Classifying North Atlantic Tropical Cyclone Tracks by Mass Moments

JENNIFER NAKAMURA, UPMANU LALL, YOCHANAN KUSHNIR AND SUZANA J. CAMARGO

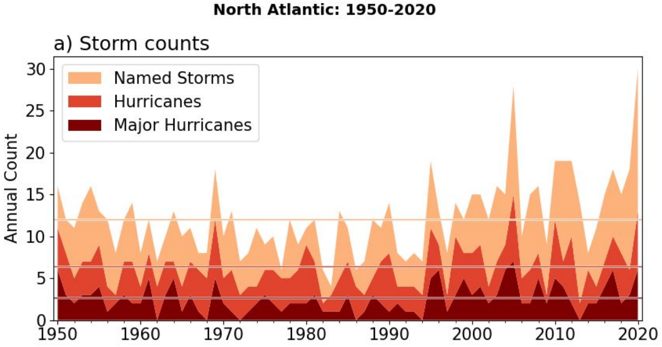
Paper walkthrough and Colab notebook summary

Abstract of the Paper

A new method for classifying tropical cyclones or similar features is introduced. The cyclone track is considered as an open spatial curve, with the wind speed or power information along the curve considered to be a mass attribute. **The first and second moments** of the resulting object are computed and then used to classify the historical tracks using standard clustering algorithms. Mass moments allow the whole track shape, length, and location to be incorporated into the clustering methodology. Tropical cyclones in the North Atlantic basin are clustered with **K-means by mass moments**, producing an optimum of six clusters with differing genesis locations, track shapes, intensities, life spans, landfalls, seasonal patterns, and trends.

Motivation

- Hurricanes are dangerous, destructive. Understanding them supports predictions and, in turn, human preparedness.
- Hurricanes numbers vary substantially from year to year.
  - If we knew better the factors causing this variability, we should be able to make more accurate, more detailed predictions
    - If it could be determined several months in advance that a major hurricane will hit Miami, people can make preparations to limit loss of life and property.
  - Various groups make predictions several months in advance, but no one gets it quite right. And we cannot predict landfall until a few days in advance. Due lacking physical knowledge, and because of stochasticity.
- How will hurricanes change with climate change? Another big question. (Gabriel Vecchi lecture, 1/21/22, LDEO Colloquium)



https://www.ncdc.noaa.gov/sotc/tropical-cyclones/202013

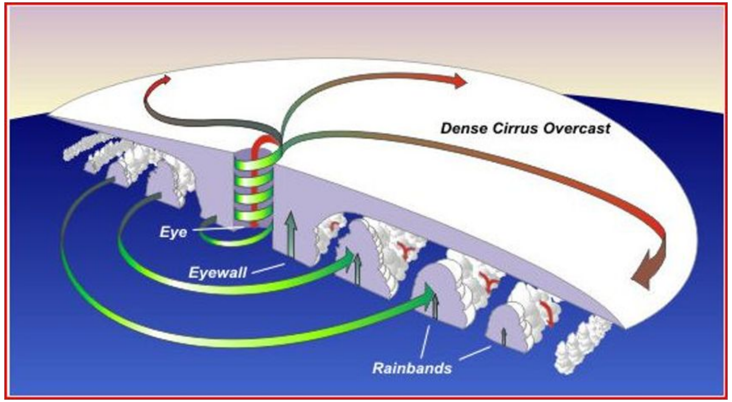
What is a hurricane?

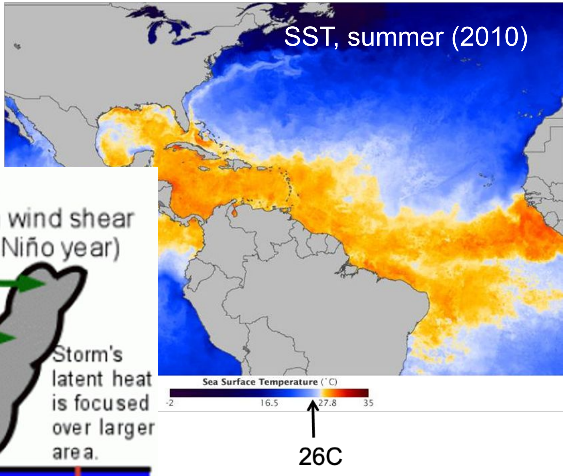
Development - two reinforcing processes of upward and inward motion

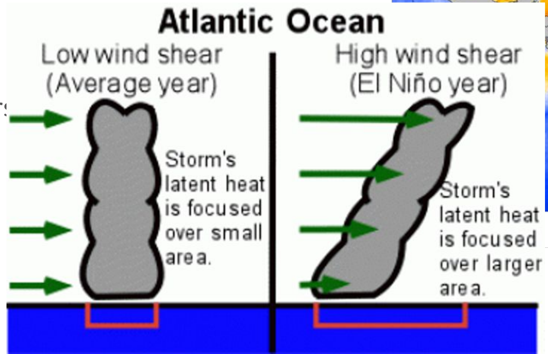
- Upward: Convection over the warm ocean, extracting latent heat from water
  - Latent heat is released in rising air, which makes the air more buoyant, thus encouraging more upward convection
- Inward: Air flows in toward the central Low (turning to right due to Coriolis)
  - Moist air rising at center pulls more air toward center

Ingredients to form a hurricane

- Sea surface temperature (SST) > 26 C
- High humidity in mid-troposphere
- Winds that change little with altitude = low wind shear
- Atmospheric conditions to produce low pressure centers



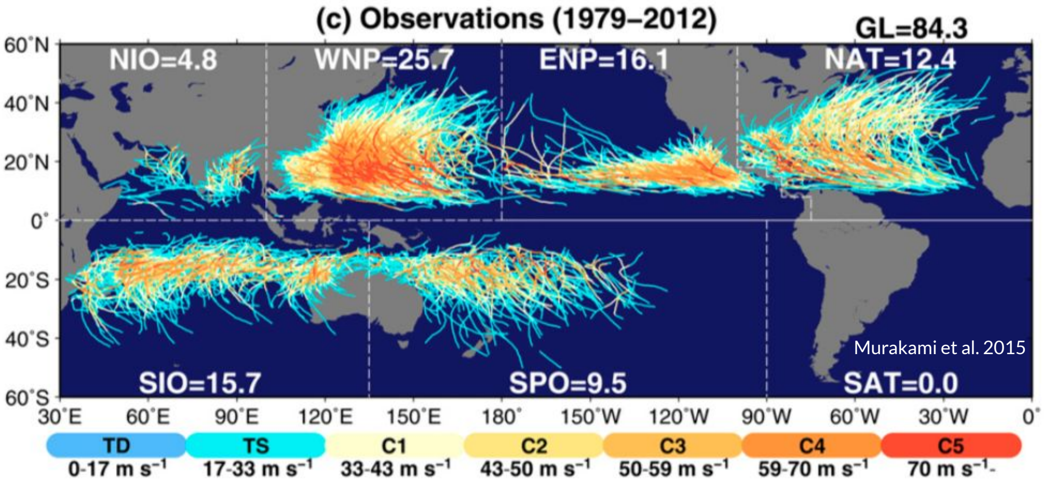




[Short video from National Geographic](https://www.youtube.com/watch?v=zP4rgvu4xDE)

https://www.youtube.com/watch?v=zP4rgvu4xDE

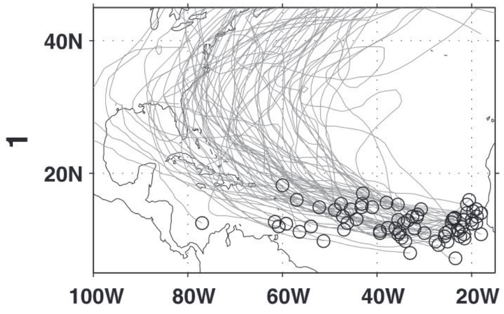
Hurricane tracks



- Hurricane tracks indicate where storm forms and where it dies
- Tracks connect the physical environment where the storm forms, through the region where it develops, and then to the (potential) landfall region where impacts occur
- Predicting storm impacts is strongly linked to understanding these tracks

- Nakamura et al. (2009) cluster North Atlantic hurricane tracks and assess the characteristics of the clusters

Cluster 1 of Nakamura et al. (2009) is most dangerous



highest winds

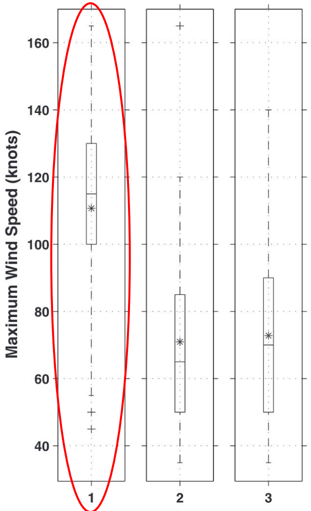


Figure 5: Max wind

longest life

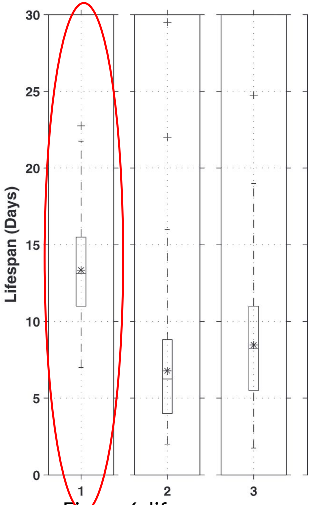


Figure 6: lifespan

Typically category 3 or 4

TABLE 3. Mean and median category of the maximum wind speed in each cluster and in all tropical cyclones.

Cluster	Mean	Median
1	3	4
2	1	1
3	1	1
4	1	TS
5	2	2
6	1	TS
All	1	1

~half make landfall

TABLE 4. Number of tropical cyclones (N) and landfalls and the landfall percentage for each cluster and for all tropical cyclones.

Cluster	N	Landfalls	Landfall percentage
1	70	32	46
2	157	35	22
3	94	0	0
4	92	26	28
5	43	14	33
6	174	148	85
All	630	255	40

What else can be explored?

- Assess how winds speeds evolve as hurricanes move across the ocean.
- Are hurricanes becoming more numerous over time? How does your analysis through 2021 compare to Nakamura et al. 2009 through 2006?
- Add data of climate indices such as El Nino. Quantify the impact on hurricane numbers in the clusters.
- Find data on historical costs of hurricanes making landfall in the US. Do the clusters have any predictive power?
- Get data for another region. Compare clusters across regions.

- ….. and much, much more! Reading the literature to understand the state of the field and open questions

Abstract of the Paper

A new method for classifying tropical cyclones or similar features is introduced. The cyclone track is considered as an open spatial curve, with the wind speed or power information along the curve considered to be a mass attribute. **The first and second moments** of the resulting object are computed and then used to classify the historical tracks using standard clustering algorithms. Mass moments allow the whole track shape, length, and location to be incorporated into the clustering methodology. Tropical cyclones in the North Atlantic basin are clustered with **K-means by mass moments**, producing an optimum of six clusters with differing genesis locations, track shapes, intensities, life spans, landfalls, seasonal patterns, and trends.

From Hurricane Data to Features

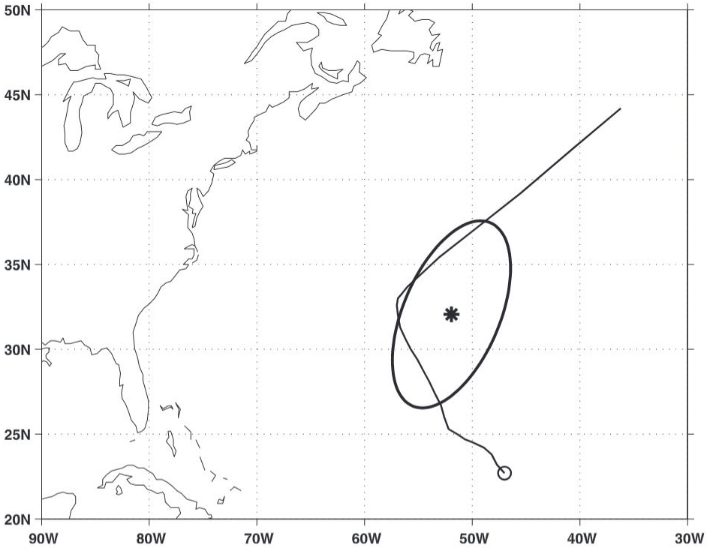


FIG. 1. Examples of the centroid (asterisk) and variance ellipse for the storm track. The track start is marked with a circle.

$$\mathbf{M1} = \frac{1}{A}\int w(\mathbf{r})\mathbf{r}\,dx\,dy = \frac{1}{\sum_{i=1}^{n} w(r_i)} \sum_{i=1}^{n} w(\mathbf{r}_i)\mathbf{r}_i, \qquad (1)$$

$$\mathbf{M2} = \frac{1}{A}\int w(\mathbf{r})(\mathbf{r} - \mathbf{M1})^2\,dx\,dy$$

$$= \frac{1}{\sum_{i=1}^{n} w(r_i)} \sum_{i=1}^{n} w(\mathbf{r}_i)(\mathbf{r}_i - \mathbf{M1}_i)^2, \qquad (2)$$

Why clustering?

- finding subgroups in a data set
- assess concentrations in spatial characteristics such as track shape, genesis location, intensity, life span, seasonality, and landfall.
- identify each cluster’s association with other factors

Clustering (or classification)

- Not that classification (as in *supervised learning*)
- Clustering is an *unsupervised learning* task
- Similarity is central to the task of clustering.
    - "we must define what it means for two or more observations to be similar or different."
    - Distance, dissimilarity
    - Correlation, measures of association
- Learning goal of a clustering analysis
    - Identify groups to
        - maximize within-group similarity
        - minimize between-group similarity

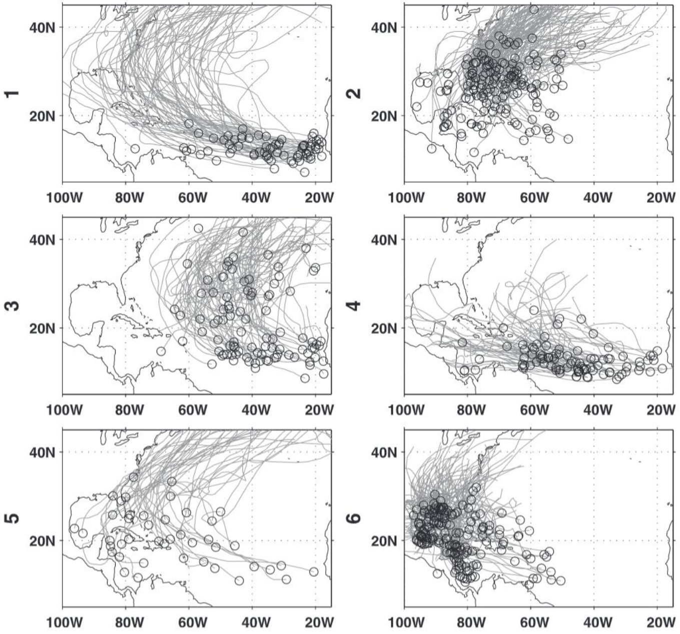


FIG. 4. Genesis location (circle) and the track (line) for the six clusters.

K-means algorithm

Algorithm 12.2 *K-Means Clustering*

1. Randomly assign a number, from 1 to K, to each of the observations. These serve as initial cluster assignments for the observations.

2. Iterate until the cluster assignments stop changing:

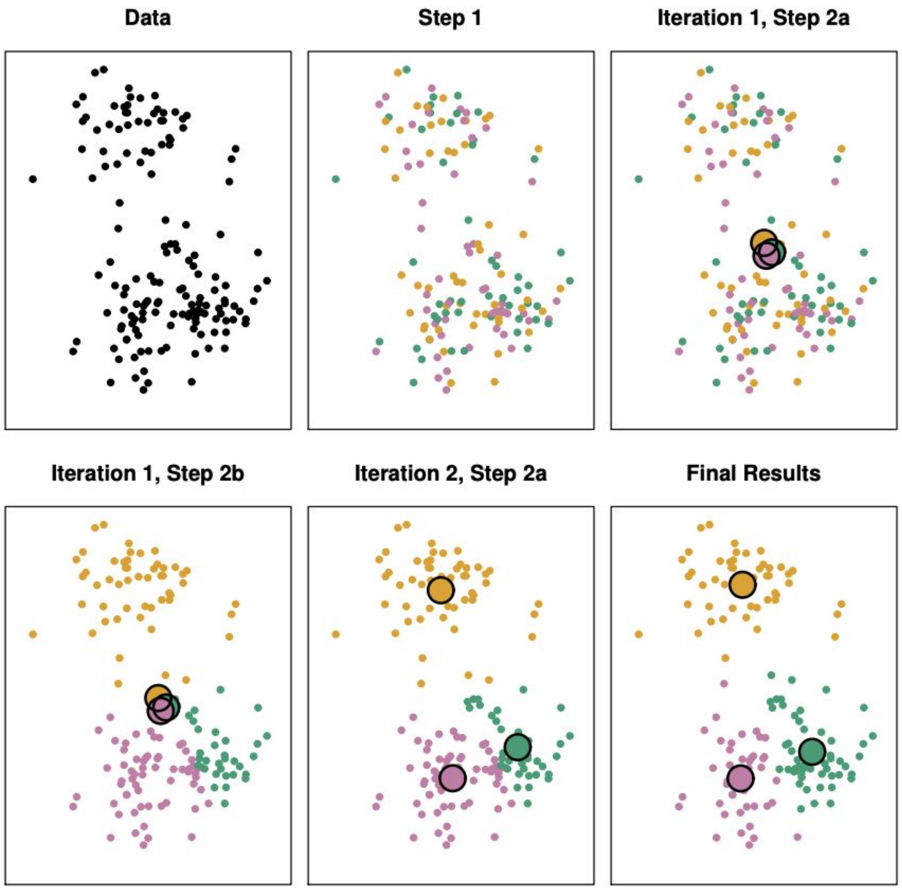
   (a) For each of the K clusters, compute the cluster *centroid*. The kth cluster centroid is the vector of the p feature means for the observations in the kth cluster.

   (b) Assign each observation to the cluster whose centroid is closest (where *closest* is defined using Euclidean distance).

K-means algorithm

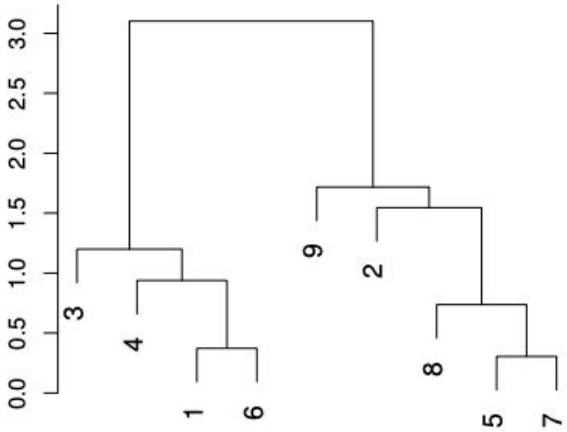
- n observations, each with a p-dimensional feature.
- minimizes within cluster variances (sum of squared distances but not Euclidean distances).
- Known as a ***centroid***-based method of clustering.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, p. 18). New York: springer.

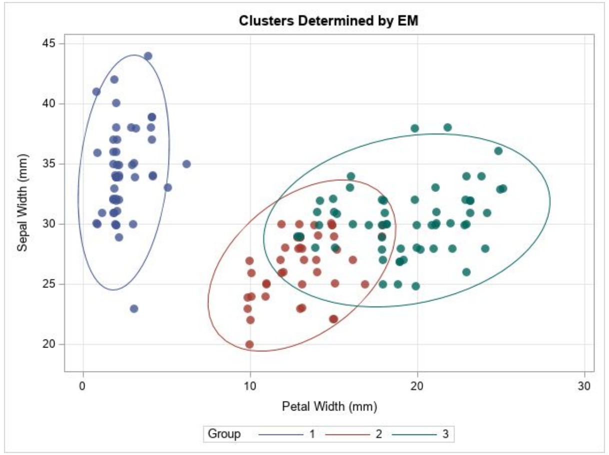


Other types of clustering analysis

Hierarchical clustering



Model-based clustering (mixture models)



Clustering Hurricanes using K-Means

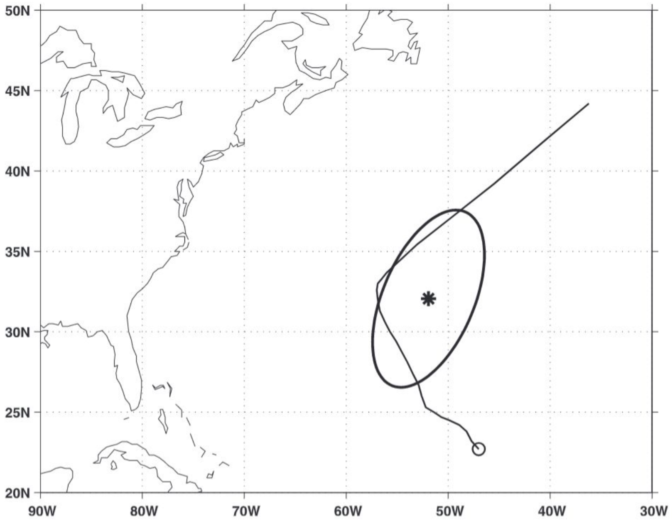


FIG. 1. Examples of the centroid (asterisk) and variance ellipse for the storm track. The track start is marked with a circle.

- Convert each hurricane track (unequal length) to a 5-dimensional vector
- Suppose that X and Y stand for longitude and latitude as random variables of a tracks, then the vector can be expressed as:

[E[X], E[Y], Var[X], Var[Y], Cov{X,Y}]

How to choose K?

Silhouette values Si (for i points in total) are used in determining the optimal number of clusters (where ai is the average distance from the ith point to the other points within the cluster and bi is the average distance from the ith point to points in another cluster (*Kaufman and Rousseeuw 1990*).

$$S_i = \frac{\min(b_i) - a_i}{\max[a_i, \min(b_i)]},$$

The paper suggests 6 be a good number of clusters, and analysis was done accordingly.

In the notebook, however, **elbow method** is used (which uses inertia rather than silhouette values as criterion) to explore different ways of clustering.

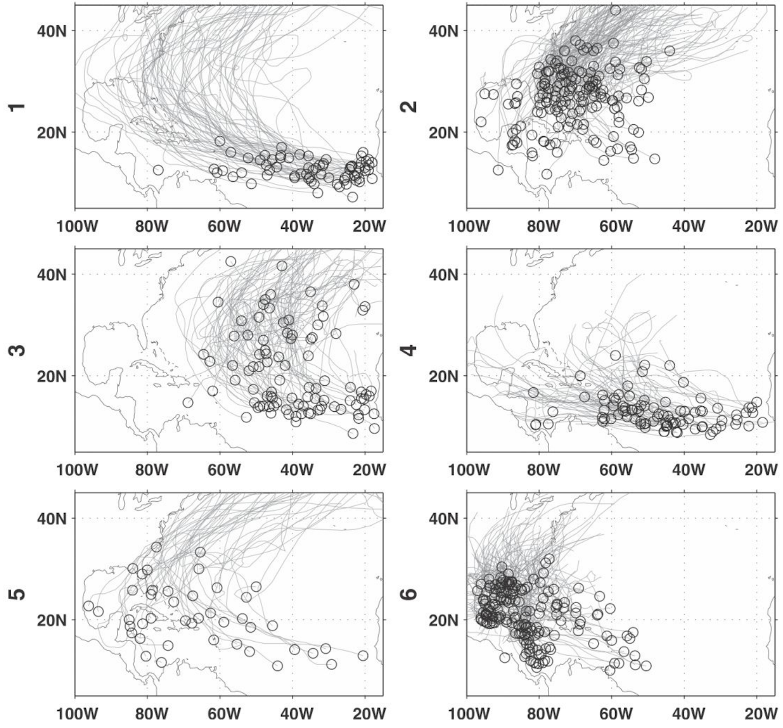
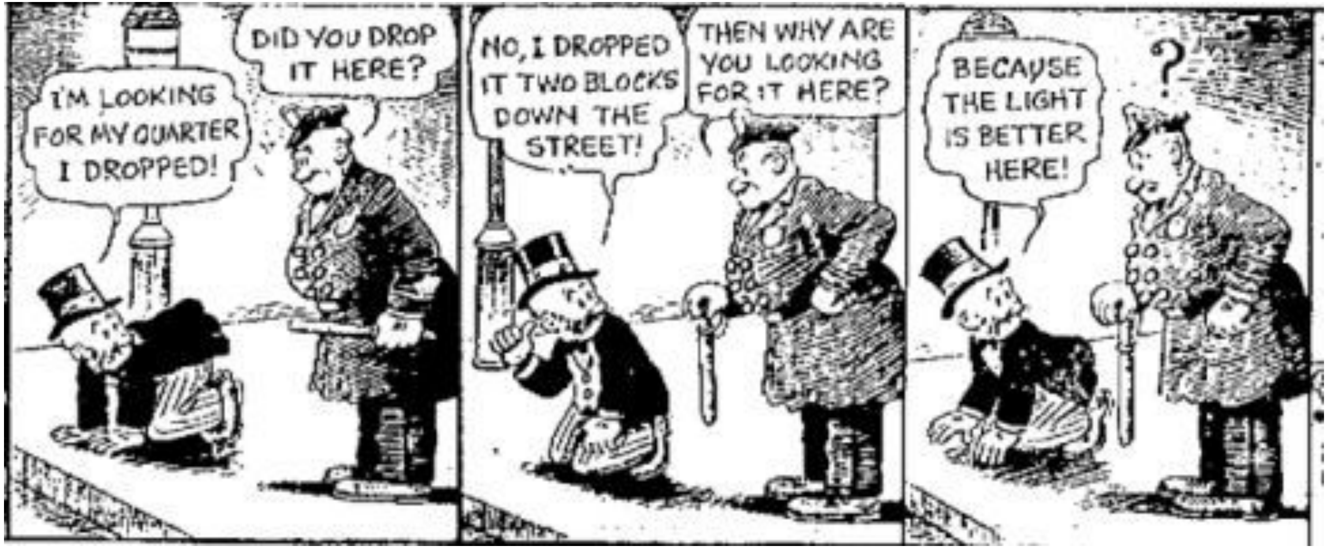


FIG. 4. Genesis location (circle) and the track (line) for the six clusters.

EDAV: look where you have light and bring in new light where you want to look!



https://axispraxis.wordpress.com/2016/03/24/the-streetlight-effect-a-metaphor-for-knowledge-and-ignorance/

Now let's look at the Google Colab Notebook

Clustering Results (continued...)

TABLE 2. Mean and range of tropical cyclone genesis locations for each cluster.

Cluster	Mean x genesis (°W)	Mean y genesis (°N)	Genesis x range (°W)	Genesis y range (°N)
1	34.60	12.61	77–18	7.2–18.2
2	70.50	26.64	96–44	11.7–44
3	39.35	20.76	68.8–17.4	8.6–42.5
4	46.83	12.76	81.6–14	8.4–24
5	65.89	20.67	96.2–20.5	10.9–34.3
6	83.16	20.77	97–50.5	10–32

Centroid information

TABLE 4. Number of tropical cyclones (N) and landfalls and the landfall percentage for each cluster and for all tropical cyclones.

Cluster	N	Landfalls	Landfall percentage
1	70	32	46
2	157	35	22
3	94	0	0
4	92	26	28
5	43	14	33
6	174	148	85
All	630	255	40

In the paper, landfall is one of many analyses done about the clustering results. It’s defined as the whether a hurricane had eventually made it to the land (causing more damage) before it died.

What does the Colab notebook cover?

Colab notebook: almost exactly the same as Jupyter Notebook except that it enables and facilitates sharing and collaboration among Google account users. It also has lots of built-in packages, which should need setup in a regular Jupyter Notebook.

The notebook contains helper codes for visualization, method for clustering, and example analysis.