Hurricane Catastrophe Parametric Model Design

Columbia University Capstone Project
in Collaboration with Guy Carpenter & Co.

Crystal Tian
Ye Yuan
Yue Wang

I. Introduction

Natural catastrophes can bring disruption to the human community where its economic and social system build upon natural resources. When catastrophic events happen, catastrophic insurance could help the local community recover and rebuild their home and business. Most of the time, due to high unperceivable damages and slow process of providing financial recovery, catastrophe insurance would encounter under-coverage and low penetration. In this project, we conveyed the tropical cyclone risk insurance design by Guy Carpenter & Company into an experimental study on hurricane insurance optimization for Jamaican area, where we want to reduce the gap between insurance payment actual loss based on few measurable features on historical hurricane events.

Specifically, we apply a matrix optimization method referred to as 'cat-in-a-box', where the geographical and wind information of each hurricane track are identified over a set of "map cells", or boxes on the affected Earth's surface. In such way, we can partition the total loss of a hurricane event into different "cells" on the affected map domain. The project consists of two parts: a mapping modeling pipeline that transforms the hurricane data into desired matrix form, and the optimization algorithms. The modeling pipeline and the algorithms should be applicable to different designated map domains to test the resilience of the optimization model. Within the data modeling pipeline, all historical hurricane track events go through filters that 1) separate track points into six damage categories based on wind features, 2) identify and retain only tracks pass through affected area, 3) auto-fulfill data points between two nonadjacent track points in terms of map cells in one event, and, 4) configure prepared dataset into the parametric matrix.

Next, we implement both constrained and unconstrained optimization algorithms, aiming to reduce the volatility of payments under uncertainty. The parametric design problem is a complex, large-scale combinatorial problem. In this paper we propose several optimization algorithms to solve the problem in reasonably low computing time.

Experimental results on a case-study of interest illustrate the computational limits and solution quality obtained with the proposed algorithms.

II. Methodology

"Cat-in-box" Map

Following the "Cat-in-box" scheme already adopted in the development of seismic parametric models (Franco et al, 2019), we divided the region affected by hurricanes into an ordered grid of cells, or boxes (see Fig.2). We will be calculating the total loss of a hurricane track by partitioning the loss based on geographical and wind features into each "box".

Provided with a stochastic set of hurricane track events data, we overlay the hurricane tracks onto the gridded map and eliminate those never passed through the affected region. Each hurricane track contains multiple time-ordered track points, their coordinates, and the associated wind speed of each track point. We will then transform the physical wind speed into 6 categories based on NOAA standards. For the rest of this report, we will refer to the six wind categories as cat0, cat1, cat2, cat3, cat4, and cat5. The higher the wind speed category, larger the damages will be brought[1].

We consider the wind speed category constant along the hurricane track until a different value is provided at a track point. Therefore, for each stochastic event, we can associate to each map cell a hurricane wind speed category, determined by the maximum wind speed category within that cell.

Binary Parametric Matrix, A

After determining the maximum wind category for each map cell for a given hurricane track, we can then construct the parametric model for our optimization. The idea is breaking down the total loss of a given hurricane into a binary matrix, A, and a loss parameter vector, X. The binary matrix A is constructed in a way that we can infer the maximum wind category in a given cell for any hurricane track. The loss parameter vector, X, on the other hand, will be representing the loss contribution of each hurricane category in each cell.

[1] NOAA Hurricane Category Standard: Category 0: wind speed less than 74 mph; Category 1: wind speed between 74-95 mph; Category 2: wind speed between 96-110 mph; Category 3: wind speed between 111-129 mph; Category 4: wind speed between 130-156 mph; Category 5, wind speed equals 157 mph or higher.

More specifically, the category matrix *A* is a t*n matrix, where t equals the total number of hurricane tracks passed through the affected area, and n equals 6*the total number of cells. For example, for an affected area with 35 cells and 10 hurricane tracks passed through, matrix *A* will be a 10-row, 180-column matrix. Every 6 columns denoting the 6 potential wind category for each of the 35 cells, with at least one column taking value 1.

Therefore, for a hurricane track, we can estimate the total loss *L'* and the calculation is shown below as the sum of loss contribution from the cells intersected by the hurricane track. Here, $A_{t,ij}$ refers to the *ith* cell on track *t,* and takes value one if the maximum category in cell *i* is category *j*, and zero otherwise. X_{ij} is the loss contribution value of cell *i*, category *j*. As a result, we will be only capturing loss contribution from intersected cells, as the matrix *A* will take value zero and null the loss contribution from uncrossed cells.

$$L_t' = \sum_{i=1}^{all\ cells} A_{t,ij} * X_{ij}$$

Take an example in Figure 1. Let us consider slicing the affected region into an ordered grid map formed by 35 cells. The red line is a hurricane event track mapped onto the grid map. On the hurricane track, we can identify several event point that segregate the track and each event point is labeled with a wind category as constructed above. For a segment on the hurricane track, we will consider the wind category consistent with its starting category level until it hits another wind category. Also, to simplify our calculation, we assume the line between two event points straight. Therefore, the estimated total loss for this hurricane track can be constructed as below:

$$L' = X_{1,0} + X_{6,0} + X_{11,1} + X_{22,2} + X_{27,2} + X_{32,3} + X_{33,2} + X_{34,0} + 0 * \sum_{i=1}^{all\ unintersected\ cell} X_{i,j}$$

In this example, the *A* matrix assigns value one to cells that are intersected by the hurricane track, and zero otherwise. Therefore, we are only capturing the loss contribution in *X* from the intersected cells.

Figure 1.

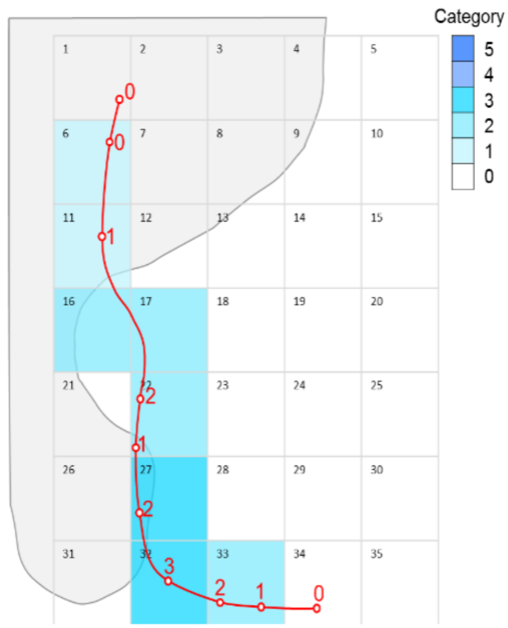


Figure 2: Example of a hurricane parametric model design.

From Parametric to Optimization

After constructing the parametric model to estimate hurricane loss using the "cat-in-box" scheme, we can start to build the optimization model. Recall that the goal of the insurance optimization is to reduce the gap between actual hurricane damages and insurance payment, which, in our case, is the estimated hurricane loss. Following this ideology, the optimization function is formulated as:

$$min \sum_{i=1}^{all\ tracks} |L_i - A_i * X|$$

Here, L_i refers to the actual loss of ith hurricane track, and $A_i * X$ is the estimated loss constructed as described above. Loss are all calculated in unit us dollar. In our project, the actual loss that we use as benchmark is generated by Guy Carpenter based on the maximum wind speed of a given the hurricane tracks. Specifically, the loss is uniformly sampled for each track based on the max hurricane category per track, according to the scheme as follows.

- If the max category = 0, actual loss randomly selected between [0, 1K]
- If the max category = 1, actual loss randomly selected between [1K, 10K]
- If the max category = 2, actual loss randomly selected between [10K, 100K]
- If the max category = 3, actual loss randomly selected between [100K, 1M]

- If the max category = 4, actual loss randomly selected between [1M, 50M]
- If the max category = 5, actual loss randomly selected between [50M, 1B]

III. Data Preparation

The goal of our modeling pipeline is to transform stochastic hurricane events data into the parametric form that we will use in the optimization model. The stochastic hurricane events dataset contains a large number of simulated hurricane tracks, equivalent to 10,000 years of data. Each hurricane is composed by a variable number of hurricane event points, and for each event point, the following information is provided:

- date: date and time of the event point
- x: longitude of the eye of the cyclone (degree);
- y: latitude of the eye of the cyclone (degree);
- mslp: minimum sea level pressure (mb);
- mws: maximum wind speed (kt);
- rmw: radius of maximum wind speed (km).

The dataset covers the North Atlantic Ocean only and it was developed based on data collected for hurricanes occurred in the North Atlantic Ocean from 1930 to 2017. There is a definitive relationship between the hurricane category level and its maximum wind speed, that when maximum wind speed is in a certain interval, the category level is assigned accordingly.

Since our dataset covers rather a large area in North Atlantic Ocean, we decided to narrow down the interested domain to around Jamaica whose center coordinates are [-77˚ W, 18˚ N], and take experiments on three types of domain size around Jamaica. We define three types of square domains around Jamaican center as follows:

- [-77˚ ± 6˚ W, 18˚ ± 6˚ N]
- [-77˚ ± 8˚ W, 18˚ ± 8˚ N]
- [-77˚ ± 10˚ W, 18˚ ± 10˚ N]

Then, for each domain sizes, we divide the region of interest in an ordered grid of cells according to the “cat-in-box” design. Following the procedure described in the previous section, we convert the dataset into a big matrix A based on its wind and geographical features. Each row of matrix A represents a hurricane track and each column is a binary entry that equals 1 if the hurricane is defined as a certain wind category in a certain cell. Every six columns will represent the six hurricane category in one “map cell”. For instance, if a track is defined as Category 0 in cell 1, the first 6 entries of this row would be (1,0,0,0,0,0); if the track doesn't pass through this cell, then the entries would be (0,0,0,0,0,0).

Simply converting all the hurricane points given in the dataset is not sufficient since it is possible that two adjacent hurricane event points might not land in the adjacent cells. We take the segment of connection between two adjacent hurricane event points and intersects with grid lines of cells. If there is no intersection, then the two hurricane event points are in the same cell; if there is only one intersection, then the two hurricane event points are in adjacent cells; if there are more than one intersection, then the two hurricane event points are not in adjacent cells. In this occasion, we generate two dummy hurricane event points for each intersection along the connection line where the dummy steps are extremely close to the intersection (0.0001 in degree).

Finally, we obtain the *A* matrix where each row represents a hurricane track, and every six column represents the maximum wind category of a certain map cell. We also produce three actual loss table containing only tracks passed through our three types of interested domain. After these steps, we are ready for the optimization part.

IV. Optimization Result

In order to test out the model, we test two types of optimization tools in MATLAB. Both of the optimization algorithms have the same optimization objective as to minimize the total prediction errors for all tracks between the actual loss, L, and predicted loss, A*X. The final output of our optimization would be the *X* matrix, which represent the optimal loss contribution for each map cell by wind category.

Optimization Objective:

$$min \sum_{i=1}^{all\ tracks} |L_i - A_i * X|$$

Our first model is the unconstrained optimization tool in MATLAB: fminunc. Using an unconstrained method, we ignore the potential constraint on loss parameters for each cell, where loss should be increasing as the category increases. However, this unconstrained algorithm has an advantage as a preliminary model in computational speed.

With a 6*6 domain, the fminunc algorithm reached the local minimum after less than 500 iterations, with 15% of loss decreased. To back-test the consistency of fminunc, we create a pseudo loss for each track that equals to the multiplication of matrix *A* and our result *X*, and applied it back to fminunc model. Ideally, we would expect a minimum objective value (estimation error) of zero, since the optimal *X* should account for all the pseudo loss. In reality, the result was promising, as the minimized estimated error decrease by over 95%. Hence we expand the domain to 8*8 and 10*10, the loss reduction is comparatively at the same level as that of domain 6*6.

	6*6 Domain	8*8 Domain	10*10 Domain
Minimum Objective Value	7.4e+11	7.8e+11	7.9e+11
Time Consumption	40 mins	3 hours	10 hours

Our second model is the unconstrained optimization tool in MATLAB: fmincon, where we take into consideration the constraint on loss parameters in output *X*, where higher hurricane wind category will lead to greater loss. We impose a simple constraint on the loss contribution of each category within each map cell, constructed as:

$$x_0 <= x_1 <= x_2 <= x_3 <= x_4 <= x_5$$

where x_i refers to the loss contributions corresponding to category i for a given cell

Testing the constrained algorithm on a 6*6 domain, we observe a local minimum of our objective value of estimation error at 8.7e+11. This minimum objective value is much higher than that of the unconstrained algorithm. However, the back-testing result shows that constrained optimization works well in explaining itself, where the pseudo loss decreased by over 91%.

V. Interpretation and Evaluation

Output Loss Parameter Pattern

Since we have solved for *X*, the output vector, we want to explore the pattern of the loss contribution of the parameters in *X*. By comparing the loss contribution for wind categories 0-4 within the same cell, we found out that the incremental increase of the loss contribution has an exponential growth. We applied the exponential model on the loss parameters of categories 0-4 and find an R-squared value as high as 0.902.

$$x_c = \beta_1 \cdot \beta_2^c$$

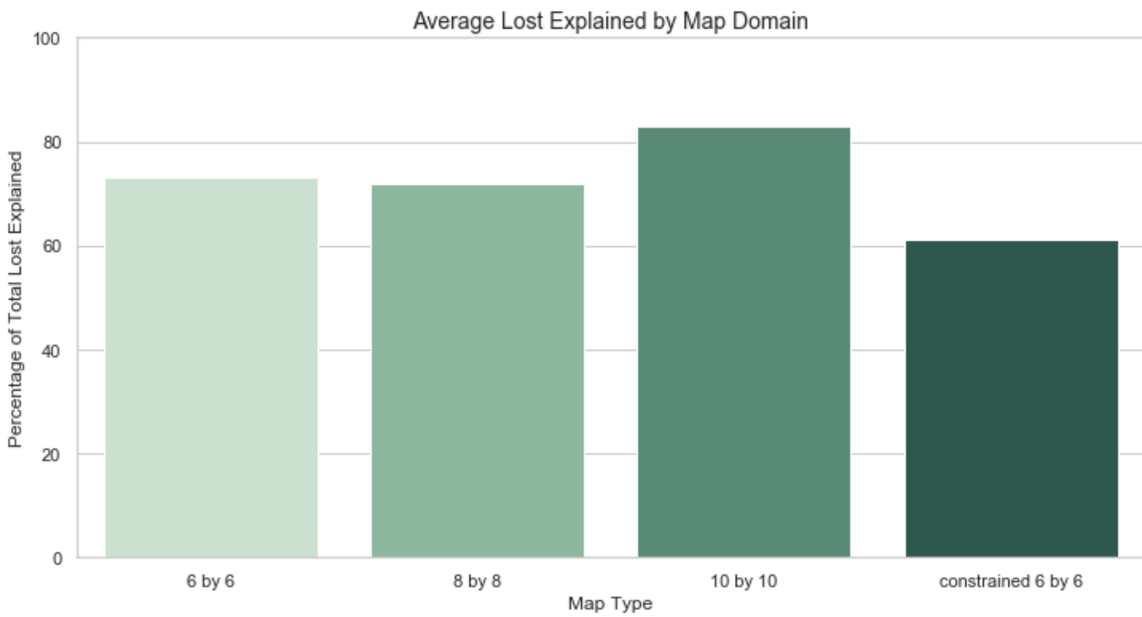
where x_c is the loss of a cell at category $c \in [0, 4]$, Here, cell loss for category 5 is excluded because it falls way off the fitting line, which can be explained by having too few data entry of category 5 in the entire dataset. Thus, it is highly possible that there is a relationship of loss and category level from 0 to 5 described as above. Such relationship found among the loss parameters output from the unconstrained algorithms shed light on other potential constraints that we can use in our constrained optimization. However, the pattern of the loss contribution could also be affected by the underlying distribution of our actual loss, which, in our case, is exponentially generated.

Actual Loss Explained

Recall that our ultimate goal is to approximate the hurricane loss and minimize the gap between insurance payment and actual damages amount. Therefore, we back-test our optimal X outputs by looking at the average loss explained per track. By simply multiplying the parametric matrix A and our output X, we calculated the estimated loss, L', for each hurricane track. Then we divide L' by the actual loss, L, for each track to get the percentage loss explained. Since underpayment will be our major concern compared to overpayment, we assign 100 percent for tracks that has higher estimated loss than actual loss.

Below is a summary of the average loss explained for our experiments. The unconstrained results perform better overall and account for more than 70% of the average loss. The constrained result, on the other hand, still explained 60% loss per track. These results confirms that our optimization works well in approximating the hurricane loss on average. Furthermore, we would expect a better estimation accuracy as we expand the interested domain size.

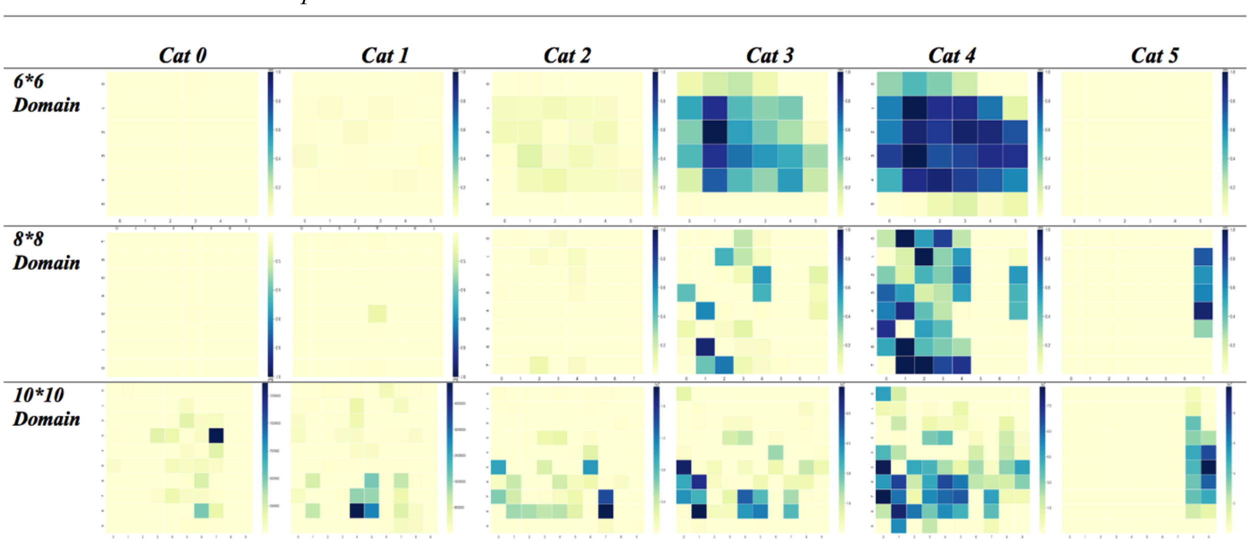
Table 2. Average Loss Explained



Loss Distribution on the Heat Map

In order to further explore the distribution of loss explained in terms of location and wind categories, we reshaped the output loss parameter, *X,* onto the affected domain for each 5 wind categories. Below is a colored heatmap of the unconstrained output *X,* which shows all the loss contribution explained by each category and by three domain sizes. In the colored map, each square cells represent 1-degree^2 square on the map, and the darker the color, the more damages are explained by that cell for that wind category.

Table 3. X Matrices Heatmap



We can find from the graphs that, for all domain type, as wind level category increases, the cells become darker and more damages are explained by the center and left-bottom areas in the domain. One outlier is the 5^{th} category which has much lower loss contribution, since the unconstrained optimization allocate it with less weight explaining the loss due to its low frequency.

Also, if take a vertical comparison, we can find that as we increase the affected domain, loss contribution are diluted and explained by the surrounding cells. In the 10*10 domain, the distribution of loss also moves towards the left-bottom area of the map with a smoother degrading color. This complies more with the real-world case, where the hurricane tracks mostly passed through the left-bottom areas of the Jamaican domain.

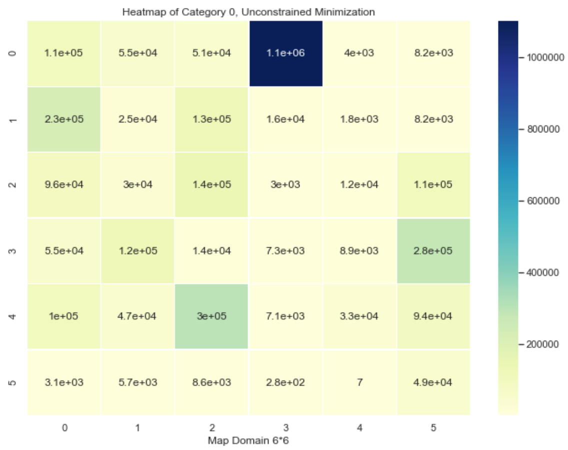
VI. Conclusion

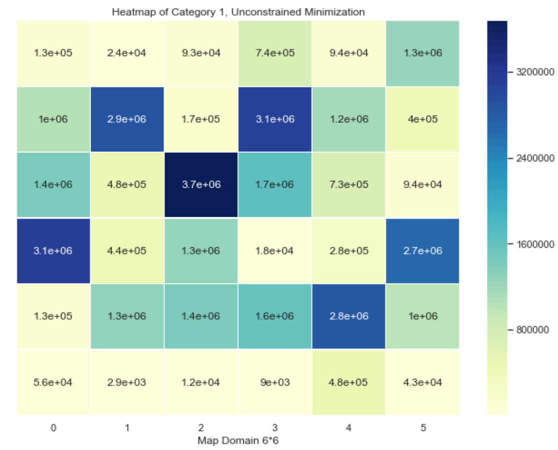
In this project, we extend the tropical cyclone risk design by Guy Carpenter into hurricane insurance optimization. Using the “cat-in-box” scheme, we were able to partition the hurricane loss based on hurricane events location and wind category features. We built a complete modeling pipeline that automatically transformed hurricane events features into the desired parametric form, and built two optimization models to estimate the optimal hurricane payments. By assessing the results performance under different domain size, we can explore the model resilience to risk and accuracy.

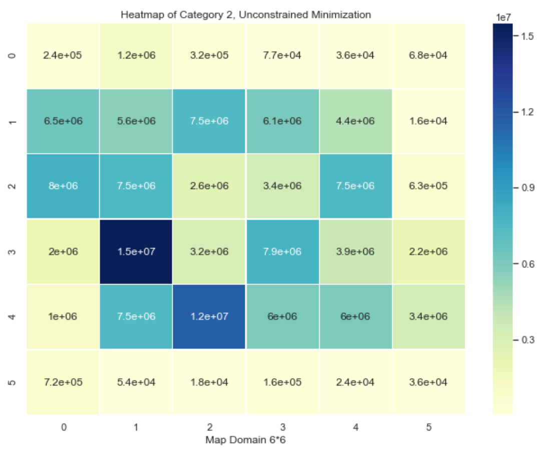
The results of the our optimization models confirms that the “cat-in-box” parametric design works well in our experimental project to estimate the optimal hurricane payment by setting the real hurricane loss as a benchmark. Aiming at minimizing the difference between actual loss and estimated hurricane loss, both the unconstrained and constrained algorithms explain loss per track on average at least 60 percent. As we expand the interested domain size around Jamaican area, the unconstrained optimization explained average loss over 80%. One takeaway is that other than the optimization mechanism, the size of the affected area also influence the optimization accuracy.

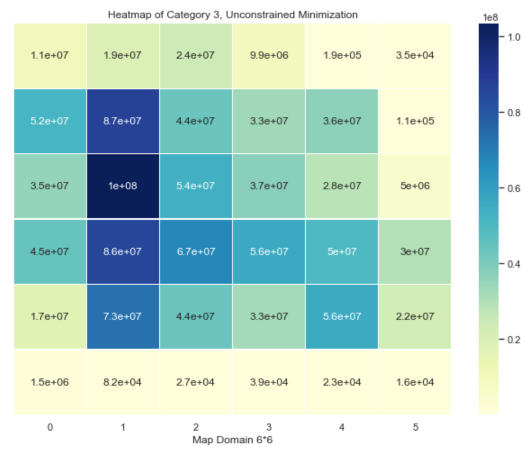
On the other hand, there are still some limitations in our experiment and thus room for future development. Rather than using real world actual loss, we use a synthetic actual loss data. However, this does not affect the explanation power of our optimization model which is capable for extension on real-world damages data. Also, we could improve our explanatory accuracy by incorporating more precise constraints on the loss contribution of each wind category.

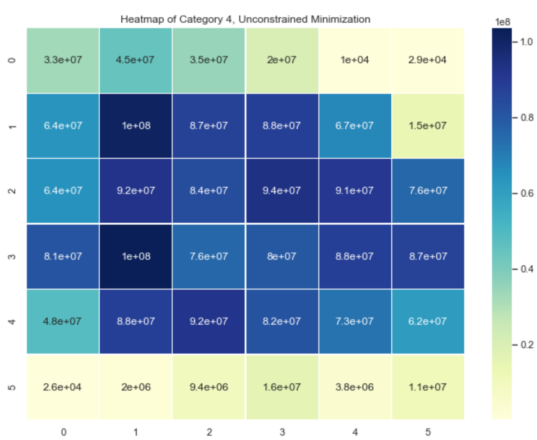
Appendix II: Unconstrained Heatmap on 6*6 Domain

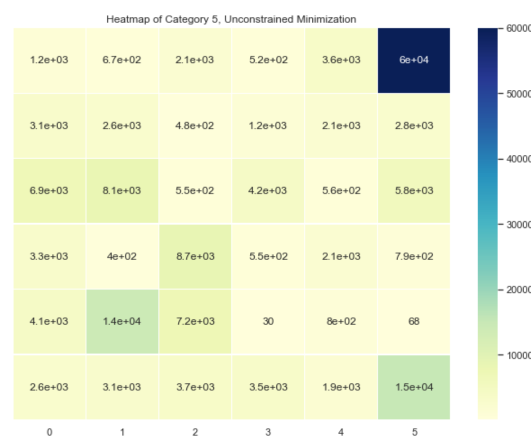




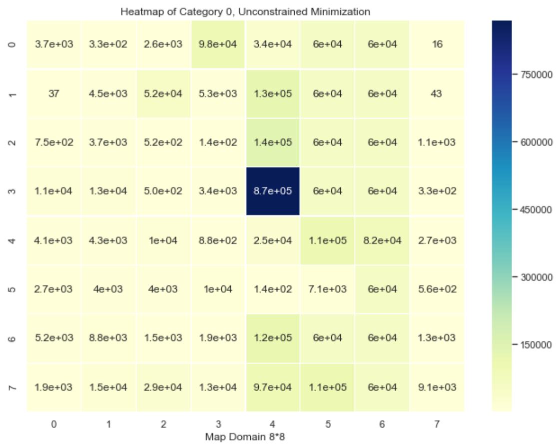


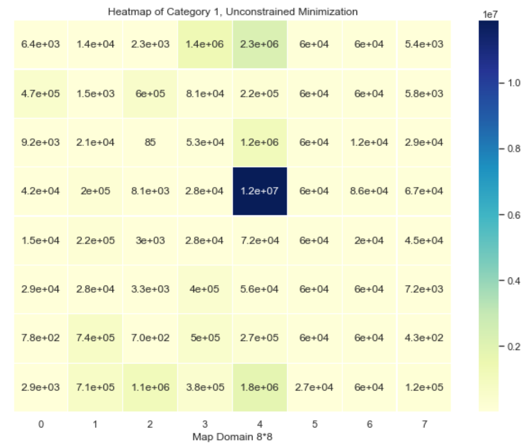


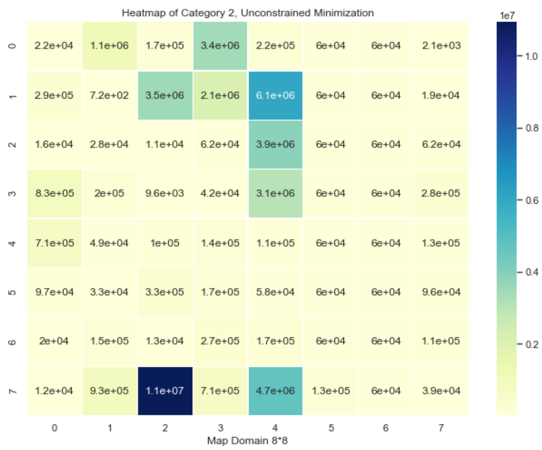


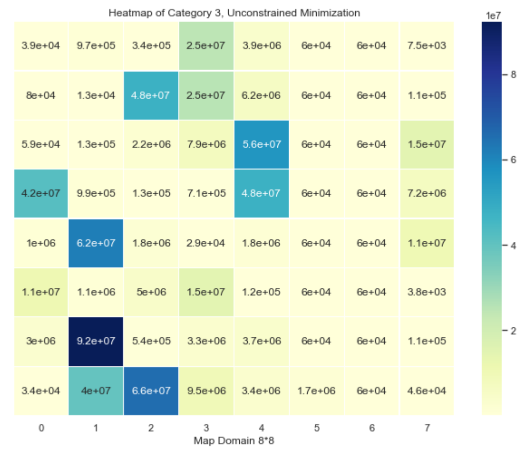


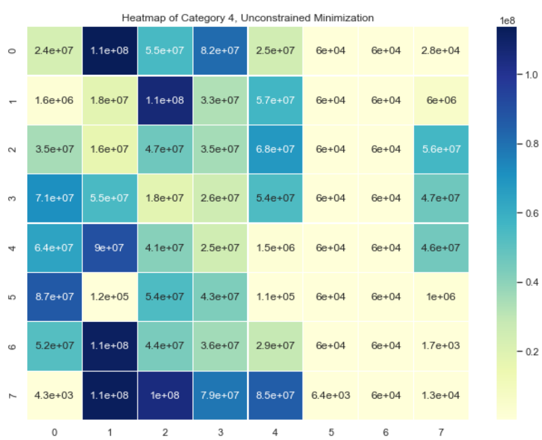
Appendix III: Unconstrained Heatmap on 8*8 Domain

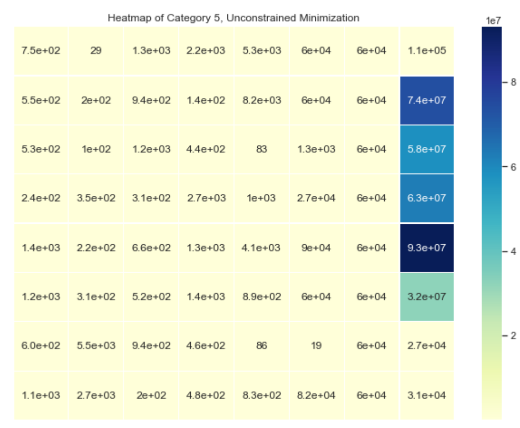












Appendix IV: Unconstrained Heatmap on 10*10 Domain