CENTRE DE RECERCA MATEMÀTICA

The influence of sea surface temperature on tropical-cyclone intensity

Alfredo Hernández Cavieres

Supervised by: Álvaro Corral & Isabel Serra

Outline

1. Introduction
    1.1 Objectives
    1.2 Tropical-cyclones
    1.3 Hypothesis
2. Regression analysis
    2.1 Comparing two populations
    2.2 Bootstrapping the coefficients
    2.3 Permutation tests
3. Geographical analysis
    3.1 New variables
    3.2 Analysis
4. Conclusions

References

Introduction

Background and motivation

For the North Atlantic (N. Atl.) and Northeast Pacific (E. Pac.) basin, it has been shown [1] that the probability distribution of the so-called power-dissipation index (PDI, a rough estimation of released energy) is affected by the annual and basin-wide averaged sea surface temperature (SST), displacing towards more extreme values on warm years (high-SST).

As the PDI integrates (cubic) wind speed over tropical-cyclone (TC) lifetime, it is an open question where the PDI increase comes from (higher speed, longer lifetime, or both).

Objectives

- Describe the influence of SST on tropical-cyclones occurrences.
- Understand the displacement towards more extreme PDI values on warm years (high-SST).
- Develop a statistical method to compare the joint distributions of the two populations (low-SST and high-SST).

Tropical-cyclones

To characterise a TC one needs to define a physically relevant measure of released energy. The released energy of each TC is summarised as

$$PDI = \sum_t v_t^3 \Delta t. \tag{1}$$

- The raw hurricane best track data (HURDAT2) is provided by the National Hurricane Center.
- We limit this study to the satellite era (1966–2016).

Higher SSTs $\Leftrightarrow$ increased water vapour in the lower troposphere [2] $\Rightarrow$ separation of years by SST.

Separation by SST

Then, the hurricane observational data is classified into occurrences in low-SST and high-SST years depending on whether they are lower or greater than

$$\langle \mathsf{SST} \rangle = \sum_{y \in \Upsilon} \frac{\mathsf{SST}(y)}{\Upsilon}, \tag{2}$$

where $\mathsf{SST}(y)$ is the mean SST of the year y, and Υ is the total number of years studied.

- The SST data (HadISST1) is provided by the Met Office Hadley Centre.

Data overview

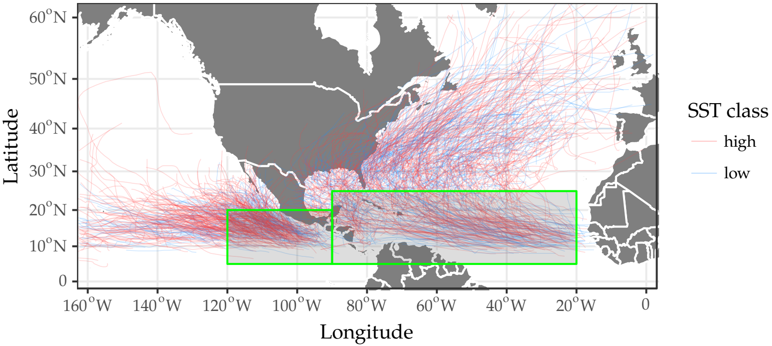


Figure: Tropical-cyclones best tracks for the North Atlantic and Northeast Pacific Oceans

Data package

Unified data sets for the N. Atl. and E. Pac. basins are available in a R package called `HurdatHadISSTData` [3].

The data sets in the `HurdatHadISSTData` package are:

- `tc.pdi.natl` – Data set for the North Atlantic basin.
- `tc.pdi.epac` – Data set for the Northeast Pacific basin.
- `tc.pdi.all` – Data set for both basins.

storm.id <chr>	storm.name <chr>	n.obs <int>	storm.duration <dbl>	storm.pdi <dbl>	max.wind <int>	mean.wind <dbl>	mean.sq.wind <dbl>	storm.year <int>	basin <chr>	sst <dbl>	sst.norm <dbl>	sst.class <chr>
AL011966	ALMA	42	252	34632626747	110	56.4	3750	1966	NATL	27.6	0.998	low
AL021966	BECKY	9	54	3413930334	65	46.1	2353.	1966	NATL	27.6	0.998	low
AL031966	CELIA	36	216	7839872104	70	35.4	1488.	1966	NATL	27.6	0.998	low
AL041966	DOROTHY	37	222	21340832518	75	54.9	3211.	1966	NATL	27.6	0.998	low
AL051966	ELLA	26	156	4646503652	45	37.7	1487.	1966	NATL	27.6	0.998	low
AL061966	FAITH	69	414	120569417711	110	79.0	6722.	1966	NATL	27.6	0.998	low

Table: Excerpt of the North Atlantic data set

Data visualisation

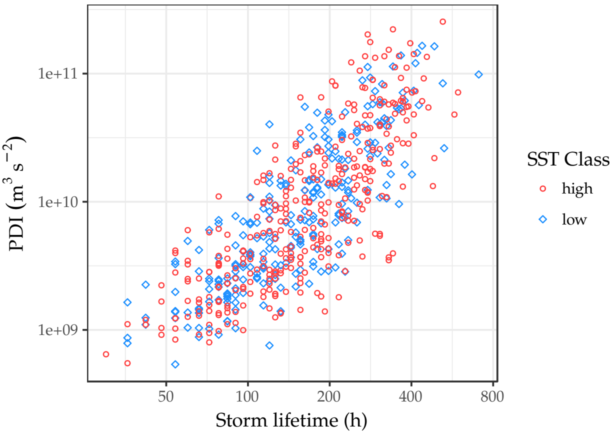


Figure: Joint distribution $f(PDI, \text{lifetime})$ of the hurricane observations for the North Atlantic basin

Hypothesis

Hypothesis statement

That SST does not directly affect the maximum wind speed of a TC: storms of equal lifetime should, in theory, have the same wind speed and PDI, and have the same joint distribution:

$$f(Y \mid X = x)_{\text{low}} = f(Y \mid X = x)_{\text{high}}. \tag{3}$$

The physical reasoning behind this is that once the cyclone is activated, the wind speed should not depend on its underlying SST.

Regression analysis using resampling methods

Linear regression

Description of the populations

$$Y = \beta_0 + \beta_1 X + \epsilon. \tag{4}$$

Regression coefficient estimates:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \qquad \widehat{\mathrm{se}}(\hat{\beta}_0) = \sqrt{\sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)},$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \qquad \widehat{\mathrm{se}}(\hat{\beta}_1) = \sqrt{\frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}.$$

$$R^2 = 1 - \frac{\mathrm{RSS}}{\mathrm{TSS}} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}, \qquad \hat{\sigma} = \sqrt{\frac{\mathrm{RSS}}{n-2}} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-1}}.$$

Linear regression

Description of the populations

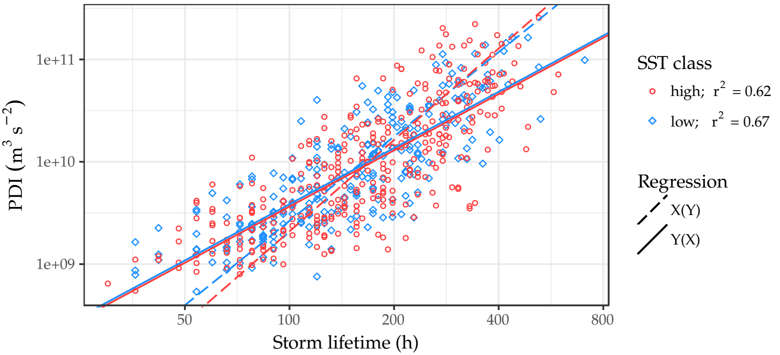
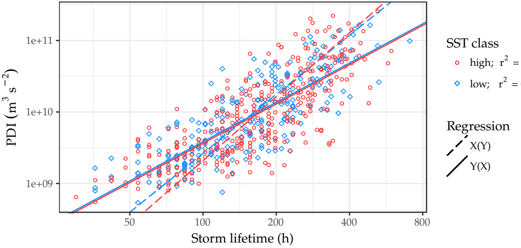


Figure: Scatterplot of the joint distribution and regression analysis for the PDI and lifetime of storms for the North Atlantic basin

Linear regression

Description of the populations



X	Y	SST class	$\hat{\beta}_0$	$\hat{\beta}_1$	R^2
lifetime	PDI	Low	-1.44 ± 0.16	0.37 ± 0.02	0.67
		High	-1.14 ± 0.14	0.34 ± 0.01	0.62
PDI	lifetime	Low	5.94 ± 0.18	1.82 ± 0.08	0.67
		High	5.91 ± 0.17	1.83 ± 0.08	0.62

Table: Linear regression coefficients obtained performing OLS on the North

Comparing the two populations

It seems straightforward to compare the coefficient estimates directly:

$$T^{(1)} = \left|\hat{\beta}_{0,h} - \hat{\beta}_{0,l}\right|, \quad T^{(2)} = \left|\hat{\beta}_{1,h} - \hat{\beta}_{1,l}\right|, \quad T^{(3)} = \left|R_h^2 - R_l^2\right|. \tag{5}$$

Polko-Zając [4] proposes alternative statistics that not only consider the nominal value of the coefficient estimates, but take into account their standard errors as well:

$$T^{(4)} = \frac{\left|\hat{\beta}_{0,h} - \hat{\beta}_{0,l}\right|}{\widehat{\text{se}}(\hat{\beta}_{0,h} - \hat{\beta}_{0,l})}, \quad T^{(5)} = \frac{\left|\hat{\beta}_{1,h} - \hat{\beta}_{1,l}\right|}{\widehat{\text{se}}(\hat{\beta}_{1,h} - \hat{\beta}_{1,l})}, \quad T^{(6)} = T^{(4)} + T^{(5)}. \tag{6}$$

Assumptions for the model

- Normality of the residuals.
- Homoscedasticity of the residuals.
- Independence of the residuals.

The standard errors, confidence intervals, and hypothesis tests associated with the linear model rely upon these being true.

Assumptions for the model

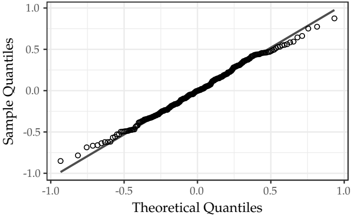
Diagnostic plots:

- Q-Q plot of the residuals.
- Residuals vs fitted values plot.

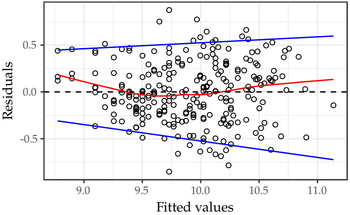
Statistical hypothesis tests:

- Lilliefors test: normality.
- Correlation test: independence.
- Breusch–Pagan test: homocedasticity.

Assumptions for the model



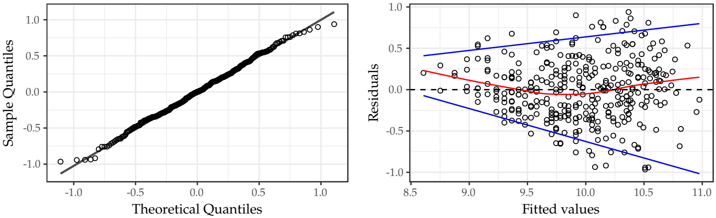
(a) Q-Q plot (low SST)



(b) Residuals vs fitted (low SST)

Figure: Diagnostic plots to analyse the residuals for the North Atlantic basin

Assumptions for the model



(a) Q-Q plot (high SST) (b) Residuals vs fitted (high SST)

Figure: Diagnostic plots to analyse the residuals for the North Atlantic basin

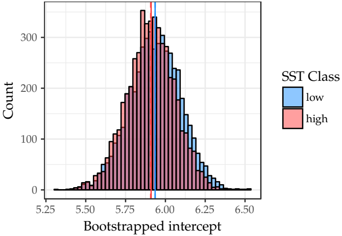
The bootstrap

We need a more robust linear model to be able to infer statistical properties of the data using linear regression as the underlying model.

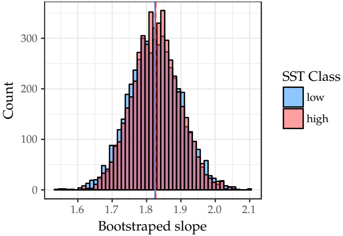
Algorithm 1: Resampling the cases using bootstrap

for $r \leftarrow 1$ **to** R **do**
- (i) Sample $i_1^*, \ldots, i_n^*$ randomly with replacement from $\{1, 2, \ldots, n\}$.
- **for** $j \leftarrow 1$ **to** n **do**
  - (ii) Set $x_j^* = x_{i_j^*}$, $y_j^* = y_{i_j^*}$
- (iii) Fit OLS regression to $(x_i^*, y_i^*), \ldots, (x_n^*, y_n^*)$.
- (iv) Calculate estimates of $\hat{\beta}_{0,r}^*$ and $\hat{\beta}_{1,r}^*$.

Bootstrapping the coefficients



(a) Histogram of the bootstrapped intercept



(b) Histogram of the bootstrapped slope

Figure: Resampled slopes and intercepts obtained by bootstrapping for the North Atlantic basin data for the PDI(lifetime) regression model

Results

X	Y	SST class	$\hat{\beta}_0^*$	$\hat{\beta}_1^*$	R^{2*}
lifetime	PDI	Low	-1.44 ± 0.15	0.36 ± 0.02	0.67 ± 0.03
		High	-1.15 ± 0.14	0.34 ± 0.01	0.62 ± 0.03
PDI	lifetime	Low	5.91 ± 0.17	1.84 ± 0.08	0.67 ± 0.03
		High	5.90 ± 0.15	1.83 ± 0.07	0.61 ± 0.03

Table: Linear regression coefficients obtained performing bootstrap on the North Atlantic basin data

X	Y	$T^{(1)}$	$T^{(2)}$	$T^{(3)}$	$T^{(4)}$	$T^{(5)}$	$T^{(6)}$
lifetime	PDI	0.289	0.027	0.048	1.404	1.324	2.727
PDI	lifetime	0.007	0.007	0.054	0.031	0.065	0.095

Table: Value of the studied statistics for North Atlantic basin data set using bootstrap

The permutation test

We need to properly quantify the statistical significance of evidence against (or in favour of) the hypothesis.

Algorithm 2: Permutation test for comparing two populations

(i) Define the null hypothesis, H_0, and the alternative.
(ii) Consider a test statistic that compares the populations which is large (small) if the null hypothesis is not true, and small (large) if it is true.
(iii) Calculate the true statistic of the data, T.
for $r \leftarrow 1$ **to** R **do**
    (iv) Create a new data set consisting of the data, randomly rearranged. Exactly how it is rearranged depends on the null hypothesis.
    (v) Calculate the statistic for this new data set, T^*.
    (vi) Compare the statistic T^* to the true value, T.
(vii) If the true statistic is greater (lower) than 95% of the random values, then one can reject the null hypothesis at $p < 0.05$.

Results

X	Y	$T^{(1)}$	$T^{(2)}$	$T^{(3)}$	$T^{(4)}$	$T^{(5)}$	$T^{(6)}$
lifetime	PDI	0.146	0.121	0.711	0.137	0.117	0.128
PDI	lifetime	0.870	0.806	0.705	0.864	0.795	0.830

Table: List of p-values of the bootstrap-powered permutation test for the North Atlantic basin data

This indicates a strong evidence in favour of the null hypothesis being true.

Geographical analysis

New variables

We focus on the geographical **genesis location** of a tropical-cyclone, as well as its **death location**.

Additionally, we calculate the total travelled distance, or **path length**, of each hurricane by means of the *spherical law of cosines*.

Path length and duration

We define mean forward speed is calculated as

$$\langle v_f \rangle = \frac{d}{\text{lifetime}}. \tag{7}$$

We think this variable may be an intermediary variable to relate the storm path length and its PDI, and expect a displacement to higher speeds for high-SST years.

Path length and duration

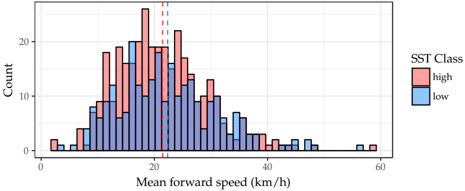


Figure: Mean forward speed histogram for the North Atlantic basin

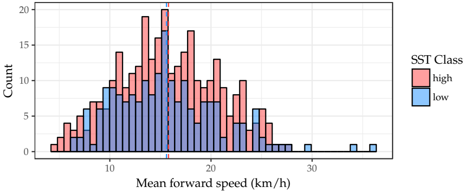
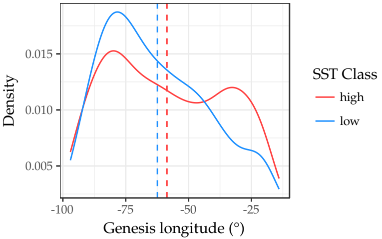
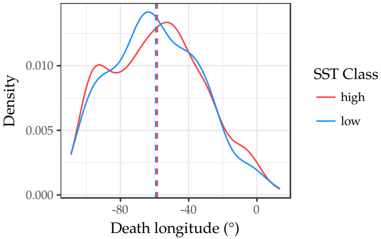


Figure: Mean forward speed histogram for the Northeast Pacific basin

Location



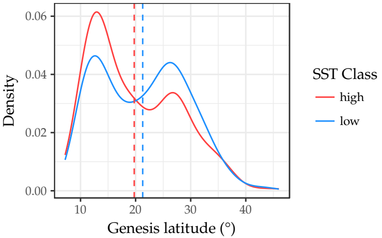
(a) Distribution of the genesis longitude



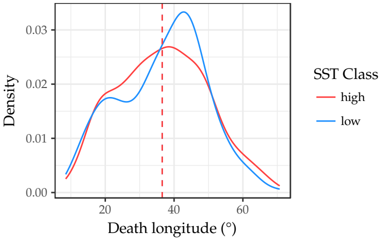
(b) Distribution of the death longitude

Figure: Longitude distributions of storms for the North Atlantic basin

Location



(a) Distribution of the genesis latitude



(b) Distribution of the death latitude

Figure: Latitude distributions of storms for the North Atlantic basin

Location

Major difference between low-SST and high-SST years is the location of the genesis:

- N. Atl: genesis $\longrightarrow$ SE, death is not displaced.
- E. Pac: genesis $\longrightarrow$ SE, death $\longrightarrow$ NE.

Conclusions

Our conclusions are compatible with the view of tropical cyclones as an activation process, in which, once the event has started, its intensity is kept in critical balance between attenuation and intensification (and so, higher SST does not trigger more intensification).

Further analysis shows that the longer lifetimes are mainly due to a shift to South-East of the TC genesis point.

References I

[1] Á. Corral, A. Ossó and J. E. Llebot. Scaling of tropical-cyclone dissipation. *Nature Physics*, 6(9):693–696, 2010. ISSN: 1745-2473. DOI: `10.1038/nphys1725`. URL: `http://www.nature.com/doifinder/10.1038/nphys1725`.

[2] K. Trenberth. Uncertainty in Hurricanes and Global Warming. *Science*, 308(5729):1753–1754, 2005. DOI: `10.1016/j.enpol.2008.08.013`. URL: `http://science.sciencemag.org/content/sci/308/5729/1753.full.pdf`.

[3] A. Hernández. Tropical-Cyclones (HurdatHadISSTData Package) on GitLab. URL: `https://gitlab.com/aldomann/hurdat-hadisst-data`.

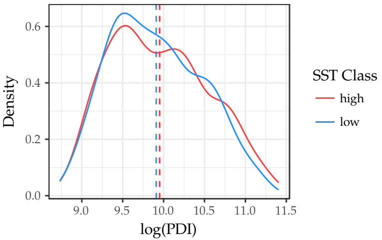
References II

[4] D. Polko-Zając. Application of the permutation test for comparing regression models. In *10th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena*, 2016.

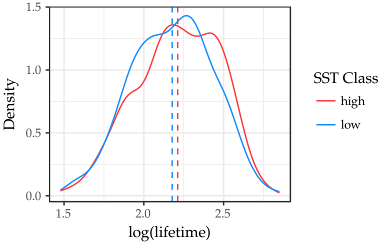
[5] A. C. Davison and D. V. Hinkley. *Bootstrap methods and their application*. Cambridge University Press, Cambridge, 1997. ISBN: 9780511802843. DOI: `10.1017/CBO9780511802843`. URL: `http://ebooks.cambridge.org/ref/id/CBO9780511802843`.

Questions?

Backup slides



(a) Marginal distribution of the PDI



(b) Marginal distribution of the lifetime

Figure: Marginal analysis for the variables of the joint distribution for the North Atlantic basin data

Data	Mean($\log(PDI)$)	Mean($\log$(lifetime))	Mean(PDI) (10^9 m^3 s^{-2})	Mean(lifetime) (h)
Low-SST	9.91 ± 0.04	2.18 ± 0.02	22.76 ± 1.95	192.94 ± 5.74
High-SST	9.95 ± 0.03	2.21 ± 0.01	18.56 ± 1.74	176.92 ± 6.49

Table: Statistical summary for the low-SST and high-SST subsets of the marginals distributions for the North Atlantic basin data

```
> cor.test(residuals(natl.fit.low), fitted(natl.fit.low))

 Pearson's product-moment correlation

data:  residuals(natl.fit.low) and fitted(natl.fit.low)
t = -1.1943e-14, df = 253, p-value = 1
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.1228426  0.1228426
sample estimates:
          cor
-7.508364e-16
```