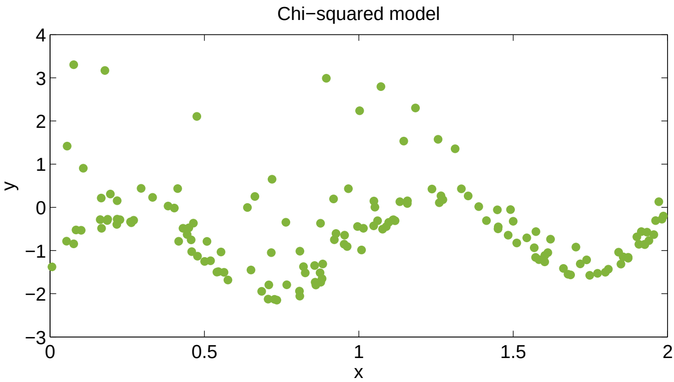
$$\mathrm{Var}[t|\mathbf{x},x] = c(x,x)^T - c(\mathbf{x},x)^T \left[c(\mathbf{x},\mathbf{x}) + \tilde{\Sigma} \right]^{-1} c(\mathbf{x},x)$$

- In our framework, the “variance” cannot be interepreted in the usual sense as we did not define a proper likelihood for y

Some results on synthetic data

- We look at different examples to test the Quantile GP (QGP) regression
- We compare against spline regression (Koenker, 2006 – R package `quantreg`)
- We look at 5 quantiles independently: $\tau \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$

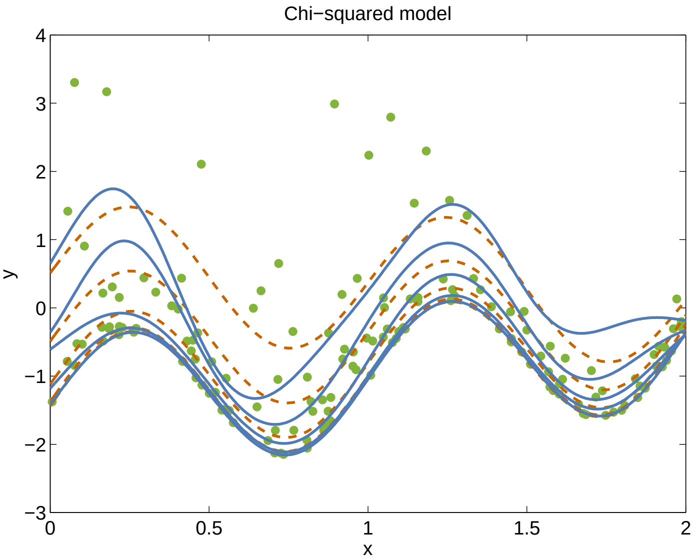
Example 1: Heteroscedastic $\mathcal{X}^2$ model



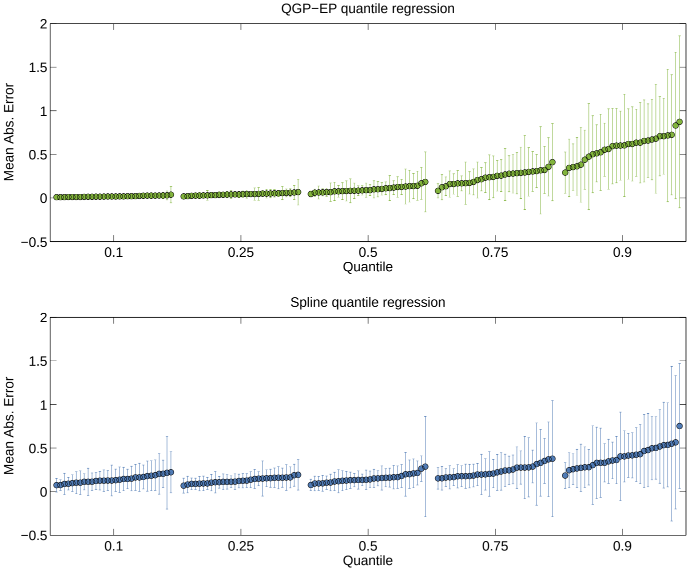
$$p(y|x) = \sin(2\pi x) + \sigma(x)(\mathcal{X}_1^2 - 2) \quad (3)$$

$$\sigma(x) = \sqrt{\frac{2.1 - x}{4}} \quad (4)$$

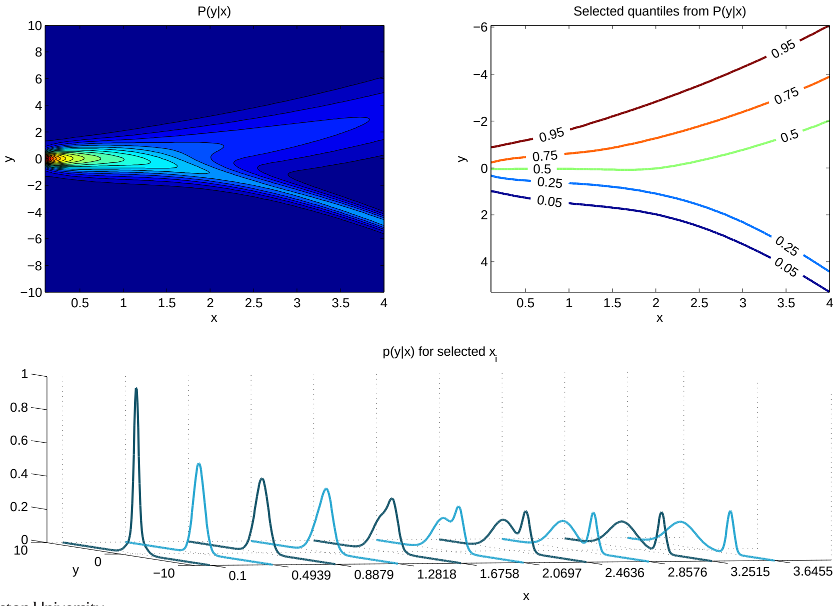
Example 1: Heteroscedastic χ^2 model



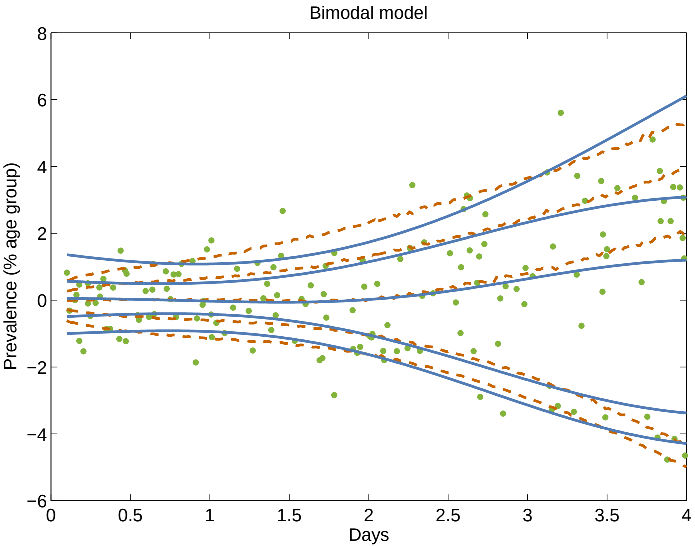
Example 1: Heteroscedastic $\mathcal{X}^2$ model



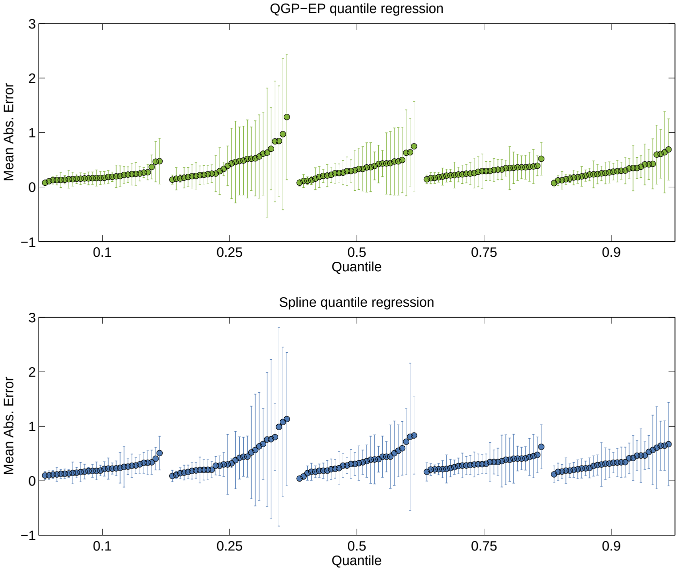
Example 2: Bimodal model



Example 2: Bimodal model

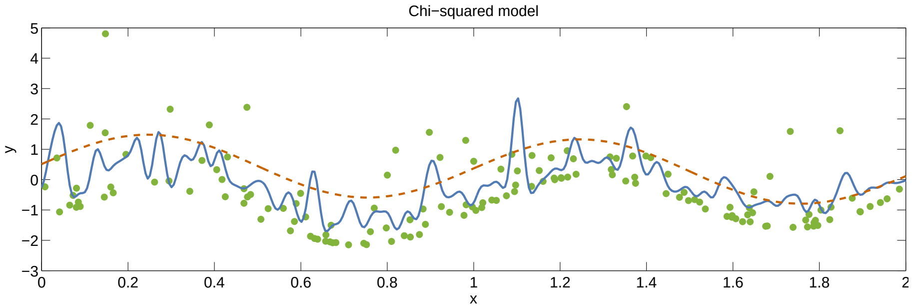


Example 2: Bimodal model



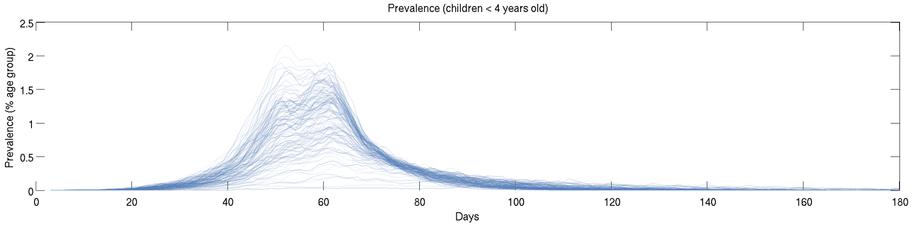
Aston University

Some remarks



- Quantiles in regions on low probability are difficult to estimate
- It is possible to use regularization to prevent rough estimates
- Looking at multiple quantiles jointly might also help in that case

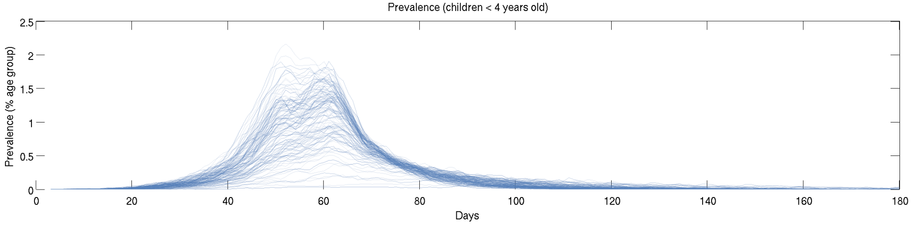
FluTe epidemic simulation model



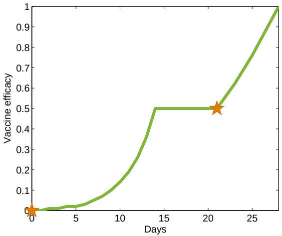
- The FluTe model[7] is a stochastic epidemic simulator
- Incorporates US statistical data on population, travel, etc.
- Transmission of disease based on interactions (home, work, school, trips...)
- Publicly available

[7] Chao et al. (2010)

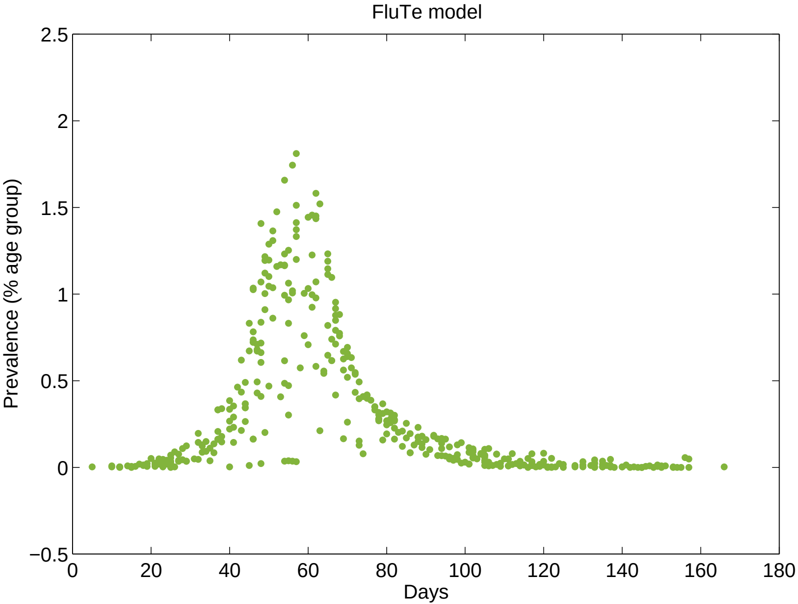
Scenario description



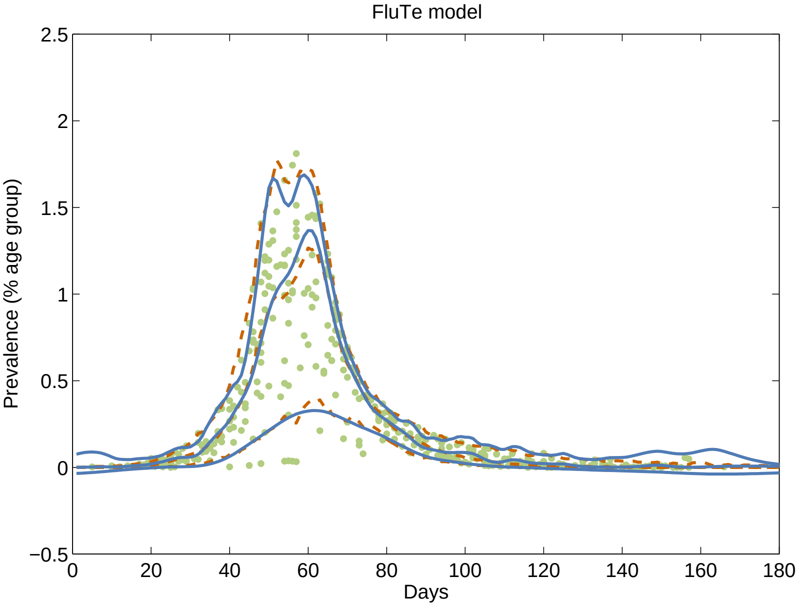
- Mass vaccination (70%) of the Seattle population
- Action taken when 1 person in 1M is infected
- 30 day period to set up vaccination program
- 2-stage vaccine with boost after 21 days
- We look at risk for < 4 year old



Training data



Quantile interpolation



Conclusions

- We introduced a quantile regression method based on the Gaussian process model
- Parameter estimation is performed using the Expectation Propagation algorithm
- The method seems to perform on par with existing methods (splines)
- There are some limitations:
  - Lack of uncertainty estimate on predictions
  - Sensitivity to training data
  - Need for regularisation
- If you have a stochastic model you'd like to try this on, please get in touch!

References

Dennis L. Chao, M. Elizabeth Halloran, Valerie J. Obenchain, and Ira M. Longini, Jr. Flute, a publicly available stochastic influenza epidemic simulation model. PLoS Comput Biol, 6(1): e1000656, 01 2010.

R. Koenker. Quantile Regression. Cambridge University Press, 2005.

T. Minka. Expectation propagation for approximate bayesian inference. In In Proceedings UAI, pages 362–369, 2001.

E. Pardoiguzquiza and P. Dowd. Multiple indicator cokriging with application to optimal sampling for environmental monitoring. Computers & Geosciences, 31(1):1–13, 2005.

N. Quadrianto, K. Kersting, M.D. Reid, T.S. Caetano, and W.L. Buntine. Kernel conditional quantile estimation via reduction revisited. In 2009 Ninth IEEE International Conference on Data Mining, pages 938–943. IEEE, 2009.

M. A. Taddy and A. Kottas. A Bayesian Nonparametric Approach to Inference for Quantile Regression. Journal of Business and Economic Statistics, 28(3):357–369, 2010.

K. Yu and R. A. Moyeed. Bayesian quantile regression. Statistics & Probability Letters, 54: 437–447, 2001.

Thank you!