Extracting Dense Regions from Hurricane Trajectory Data

Praveen Kumar Tripathi *Madhuri Debnath* *Ramez Elmasri*
Department of Computer Science, University of Texas Arlington

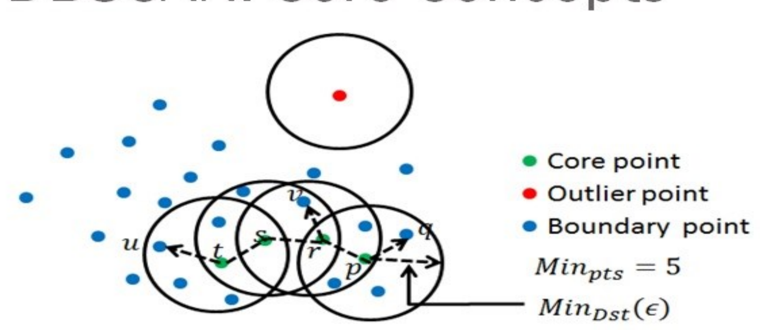
Abstract:

Weather data is a classic example of spatio temporal data, with time and space as two of its key attributes. Clustering has been one of the key techniques used for analyzing the storm trajectories. Our work considers a hurricane trajectory as a set of points. We focus on the point data for finding the regions that are hot spots for the storms. We use DBSCAN algorithm. We use different set of attributes, first only the spatial (longitude and latitude), then the combination of spatial with the non spatial attributes (wind speed and time). The results show the impact of the respective non-spatial attributes on the spatial attributes during clustering and hence the identified dense regions. We use quality measures to validate our clusters.

Contributions:

- used DBSCAN to do obtain the clusters considering trajectory data as point data.
- Evaluated the impact of non spatial attributes, viz., wind speed and time on the spatial clustering results.
- Identified the dense regions (clusters), as the hot spots for the regions corresponding to the storm starting spots, storm landing spots and the spots prone to storm activity.
- Proposed a relative temporal framework to incorporate the time information for the analysis of storms using temporal – DBSCAN

DBSCAN: Core Concepts



q is directly density reachable from p.
v is density reachable from p (through core point r)
u and q are density connected (through core points s or r)

Cluster Definition: A cluster C is a subset of objects satisfying the following:

- **Connected:** $\forall p, q \in C, p$ and q are density connected
- **Maximal:** $\forall p, q$, if $p \in C$, and q is density reachable from p, then $q \in C$.

Experiments and Analysis:

Data set :

- Temporal Coverage : 1950 – 1999 (50 years) (15319 data points, 496 trajectories)
- Spatial Coverage : North Atlantic
- Interval : 6 hourly (0000,0600,1200,2400)
- Source : http://weather.unisys.com/hurricane/atlantic

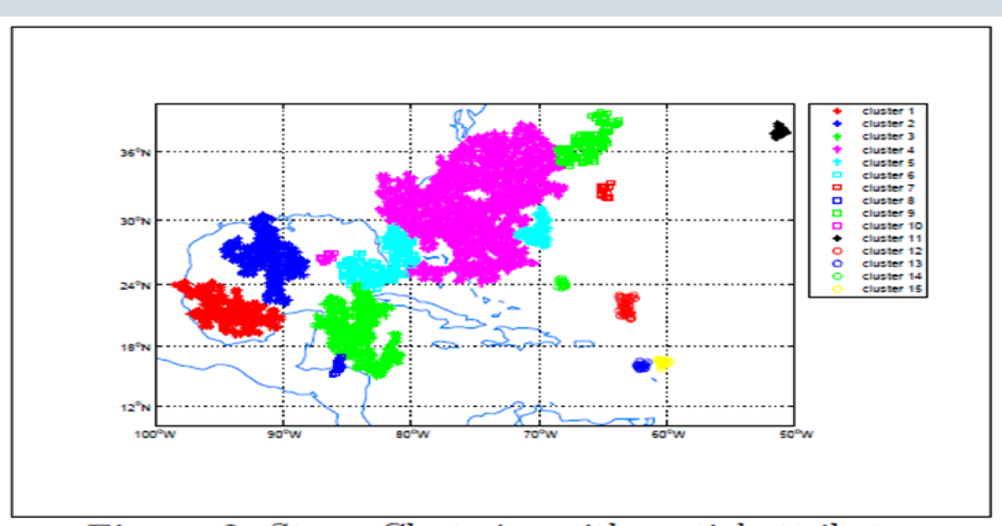


Figure 2: Storm Clustering with spatial attributes

Table 1: Storm clustering analysis, on Spatial clustering , $Min_{Pts} = 10$, $Min_{Dst} = 35\ km$

Storm ID	$Storm_{rank}$ (#traject.)	$Storm_{rank}$ (#data points)	$\#traject.(Cluster_i)$	$\#DataPts.(Cluster_i)$	$Color(Symb.)$
1	6	4	53	361	Red (star)
2	3	3	68	371	Blue (star)
3	2	2	80	471	Green (star)
4	1	1	185	1673	Magenta (star)
5	7	7	37	76	Cyan (star)
6	5	5	58	244	Cyan (square)
7	10	9	16	24	Red (square)
8	4	6	60	144	Blue (square)
9	9	11	18	20	Green (square)
10	14	10	9	22	Magenta (square)
11	8	8	27	40	Black (star)
12	13	14	12	15	Red (circle)
13	11	15	14	15	Blue (circle)
14	15	12	9	20	Green (circle)
15	12	13	14	17	Yellow (circle)

Figure 5: Storm starting clusters

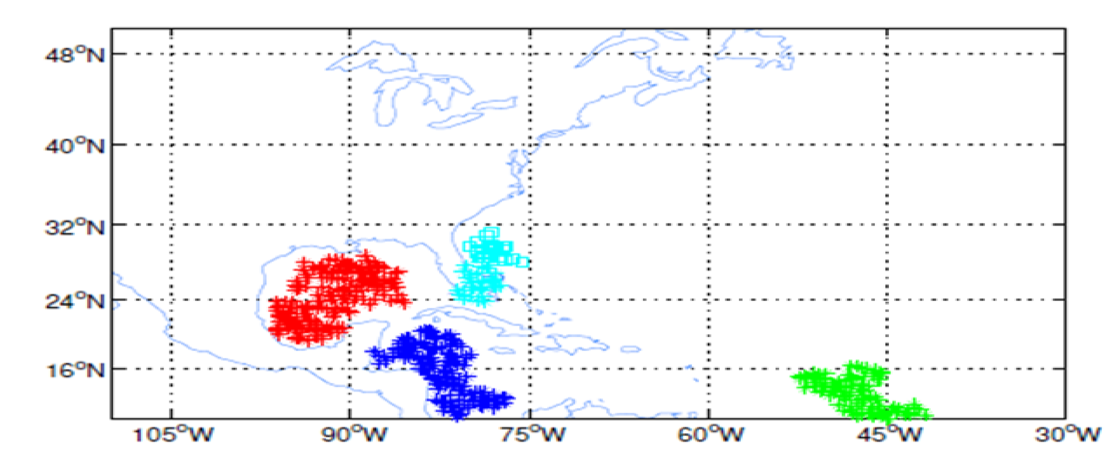
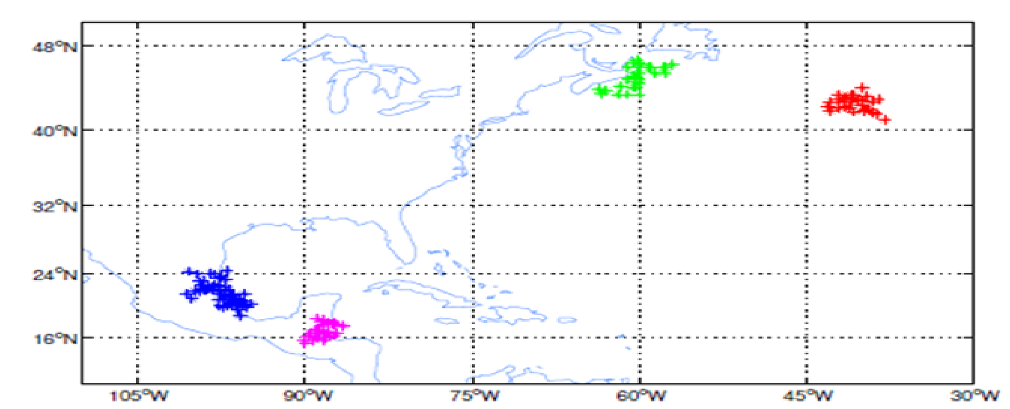


Figure 6: Storm Landing clusters



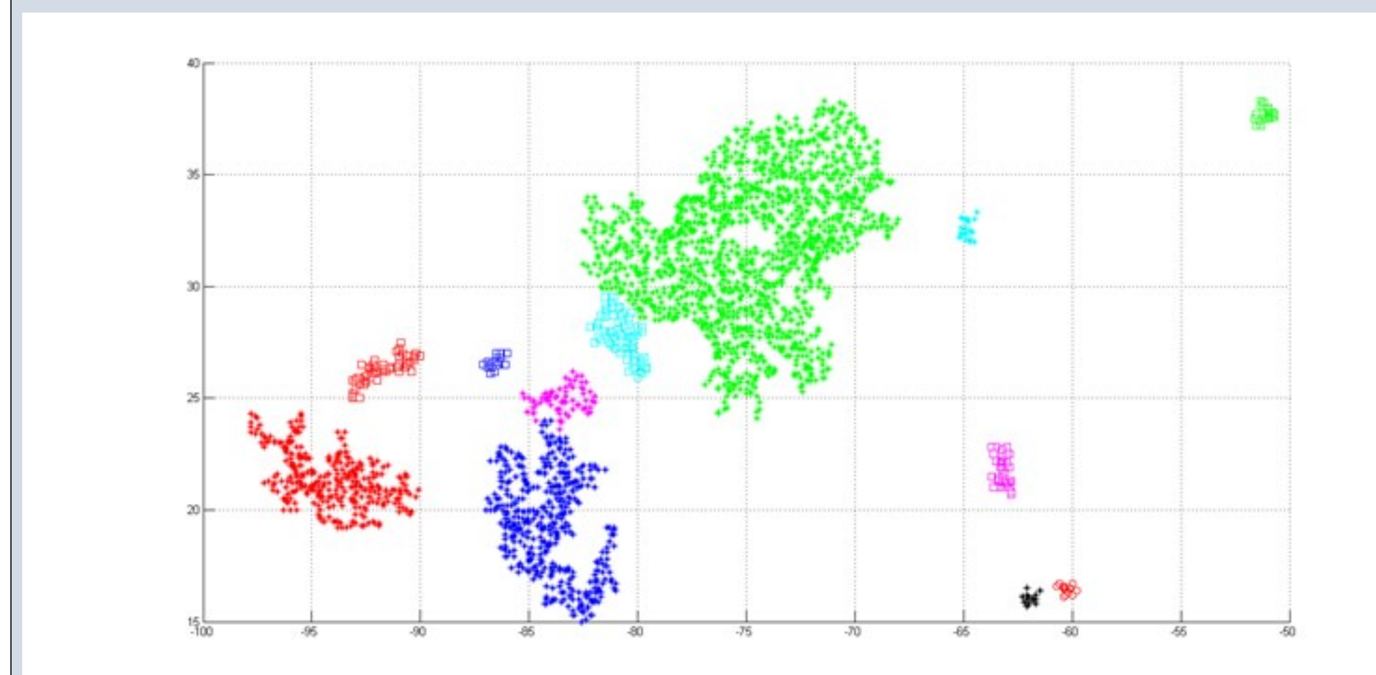
[2] Quality measure $= \sum_{i=1}^{num_{clus}} \left(\frac{1}{2C_i} \sum_{x \in C_i} \sum_{y \in C_i} dist(x,y)^2 \right) + \frac{1}{2|N|} \sum_{w \in N} \sum_{z \in N} dist(w,z)^2$ (2)

Table 3: Qualitative measure of clustering results

$Min_{DstW speed}(mph)$	$Q_{W speed}(mph^2)$	$Q_{spatial}(km^2)$
20	4.92E+04	3.35E+04
30	1.49E+05	8.62E+04
40	1.15E+06	2.56E+05
50	1.43E+06	2.96E+05
60	1.83E+06	3.46E+05
70	2.20E+06	4.11E+05
80	2.41E+06	4.73E+05
90	2.53E+06	5.18E+05
100	2.55E+06	5.18E+05

Table 2: Storm Clustering analysis, Impact of Non spatial attribute $Min_{Pts} = 10$, $Min_{Dst} = 35\ km$

$Min_{DstW speed(mph)}$	$Mean(Std(Cluster\_i))$	(#clusters)
20	13.5669	5
30	16.222	6
40	22.4366	9
50	23.2284	9
60	25.458	12
70	25.9649	14
80	26.3079	14
90	27.6229	15
100	27.6691	15



References:

1.Martin Ester , Hans-peter Kriegel , Jörg S , Xiaowei Xu, A density-based algorithm for discovering clusters in large spatial databases with noise , Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96), pp. 226–231

2. Birant, D., and Kut, A. ST-DBSCAN: An algorithm for clustering spatial-temporal data. Data Knowledge Engineering , 2007 pp 208-221

3.Gil Lee and Han, J. Trajectory Clustering: A partition-and-group framework. In SIGMOD 2007, pp 593-604.

4. Gudmundsson, J, Thom A, Vahrenhold J, Of motifs and goals : mining trajectory data. In Proceedings of the 20th International Conference on Advances in Geographic Information Systems, SIGSPATIAL'12 129-138.

5. Gil Lee, J Han , X Li and H Gonzalez, Traclass : Trajectory Classification using hierarchical region-based and trajectory-based clustering VLDB 2008.

6..Vieira M., Bakalov P., and Tsotras V., On-Line Discovery of Flock Patterns in Spatio-Temporal data, Proceedings of 17th ACM SIGSPATIAL International conference, GIS'09, pp. 286-295

7..Jeung H., Yiu M., Zhou X. Jensen C., and Shen H., Discovery of convoy in Trajectory databases. PVLDB, 2008.

8..Kalnis P., Mamoulis, and Bakiras S., On Discovering Moving Clusters in Spatio-Temporal data, In SSTD 2005, pp. 364-381.