

# Luminosity

Cameron, M.      Rosales, V.      Westermann, J.P.

July 30, 2017

**Contents**

**List of Figures**

**List of Tables**

Notes (Please Read):

- `print(df.to_latex())` will give a latex export of the dataframe in the jupyter notebook, you can use this to quickly copy paste into latex
- `plt.savefig(<path>)` will allow you to save your figure as a png so it can be used in the document/presentation
- Make sure to commit and pull as much as possible to avoid merge errors
- Don't edit the styles of this document yet, will do that at the end

## 1 Introduction

### 1.1 Summary

TODO Will do this last

### 1.2 Literature Review

#### 1.2.1 Luminosity-based Approach

TODO Michael: Try to accumulate as much as possible. We have such a long list of papers anyway.

#### 1.2.2 Natural Disaster Economics

TODO Viviana: obviously, as the expert.

## 2 Data

### 2.1 Data Description

TODO Micheal: Describe what the data looks like, how many observations there are, where we got it, who else has used it etc.

## 2.2 Data Preprocessing

**Volume** While the number of observations is extremely low, as only annual images are available from 1992 to 2013, the dimensionality of each observation (image) is considerable. With a size of 16801 by 43201, every image contains 725820001 pixels in total, which results in more than 700MB of disk-space required for only one image in uncompressed format. This also means that computations on the entire dataset are not possible with common personal computing architecture.

**Diverse Sources** To aggregate meaningful information to use in the model, such as economic indicators from the world bank (gdp, import, export, inflation, etc.), we need to join data from a lot of different sources on different aggregation levels.

### 2.2.1 QGIS

TODO Viviana

### 2.2.2 Python Architecture

To enable exploration and flexible modelling of the data even on a platform where memory and computing power are quite restricted, we built a supporting architecture in python for loading the images and aggregating the features. To quickly load entire time series of the most important places, the sections of the image representing the biggest cities and a 150x150 area around them are cut out and stored as separate time series. This allows for more easily training predictive models for the luminosity. Additionally, functions were created to easily load a 'section' time series (the series of images for one particular subsection of the world's image), given coordinates and the desired size of the image. The stored data is no longer in the 'tif' format provided at download. All images have already been passed into arrays and stored as compressed numpy arrays. That reduces the required storage space by an order of magnitude. The functions to preprocess and load the data can be found in the GitHub repository.

### 2.2.3 Denoising

## 3 Modelling

### 3.1 Disaster Impact Models

Exploratory analysis and some basic models are a first step to assessing the impact of natural disasters on the luminosity time series. For this, we need to make a modelling decision regarding how the disaster (represented only as a single point location) can be geospatially associated with pixel luminosity values. TODO Jonas, literature on this?

#### 3.1.1 Linear Distance-based Modelling

One practical mathematical choice is a function decaying with distance affecting areas or even individual pixels on the grid. The advantage of this method is that it is simple to explain, easily tuneable and leaves a lot of flexibility for modelling. Additionally, it captures the notion that areas in the vicinity of a disaster event are more likely to be affected than those further away by default. We trialed this approach using the luminosity data in combination with TODO (NGOAAASDF Name?). We extracted 150x150 squares from the satellite image of the entire planet in a similar fashion as a convolutional neural network would apply to an image. Then we filtered the top  $t\%$  most luminous 'subimages', as we can generally disregard those parts of the world that contain no lights (oceans, deserts, etc.) and calculate a disaster coefficient based on a function decaying linearly with distance. The  $t$  percentage threshold can be tuned in order to keep the data small enough to process. This is necessary as in practice, means calculating the distance between every section of the image's center and every earthquake and then applying a linear function of the distance and the magnitude for every earthquake that occurs in a given year, then summing up these effects to generate a disaster impact variable for that observation. This results in a high computational load as for every section that we extract from the image, we need to compute the euclidian distance to every single one of the 25000 earthquakes, making in almost infeasible on a common laptop. However, already only the luminosity growth between beginning and end of the image time series and the disaster coefficient show promising qualities of the data when plotted against one another. However, this approach makes some strong assumptions about the

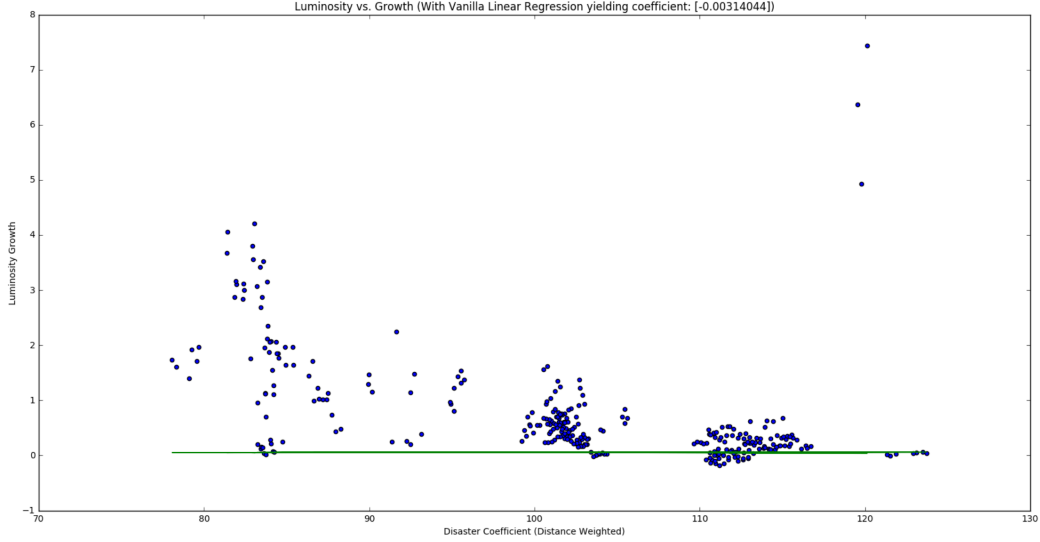


Figure 1: Luminosity Growth 1992-2013 plotted against a linearly decaying disaster coefficient for 150x150 image sections.

nature of natural disasters that don't hold in reality. An important factor in how much impact a disaster has on a region are geographical features: An earthquake will affect different areas differently based on their rockbed and geological consistency while e.g. storms and floods depend strongly on the topography. Nevertheless, the calculation of such a disaster coefficient yields promising results and suggests there is a relationship in the data between disaster coefficient and the growth in a location, proxied by luminosity.

### 3.1.2 Recorded Disaster Impact-based Modelling

Another possible approach, that is commonly applied in Economics, is to use institutional data on which cities were affected by an earthquake. Using the TODO datasets we can obtain records of cities that were affected by a given earthquake in a given year. Matching this with a simple list of the world's largest 50000 cities, we can obtain a panel dataset that associates earthquake damage without any measure of magnitude to cities, some affected, others not. We can then

## **3.2 Panel Model**

### **3.2.1 Region-based Panel**

TODO Viviana

### **3.2.2 Section-based Panel**

TODO Jonas

### **3.2.3 Dynamic Panel**

TODO Viviana: Describe here how the model that you are using is constructed, where you got it, etc.

## **4 Results**

### **4.1 Case Analysis**

TODO Micheal: This is where your case analysis for different places goes, try to add some statistical tests etc. if possible. E.g. distribution of light one year vs the next compared to overall time series distribution changes (shocks).

### **4.2 Modelling Results**

TODO Viviana: Describe the results of the regression here, significant values and what those values mean.

### **4.3 Conclusions**

TODO Will do this just before the summary

### **4.4 Outlook**

TODO Jonas

