Modeling for Seasonal Marked Point Processes: An Analysis of Evolving Hurricane Occurrences

Adam Peterson

UM Dept. Biostatistics

February 19th, 2019

Motivation

- ICAT Data:
  - Occurrence of (239 total) Hurricanes in time
- Scientific Question:
  - Hurricane damage due more to frequency or strength?

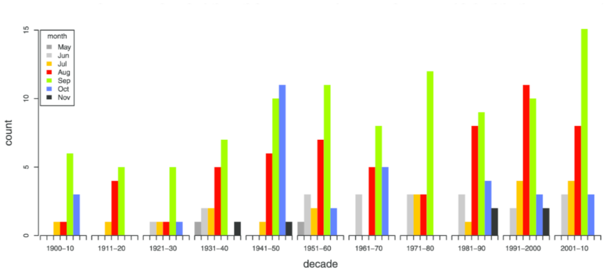


FIG. 2. *The number of hurricanes within one season aggregated by decades. In each decade, the number of hurricanes is grouped by months.*

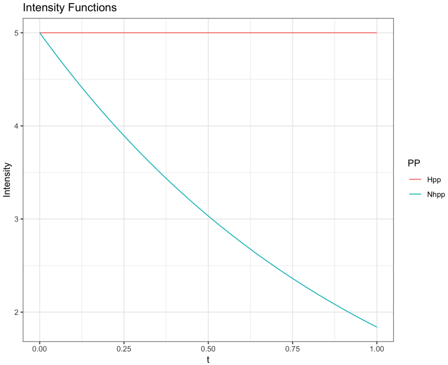
Agenda

Start with simple model, iteratively build in more complexity

1. Estimate "density" of hurricane occurrence non-parametrically
2. Incorporate hurricane time-correlation
3. Incorporate "Marks"
    - e.g. hurricane damage, windspeed

An Aside: Poisson Processes

- Homogenous Poisson Process: $N_1(t) \sim$Poisson(λ)
- Non Homogenous Poisson Process: $N_2(t) \sim Poisson(\lambda(t))$



Notation and Naive Model

- t_i the time at which a hurricane "occurs"
  - these are rescaled to be within $[0, 1]$
- Factorize $\lambda(t)$:
  - $\lambda(t) = \gamma f(t)$
  - $\gamma = \int_0^T \lambda(t)dt < \infty$

$$p(\{t_i\}_{i=1}^n|\gamma, f(\cdot)) \propto \exp(-\gamma)\gamma^n \prod_{i=1}^n f(t_i)$$

- Goal: estimate $f(\cdot), \gamma$:
- γ - conjugacy through gamma prior or improper prior + MCMC
- $f(\cdot)$ - Dirichlet Process(DP) mixture model

An Aside: DP Mixture models

- Dirichlet Process is a distribution over distributions
- "Non-Parametric" Bayesian way to model densities
- Visualizations help

Naive Model continued

$$t_i \mid G, \tau \sim f(t_i \mid G, \tau) = \int_0^1 \text{Beta}(t_i \mid \mu\tau, (1-\mu)\tau) dG(\mu)$$

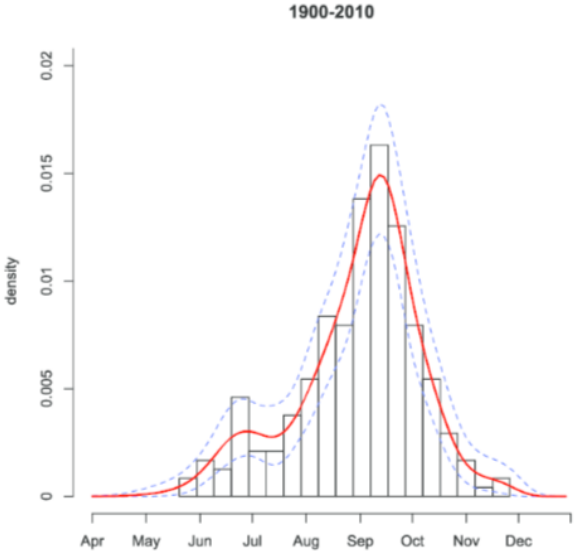
$$G(\mu) = \sum_{j=1}^{\infty} w_j \delta_{\mu_j}(\mu) \quad \mu_j \sim G_0 = \text{Unif}(0,1)$$

$$z_j \overset{iid}{\sim} \text{Beta}(1,\alpha); \qquad w_1 = z_1$$

$$w_j = z_j \prod_{r=1}^{j-1} (1 - z_r), \quad j \geq 2;$$

Estimation via Gibbs sampling

Naive Model Results



Increasing Model Complexity: Temporal Correlation

- How to incorporate correlation?
  - Use a latent variable!
- How and why?
  - We now want the intensities to be defined hierarchically, so that each year(s) has it's own Poisson intensity and there is a sharing of information (shrinkage) between years.

Increasing Model Complexity: Temporal Correlation

- $t_{i,k}, i = 1, ..., n_i; k = 1, ..., K$
- We now model $\{\lambda_k(t) = \gamma_k f_k(t) : k \in \mathcal{K}\}$

$$f_k(t) \equiv f(t|G_k, \tau) = \int_0^1 \text{Beta}(t|\mu\tau, (1-\mu)\tau)dG_k(\mu)$$

$$G_k(\mu) = \sum_{j=1}^{\infty} w_j \delta_{\mu_{j,k}}(\mu) \qquad \mu_{j,k} \sim PBAR(b, a, \rho)$$

- Weights - same across seasons
- Parameters - μ - are not.
- γ_k details...

Model 2: Application

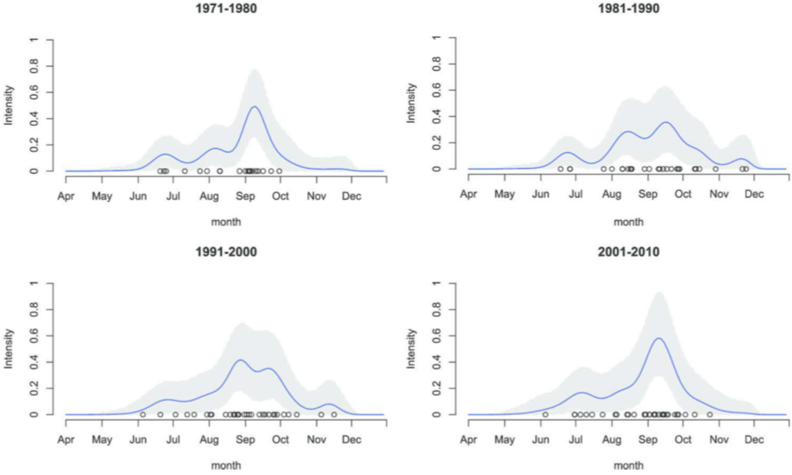


FIG. 5. *Posterior mean estimates (solid line) and 95% intervals (gray bands) of the hurricane intensity during 1971–2010. Points along the horizontal axis correspond to the observations.*

Model 2: Application

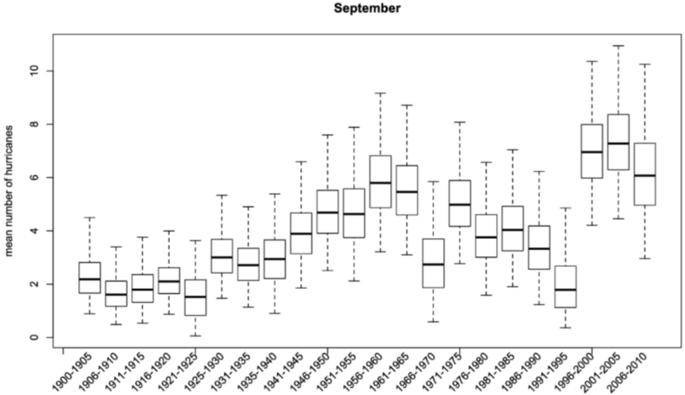


FIG. 4. *Boxplots of posterior samples for the average number of hurricanes in the month of September across five-year periods from 1900 to 2010.*

DDP-AR (Model 3): Incorporate Marks

- $\{t_{i,k}, y_{i,k}, z_{i,k}\}$
- In this setting y is the Maximum wind speed and z the associated (standardized) damage

$$f_k(t, y, z) = \int \mathrm{Beta}(t|\mu_k\tau, (1-\mu_k)\tau) N(y|\nu_k, \sigma^2) N(z|\eta_k, \xi^2) dG_k(\mu, \nu, \eta)$$

$$G_k = \sum_{j=1}^{N} w_j \delta(\mu, \nu, \eta)$$

- Use AR(1) process to impose dependence on kernel means across decades.
  - $\eta_{j,k} \mid \eta_{j,k-1} \sim N(\beta\nu_{j,k-1}, \sigma^2)$

Model 3: Inference

- Hypothesis - conditional on a certain time period, what is the distribution of the maximum wind speed?

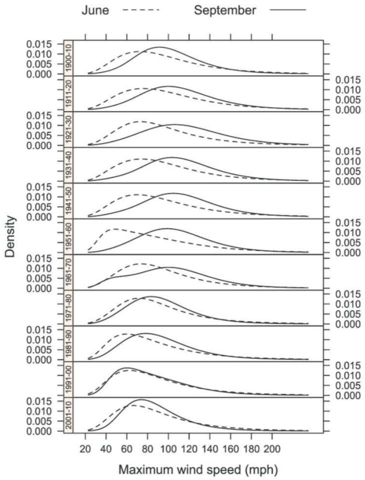
$$f_k(y|t, G_k) = \frac{f_k(y, t|G_k)}{f_k(t|G_k)} = \frac{\sum_{j=1}^{N} w_j \mathrm{Beta}(t|\mu_{j,k}\tau, (1-\mu_{j,k})\tau) N(y \mid \nu_{j,k}, \sigma^2)}{\sum_{j=1}^{N} w_j \mathrm{Beta}(t|\mu_{j,k}\tau, (1-\mu_{j,k})\tau)}$$

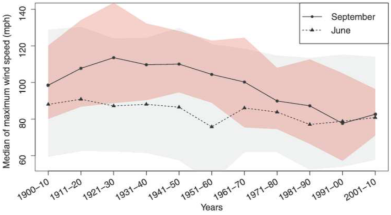
$$f_k(y|t, G_k) = \sum_{j=1}^{N} w_{j,k}^*(t) N(y|\nu_{j,k}, \sigma^2)$$

Model 3: Application

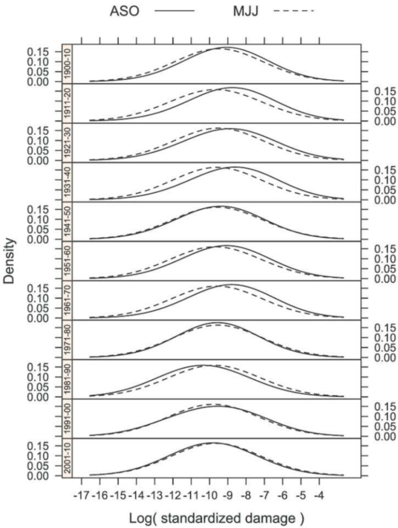
- DDP-AR model was fit to the full ICAT dataset, using conjugate priors and a Gibbs sampler variant to estimate the parameters of interest.

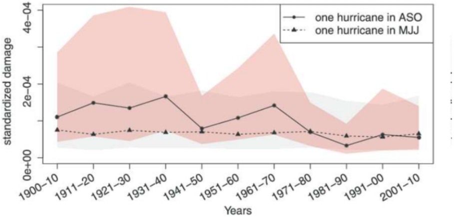
Model 3: Results



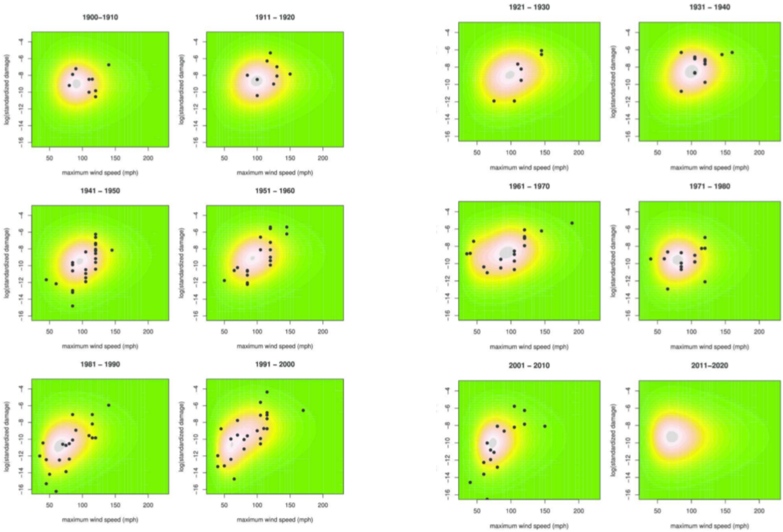


Model 3: Results (cont'd)





Model 3: Results (cont'd)



Final Remarks

- Model fit can be checked via Q-Q plot with Unif(0,1) variates
- Extensions include modifying assumptions, incorporating covariate information, and more.