Bayesian spatial quantile regression

Brian J. Reich and Montserrat Fuentes
North Carolina State University

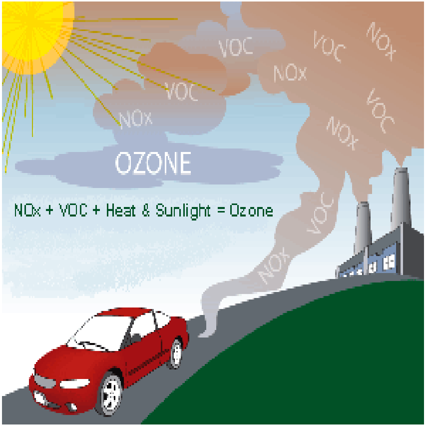
and

David B. Dunson
Duke University

E-mail:reich@stat.ncsu.edu

Tropospheric ozone

- Tropospheric ozone has been linked with several adverse health effects and is regulated by the EPA.
- Ozone is primarily a secondary pollutant.



Meteorology/ozone

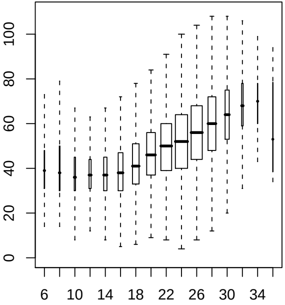
- Ozone has a complicated relationship with meteorology.
- Ozone is usually the highest on hot sunny days with low wind.
- Due to the strong dependence on weather conditions, ozone levels may be sensitive to climate change.
- There is great interest in studying the potential effect of climate change on ozone levels, and how this change may affect regulation and public health.

Our objectives are to:

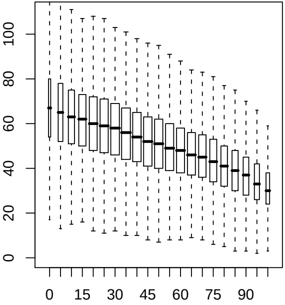
1. Build a statistical model for daily ozone as a function of daily weather.
    - Relationships are nonlinear and not restricted to the mean
    - Bayesian spatial quantile regression
2. Inspect spatial and temporal trends in daily ozone
    - Ozone is generally decreasing
    - Compare trends in the center and tails, and across space
3. Use our statistical model to investigate the effects of climate change on ozone
    - Generate replications under different climates
    - Discuss changes in the median ozone and the probability of non-compliance.

Quantile regression for ozone

- Ozone often exhibits complicated relationships with weather.
- Also, the EPA currently regulates the fourth highest value of the year (99^{th} quantile) so it is important to accurately model the entire distribution.



Temperature (C)



Cloud cover (%)

Density regression

- Let $f(y|\mathrm{x})$ be the density of ozone given covariates x (weather, lat, long, etc.).
- There are several NP Bayes models for the conditional distribution (Gelfand et al., 2005; Griffin and Steel, 2006; Reich and Fuentes, 2007; Dunson and Park, 2008).
- Many of these models are infinite mixtures with mixture probabilities that depend on x.
- Although these models are quite flexible, one drawback is the difficulty in interpreting the effects of each covariate, for example, whether there is a statistically significant time trend in the distribution's upper tail probability.

Quantile regression

- As a compromise between fully-general Bayesian density regression and the usual additive mean regression, we propose a Bayesian spatial quantile regression model.

- Quantile regression models the distribution's quantiles as additive functions of the predictors.

- This additive structure permits inference on the effect of individual covariates on the response's quantiles.

- The additive structure also permits density regression for high dimensional x.

Quantile regression

- Let y_i be the data and $\mathrm{X}_i = (X_{i1}, ..., X_{ip})'$ be the covariates.
- Quantile regression models y_i's conditional density via its quantile function (inverse CDF) $q(\tau|\mathrm{X}_i, \mathrm{s}_i)$, defined so that

$$P\{y_i < q(\tau|\mathrm{X}_i)\} = \tau \in [0, 1].$$

- We model $q(\tau|\mathrm{X}_i)$ as

$$q(\tau|\mathrm{X}_i) = \mathrm{X}_i'\boldsymbol{\beta}(\tau)$$

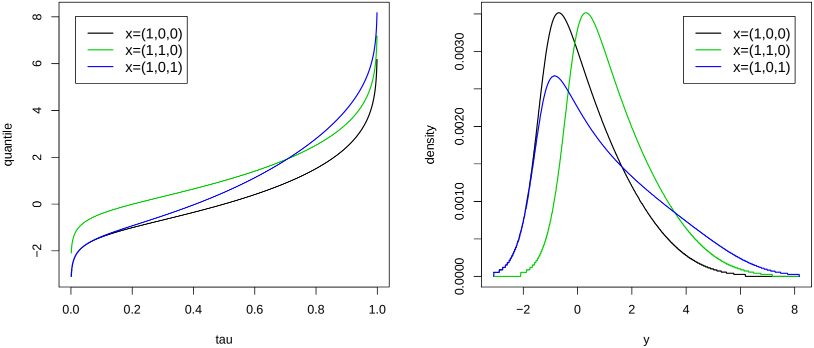
  where $\boldsymbol{\beta}(\tau) = (\beta_1(\tau), ..., \beta_p(\tau))'$ are the coefficients for the τ^{th} quantile level.

- For example, $\beta_j(0.5)$ and $\beta_j(0.99)$ are the effects of the j^{th} covariate on the median and upper tail probability, respectively.

Example

$$q(\tau|\mathrm{X}_i) = (\tau+1)\Phi^{-1}(\tau)X_{i1} + X_{i2} + 2\tau^2 X_{i3}$$

- $\beta_1(\tau) = (\tau+1)\Phi^{-1}(\tau)$ (Φ is the normal CDF)
- $\beta_2(\tau) = 1$
- $\beta_3(\tau) = 2\tau^2$



Classical quantile regression

The usual quantile regression estimate is

$$\hat{\boldsymbol{\beta}}(\tau_k) = \arg\min_{\boldsymbol{\beta}} \sum_{y_i > \mathrm{X}_i'\boldsymbol{\beta}} \tau_k |y_i - \mathrm{X}_i'\boldsymbol{\beta}| + \sum_{y_i < \mathrm{X}_i'\boldsymbol{\beta}} (\tau_k - 1)|y_i - \mathrm{X}_i'\boldsymbol{\beta}|.$$

- Different quantiles are analyzed separately.
- There is no model for the quantile process.
- We propose a model-based approach to analyze all quantile levels simultaneously, and extend this to the spatial setting.

Bayesian model with no covariates

- First assume $p = 1$ and $X_{i1} = 1$ so $q(\tau|\mathrm{X}_i) = \beta(\tau)$.

- Let

$$\beta(\tau) = \sum_{m=1}^{M} B_m(\tau)\alpha_m,$$

  where $B_m(\tau)$ is a known basis function of τ and α_m are unknown coefficients that determine the shape of the quantile function.

- We use Bernstein basis polynomials

$$B_m(\tau) = \binom{M}{m} \tau^m (1-\tau)^{M-m}.$$

Model with no covariates

- The process $\beta_1(\tau)$ must be constructed so that $q(\tau)$ is nondecreasing in τ.
- An attractive feature of these basis functions is that if $\alpha_m \geq \alpha_{m-1}$ for all $m > 1$, then $\beta_1(\tau)$ is an increasing function of τ.
- We reparameterize to $\delta_1 = \alpha_1$ and $\delta_m = \alpha_m - \alpha_{m-1}$ for $m = 2, ..., M$.
- We ensure the quantile constraint by introducing latent unconstrained variable δ_m^* and taking $\delta_1 = \delta_1^*$ and

$$\delta_m = \begin{cases} \delta_m^*, & \delta_m^* \geq 0 \\ 0, & \delta_m^* < 0 \end{cases}$$

for $m > 1$.

Centering distribution

- The δ_m^* have independent normal priors $\delta_m^* \sim \mathrm{N}(\bar{\delta}_m(\Theta), \sigma^2)$, with unknown hyperparameters Θ.
- We pick $\bar{\delta}_m(\Theta)$ to center the quantile process on a parametric distribution $f_0(y|\Theta)$, for example, a $\mathrm{N}(\mu_0, \sigma_0)$ random variable with $\Theta = (\mu_0, \sigma_0)$.
- Letting $q_0(\tau|\Theta)$ be the quantile function of $f_0(y|\Theta)$, the $\bar{\delta}_m(\Theta)$ are then chosen (ridge regression) so that
$$q_0(\tau|\Theta) \approx \left[\sum_{m=1}^{M} B_m(\tau) \sum_{l=1}^{m} \bar{\delta}_l(\Theta)\right].$$
- As $\sigma \rightarrow 0$, the quantile functions are increasing shrunk towards the parametric quantile function $q_0(\tau|\Theta)$, and the likelihood is similar to $f_0(y|\Theta)$.

Quantile regression with covariates

- Adding covariates, the conditional quantile function becomes

$$q(\tau|\mathrm{X}_i) = \mathrm{X}_i\boldsymbol{\beta}(\tau) = \sum_{j=1}^{p} X_{ij}\beta_j(\tau).$$

- We assume

$$\beta_j(\tau) = \sum_{m=1}^{M} B_m(\tau)\alpha_{jm},$$

  where α_{jm} are unknown coefficients.

- The processes $\beta_j(\tau)$ must be constructed so that $q(\tau|\mathrm{X}_i)$ is nondecreasing in τ for all X_i.

- As before, we reparameterize to $\delta_{j1} = \alpha_{j1}$ and $\delta_{jm} = \alpha_{jm} - \alpha_{jm-1}$ for $j = 2, ..., M$ to impose constraints on latent variables δ^*_{jm} .

Quantile regression with covariates

- Collecting terms with common basis functions gives

$$\mathrm{X}_i'\boldsymbol{\beta}(\tau,\mathrm{s}) = \sum_{m=1}^{M} B_m(\tau)\theta_m(\mathrm{X}_i), \tag{1}$$

  where $\theta_m(\mathrm{X}_i) = \sum_{j=1}^{p} X_{ij}\alpha_{jm}$.

- Therefore, if $\theta_m(\mathrm{X}_i) \geq \theta_{m-1}(\mathrm{X}_i)$ for all $m > 1$, then $\mathrm{X}_i\boldsymbol{\beta}(\tau)$, and thus $q(\tau|\mathrm{X}_i)$, is an increasing function of τ.

- To specify our prior for the α_{jm} to ensure monotonicity, we assume that $X_{i1} = 1$ for the intercept and the remaining covariates are suitably scaled so that $X_{ij} \in [0,1]$ for $j > 1$.

- Since the constraints are written in terms of the difference between adjacent terms, we reparameterize to $\delta_{j1} = \alpha_{j1}$ and $\delta_{jm} = \alpha_{jm} - \alpha_{jm-1}$ for $j = 2, ..., M$.

Quantile regression with covariates

- We ensure the quantile constraint by introducing latent unconstrained variable $\delta_{jm}^* \sim \mathrm{N}(\bar{\delta}_{jm}(\Theta), \sigma_j^2)$ and taking

$$\delta_{jm} = \begin{cases} \delta_{jm}^*, & \delta_{1m}^* + \sum_{j=2}^{p} I(\delta_{jm}^* < 0)\delta_{jm}^* \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

- We center the intercept curve on a parametric quantile function $q_0(\Theta)$. The remaining coefficients have $\bar{\delta}_{jm}(\Theta) = 0$ for $j > 1$.

- We have assumed that the quantile process is a linear function of the covariates, simplifying interpretation.

- Transformations of the original predictors such as interactions or basis functions can be added to give a more flexible model.

Bayesian spatial quantile regression

- We let the quantile functions vary spatially by letting

$$\beta_j(\tau, \mathrm{s}) = \sum_{m=1}^{M} B_m(\tau)\alpha_{jm}(\mathrm{s}),$$

  where $\alpha_{jm}(\mathrm{s})$ are spatially-varying basis function coefficients.

- As before, we transform to $\delta^*_{jm}(\mathrm{s})$ and enforce monotonicity constraints at each spatial location.

- The $\delta^*_{jm}(\mathrm{s})$ are independent (over j and m) Gaussian spatial processes with $E(\delta^*_{jm}(\mathrm{s})) = \bar{\delta}_{jm}(\Theta)$ and

$$Cov(\delta^*_{jm}(\mathrm{s}), \delta^*_{jm}(\mathrm{s}')) = \sigma_j^2 \exp\left(-||\mathrm{s} - \mathrm{s}'||/\rho_j\right).$$

- Where σ_j^2 is the variance of $\delta^*_{jm}(\mathrm{s})$ and ρ_j determines the range of the spatial correlation function.

Simulation study

- For each of the $S = 50$ simulated data sets we generate $n = 20$ spatial locations s_i uniformly on $[0,1]^2$.

- The $p = 3$ covariates are generated as $X_1 \equiv 1$ and $X_{i2}, X_{i3} \overset{iid}{\sim}$ U(0,1), independent over space and time.

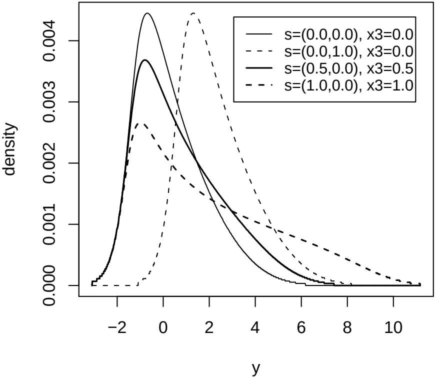
- The true quantile function is

$$q(\tau|\mathrm{X}_i, \mathrm{s}_i) = 2s_{i2} + (\tau + 1)\Phi^{-1}(\tau) + \left(5s_{i1}\tau^2\right) X_{i3},$$

- which implies that
    - $\beta_1(\tau, \mathrm{s}_i) = 2s_{2i} + (\tau + 1)\Phi^{-1}(\tau)$
    - $\beta_2(\tau, \mathrm{s}_i) = 0$
    - $\beta_3(\tau, \mathrm{s}_i) = 5s_{1i}\tau^2$

Simulation study

- The mean is higher in the north.
- X2 is null ($\beta_2(\tau, s_i) = 0$)
- X3 increases the probability of extreme events, especially in the west ($\beta_3(\tau, s_i) = 5s_{1i}\tau^2$).



Simulation study

- We compare three methods: usual quantile regression using the quantreg package in R, our Bayesian quantile regression model without spatial covariance, and our full spatial model.
- For each simulated data set and each method we compute point estimates and 90% intervals for $\beta_j(\tau_k, \mathrm{s}_i)$ for $j = 1, 2, 3$, $\tau_k \in \{0.05, 0.10, ..., 0.95\}$, and all s_i.
- We compare methods using mean squared error averaged over space, quantile levels, and simulated data set.
- Specifically, the MSE for the j^{th} quantile function is

  $$MSE_j = \frac{1}{SnK} \sum_{sim=1}^{S} \sum_{i=1}^{n} \sum_{k=1}^{K} \left(\hat{\beta}_j(\tau_k, s_i)^{(sim)} - \beta_j(\tau_k, s_i) \right)^2,$$

  where $\hat{\beta}_j(\tau_k, s_i)^{(sim)}$ is the estimate for simulation *sim*.
- Coverage probability is computed similarly.

Priors

- We use $M = 10$ knots.
- We use vague yet proper priors $\sigma_j^2 \sim$ InvG(0.1,0.1) and $\rho_j \sim$ Gamma(0.06,0.75).
- The prior for the ρ_j is selected so the effective range $-\rho_j \log(0.05)$, i.e., the distance at which the spatial correlation equals 0.05, has prior mean 0.25 and prior standard deviation 1.
- The centering distribution f_0 was taken to be skew-normal with location $\mu_0 \sim$ N(0,10^2), scale $\sigma_0^2 \sim$ InvGamma(0.1,0.1), and skewness $\psi_0 \sim$ N(0,10^2).
- We conducted a sensitivity analysis and the results do not appear to be sensitive to these priors.

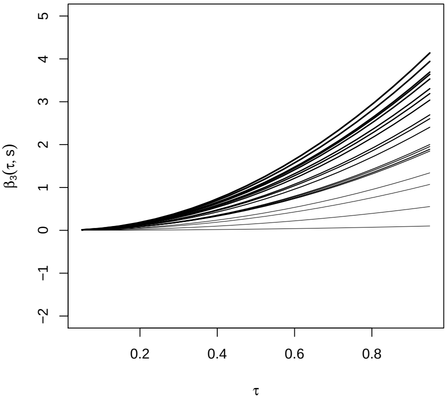
Results

Method	MSE			Coverage		
	β_1	β_2	β_3	β_1	β_2	β_3
Quantreg	0.51	0.99	1.02	0.89	0.89	0.88
Bayes - nonspatial	0.31	0.21	0.62	0.82	0.93	0.85
Bayes - spatial	0.13	0.12	0.27	0.86	0.93	0.88

- Our model has small MSE and good frequentist coverage probability (nominal level is 0.9).
- Borrowing strength across space reduces MSE.
- The posterior mean of $\beta_3(\tau)$ for a representative data set is plotted on the next slides.

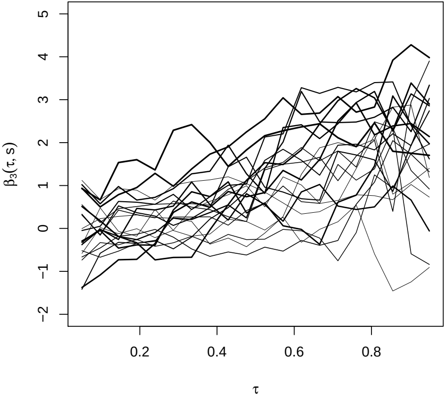
True values

Plot of $\beta_3(\tau, \mathrm{s}_i) = 5s_{1i}\tau^2$; each line is the true curve for one spatial location.



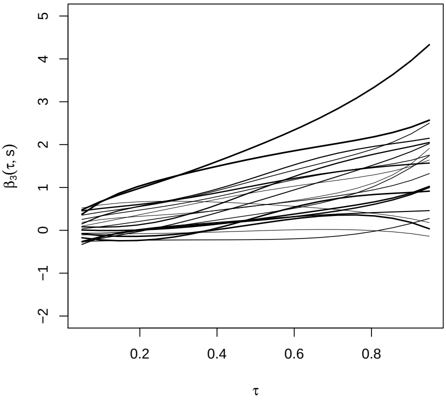
Usual quantile regression estimate

Usual quantile regression does not pool information across quantile level or spatial location.



Our spatial Bayesian estimate (posterior mean)

Our method pools information across quantile level and spatial location.



Analysis of Atlanta ozone data

- We begin by comparing several models using only data from the 12 stations in the Atlanta area.
- We analyze the maximum daily 8-hour average ozone from the summers of 1997-2005.
- There are four predictors: year (linear trend) and the daily average temperature, cloud cover, and wind speed.
- We also include quadratic terms and the interaction between temperature and cloud cover.
- We used the same priors as in the simulation study. This prior for the spatial ranges gives prior 95% intervals (0.00,0.99) and (0.00,0.80) for the correlation between the closest and farthest pairs of points, respectively.

Model comparisons

- We compare our model with the fully-Gaussian spatial model with spatially-varying coefficients,

$$y(\mathrm{s},t) = \sum_{j=1}^{p} X_j(\mathrm{s},t)\beta_j(\mathrm{s}) + \mu(\mathrm{s},t) + \varepsilon(\mathrm{s},t), \tag{2}$$

  where $\beta_j(\mathrm{s})$ and $\mu(\mathrm{s},\cdot)$ are spatial Gaussian processes.

- To compare models we randomly removed all observations for 10% of the days ($N = 910$ total observations), and compute:
  - root mean squared prediction error ($RMSE$)
  - mean absolute prediction deviation (MAD)
  - coverage probability of 95% intervals

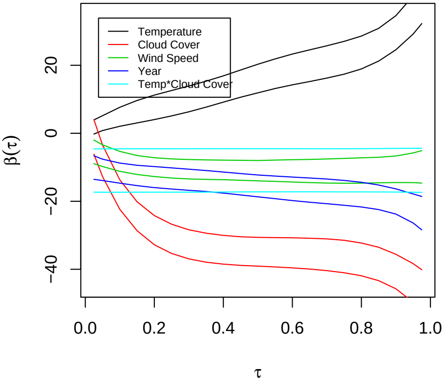
Model comparisons

Model	Covariates	Coverage Prob	RMSE	MAD
Gaussian	Linear	0.965	14.3	9.34
Gaussian	Quadratic	0.969	14.2	9.07
QR - full	Linear	0.987	13.2	7.87
QR - full	Quadratic	0.986	13.0	7.64

- Our model gives the correct coverage probability and more accurate predictions than the Gaussian model.
- Quadratic terms improve prediction.

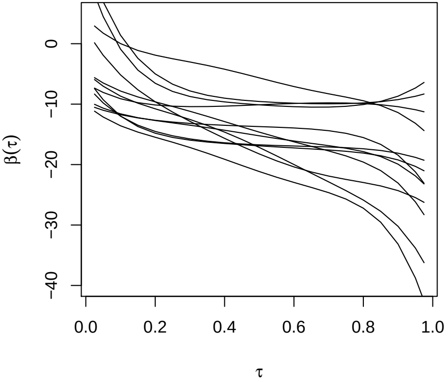
Posterior 95% intervals for the main effects

- Temperature and cloud cover have strong effects on the upper tail.
- There is a negative time trend, especially in the upper tail.



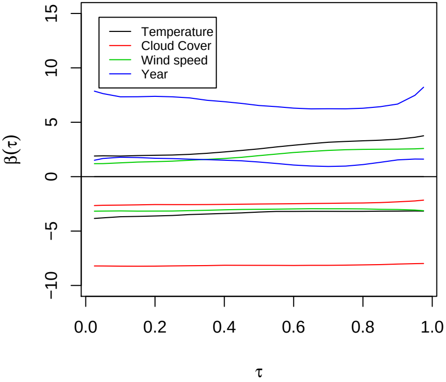
Posterior mean time trend by spatial location

Each line is the quantile function for a location. There is some spatial variability in the linear time trend.



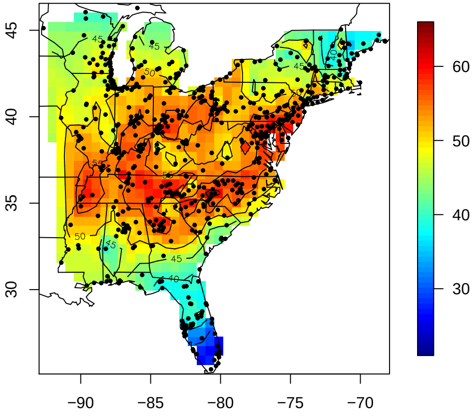
Posterior 95% intervals for the quadratic effects

Cloud cover and year have significant quadratic terms, although they are much weaker than the linear terms.



Eastern US data

We analyze monitoring data for 631 stations in the Eastern US for the summers of 1997-2005. In this plot the color is the median daily ozone (ppb) and the points are the stations.



Approximate method

- This data set has over 450,000 observations.
- Running our full model is not feasible for this data set.
- To approximate the full Bayesian analysis, we propose a two-stage approach.
- We first perform separate quantile regression at each site for a grid of quantile levels to obtain estimates of the quantile process and their asymptotic covariance.
- In a second stage, we analyze these initial estimates using the Bayesian spatial model for the quantile process.

Approximate method

- Let

  $$\hat{\boldsymbol{\beta}}(\mathsf{s}_i) = \left[\hat{\beta}_1(\tau_1,\mathsf{s}_i),...,\hat{\beta}_1(\tau_K,\mathsf{s}_i),\hat{\beta}_2(\tau_1,\mathsf{s}_i),...,\hat{\beta}_p(\tau_K,\mathsf{s}_i)\right]'$$

  be the estimates from usual quantile regression (separate by site and quantile level).

- Define its asymptotic covariance as $\text{Cov}\left(\hat{\boldsymbol{\beta}}(\mathsf{s}_i)\right) = \Sigma_i$

- We fit the model

  $$\hat{\boldsymbol{\beta}}(\mathsf{s}_i) \sim \text{N}\left(\boldsymbol{\beta}(\mathsf{s}_i), \Sigma_i\right),$$

  where the elements of $\boldsymbol{\beta}(\mathsf{s}_i)$ are functions of Bernstein basis polynomials with priors as in the full Bayesian model.

- We use a variation of the parametric bootstrap for interval estimates.

Approximate method

- This approximation provides a dramatic reduction in computational time because the dimension of the response is reduced from the number of observations at each site to the number of quantile levels in the approximation

- Also, the approximate model gives conjugacy for the δ^*_{jm}(s) which comprise the quantile functions (they have truncated Gaussian full conditionals).

- We fit this approximation to the simulated data, Atlanta data, and finally the entire Eastern US data set.

Results for the simulation study

	MSE			Coverage		
Method	β_1	β_2	β_3	β_1	β_2	β_3
Quantreg	0.51	0.99	1.02	0.89	0.89	0.88
Bayes - nonspatial	0.31	0.21	0.62	0.82	0.93	0.85
Bayes - spatial	0.13	0.12	0.27	0.86	0.93	0.88
Bayes - Approx	0.09	0.13	0.23	0.86	0.91	0.88

The approximate model has small MSE and good coverage probabilities.

Model comparisons for the Atlanta data

Model	Covariates	Coverage Prob	RMSE	MAD
Gaussian	Linear	0.965	14.3	9.34
Gaussian	Quadratic	0.969	14.2	9.07
QR - full	Linear	0.987	13.2	7.87
QR - full	Quadratic	0.986	13.0	7.64
QR - approx	Linear	0.959	13.1	7.76
QR - approx	Quadratic	0.947	12.9	7.62

- The prediction intervals have the correct coverage probability for the approximate model.

- For these data the approximate method actually gives smaller *MSE* and *MAD* than the full model.

Eastern US data

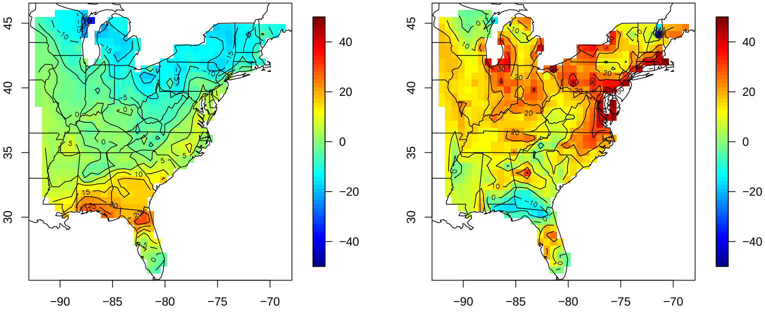
- We fit the approximate model on the Eastern US data using quadratic terms.
- To justify that this model fits well, we randomly removed observations for 2% of the days ($N = 12,468$ observations).
- The out-of-sample coverage probability of the 95% prediction intervals was 93.4%.

Effect of temperature on the 95% quantile

Let X_1 be temperature and X_2 be cloud cover. The plot below gives the posterior mean of

$$X_1\beta_1(\tau,s) + X_2\beta_2(\tau,s) + X_1^2\beta_3(\tau,s) + X_2^2\beta_4(\tau,s) + X_1X_2\beta_5(\tau,s),$$

with $\tau = 0.95$ and cloud cover fixed at 10% for two temperatures.

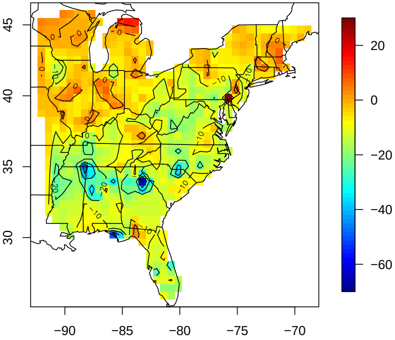


20^o (C)

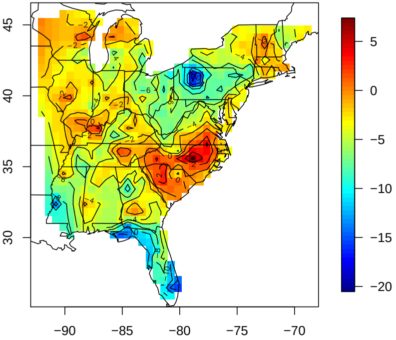
30^o (C)

Spatial distribution of the linear time trend

- The decreasing trend in ozone is the largest in the South.
- In the Carolinas the median is decreasing faster than the 95% quantile.



$\beta_j(0.95, s)$



$\beta_j(0.95, s)$-$\beta_j(0.50, s)$

Forecasting ozone levels under different climates

- Using the estimate of the conditional density of daily ozone, we simulate several realizations of the ozone process to forecast yearly summaries under different climate scenarios.

- These simulations vary temperature, wind speed, and cloud cover and assume all other factors (emissions, land-use, etc.) are fixed.

- Certainly other factors will change in the future (for example emissions may decline in response to new standards) so these projections are not meant to be realistic predictions.

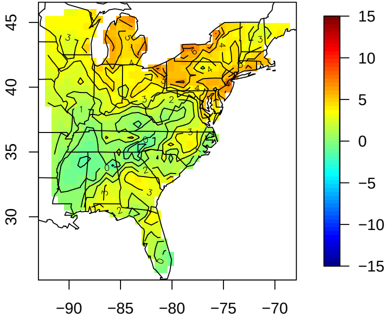
- Rather, they are meant to isolate the effect of climate change on future ozone levels.

Forecasting ozone levels under different climates

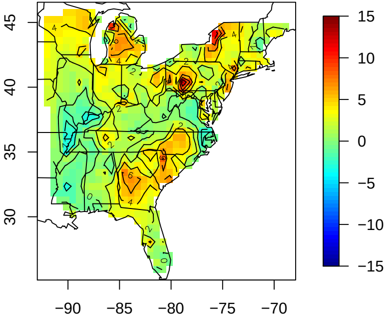
- For each climate scenario we simulate 50 realizations of the daily ozone process over a five year span.
- For each replication and spatial location we compute the median ozone and whether the location is in compliance with the EPA standard.
- We present the mean over the 50 realizations.
- The three climate scenarios are
  1. Current climate: we use the observed X_i.
  2. Warmer climate: we use observed X_i but increase the daily average temperature by 2^oC every day at every location.
  3. Future climate: we use 2041-2045 modeled meteorology from the Geophysical Fluid Dynamics Laboratory's (GFDL) deterministic atmospheric computer model (AM2.1).

Projected change in yearly median

The largest changes are predicted in the Northeast and Great Lakes Regions.



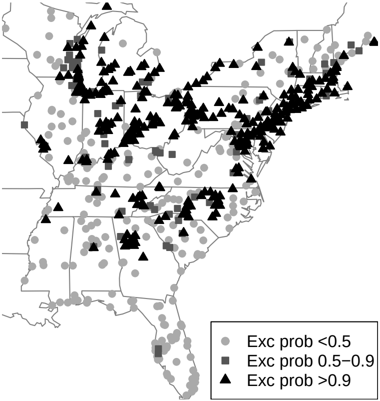
Shifted temp



2041-2045 met

Projected exceedence probabilities

Below are maps of the probability that the three-year (2041-2043) average of the fourth-highest daily maximum 8-hour average ozone concentrations exceeds 75 ppb under the future climate scenario.



Future work

- Our simulations show under-coverage for the regression coefficients (not prediction) if there is residual spatial/temporal correlation. Copula?
- Can we derive a non-linear quantile regression model that is invariant to a class of transformations?
- How to account for informative sampling?
- These projections are also made with deterministic (rather than statistical) models. Both have advantages. It would be interesting to compare them.
- Quantile regression could also be used for a comprehensive model validation/calibration.