Global Cyclone Detection and Tracking using Multiple Remote Satellite Data

Ashit Talukder, Shen-Shyang Ho, Timothy Liu, Wendy Tang, Andrew Bingham, Eric Rigor
Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Ave, MS: 300-123
Pasadena, CA 91109
Email: *FirstName.LastName*@jpl.nasa.gov

Abstract

Abstract-Current techniques for cyclone detection and tracking employ NCEP (National Centers for Environmental Prediction) models from in-situ measurements. This solution does not provide global coverage, unlike remote satellite observations. However it is impractical to use a single Earth orbiting satellite to detect and track events such as cyclones in a continuous manner due to limited spatial and temporal coverage. One solution to alleviate such persistent problems is to utilize heterogeneous sensor data from multiple orbiting satellites. However, this solution requires overcoming other new challenges such as varying spatial and temporal resolution between satellite sensor data, the need to establish correspondence between features from different satellite sensors, and the lack of definitive indicators for cyclone events in some sensor data.

In this NASA Applied Information Systems Research (AISR) funded project, we describe an automated cyclone discovery and tracking approach using heterogeneous near real-time sensor data from multiple satellites. This approach addresses the unique challenges associated with mining, data discovery and processing from heterogeneous satellite data streams. We consider two remote sensor measurements in our current implementation, namely: the QuikSCAT wind satellite data, and the merged precipitation data using TRMM and other satellites. More satellites will be incorporated in the near future and our solution is sufficiently powerful that it generalizes to multiple sensor measurement modalities. Our approach consists of three main components: (i) feature extraction from each sensor measurement, (ii) an ensemble classifier for cyclone discovery, and (iii) knowledge sharing between the different remote sensor measurements based on a linear Kalman filter for predictive cyclone tracking. Experimental results on historical hurricane datasets demonstrate the superior performance of our automated approach compared to previous work. Results of our cyclone detection and tracking technology using our knowledge sharing approach is discussed and is compared with the list of cyclones reported by the National Hurricane Center for a specific year. The performance quality of our automated cyclone detection solution is found to closely match the manually created database of cyclones from the National Hurricane Center in our initial analysis.

1. Introduction

Tropical and extra-tropical cyclones are important components of the Earth climate system that exhibit variability at different temporal and spatial scales. Cyclone landfall causes great devastation, incurs fatality, and affects people's livelihood. To identify and track tropical weather system, the Tropical Prediction Center/National Hurricane Center (TPC/NHC) uses conventional surface and upper-air observations and reconnaissance aircraft reports [1], and these are concentrated in the North American coasts and in Japan/Europe to some degree. Coverage on a global basis, especially in under-developed and developing nations such as large portions of Asia, Africa is limited or lacking which results in disastrous consequences in many of these regions. In recent years, some studies have used satellite images that are manually retrieved and analyzed to improve the accuracy of cyclone tracking; this procedure is currently slow, tedious, involves coverage of only local regions in North America, and requires close analysis by teams of experts.

In this paper, we describe a novel automated global cyclone discovery and tracking approach on a truly global basis using near real-time (NRT) (and historical) sensor data from multiple satellites. Our current implementation employs two types of satellite sensor measurements, namely: the QuikSCAT wind satellite data, and the merged precipitation data using TRMM and other

satellites. We address the challenges of mining heterogeneous data from multiple orbiting satellites at different spatial and temporal resolutions (see Figure 1). In particular, knowledge sharing between the heterogeneous sensor measurements addresses the problem where some sensor measurements lack definitive indicator for cyclone events, and the spatial and temporal resolutions differ for different sensor. For instance, one cannot confidently identify cyclone based on TRMM precipitation data alone even though it has a finer temporal resolution than QuikSCAT. Through our knowledge sharing methodology, QuikSCAT wind data provides information to the TRMM precipitation data about the likely cyclone location so that the TRMM detector can focus its search on some local region and reduces false alarm for cyclone detection on TRMM data measurements (see Figure 2).
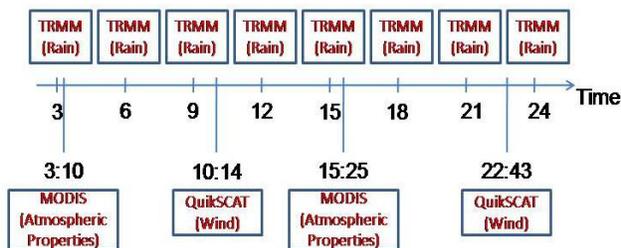
Figure 1. Data availability timeline from TRMM (3B42 data), QuikSCAT (L2B data) and Aqua (MODIS) on 18 Aug 2007 for Hurricane Dean

Peta-bytes of Earth science remote sensor measurements acquired by the NASA satellites are publicly available for analysis and knowledge discovery. These data consist of both archived historical (science products) unlabeled data and near real-time (NRT) data streams, much of which is also not analyzed. There are a number of challenges pertaining to mining data from orbiting satellites. For example, each orbiting satellite (such as QuikSCAT, AVHRR, MODIS) typically cannot monitor a region continuously and the measurements are instantaneous. While these challenges cannot be completely overcome, one can minimize their effects by using data from multiple satellites. However, different satellites provide different measurements. Moreover, different satellite sensors acquire measurements at different spatial and temporal resolutions. These problems make mining heterogeneous data from multiple orbiting satellites extremely challenging, and remains as of now primarily an unsolved problem.
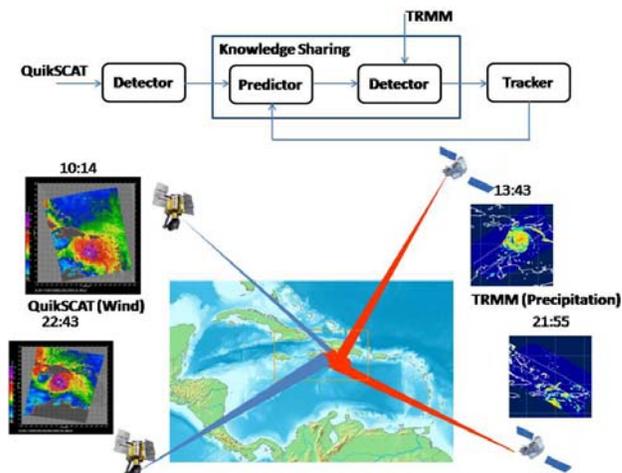
Figure 2. Utilizing the QuikSCAT and the TRMM satellites for cyclone tracking via knowledge sharing

Besides the challenges posed by mining heterogeneous remote satellite data, in general there are challenges related specifically to the problem of detection and tracking of cyclones. First, cyclone events are dynamic in nature i.e., they evolve rapidly in shape and size over time. Second, there is a lack of annotated negative (non-cyclone) examples by experts; this makes training of classifiers for cyclone detection a difficult one. Third, a single satellite sensor may miss a cyclone event due to a pre-defined orbiting trajectory. Our approach addresses and provides an effective solution to each of these challenges and we demonstrate its effectiveness on some recent hurricane events.

The paper is organized as follows. Section 2 provides a brief review on previous work on cyclone detection and tracking. Section 3 describes the data used in our implementation for cyclone discovery and tracking. In Section 4, we described our approach for cyclone discovery using an ensemble classifier and knowledge sharing between QuikSCAT and TRMM data for cyclone tracking. In Section 5, extensive experimental results to validate our proposed approach using historical tropical cyclone occurrences are presented.

2. Previous Work
No solution currently exists that uses heterogeneous sensor measurements to *automatically* detect and track cyclones. In a few partially successful studies, visible and infrared images from geo-stationary satellites are analyzed manually, together with other data sources, using the Dvorak technique [2] to classify the tropical cyclone development stage. Different intensity and track forecast models are computed based on the identified hurricane location and related information. The models are analyzed manually to eliminate the unlikely

predictions. Automated tropical cyclone forecasting system provides an organized framework for forecaster to access information such as cyclone data and numerical weather prediction (NWP) model data has been developed [3, 4]. Forecasters, however, have to make their own conclusions based on the available information. These work focus on detecting and tracking hurricanes that are likely to landfall only in North America, and they involve human interference and decisions.

Prior techniques proposed for automated storm or cyclone identification and tracking use aerial reconnaissance aircraft data and local radar data that have limited coverage and do not measure parameters on a global scale. An improved algorithm for the Weather Surveillance Radar, 1988, Doppler (WSR-88D) has been proposed for storm identification and tracking [5]. Sinclair [6] noted the importance of good features for cyclone identification and proposed a variant of the vorticity feature. Lee and Liu [7] proposed an automated approach for the Dvorak technique using an elastic graph dynamic link model based on elastic contour matching. Lakshmanan et al. [8] proposed a hierarchical K-means clustering method to identify storms and their motions at different scales. These approaches are concerned with storm or cyclone tracking which require manually locating the initial cyclone tracks and tedious data retrieval.

There are existing and developing web-based information systems which systematically archive satellite measurements for hurricanes[1] or (more generally) tropical cyclones[23] for scientific purposes. However, these information systems are based on track information from TPC/NHC. Two products of JAXA (Japan Aerospace Exploration Agency) related to our research are the "AMSR-E Typhoon Real-Time Monitoring" for the Western Pacific region and a global real-time monitor using the TRMM satellite[4]. Again, these products involve human detection and tracking. One interesting development in event monitoring is the Autonomous Science-craft Experiment (ASE) which automatically prioritizes and schedules observations on regions of interest [9]. Currently, this technology is used for NRT

monitoring of events such as volcano activities[5] and floods[6].

3. Data Description

In this section, we describe the two types of remote sensing data used in our cyclone discovery and tracking implementation: QuikSCAT wind data from a polar orbiting satellite (Section 3.1), and the merged high quality/infrared (HQ/IR) precipitation data from the TRMM orbiting satellite and other Geostationary Operational Environmental Satellites (GOES) (Section 3.2).

3.1. QuikSCAT Wind Data

The QuikSCAT (Quick Scatterometer) mission provides important high quality ocean wind data set. QuikSCAT is a polar orbiting satellite with 1800 km wide measurement swath on the Earth surface. Generally, this results in twice per day coverage over a given geographic region. The specialized microwave radar (SeaWinds instruments) on the QuikSCAT satellite measures wind speed and direction under all weather and cloud conditions over Earth oceans. Near real-time wind data is available to weather forecasting agencies from NOAA within three hours of observation. The ocean wind vectors in the measurement swaths have a spatial resolution of 12.5 and 25 km. The ocean wind data is used for global weather forecasting and modeling. It is also used to understand environmental phenomena such as El-Niňo, tropical cyclones, and the effects of winds on ocean biology.

The SeaWinds Processing and Analysis Center (SeaPAC) at JPL is responsible for the reception of the telemetry data from the satellite, raw data processing and analyzing. The processed data is then delivered to the Physical Oceanography Distributed Active Archive Center[7] (PO.DAAC) for public distribution. More information about QuikSCAT science data product is found in [10].

Recent research showed that QuikSCAT data is useful for early identification of tropical depression [11] and early detection of tropical cyclones [1, 12]. Moreover, QuikSCAT data has been used in the three-dimensional variational data assimilation technique for better cyclone track and intensity forecasting [13]. Our recent work [14] showed the feasibility of using QuikSCAT wind measurements for automated cyclone identification.

3.2. Precipitation Data from TRMM satellite

[1] http://disc.sci.gsfc.nasa.gov/hurricane/HurricaneArchiveGallery.html (only North America hurricanes)

[2] http://tropicalcyclone.jpl.nasa.gov/hurricane/main.jsp

[3] http://sharaku.eorc.jaxa.jp/TYP_DB/index_e.shtml

[4] http://www.eorc.jaxa.jp/TRMM/NRTtyphoon/index_j.htm (Japanese Only)

[5] http://modis.higp.hawaii.edu/

[6] http://www.dartmouth.edu/~floods/index.html

[7] http://podaac.jpl.nasa.gov/

The Tropical Rainfall Measurement Mission (TRMM) is a joint mission between NASA and the Japan Aerospace Exploration Agency (JAXA) designed to monitor and study tropical rainfall[8]. The TRMM satellite carries five remote sensing instruments onboard, namely: Precipitation Radar (PR), TRMM Microwave Imager (TMI), Visible Infrared Scanner (VIRS), Clouds and Earth Radiant Energy Sensor (CERES), and Lightning Imaging Sensor (LIS).

TRMM satellite orbits between 35 degrees north and 35 degrees south of the equator. It takes measurements between 50 degrees north and 50 degrees south of the equator. The real-time processing and post-processing of the TRMM science data is performed by the TRMM Science Data and Information System (TSDIS). All TRMM products are archived and distributed to the public by the Goddard Distributed Active Archive Center (GES DISC DAAC)[9].

The (Level) 3B42 TRMM data product used in this paper is produced using the combined instrument rain calibration algorithm using an optimal combination of (Level) 2B-31 data (vertical hydrometeor profiles using PR radar and TMI data), (Level) 2A-12 data (vertical hydrometeor profiles at each pixel from TMI data), SSMI (Special Sensor Microwave/Imager[10]), AMSR (Advanced Microwave Scanning Radiometer on board the Advanced Earth Observing Satellite-II (ADEOS-II) [11]) and AMSU (Advanced Microwave Sounding Unit on NOAA geostationary satellites) precipitation estimates, to adjust IR estimates from geostationary IR observations. Near-global estimates are made by calibrating the IR brightness temperatures to the precipitation estimates. The 3B-42 data quantifies rainfall for 0.25°×0.25° degree grid boxes every 3 hours and the precipitation measurements range from 0.0 to 100mm/hr.

4. Heterogeneous remote satellite-based detection and tracking approach
In Section 4.1, we describe the QuikSCAT features used in our ensemble approach for cyclone detection described in Section 4.2. Our cyclone tracking solution based on knowledge sharing between heterogeneous TRMM and QuikSCAT data is described in Section 4.3.

[8] http://trmm.gsfc.nasa.gov/

[9] http://disc.sci.gsfc.nasa.gov/

[10] http://www.ncdc.noaa.gov/oa/satellite/ssmi/ssmiproducts.html

[11] http://sharaku.eorc.jaxa.jp/AMSR/index_e.htm

4.1. QuikSCAT Feature Selection
In our automated cyclone identification and tracking approach, features which characterize and identify a cyclone are selected and extracted from the QuikSCAT satellite data. We utilize the QuikSCAT Level 2B data which consists of ocean wind vector information organized by full orbital revolution of the satellite as it is very similar to the NRT wind data. One satellite full polar orbiting revolution takes about 101 minutes. The Level 2B data are grouped by rows of wind vector cells (WVC) which are squares of dimension 25 km or 12.5 km. A complete coverage of the earth circumference requires 1624 WVC rows at 25 km spatial resolution, and 3248 rows at 12.5 km spatial resolution. The 1800 km swath width amounts to 72 25 km WVCs or 144 12.5 km WVCs. Occasionally, the measurements lie outside the swath. Hence, the Level 2B data contains 76 WVCs at 25 km spatial resolution and 152 WVCs at 12.5 km spatial resolution to accommodate such instances.

There are 25 fields in the data structure for the Level 2B data. We are, however, only interested in the latitude, longitude, and the most likely wind speed and direction for the WVCs. The fields that are of interest to us are summarized in the Table 1.

After the Level 2B data is received, it needs to be interpolated on a uniformly gridded surface. This is due to the non-uniformity in the measurements taken by the QuikSCAT satellite on a spherical surface. The nearest neighbor rule is used for this pre-processing procedure for both wind speed and direction.

Table 1

Field	Unit	Minimum	Maximum
WVC latitude	Deg	-90.00	90.00
WVC longitude	Deg E	0.00	359.99
Selected speed	m/s	0.00	50.00
Selected direction	Deg from North	0.00	359.99

Histograms are constructed to estimate the underlying probability density of the wind speed (WS) and wind direction (WD) within a predefined bounding box extracted from a QuikSCAT image. Let $WS(i,j)$ and $WD(i,j)$ be the wind speed and wind direction at location (i,j). One defines the direction to speed ratio (DSR) at (i,j) as

$$DSR(i,j) = \frac{WD(i,j)}{WS(i,j)}.$$

When there is a strong wind with wind circulation, the DSR at a wind vector cell (WVC) will be small. In particular, a histogram constructed to estimate the

underlying probability density of DSR in a region will have a skewed distribution towards the smaller value. When there is weak (or no) wind with no circulation, DSR histogram does not have the skewed characteristics. We use a bin size of 4, 30, and 5 for WS, WD, and DSR, respectively according to [14]. One notes that there is a marked difference between a cyclone event and a non-cyclone event in their WS, WD and DSR estimated probability density using histogram. These histogram features are helpful in discriminating between the two events. When a region contains a cyclone, the WS histogram shows a density estimate that skewed towards the larger values. Furthermore, WD histogram shows a "near uniform" distribution.

According to the National Oceanic and Atmospheric Administration (NOAA), a cyclone is defined to be a "warm-core non-frontal synoptic-scale" system, with "organized deep convection and a closed surface wind circulation about a well-defined center". To discriminate between cyclone and non-cyclone events based on this circulation property, we use two additional features: (i) a measure of relative strength of the dominant wind direction (DOWD) [14], and (ii) the relative wind vorticity (RWV).

Let $u(i,j)$ and $v(i,j)$ be the u-v components of the wind direction $WD(i,j)$ at location (i,j) with $1 \leq i \leq$ m and $1 \leq j \leq$ n. One constructs a (m× n)-by-2 matrices M of the form

$$\begin{pmatrix} u(1,1) & v(1,1) \\ \vdots & \vdots \\ u(1,n) & v(1,n) \\ u(2,1) & v(2,1) \\ \vdots & \vdots \\ u(m-1,n) & v(m-1,n) \\ u(m,1) & v(m,1) \\ \vdots & \vdots \\ u(m,n) & v(m,n) \end{pmatrix}$$

Let λ_1 and λ_2 be the eigenvalues of matrix M such that $\lambda_1 < \lambda_2$. The eigenvalue ratio of a bounding box B of dimension m by n is

$$ER_B = \frac{\lambda_2}{\lambda_1}$$

ER_B is used to quantify the relative strength of the dominant wind direction (DOWD) [14] within the region of interest (box) B. If there is circulation (i.e. in regions with a cyclone), ER_B will be near to 1. If the wind is unidirectional (regions that do not have a storm or a cyclone), λ_2 will be much greater than λ_1. As a result, ER_B is much larger.

The relative wind vorticity (RWV) [15] at location (i,j) is

$$RWV_{(i,j)} = \left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \right) = \frac{1}{2d} \left(v_{(i+1,j)} - v_{(i-1,j)} - u_{(i,j+1)} + v_{(i,j-1)} \right)$$

where u and v are the two wind vector components in the west-east and south-north directions, and d is the spatial distance between two adjacent QuikSCAT measurements in a uniformly gridded data. One notes that wind vorticity has been used for cyclones analysis [6, 12].

4.2. Ensemble Classifier for Cyclone Detection
Ensemble methods are learning algorithms that make predictions on new observations based on a majority vote from a set of classifiers or predictors. We build an ensemble classifier to identify cyclones in QuikSCAT images. First, regions likely to contain a cyclone are localized based on wind speed. Then, regions that have an area below a threshold are removed. Five classifiers based on features extracted from the QuikSCAT training data are constructed to identify the cyclones. Two classifiers are simple thresholding classifier based on the *DOWD* and the *RWV* features. The other three classifiers are support vector machine (SVM) [18] using histogram features for *WS* and *WD,* and *DSR* similar to [14]. The classification decision is based on a majority vote among the five classifiers. Figure 3 shows the ensemble classifier design.
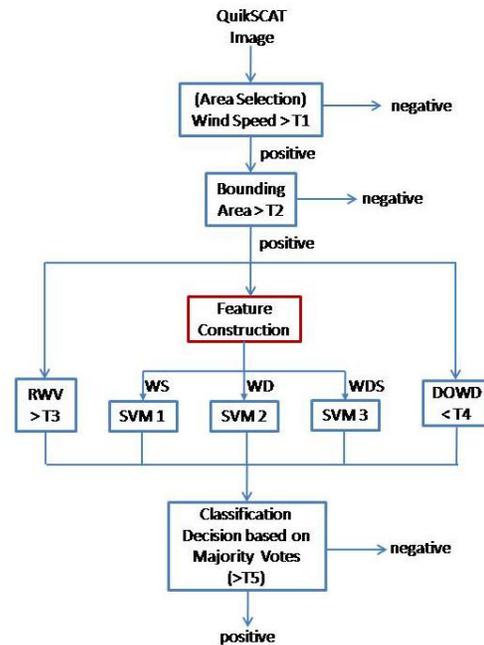

Figure 3. Ensemble Classifier (Cyclone Discovery Module)

4.3. Knowledge Sharing between TRMM and QuikSCAT data for Cyclone Tracking
Our multi-sensor knowledge-sharing solution leverages the strength of each remote sensor type. QuikSCAT has excellent information for accurate cyclone detection but lacks sufficient temporal resolution (each pass-through is repeated every 12 hours). TRMM on the other hand has excellent temporal resolution of 3 hours, but lacks good discriminative ability for accurate cyclone detection.

Therefore, we employ QuikSCAT for cyclone detection (every 12 hours), and TRMM data for tracking (every 3 hours) based on knowledge passed through by the cyclone detector classifier from QuikSCAT. This solution therefore ensures a high detection rate for cyclones while maintaining a fine temporal resolution during cyclone tracking. Our automated cyclone tracking using knowledge sharing is shown in Figure 4. Initially, QuikSCAT data is retrieved from the database or from real-time streaming information, and is input into the cyclone discovery module (Figure 3) to locate/identify possible cyclones. The cyclone location is then used to predict the regions that are likely to contain a cyclone at the next incoming data stream retrieved using a linear Kalman filter predictor. If the next data stream is the 3B42 TRMM data, a constrained search is carried out around the region most likely to contain the cyclone as identified by the Kalman Filter predictor. This constrained tracking via the Kalman Filter predictor is especially important for the 3B42 TRMM precipitation data as it is not a definitive indicator of cyclones and is susceptible to high false alarms. The estimated search region localizes the region that is most likely to contain cyclone based on past cyclone tracks and hence the incidence of false alarms is minimized by a large margin. A cyclone is localized by applying a threshold to the TRMM precipitation rate measurement (T6 = 0). After a cyclone is located in the TRMM data, the Kalman filter measurement update ("correction") is applied to obtain an estimate of the new state vector or the predicted location of the cyclone in the next TRMM (or QuikSCAT) observation cycle after 3 hours.
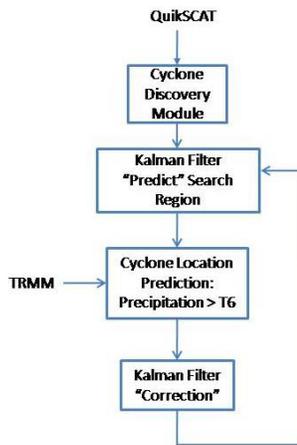
Figure 4. Knowledge sharing between TRMM and QuikSCAT data for Cyclone Tracking
The system equations used in the Kalman filter are

$$x_{k+1} = A_{k+1}x_k + w_k$$
$$z_k = H_k x_k + v_k$$

where x_{k+1} is the state vector at time instance k+1, z_k is the observation vector at time instance k, A_{k+1} is the state transition matrix, H_k is the observation matrix, w_k and v_k

are Gaussian noise at time instance k. The matrix form of the above system equations are as follows.

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \\ \Delta x_{k+1} \\ \Delta y_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \Delta t_{k+1} & 0 \\ 0 & 1 & 0 & \Delta t_{k+1} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_k \\ y_k \\ \Delta x_k \\ \Delta y_k \end{bmatrix} + w_k$$

$$\begin{bmatrix} \hat{x}_k \\ \hat{y}_k \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_k \\ y_k \\ \Delta x_k \\ \Delta y_k \end{bmatrix} + v_k$$

where Δt_{k+1} is the time difference between the next satellite image and the current satellite image. This is a known parameter between two consecutive TRMM satellite images (3 hours), and between a current QuikSCAT image and the next TRMM satellite image. As mentioned earlier, since the spatial resolution varies for different satellite data, we use the longitude and latitude coordinates as the fixed x-y reference frame for the tracking computation.

An important novel contribution of our solution for knowledge sharing via prediction is the modeling of the cyclone predictor and tracker that takes into account the widely varying spatial characteristics of cyclones.

Cyclones are dynamic events and their size evolves rapidly over time. Typical tracking and prediction techniques use the center of an object as the single point to track and predict over time. This model works well for rigid objects that do not change shape with time. However, modeling and predicting the evolution of a cyclone in space over time using only the cyclone center will be grossly inadequate since cyclones often increase in size as they evolve from a depression to a storm to a hurricane, and then decrease rapidly in size after hitting landfall. We therefore model the cyclone as a four-dimensional state vector that is described by the maximum and minimum latitude/longitude of the bounding box spanned by the cyclone.. Our hypothesis is that the bounding box that is described by the (x,y) spatial span of the cyclone evolves linearly in space over time. We expand (or contract) the estimated bounding box based on the estimated Kalman error covariance to define a search region for the cyclone in the TRMM image. This modeling approach significantly improves the quality of knowledge sharing between heterogeneous satellites as compared to using a predictor/tracker using only the center coordinates of the cyclone.

5. Experimental Results
In Section 5.1, experimental results on classification for both preprocessed cyclone/non-cyclone images to Cyclone Discovery Module (CDM). In Section 5.2, the experimental results show that the CDM is robust and works on QuikSCAT swath satellite data. In Section 5.3, the feasibility of the knowledge sharing between two

different satellite data for cyclone tracking is demonstrated.

5.1. Identification of Preprocessed Cyclone Images from North Atlantic and Gulf Region in 2003

Our training data consists of 191 QuikSCAT images of tropical cyclones (i.e. tropical depression, tropical storms, and hurricanes) occurring in the North Atlantic Ocean in 2003. We also randomly collected 1833 unlabeled examples from four days in 2003 when no tropical cyclone occurred. These examples, labeled as negative examples, are included into the training set. Our testing set consists of 84 cyclone events in the North Atlantic Ocean in 2006 and 1822 non-cyclone events, collected from four days in the same year when there was no tropical cyclone.

Table 2 shows the performance for the various classification systems on the testing examples. The DOWD classifier [14], RWV classifier, and the Cyclone Discovery Module (CDM) are compared to the cyclone identification system (CIS) and the SVM ensemble proposed in [14]. The SVM ensemble uses identical parameters to the SVM classifiers in the CDM. The CIS parameters are found in [14]. The DOWD and RWV classifiers use the thresholds we determined earlier. The parameters for CDM are set as follows: T1=12m/s, T2=400 pixels, T3=1.510, T4=1.958, and T5 = 2.

From Table 2, one sees that CDS is a significant improvement from the CIS [14]. Moreover, RWV classifier by itself is also a powerful classifier with both a high true positive rate (TPR) and a extremely favorable true negative rate (TNR) compared to the other classifiers. However, one notes that for RWV classifier to achieve the TPR of CDS (by lowering the threshold value), its TNR becomes less than 0.7.

Table 2. Comparison of various classifier on the testing data (TPR: True Positive Rate; TNR: True Negative Rate; FPR: False Positive Rate; FNR: False Negative Rate)

	Cyclone Discovery Module (CDM)	SVM Ensemble [14]	RWV	DOWD [14]	CIS [14]
TPR	0.9167 (77)	0.8810 (74)	0.8690 (73)	0.8452 (71)	0.7262 (61)
TNR	0.7607 (1386)	0.7261 (1323)	0.8562 (1560)	0.4232 (771)	0.5521 (1006)
FPR	0.2393 (436)	0.2739 (499)	0.1438 (262)	0.5768 (1051)	0.4479 (818)
FNR	0.0833 (7)	0.1190 (10)	0.1310 (11)	0.1548 (13)	0.2738 (23)

5.2. Identification and Tracking of Tropical Cyclones occurring in North Atlantic and Gulf Region in 2005

2141 QuikSCAT L2B swaths between latitude 5N and 60N, and between longitude 0W and 100W (North Atlantic Ocean) in 2005 are collected to test our proposed Cyclone Discovery Module (CDM). We also collected the best-track data from the National Hurricane Center (NHC) to validate our experiment results.

The overall result is as follows.
1. All 26 tropical cyclones reported by NHC are detected.
2. 1 post-season NHC identified subtropical storm is detected.
3. 2 out of 3 tropical depressions (that did not developed further) reported by NHC are detected.

We note that
1. CDM picked up earlier signs of Hurricane Maria 3 days before NHC.
2. CDM picked up earlier signs of Hurricane Vince, when it is an extra-tropical storm, 3 days before NHC [Atlantic Tropical Weather Outlooks (NOAA) did not discuss the non-tropical precursor disturbance to Vince until it had begun to acquire subtropical characteristics] (see Figure 5). Hurricane Vince is the first known tropical cyclone to reach the Iberian Peninsula according to NHC.
3. One earlier weather system that may be related to Tropical Storm Lee which deserves further investigation.
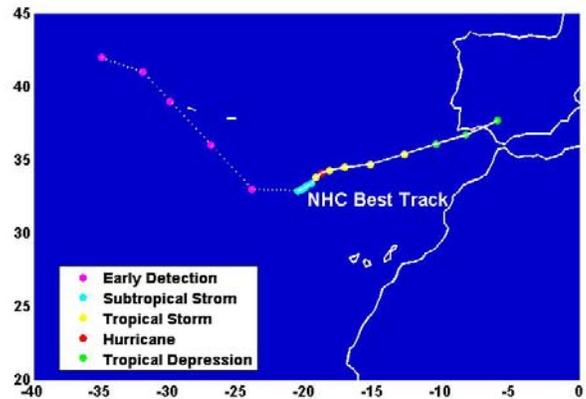
Figure 5. CDM detects Hurricane Vince when it is still an extra-tropical storm.

5.3. Tracking 2007 Hurricane Gonu in the Indian Ocean using QuikSCAT and TRMM data via knowledge sharing

Figure 6 demonstrates the feasibility of tracking methodology using both Level 2B QuikSCAT data and 3B42 TRMM data for Hurricane Gonu (reaching Category 5 wind speed level) in the North Indian Ocean in 2007 for two days. It is the strongest tropical cyclone since record keeping begun in 1945 for the North Indian Ocean and the Arabian Sea. Hurricane Gonu is an interesting weather event as tropical cyclones developed in the Arabian Sea very rarely exceed the tropical storm intensity, i.e. becoming a hurricane.
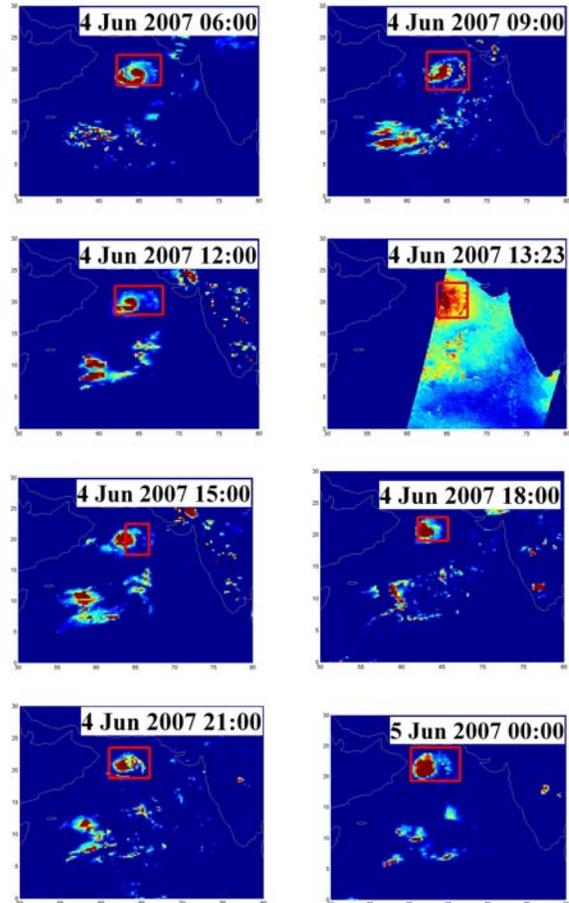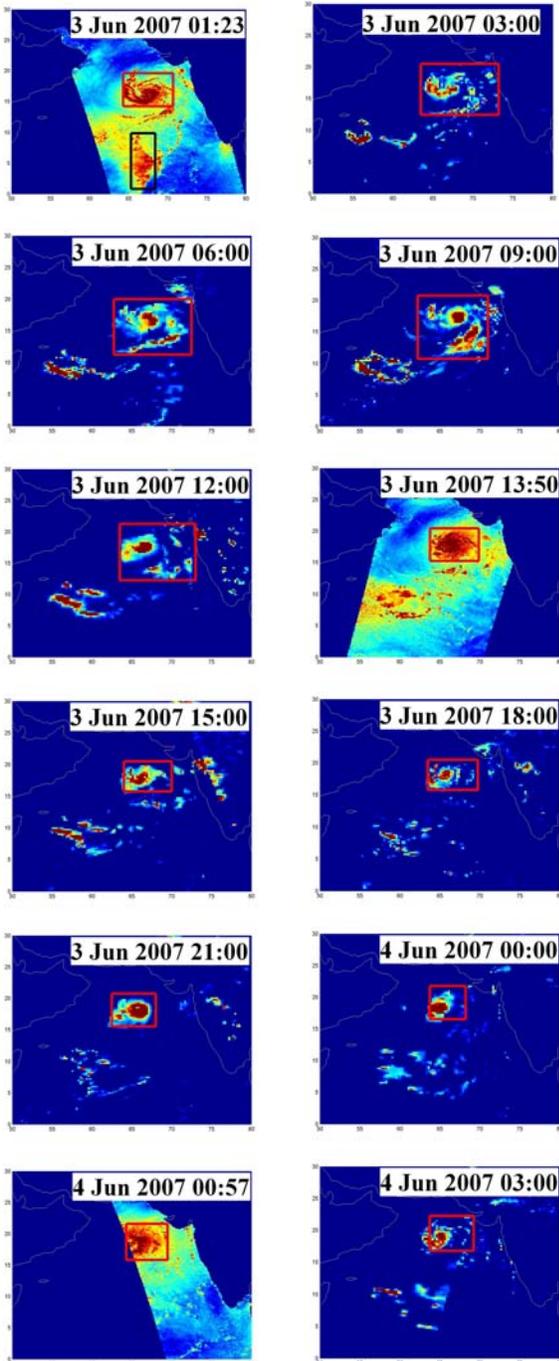

Figure 6. Two days tracking of Hurricane Gonu in 2007 using the QuikSCAT and the TRMM data. A red box bounds the predicted cyclone. The region bounded by a black box is a region correctly identified as a non-cyclone region.

6. Conclusion

Autonomous knowledge discovery from massive heterogeneous satellite data is extremely desirable for advance scientific understanding of the global climate, environmental science, space science, and Earth science. Yet, conventional methods cannot handle such massive unlabeled high-dimensional heterogeneous data. These data remain largely unexplored and under-utilized due to the lack of human resources to manually analyze such data using science experts, inadequate data mining techniques to process these data. Our solution provides a novel, first-of-a-kind solution to heterogeneous satellite data mining and knowledge sharing for event detection and tracking from near-real-time data streams and from massive historical science data sets.

7. Acknowledgements

8. References

[1] R. J. Pasch, S. R. Stewart, and D. P. Brown. Comments on "Early detection of tropical cyclones using seawinds-derived vorticity". *Bulletin of the American Meteorological Society*, 85(10):1415-1416, 2003.

[2] V. F. Dvorak. Tropical cyclone intensity analysis using satellite data. NOAA Tech. Rep. NESDIS 11, 1984.

[3] R. J. Miller, A. J. Schrader, C. R. Sampson, and T. L. Tsui. The Automated Tropical Cyclone Forecasting System (ATCF). *Weather and Forecasting*, (5):653-660, 1990.

[4] C. R. Sampson and A. J. Schrader. The Automated Tropical Cyclone Forecasting System (Version 3.2). *Bulletin of the American Meteorological Society*, 81(6):1231-1240, 2000.

[5] J. T. Johnson and et al. The storm cell identification and tracking algorithm: An enhanced WSR-88D algorithm. *Weather and Forecasting*, 13(2):263-276, 1998.

[6] M. R. Sinclair. Objective identification of cyclones and their circulation intensity, and climatology. *Weather and Forecasting*, 12(3):595-612, 1997.

[7] R. S. T. Lee and J. N. K. Liu. An elastic contour matching model for tropical cyclone pattern recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, Part B, 31(3):413-417, 2001.

[8] V. Lakshmanan, R. Rabin, and V. DeBrunner. Multiscale storm identification and forecast. *Atmospheric Research*, 67:367-380, 2003.

[9] S. Chien and et al. Using Autonomy Flight Software to Improve Science Return on Earth Observing One, *Journal of Aerospace Comp., Inform., and Comm.*, April 2005.

[10] T. Lungu and et al. *QuikSCAT Science Data Product User's Manual*. 2006.

[11] K. B. Katsaros, E. B. Forde, P. Chang, and W. T. Liu. Quikscat's seawinds facilitates early identification of tropical depressions in 1999 hurricane season. *Geophysical Research Letters*, 28(6):1043-1046, 2001.

[12] R. J. Sharp, M. A. Bourassa, and J. J. O'Brien. Early detection of tropical cyclones using seawinds-derived vorticity. *Bulletin of the American Meteorological Society*, 83(6):879-889, 2002.

[13] X. Liang, B. Wang, J. C. Chan, Y. Duan, D. Wang, Z. Zeng, and L. Ma. Tropical cyclone forecasting with model-constrained 3D-Var. ii: Improved cyclone track forecasting using AMSU-A, QuikSCAT and cloud-drift wind data. *Q.J.R. Meterol. Soc.*, 133:155-165, 2007.

[14] S.-S. Ho and A. Talukder, Automated Cyclone Identification from Remote QuikSCAT Satellite Data, *IEEE Aerospace Conference*, 2008.

[15] L. Wang, K.-H. Lau, C.-H. Fung, and J.-P. Gan. The relative vorticity of ocean surface winds from the QuikSCAT satellite and its effects on the geneses of tropical cyclones in the South China Sea. *Tellus*, 59A:562-569, 2007.

[16] Paul Chang and Zorana Jelenak, *NOAA Operational Satellite Ocean Surface Vector Winds Requirements Workshop Report*, 2006.

[17] Chapelle, O., Zien, A., and Schölkopf, B. *Semi-supervised learning*. MIT Press, 2006.

[18] V. Vapnik. *The Nature of Statistical Learning Theory*, Springer-Verlag, 1995.

[19] Rich, Caruana. Multitask Learning. *Machine Learning*, no. 28, 41-75, 1997.