

# BDA Project: Hurricane forecasting in Stan

José Miguel Ramírez & Jonas Lindblad  
Aalto University



# Hurricane introduction

- ▶ Destructive storms occurring in the late summer and fall in the northern hemisphere's tropical region.



# Hurricane introduction

- ▶ Destructive storms occurring in the late summer and fall in the northern hemisphere's tropical region.
- ▶ Classified by their wind intensity at the eye wall.



# Hurricane introduction

- ▶ Destructive storms occurring in the late summer and fall in the northern hemisphere's tropical region.
- ▶ Classified by their wind intensity at the eye wall.
- ▶ They can cause extreme levels of flooding and destroy many buildings.



# Hurricane introduction

- ▶ Destructive storms occurring in the late summer and fall in the northern hemisphere's tropical region.
- ▶ Classified by their wind intensity at the eye wall.
- ▶ They can cause extreme levels of flooding and destroy many buildings.
- ▶ Monetary damages and loss of lives increase with an almost exponential character as a function of storm intensity.



# Hurricane forecasting basics

- ▶ Forecasters predict two quantities: track and intensity

# Hurricane forecasting basics

- ▶ Forecasters predict two quantities: track and intensity
- ▶ Two kinds of models: *dynamical* and *statistical*



# Hurricane forecasting basics

- ▶ Forecasters predict two quantities: track and intensity
- ▶ Two kinds of models: *dynamical* and *statistical*
  - ▶ Dynamical: simulate the laws of physics

# Hurricane forecasting basics

- ▶ Forecasters predict two quantities: track and intensity
- ▶ Two kinds of models: *dynamical* and *statistical*
  - ▶ Dynamical: simulate the laws of physics
  - ▶ Statistical: estimate based on historical data

# Hurricane forecasting basics

- ▶ Forecasters predict two quantities: track and intensity
- ▶ Two kinds of models: *dynamical* and *statistical*
  - ▶ Dynamical: simulate the laws of physics
  - ▶ Statistical: estimate based on historical data
- ▶ Dynamical vs. statistical: good at long- and short-range forecasts respectively

# Hurricane forecasting basics

- ▶ Forecasters predict two quantities: track and intensity
- ▶ Two kinds of models: *dynamical* and *statistical*
  - ▶ Dynamical: simulate the laws of physics
  - ▶ Statistical: estimate based on historical data
- ▶ Dynamical vs. statistical: good at long- and short-range forecasts respectively
- ▶ ... but for hurricane forecasts short-range is usually more interesting

# Hurricane forecasting basics

- ▶ Forecasters predict two quantities: track and intensity
- ▶ Two kinds of models: *dynamical* and *statistical*
  - ▶ Dynamical: simulate the laws of physics
  - ▶ Statistical: estimate based on historical data
- ▶ Dynamical vs. statistical: good at long- and short-range forecasts respectively
- ▶ ... but for hurricane forecasts short-range is usually more interesting
- ▶ *Rapid intensification*: forecasted better by dynamical models

# Hurricane forecasting basics

- ▶ Forecasters predict two quantities: track and intensity
- ▶ Two kinds of models: *dynamical* and *statistical*
  - ▶ Dynamical: simulate the laws of physics
  - ▶ Statistical: estimate based on historical data
- ▶ Dynamical vs. statistical: good at long- and short-range forecasts respectively
- ▶ ... but for hurricane forecasts short-range is usually more interesting
- ▶ *Rapid intensification*: forecasted better by dynamical models
- ▶ This project: a *statistical* model for *intensity*

## Hurricane forecasting basics: the SHIPS data

The US government forecasting agency, the National Hurricane Center (NHC), uses a large number of models operationally. The models (together: the *model ensemble*) are used together with experienced meteorologists' judgment to provide the official forecast.

- ▶ Surprisingly, the best *single* (short-range) model is a multiple linear regression!

# Hurricane forecasting basics: the SHIPS data

The US government forecasting agency, the National Hurricane Center (NHC), uses a large number of models operationally. The models (together: the *model ensemble*) are used together with experienced meteorologists' judgment to provide the official forecast.

- ▶ Surprisingly, the best *single* (short-range) model is a multiple linear regression!
- ▶ The NHC regression model: Statistical Hurricane Intensity Prediction Scheme (SHIPS)



# Hurricane forecasting basics: the SHIPS data

The US government forecasting agency, the National Hurricane Center (NHC), uses a large number of models operationally. The models (together: the *model ensemble*) are used together with experienced meteorologists' judgment to provide the official forecast.

- ▶ Surprisingly, the best *single* (short-range) model is a multiple linear regression!
- ▶ The NHC regression model: Statistical Hurricane Intensity Prediction Scheme (SHIPS)
- ▶ SHIPS: ~140 covariates, many calculated from data sources more easily available to the NHC

# Hurricane forecasting basics: the SHIPS data

The US government forecasting agency, the National Hurricane Center (NHC), uses a large number of models operationally. The models (together: the *model ensemble*) are used together with experienced meteorologists' judgment to provide the official forecast.

- ▶ Surprisingly, the best *single* (short-range) model is a multiple linear regression!
- ▶ The NHC regression model: Statistical Hurricane Intensity Prediction Scheme (SHIPS)
- ▶ SHIPS: ~140 covariates, many calculated from data sources more easily available to the NHC
- ▶ SHIPS dataset: publically available with no restrictions (link: SHIPS Development)

# Hurricane forecasting basics: the SHIPS data

The US government forecasting agency, the National Hurricane Center (NHC), uses a large number of models operationally. The models (together: the *model ensemble*) are used together with experienced meteorologists' judgment to provide the official forecast.

- ▶ Surprisingly, the best *single* (short-range) model is a multiple linear regression!
- ▶ The NHC regression model: Statistical Hurricane Intensity Prediction Scheme (SHIPS)
- ▶ SHIPS: ~140 covariates, many calculated from data sources more easily available to the NHC
- ▶ SHIPS dataset: publically available with no restrictions (link: SHIPS Development)
- ▶ ... but the documentation is terrible

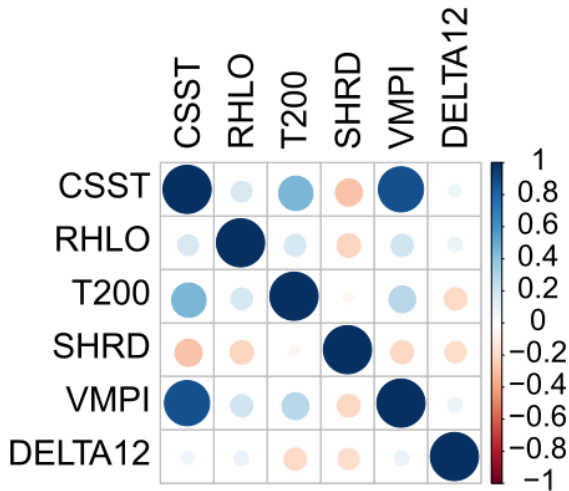
# Hurricane forecasting basics: the SHIPS data

The US government forecasting agency, the National Hurricane Center (NHC), uses a large number of models operationally. The models (together: the *model ensemble*) are used together with experienced meteorologists' judgment to provide the official forecast.

- ▶ Surprisingly, the best *single* (short-range) model is a multiple linear regression!
- ▶ The NHC regression model: Statistical Hurricane Intensity Prediction Scheme (SHIPS)
- ▶ SHIPS: ~140 covariates, many calculated from data sources more easily available to the NHC
- ▶ SHIPS dataset: publically available with no restrictions (link: SHIPS Development)
- ▶ ... but the documentation is terrible
- ▶ SHIPS: only a point estimate; our project: a predictive distribution

# Hurricane forecasting basics: the SHIPS data

## Hurricane forecasting basics: our selection



## Intensity change predictive model

## Model: limitations



Model: posterior predictive checking

Model: variables to use

Model: marginal posteriors of coefficients

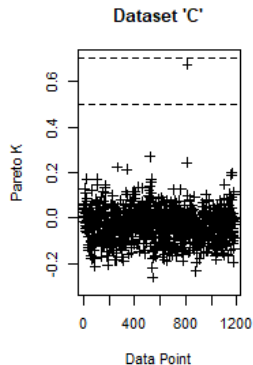
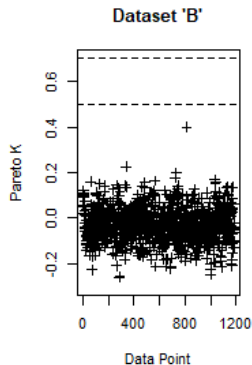
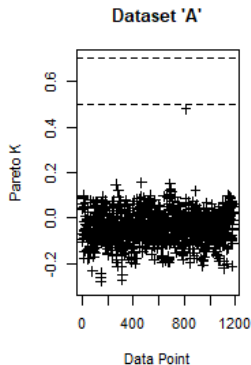
Model: margin of SHRD and CSST

# Forecasting

# Forecasting: Model Comparison (1)

Dataset comparison for the **linear regression model** (LOOCV)

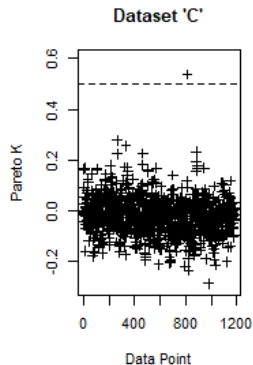
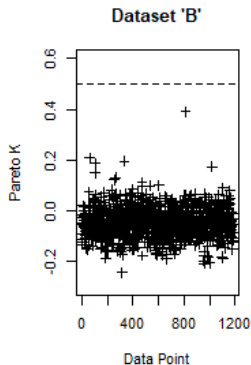
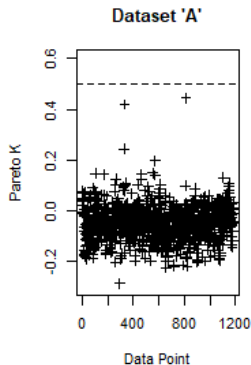
Dataset	elpd_diff	se_diff
<b>C</b>	0.0	0.0
<i>B</i>	-25.0	6.5
<i>A</i>	-27.4	6.3



## Forecasting: Model Comparison (2)

Dataset comparison for the **skewed regression model** (LOOCV)

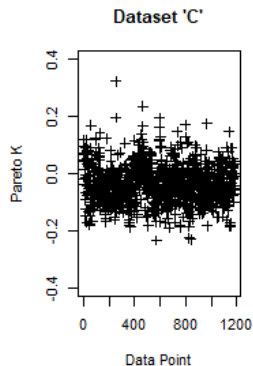
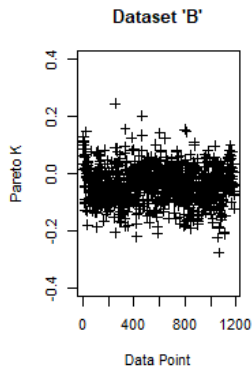
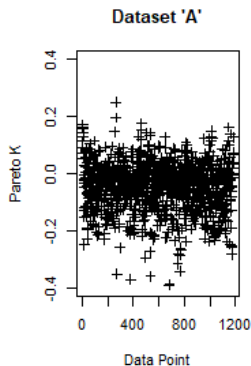
Dataset	elpd_diff	se_diff
<b>C</b>	0.0	0.0
<i>B</i>	-23.2	6.2
<i>A</i>	-28.7	6.2



# Forecasting: Model Comparison (3)

Dataset comparison for the **Changing variance model** (LOOCV)

Dataset	elpd_diff	se_diff
<b>C</b>	0.0	0.0
<i>B</i>	-32.6	8.2
<i>A</i>	-37.1	8.2





## Forecasting: Model Comparison (4)

Model comparison using the Dataset C (LOOCV)

Model	elpd_diff	se_diff
<b>Variance</b>	0.0	0.0
Skew	-176.3	27.8
Linear	-205.5	34.9

Forecasting: what about the NHC?

## Problems to solve & development ideas

- ▶ variable selection in full SHIPS dataset

## Problems to solve & development ideas

- ▶ variable selection in full SHIPS dataset
- ▶ more time series autoregressive components

## Problems to solve & development ideas

- ▶ variable selection in full SHIPS dataset
- ▶ more time series autoregressive components
- ▶ use LGEM model (will explain)

## Conclusions & contact info

## Additional information

# R Markdown

This is an R Markdown presentation. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document.



# Slide with Bullets

- ▶ Bullet 1

## Slide with Bullets

- ▶ Bullet 1
- ▶ Bullet 2

## Slide with Bullets

- ▶ Bullet 1
- ▶ Bullet 2
- ▶ Bullet 3

## Slide with R Output

```
summary(cars)
```

##	speed	dist
##	Min. : 4.0	Min. : 2.00
##	1st Qu.:12.0	1st Qu.: 26.00
##	Median :15.0	Median : 36.00
##	Mean :15.4	Mean : 42.98
##	3rd Qu.:19.0	3rd Qu.: 56.00
##	Max. :25.0	Max. :120.00

# Slide with Plot

