Computers and Geosciences 121 (2018) 53–66

Contents lists available at ScienceDirect

Computers and Geosciences

journal homepage: www.elsevier.com/locate/cageo

A framework for natural phenomena movement tracking – Using 4D dust simulation as an example

Manzhu Yu[a], Chaowei Yang[a,*], Baoxuan Jin[b,1]

[a] NSF Spatiotemporal Innovation Center, Department of Geography and Geoinformation Science, College of Science, George Mason University, Fairfax, VA, 22030, USA
[b] Yunnan Provincial Information Center for Natural Resource, Kunming, Yunnan, China

ARTICLE INFO

Keywords:
Feature tracking
Spatiotemporal feature detection
4D
Dust simulation

ABSTRACT

Natural phenomena evolve in space and time are often highly dynamic. Numerical simulations and earth observations have provided the capability to capture and study the complex evolvement of natural phenomena in a discrete fashion. It is demanding but challenging to extract events from these datasets automatically. Based on the previous research on feature identification, this research presents a movement tracking framework to analyze evolvements and dynamic movements of detected events. The framework consists of three components: feature identification, movement tracking, and track simplification. Based on the proposed framework, dust storm events are systematically detected and analyzed concerning their dynamic movements from a 4D (x, y, z, and t) simulation dataset over North Africa, the Mediterranean, and the Middle East from December 2013 to November 2014. The systematic research includes single event, multi-event, and seasonal analyses. Evaluation of the detected dust events shows that the tracked dust events align well with observations, with ~80% identification accuracy and consistency in the movement pattern. To briefly demonstrate its capability, we adopted the proposed framework to detect precipitation events from 3D (x, y, and t) precipitation observation data.

1. Introduction

Natural phenomena evolve in space and time and can be highly dynamic (Yang et al., 2011). With the improvement of numerical simulations and earth observations in spatiotemporal resolution and coverage, scientists and researchers can capture and study complex physical processes and evolution patterns in a discrete fashion. These simulations and observations can be three-dimensional (3D: x, y, and t) or four-dimensional (4D: x, y, z, and t), allowing the investigation of the movement patterns of natural phenomena in the temporal and vertical dimensions. The obtained knowledge or insights may include "where and when natural phenomena happen," "how long a natural phenomenon lasts," or "what the common transport pathway is for a natural phenomenon."

GIScience methodologies and techniques assist the understanding of dynamic geographic changes over space and time, but challenges remain in handling complex natural phenomena, especially for data with higher dimensions (Yuan, 2001; Worboys, 2005; Pultar et al., 2010). The increasing spatiotemporal resolution of simulations and earth observations has become more complicated for scientists to examine manually. Although numerical simulations and earth observations provide the spatiotemporal data source, researchers and scientists still need to develop algorithms to identify and track the movement of features (e.g., thunderstorm, hurricane, ocean eddy). Automatically identifying and tracking features are challenging; because features are moving with changing boundaries and capable of splitting and merging, and these movement patterns distribute over space and time. Therefore, providing an efficient way to detect these movement patterns is essential to the natural phenomena analysis. Besides, tracking features at different thresholds convey different information about the phenomena. It is essential to be able to track the movement of events at various thresholds efficiently.

The objectives of our research are threefold: 1) identify features based on out previous work (Yu and Yang, 2017) and introduce the tracking framework to connect the identified features in consecutive time steps; 2) apply the framework to a 4D simulation dataset; and 3) analyze the evolvements and dynamic movements of the events. Dust events are chosen as case studies to illustrate how this tracking approach can be used to represent and analyze the dynamic movements of natural phenomena. For an individual dust event, it is essential to

* Corresponding author.
E-mail address: cyang3@gmu.edu (C. Yang).

[1] Manzhu Yu co-designed the research, conducted experiments, and wrote the manuscript. Chaowei Yang formed the research idea, co-designed the research, analyzed the results and revised the manuscript. Baoxuan Jin contributed to methodology design.

https://doi.org/10.1016/j.cageo.2018.10.003
Received 7 April 2018; Received in revised form 24 August 2018; Accepted 1 October 2018
Available online 04 October 2018
0098-3004/

understand the physical process of dust up-lift from arid and semi-arid regions, transport in the air, and deposition back to the ground. Besides, one of the major topics towards mineral dust is the spatiotemporal patterns of dust transport from desert source regions (Israelevich et al., 2003; Prospero and Lamb, 2003; Borbély-Kiss et al., 2004). Dust events that originate from a specific source show region-specific patterns of transport pathways (Israelevich et al., 2003; Moulin et al., 1998). Natural phenomena tend to interact with each other during transport in the atmosphere, such as split and merge (El-Askary et al., 2002; Ammar et al., 2014; Stein et al., 2015). These complex physical processes can only be adequately addressed in a 3D or 4D environment.

Related research of 3D feature tracking is reviewed in Section 2. Section 3 introduces the movement tracking approach. Section 4 analyzes dust storm events, their dynamic changes, and transport pathways, and Section 5 evaluates the resulting dust events with visibility observations, and archived dust events in NASA Earth Observatory. Finally, Section 6 offers conclusion from this research followed by potential future work.

2. Related works

Tracking natural phenomena includes two categories of methods: centroid- and overlapping-based. Centroid-based tracking methods normally treat the features in consecutive time steps whose centroids are within a certain radius in the same track. Johnson et al. (1998) made a guess on the cell centroid locations at t_{n-1} to where they would be at t_n based on its positions at several previous times, and assigned each cell at t_n to the closest unassigned centroid within a certain search radius. Lakshmanan et al. (2009) tied projected cells within a size-based radius (given by $\sqrt{(A/\pi)}$, where A is the area of the storm, when dealing with 2D) across different time steps.

Tracking methods using overlapping mechanism require that spatial and temporal frequencies be high enough regarding the expected size and speed of the features to track (Samtaney et al., 1994). Otherwise, unlinked features need to be associated using additional information in a second iteration of tracking. Choi et al. (2009) calculated the degree of association between overlapping storm features using an inverse cost function. The degree of association reflects the size similarity and moving speed of two associated features. Han et al. (2009) treated cells at t_n that have 50% or more significant overlap with cells from t_{n-1} as first matched, while unmatched cells are associated using a global cost function or assigned a new ID. A better approach is that of Dixon and Wiener (1993), where it utilized a combined approach of areal overlapping and centroid matching. First, storms that overlap significantly at two successive times are likely to be from the same storm. Then, an optimization scheme determines the most likely match between storms identified at successive scans. The optimization selects the track paths with shorter lengths, connects storms with similar characteristics (such as size and shape), and eliminates those tracks that exceed the maximum expected speed of storm movement.

Choosing the right approach is closely related to the spatiotemporal scales of datasets. Overlap-based methods are more suitable for tracking more substantial features with higher temporal resolution, and centroid-based methods may not be suitable for features with variable sizes (Lakshmanan et al., 2009). Therefore, a tracking approach needs to be designed and developed specifically for the natural phenomenon.

3. Feature identification and tracking methodology

The procedure of constructing the four entities of the framework includes the following: 1) identification of static dust storm features; 2) track features over consecutive time steps; and 3) composition of the event through tracking results (Fig. 1a).

3.1. Data and implementation

Dust simulation outputs were obtained from BSC-DREAM8bv2.0 (Pérez et al. 2006a, 2006b; Basart et al., 2012), a dust forecast operational system with the updated version of the former Dust Regional Atmospheric Model (DREAM; Nickovic et al., 2001) maintained by the Barcelona Supercomputing Center. The simulated dust concentration data include 12 months from December 2013 to November 2014 and cover a standard latitude/longitude grid of approximately 0.3° × 0.3° resolution for the broad north African and European domain (25.7W–59.3E, 0.76S–64.3N). The temporal resolution is hourly, and each time step contains voxel number of 256 × 196 × 24 (latitude, longitude, pressure level) (Fig. 1b).

The implementation was conducted in Python (Van Rossum and Drake, 1995) as a prototype, including the feature identification (Section 3.2) and tracking algorithms (Section 3.3). Visualization was implemented with the assistance of third-party libraries, including Numpy (Van Der Walt et al., 2011) and Scikit-learn (Pedregosa et al., 2011).

3.2. Identifying static meteorological features

The identification of meteorological features at each time step is conducted using a region-grow based algorithm that integrates the idea of the region-grow algorithm (Zucker, 1976) into 3D context; the simplified pseudo-code version of the algorithm is illustrated in Fig. 2. The computational complexity of this algorithm is quadratic concerning the number of voxels at two consecutive time steps. This algorithm is based on Yu and Yang (2017). The original algorithm has a multi-thresholding approach, which facilitates the identification of multiple high-concentration substorms within a larger low-concentration system. In this research, it is simplified to a single-thresholding approach, so that the tracked dust event is of the same level of concentration. A meteorological feature is specified as a contiguous volume with a concentration/intensity value greater than a threshold (*Dth*), while its volume is greater than a threshold (*Vth*). For each identified meteorological feature, the geometry is calculated using the boundary extraction method - Marching Cubes (Lorensen and Cline, 1987). Besides, associated attributes are calculated (e.g., concentration-weighted centroid in degree, speed in degree/hour, number and position of pixels in the dust storm object, area in degree*degree, maximum and average concentration or intensity). For the 4D dust simulation data, a dust concentration threshold of 360 μg/m3 and a volume threshold of 10 voxels were used.

3.3. Tracking the linkages of features over consecutive time steps

After the feature identification strategy (Section 3.2) is applied to the meteorological data, a list of feature objects exists for each time step. The tracking algorithm associates these objects across time to track the progress of the features as they form, move, and dissipate.

3.3.1. Overlapping strategy

Since the volumetric size of dust feature varies from 10 to 500 voxels, an overlap-based method was developed with an additional check, detailed in the pseudo-code (Fig. 3). In the first overlap check, the dust features over consecutive time steps are checked to track the potential linkages, based on the assumption that meteorological features from a later time step have partial overlap with those from an earlier time step. This overlap approach performs a matching test on features extracted from one timestep with all of the features extracted from the subsequent time step, and all combinations of features from dataset t_{i+1} (for amalgamation/bifurcation). The best match is selected by minimizing the cost function defined as follows: $\beta = 100 - O(C^t, O^{t-1})$, where O is a function measuring the percentage of overlap between the candidate object C^t and the object O^{t-1}.

The second check considers the rare cases when a feature is small in size and moving fast compared to the spatial and temporal resolution of

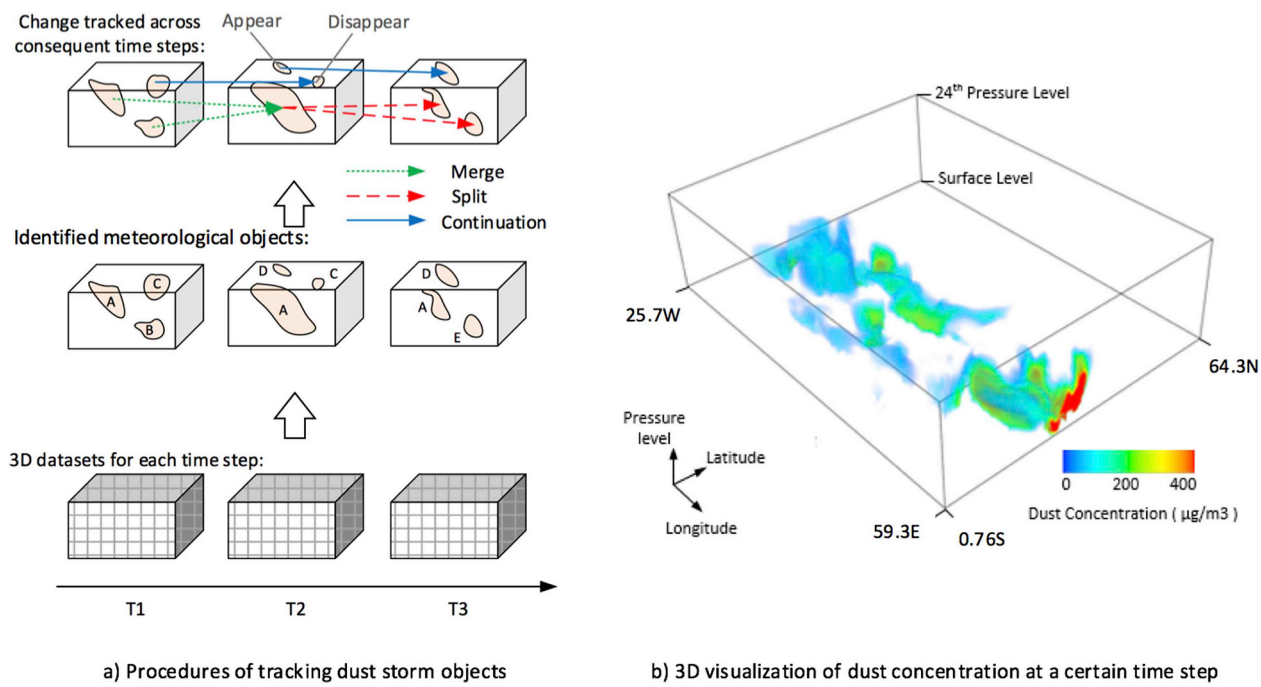

Fig. 1. a) General workflow of tracking dust storm objects based on dust simulation data; b) 3D visualization of simulated dust concentration, which covers the area of broad north African and European domain (25.7W–59.3E, 0.76S–64.3N), with 24 vertical pressure levels.

the dataset, in which case the first spatial overlap might not track the feature. Attributes of the current pair of identified features, including centers of mass and volumes, are compared; if the difference in attributes is below the threshold, the two features are considered as an additional continuation. The computational complexity of this algorithm depends on the number of features at two consecutive time steps, and the number of time step pairs.

3.3.2. Assigning objects to tracks

After checking the overlap, there are several possible cases for the linkage between t and $t+1$ as outlined below:

1) **Continuation**: A meteorological feature at t is linked to a feature at $t+1$. The feature may change location or shape across time but remains as one *object*. The two possibilities for continuation are: growth (with a growing volume or concentration) and decay (with a decreasing volume or concentration).
2) **Merge**: Multiple meteorological features at t are linked to a single feature at $t+1$. The choice is to propagate the track history of any feature at t to that at $t+1$. If not, the merged feature serves as the starting feature of a new track. Otherwise, the option becomes choosing an appropriate object at t to be the one propagating its track history to the one at $t+1$. In this study, the track history of the meteorological feature at t with the maximized curvature smoothness was propagated to the one at $t+1$, and the remaining features at t are considered at the end of their corresponding *objects*. The curvature between a candidate feature at $t+1$ and the track is defined as follows: C= $\angle(T^{t}_{\lambda,\phi}, F^{t+1}_{\lambda,\phi}) - \angle(T^{t-1}_{\lambda,\phi}, T^{t}_{\lambda,\phi})$, where $T^{t}_{\lambda,\phi}$ is the centroid point of the feature object of the track at t, and $F^{t+1}_{\lambda,\phi}$ is the centroid point of the candidate feature at $t+1$. An angle allowance from -90° to $+90^\circ$ of curvature is accepted. The smallest curvature represents the maximized curvature smoothness.
3) **Split**: A single meteorological feature at t is linked to multiple features at $t+1$. As with merging, this results in the option to assign the track history to any feature at $t+1$. Addressing this problem, the track history of the meteorological feature at t is assigned to the one at $t+1$ with the maximized curvature smoothness, and the remaining features at $t+1$ are considered as the first one of new *Objects*.
4) **Appear**: There is no meteorological feature at t that can be linked to the one at $t+1$. In this case, the object at t+1 refers to the first feature in a new *object* (i.e. birth of a new *object*).
5) **Disappear**: The meteorological feature at t has no linkage to any feature at $t+1$, in which case, the feature at t refers to the last feature in its *object* (i.e. death of this *object*).

With the above five types of linkage between identified features in consecutive time steps, individual *objects* are constructed. Each *object* starts from its appearance, ends with its disappearance, and is linked by continuation. Each *object* is associated with a unique ID, and the time step of the meteorological feature is identified and linked to its associated geometry. In addition, splitting and merging *linkages* among specific meteorological features are recorded at specific time steps. Based on the established *objects* and *linkages*, meteorological events are reconstructed by linking *objects* that have *linkages* among each other (Fig. 4). The process of how a meteorological event evolves and is transported is efficiently retrieved.

3.3.3. Track merging

The procedure described in Sections 3.3.1 and 3.3.2 produces a set of tracks that are locally optimal at each timestep. However, this approach might lead to the false assignment of an object to a track, or the premature ending of a track where a feature object has not been identified in the next timestep. To avoid these problems, we adopt the track merging strategy to find an optimal set of tracks, which extends a track which ends at timestep *t-1*, by another track, which begins at timestep t or timestep $t+1$. The connection between the first and the second track is built on the fact that the first feature object of the second track is within an adaptive search area of the last feature object of the first track (Fig. 5). The adaptive search area is defined by the adaptive search radius and an angle allowance of 90° (from -90° to $+90^\circ$). The adaptive search radius is defined as follows: $\gamma = h \times \Delta$, where h is

Algorithm 1: Feature Identification

Input: 3D meteorological dataset
Output: Labeled features, labels: 1 – N (N: total number of features)

```
function featureIdentification(Input)
features ← regionGrow(Input)
for each feature in features
    if feature.getVolume() < Vth then delete feature end if
end for
Output ← features

function regionGrow(Input)
for each voxel at location (x,y,z)
    if voxel[x][y][z].getLabel() = unlabeled and voxel[x][y][z].getIntensity() > 0 then
        growRegion(x,y,z)
        currentLabel += 1
    end if
end for

function growRegion(x,y,z)
voxel[x][y][z].setLabel(currentLabel)
q ← Queue()
q.put([x,y,z])
while q.empty() == False then
    voxelIndex ← q.get()
    for x,y,z in each 6-neighbor of voxelIndex
        checkNeighbor(x,y,z,q)
    end for
end while

function checkNeighbor(x,y,z,q)
voxelIntensity ← voxel[x][y][z].getIntensity()
voxelLabel ← voxel[x][y][z].getLabel()
if (voxelIntensity > Dth) and (voxelLabel!=currentLabel) then
    voxel[x][y][z].setLabel() ← currentLabel
    q.put([x,y,z])
end if
```

Fig. 2. Pseudo code of feature identification.

represented as the time interval and Δ refers to the velocity of the last feature object in the first track. When multiple candidates exist as the second track to be merged, only one is selected based on the maximized curvature smoothness. As illustrated (Fig. 5), when two candidates exist to merge with the first track with candidate track #1 having a smaller curvature than candidate track #1, thus candidate track #1 is selected as the second track to merge with the first one.

3.4. Track simplification

To improve visualization effects, tracks are processed to generate simplified geometry and reduce visualization clutter. Instead of working with all centroid points on the path of a meteorological event, the path is divided into a series of line segments and the line segment data are used to quantify the path and infer patterns in dust event behavior. The simplification strategy is inspired by the approaches of Adrienko and Adrienko (2011) for extracting characteristic points of trajectories, and Moy et al. (2015) for calculating step length of *C. elegans* Locomotory.

While Adrienko and Adrienko (2011) applied the algorithm to GPS datasets to reduce data volume to further aggregate multiple similar trajectories into a generalized one, the same approach is applied to describe the transport path of dust storm events quantitatively. In addition, Adrienko and Adrienko (2011) offered another parameter, *MinStopDuration*, to constrain the minimum time spent in approximately the same position to be treated as a significant stop. Conversely, in the current research it is expected that meteorological events do not have a significant stop, and since meteorological events continuously move and evolve, therefore this parameter is removed. Moy et al. (2015) used the simplification algorithm to determine the location of a turning event where "the angle between two movement segments joining three successive positional fixes is less than a critical angle."

Therefore, we extend and finalize the track simplification constraints as *TurningAngle*, *MinDistance*, and *MaxDistance*. *TurningAngle* refers to the minimum angle between the directions of consecutive trajectory segments to be considered as a significant turn. *MinDistance* is when the distance between two consecutive points is below this value, the points are treated as approximately the same position. *MaxDistance* refers to the maximum allowed distance between consecutive characteristic points extracted from the trajectory (i.e., if the trajectory has a straight segment with the length more than this value, representative points must be taken such that the distance between them do not

Algorithm 2: Feature Tracking

Input: *F* as the set of identified features at all time steps.
Output: *c* as the final connection set.

```
c ← ø
for each t = 1, …, tmax-1
  for each feature ϵ F[t]
    for each feature' ϵ F[t+1]
      geom, geom' ← feature.getGeom(), feature'.getGeom()
      if 3D_overlap(geom, geom') == True then
        c.setConnection(feature, feature')          # first round check
      end if
    end for
  end for
  F' ← F.getRemain()
  for each feature ϵ F'[t]
    for each feature' ϵ F'[t+1]
      centerOfMass, centerOfMass' ← feature.getCenter(), feature'.getCenter()
      volume, volume' ← feature.getVolume(), feature'.getVolume()
      if 3D_distance(centerOfMass, centerOfMass') < Td or |volume – volume'| < Tv then
        c.addConnection(feature, feature')          # second round check
      end if
    end for
  end for
end for
```

Fig. 3. Pseudo code for feature tracking.

exceed this value). To avoid deleting useful points, a set of tolerant parameters are used so that more points will be reserved and only the most redundant points are disgarded. A *MinDistance* of 3, a *MaxDistance* of 5, and a minimum *TurningAngle* of 150° are used to detect the representative points. The strategy of track simplification is illustrated (Fig. 6a) by marking the characteristic points in red based on the abovementioned three constraints, resulting in the simplified track (red in Fig. 6b). An example track simplification is demonstrated in Fig. 6c and d.

4. Event analysis: case studies

Dust simulation data are used as the case study to demonstrate the capability of the feature identification and tracking prototype in analyzing the spatiotemporal behavior of a meteorological event. The event analysis facilitates the study of mesoscale natural phenomena by investigating the frequent interactions of natural phenomena in both horizontal and vertical aspects, integrating volumetric components.

The origins of major dust events are the *preferential dust sources* in North Africa used in emission scheme of BSC-DREAM model (Basart et al., 2012) and including 1) Bodélé; 2) Mali; 3) Mauritania; 4) Western Sahara-Morocco; 5) Algeria-Adrar; 6) North of Algeria-Tunisia; 7) Libya desert; 8) An-Nafud desert; and 9) Rub' Al Khali desert (Fig. 7).

4.1. Single event analysis

The framework supports the analysis of a single meteorological event, which seeks information about how a specific event behaves in space and time, including the evolution of object attributes (e.g., concentration weighted centroid, volume, average concentration/intensity), and movement patterns (speed and trajectory). The obtained information helps to examine the attributes of spatiotemporal objects and how these properties change in an event lifecycle, and how the examination offers insights on mechanisms responsible for the changes.

An example of representing a specific dust event and its composition of spatiotemporal objects in 4D is illustrated (Fig. 8). The trajectory of a particular dust event (Fig. 8A) is generated based on the centroids of each static objects that belong to this event, and the dust volume

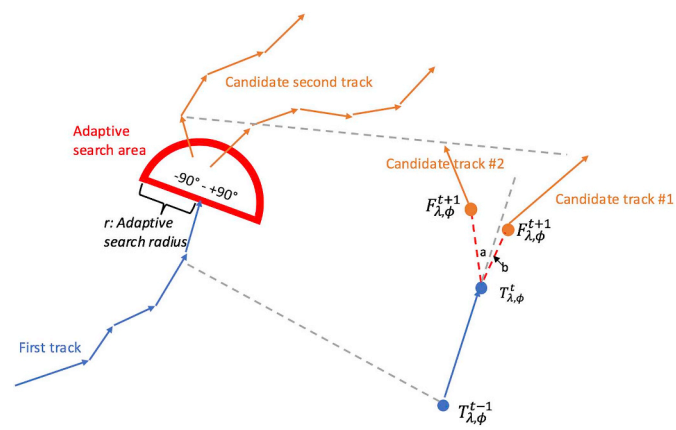

Fig. 5. Tracking merging strategy (simplified representation in 2D).

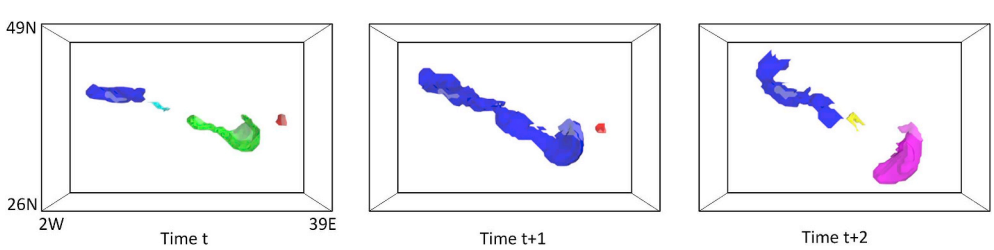

Fig. 4. Example of meteorological object merging and splitting at consecutive time steps within a subdomain. The same colors link the same event through the time steps. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

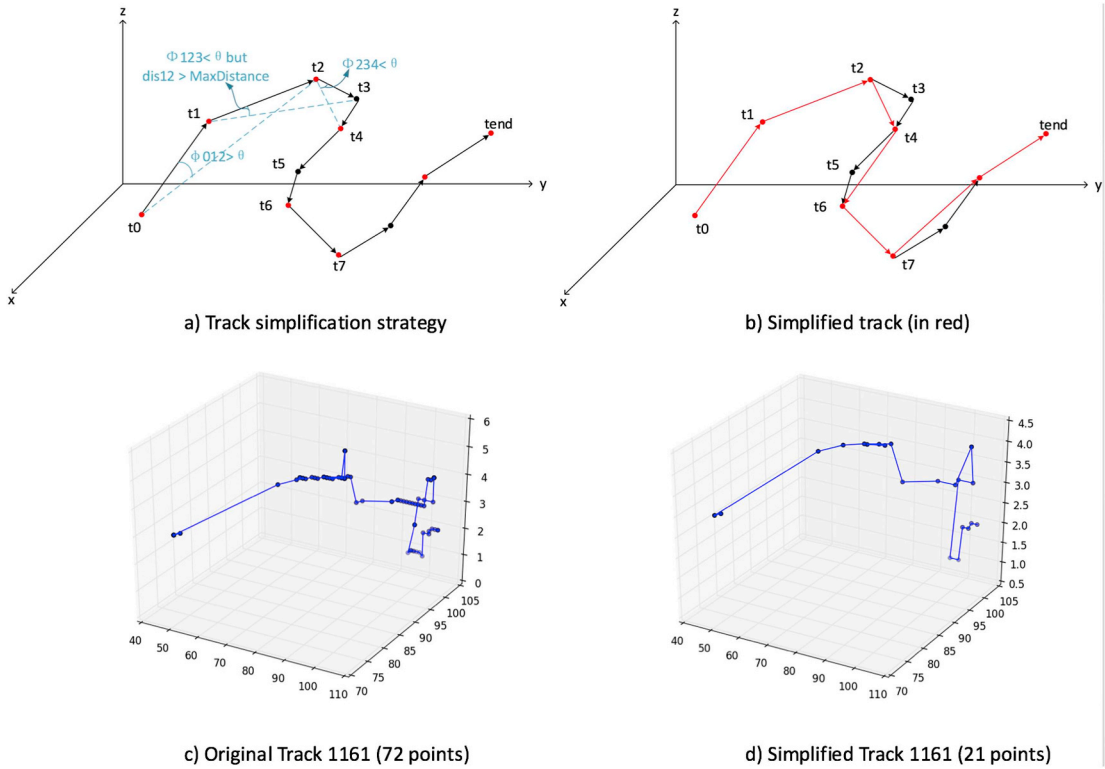

Fig. 6. Strategy of simplifying track. Fig. 6a and b represent the track simplification strategy using a synthetic example, whereas Fig. 6c and d use a detected track from the experiment data to represent the track before and after simplification.

(Fig. 8b) indicates the object shape at a particular time step. Object Ob107 is blown in a low speed from its initiation to Time Step 5 and then moves with a higher speed until it splits and generates Object Ob111 (Fig. 8e). During the lifecycle of Event E60, the volume and the average concentration of dust plumes generally degrade.

4.2. Multi-event analysis

Based on single event analysis, the framework obtains the transport pathways of selected dust events to analyze their spatiotemporal patterns. The pathways of dust events occurring during June 2014 (Fig. 9a) that originate from Libya desert with each dust object displayed in blue,

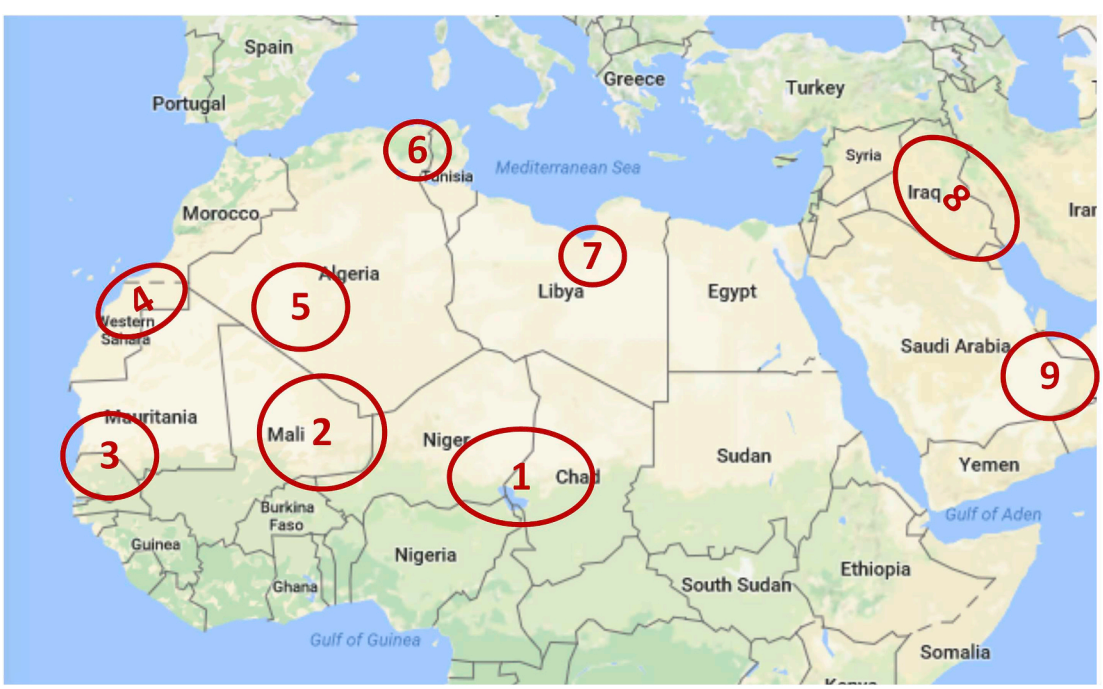

Fig. 7. Preferential dust sources in North Africa used in emission scheme of BSC-DREAM model.

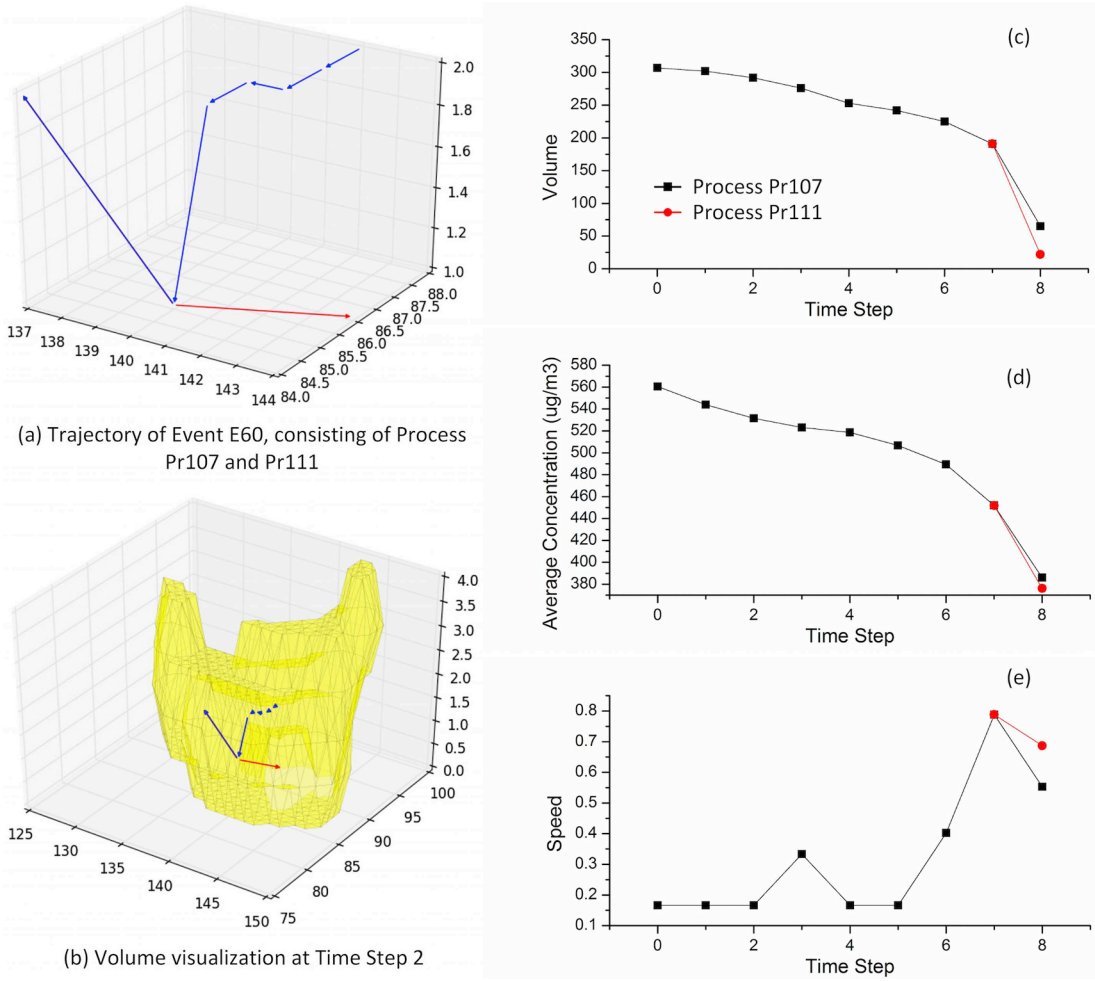

Fig. 8. Trajectory and attributes of an example event, Event E60, which consists of Objects Ob107 and Ob111. The Object Ob107 at Time Step 7 splits and generates Object Ob111.

and split and merge interactions displayed in red and green, respectively. There are 15 events originating from Libya desert compared to 65 events in the entire study area (Fig. 9b). In June the dust originating from the Libya desert was transported from the Libyan coast toward eastern Mediterranean or northwest to the western Mediterranean basin. Along the transport corridor of North of Algeria-Tunisia, evidence of merging dust plumes is illustrated in the green pathways (Fig. 9b), consistent with dust entrainment from the underlying dust source areas.

4.3. Seasonal analysis

The proposed framework further offers insight into the seasonal analysis of meteorological events.

4.3.1. Seasonal transport patterns

An example of querying dust events originating from Libya desert in the four seasons of 2014 illustrates the seasonality of the events (Fig. 10). Across the four seasons dust transport pathways follow the E-W winds but show distinct differences among seasons (Moulin et al., 1998; Rezazadeh et al., 2013). During the winter season (DJF) for dust originating from the Libya desert, transports is westward along the Libyan coast, entraining dust from North of Algeria-Tunisia and Algeria toward Mali. During the spring season (MAM) there is more complexity of dust transport. In addition to the main transport pathways from the Libyan desert west along the coast, there is evidence of additional pathways and splitting and merging interactions. It is hypothesized that this increased complexity in the dust transport pathways is a consequence of Sharav cyclones and cold fronts which occur typically in spring (Varga et al., 2014). The dynamic lifting during the cool season from cold fronts results in higher dust transport entrainment than other seasons (Fig. 11b). During the summer season (JJA) dust transport pathways are similar to that in June, and dust transport falls to its lowest levels during fall season (SON) (Fig. 11d).

4.3.2. Seasonal occurrence, dust load, duration of dust events from preferential dust sources

The specific contribution of the different source areas for the occurrence of dust events is investigated (Fig. 12). For each dust event, the source contribution is calculated from the total number of trajectories from the source regions. In the fall season, Bodélé (S1) contributes more than other source regions, whereas in summer seasons An-Nafud desert (S8) and Rub' Al Khali desert (9) contribute more. Moreover, Western Sahara-Morocco (S4) is a consistent dust contributor throughout the year except in the autumn. The monthly occurrence of dust events shows a seasonal cycle characterized by the fall minimum (September to November; 10.8%) and abroad spring-summer maximum (June to August, 62.0%). On a monthly basis the highest dust events are in February (13.1%) and June (12.9%), followed by secondary maxima in March (12.3%) and a July (11.6%).

The specific contribution of the different source areas to the total dust load of dust events is also investigated (Fig. 13). The dust load of events originating from An-Nafud and Rub' Al Khali deserts (S8 and S9) contribute to 73% during the summer season, indicating these two dust

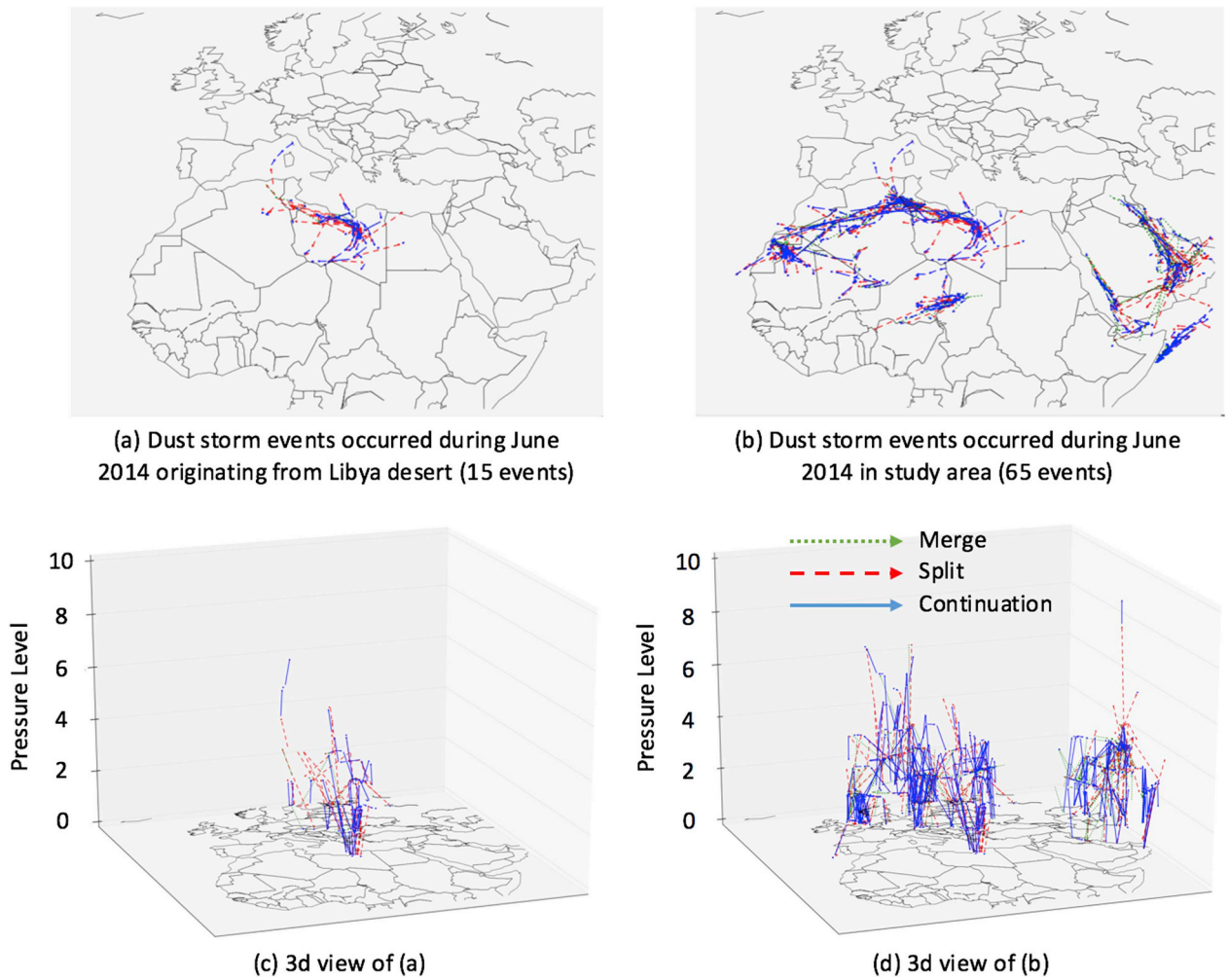

Fig. 9. Example of querying the dust events occurring during June 2014 that originate from Libya desert (a & c), and without constraining dust source (b & d). In the 3d views, vertical levels are the pressure levels.

sources are having severe dust storms during Summer 2014. Dust from Bodélé depression (S1) is the third major source of dust load and remains consistent during the four seasons. The dust load from Western Sahara-Morocco (S4) decreases from winter to fall annually, while that from Algeria-Adrar and North of Algeria-Tunisia (S5 and S6) is higher in the spring and winter and lower in the summer and fall (see Fig. 14).

During December 2013 to November 2014 the duration and frequency of the dust events were less than one day (56%), one day (24%), two days (13%), and greater than five days (1%). The frequency of dust event duration was seasonally dependent, with the mean duration decreasing from summer (1.8 days) to winter (0.9 days). This pattern is in agreement with Duchi et al. (2016).

4.3.3. Seasonal mean event locality

The mean event locality for all dust source regions was investigated by introducing a track locality to describe whether a dust storm event is moving locally or is being distributed across to a larger area. Here we are not using the simplified tracks but the original tracks resulted after Section 3.3. A long distance track or event does not necessarily transport in a large area, but can also be moving within a small system (Fig. 15). To quantify the locality L, the ratio of the total centroid moving distance to the length of the transporting bounding box over the full lifecycle of a track is used as expressed in Eq (1). A higher locality indicates that the track is transporting dust within a relatively small area (red track in Fig. 15), while a lower locality indicates that the track is transporting dust over a relatively large area (blue track in Fig. 15).

$$L_{track} = \frac{Distance\ \ (Centroid_{t0},\quad Centroid_{t1},\quad ...,\quad Centroid_{tn})}{Max\ \ (xmax - xmin,\quad ymax - ymin,\quad zmax - zmin)} \tag{1}$$

An event may consist of multiple tracks, and the locality of a dust event is expressed in Eq (2), where i represents the track number, and n is the number of tracks within this event.

$$L_{event} = \frac{\sum_{i=1}^{n} Distance_i}{Max\ \ (xmax - xmin,\quad ymax - ymin,\quad zmax - zmin)} \tag{2}$$

Locality is calculated as the proportion of distance to the area of the region that the event has traveled. A larger the locality value is associated with the event being more localized, while a smaller value indicates that the event transports dust across a larger region. Fig. 16 shows the seasonality the locality value of the events that originated from these sources. In the winter season the locality of events originating from Bodélé (S1) is larger than the that of any other source region, indicating the dust events from Bodélé (S1) are constrained within a smaller region. In addition, the mean locality values of spring and fall seasons are lower than other seasons, indicating that dust events from these are transport further. Summer season shows a larger variability of locality among the nine different dust source regions. Dust events from Western Sahara-Morocco, An-Nafud desert, and Rub' Al Khali deserts (S4, S8, and S9) tend to travel within a more constrained region.

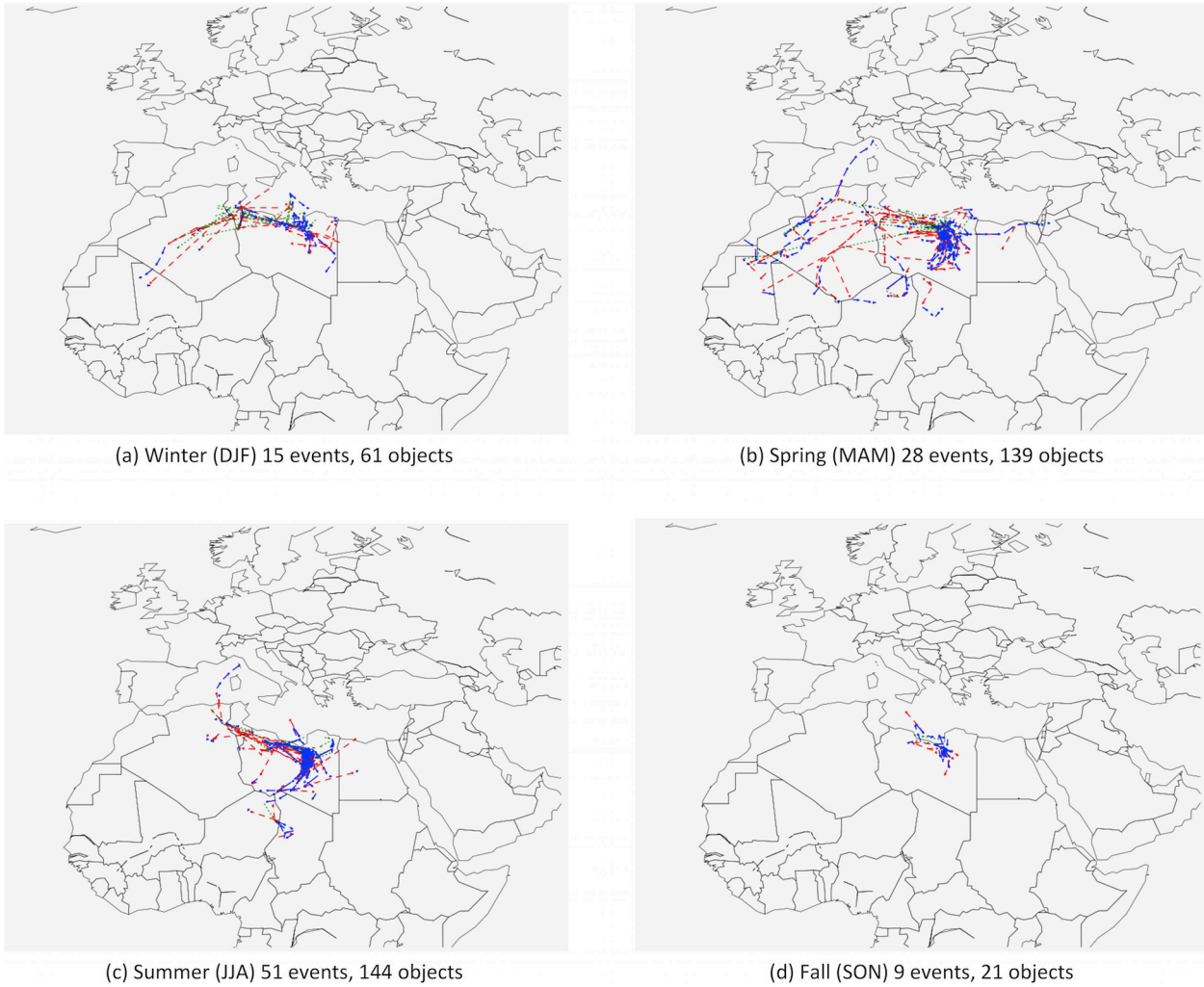

Fig. 10. Seasonal transport patterns of dust events originating from Libya desert.

5. Evaluation of detected dust events

The detected dust events are evaluated by two kinds of observation sources: station-based visibility observation; and archived dust events from NASA Earth Observatory. Visibility observation data are used to verify that the identified dust storm objects cover the suspected area of dust occurrence. The archived dust events from NASA Earth Observatory are used to verify that the transport pathways are consistent with the officially recorded situation. Herein the objective is not to evaluate the dust model but the utility of the research procedure to identify and track dust events from the model's simulation output. The validation of BSC-DREAM8bv2.0 is documented in Basart et al. (2012).

5.1. Evaluation of identified dust storm based on visibility observation

Dust feature identification is verified using the visibility observation obtained from the Integrated Surface Global Hourly Data archived at the National Climatic Data Center (NCDC) (https://data.noaa.gov/dataset/integrated-surface-global-hourly-data). This dataset is in text format, indicating the visibility value at each specific station with the station's latitude, longitude, and altitude. These weather observation stations are not regularly distributed; stations are densely distributed where it has a larger population or where the place needs to be monitored. The temporal resolution of visibility data is hourly. The visibility dataset is station-based weather observations (unit of meters) and dust conditions (visibility observation < less than 10 km). Based on World Meteorological Organization protocol (WMO, 2013), dust events are classified according to visibility into four categories: 1) dust haze, with no visibility constraints; 2) blowing dust, 1–10 km; 3) dust storm, 200 m-1km; and 4) severe dust storm, < 200 m. Since dust haze is not corresponding to a particular visibility level, we do not consider it in our following evaluation.

The procedure for investigating the utility of the procedure selects the available visibility data within the covered area of the identified dust storm object at each time step. There are multiple possibilities for verifying the results of detected dust events (Fig. 17): 1) if a station observation with low visibility (< *Threshold*) is covered by a detected dust storm object, it is considered as a 'hit'; 2) if not, it is considered as a 'miss'; and 3) if a station observation with high visibility (> *Threshold*) is covered by a detected dust storm object, it is considered as a 'false alarm'. Here the visibility *Threshold* is varied by using 200 m (covering severe dust storm only), 1 km (covering dust storm and severe dust storm), and 10 km (covering blowing dust, dust storm and severe dust storm). Based on the above three options, evaluation scores are calculated using probability of detection (POD, defined as *hits*/*hits* + *misses*) or false alarm ratio (FAR, defined as *false alarms*/*hits* + *false alarms*) (Wilks, 2006), commonly used to assess the detection/forecast results.

For each time step, the identification result is verified with available visibility observations. Fig. 18 illustrates an example of a particular time step with different types of dust events (color coded in terms of

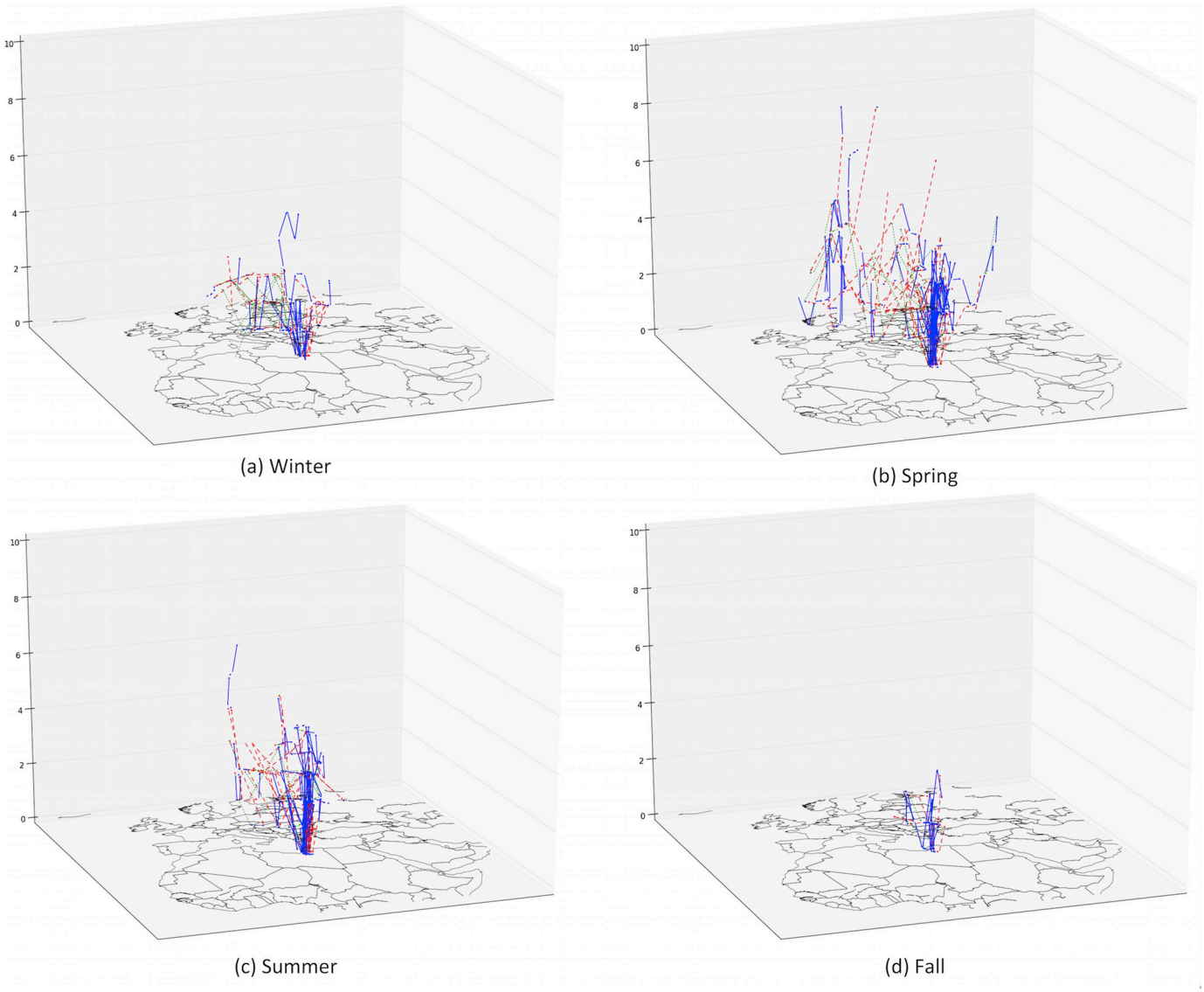

Fig. 11. 3D view of seasonal transport patterns of dust events originating from Libya desert.

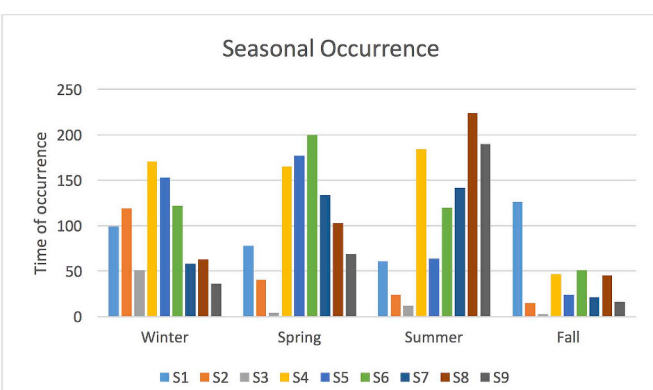

Fig. 12. Seasonality in occurrence of dust events from preferential dust sources.

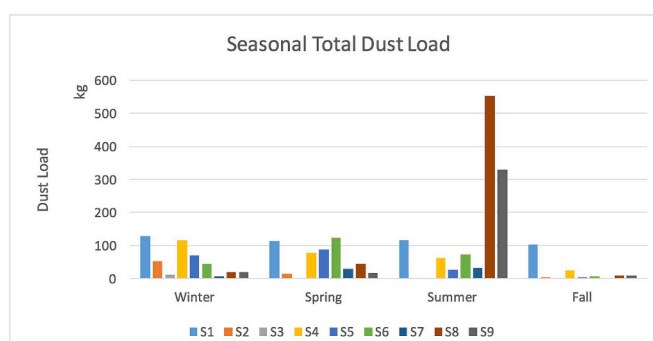

Fig. 13. Seasonal total dust load of dust events from preferential dust sources.

visibility range) and the identification result. With a visibility *Threshold* of 200 m, the POD value has a mean of 91% and a standard deviation of 13%; the FAR has a mean of 16% and the standard deviation of 9%. With a visibility *Threshold* of 1 km, the POD has a mean of 85% and a standard deviation of 5%; the FAR has a mean of 12% and a standard deviation of 8%. With a visibility *Threshold* of 10 km, the POD value has a mean of 76% and a standard deviation of 15%; the FAR has a mean of 24% and a standard deviation of 10%. Overall, the identification result achieves a relatively high POD with a low FAR, indicating the identification has a good performance.

A possible reason that the POD value is not higher is that the stations are irregularly distributed. This results in limited number of

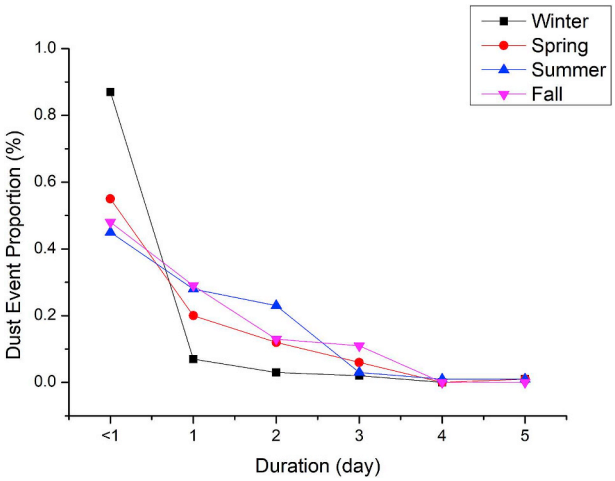

Fig. 14. Dust event duration as a function of the seasons of the year.

stations covered by the identified dust storms. In addition, not all stations provide available observation dataset for verification continuously, reducing the pool of data. Another explanation is that visibility data are observed on the surface (two dimensions), while dust events are reconstructed in three dimensions.

5.2. Evaluation of tracked dust events with NASA Earth Observatory

The transport pathways of the detected dust events are evaluated using the achieved records in NASA Earth Observatory, which records outstanding dust events with a textual description and MODIS images produced by NASA GSFC. Within the time frame of our experiment dataset, there are two events recorded - February 28, 2014 over West Africa (NASA Earth Observatory, 2014a) and June 15, 2014 over Oman (NASA Earth Observatory, 2014b). Both events have consistent transport pathways with our detection (Fig. 19). The first dust event was originated with harmattan desert wind blowing across the Sahara Desert from the northeast or east. During this dust event, dust was transported to the west towards the Cape Verde Islands in the Atlantic Ocean (Fig. 19a and b). The second event, originating with the tropical cyclone Nanauk moving northwest toward Oman's shores, transported dust from the northern border of the country toward the southwest, extending over the Arabian Sea (Fig. 19c and d).

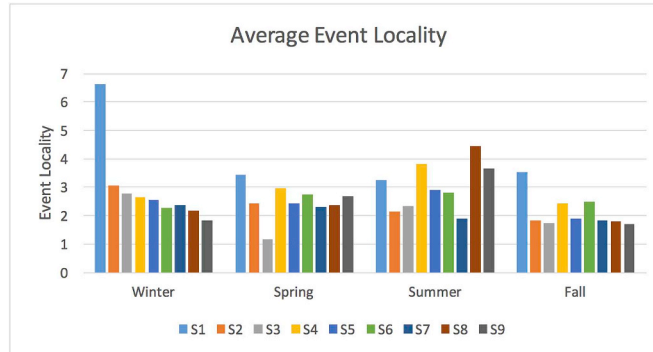

Fig. 16. Average event locality of dust events from preferential dust sources.

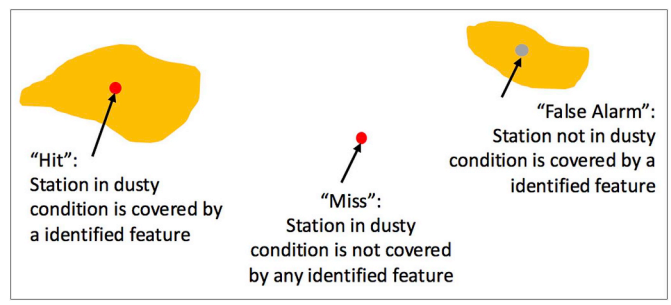

Fig. 17. Illustration of 'hit', 'miss', and 'false alarm'. Yellow polygons represent identified dust features, and points represent visibility stations. Red points represent that the stations are in dusty condition (< visibility Threshold). The gray point represents that the station is not in dusty condition (> visibility Threshold). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

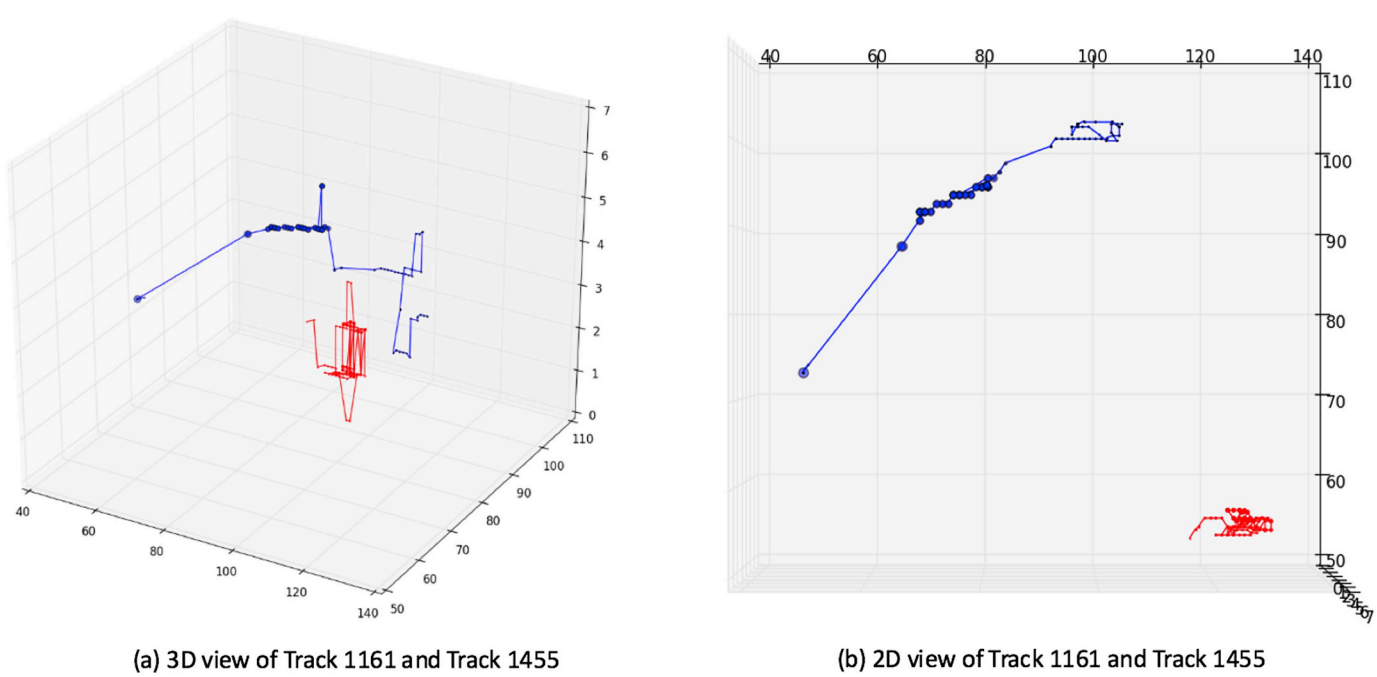

Fig. 15. Locality example. The distances of Track 1161 (blue) and Track 1455 (red) are about the same, which is 109°; but locality of Track 1161 is 0.15, while the locality of Track 1455 is 0.61. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

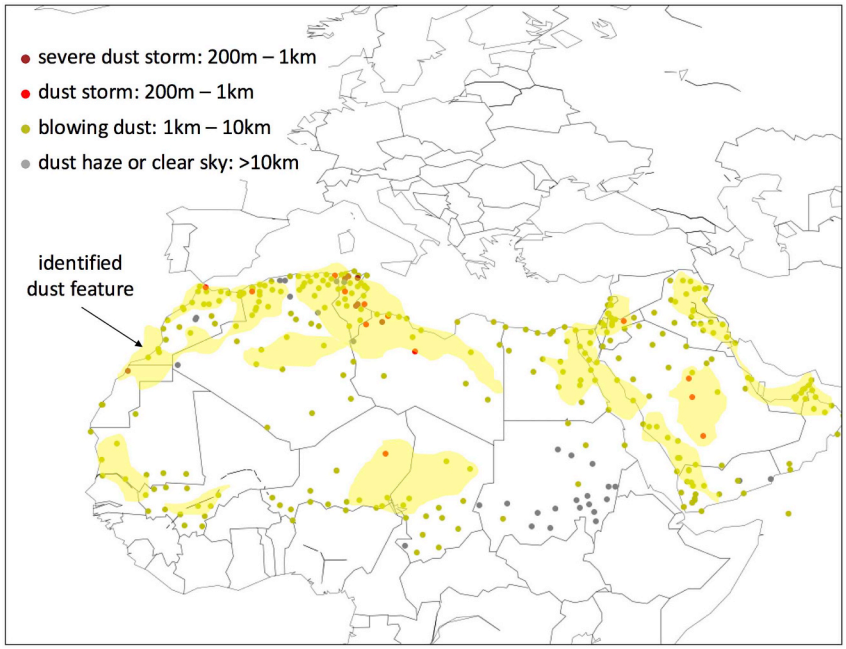

Fig. 18. Visibility values and identified dust features at April 1st, 2014, 0 UTC. Visibility stations are colored coded corresponding to different levels of dust storms: severe dust storm in brown color, dust storm in red color, blowing dust in yellow-green color, dust haze or clear sky in gray color. Yellow polygons are the identified dust features. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

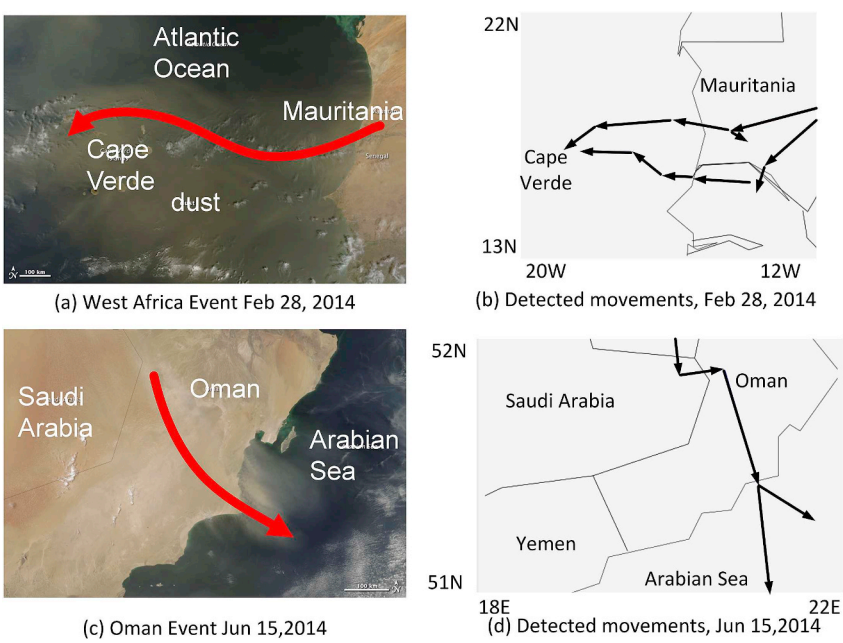

Fig. 19. (a) Transport pathways described in NASA Earth Observatory for the West Africa Event, on Feb 28, 2014; (b) Transport pathway resulted from the experiment for the same event as (a); (c) Transport pathways described in NASA Earth Observatory for the Omen Event, on Jun 15, 2014; (d) Transport pathway resulted from the experiment for the same event as (b).

5.3. Discussion

To demonstrate the capability of the framework on other meteorological events, the framework is tested using 3D precipitation dataset from TRMM Multi-satellite Precipitation Analysis.[2] The data covers the tropical area of 180°W −180°E and 50°S −50°N, with an original spatial pixel size of 0.25° *0.25°, so the total pixel number of an original dataset is 1440 × 400 (longitude, latitude). In the precipitation test, we used a 24-h dataset on 8/18/2016 with 3-hourly temporal resolution, and set the rainfall threshold of 2.5 mm/h and an area threshold of 10 pixels (Fig. 20). Fig. 20a illustrates the rainfall tracks in arrows during the day, where as Fig. 20b, c, and 20d show three zoomed-in tracks. For example, Fig. 20b represents the precipitation event starting from East India to Central India, which can be verified by the 2016 India Rainfall Statistics Report (India Meteorological Department, 2017). The event transports northwest, then split with into two tracks (red arrow), which merge together (green arrow) when the event ends.

One limitation of this approach is that we utilized a fixed threshold, i.e., 360 μg/m3, to identifying dust features for the simplicity and feasibility of tracking. However, as a dust event continues, the dust concentration that can best characterize the dust event is evolving, or changing continuously. Therefore, we will integrate fully the multi-thresholding approach, proposed by Yu and Yang (2017), with the tracking algorithm in this paper. The ultimate goal is to establish a multi-level dust event tracking, covering different levels of dust intensity and also different phases of dust evolution.

[2] https://giovanni.sci.gsfc.nasa.gov/giovanni/.

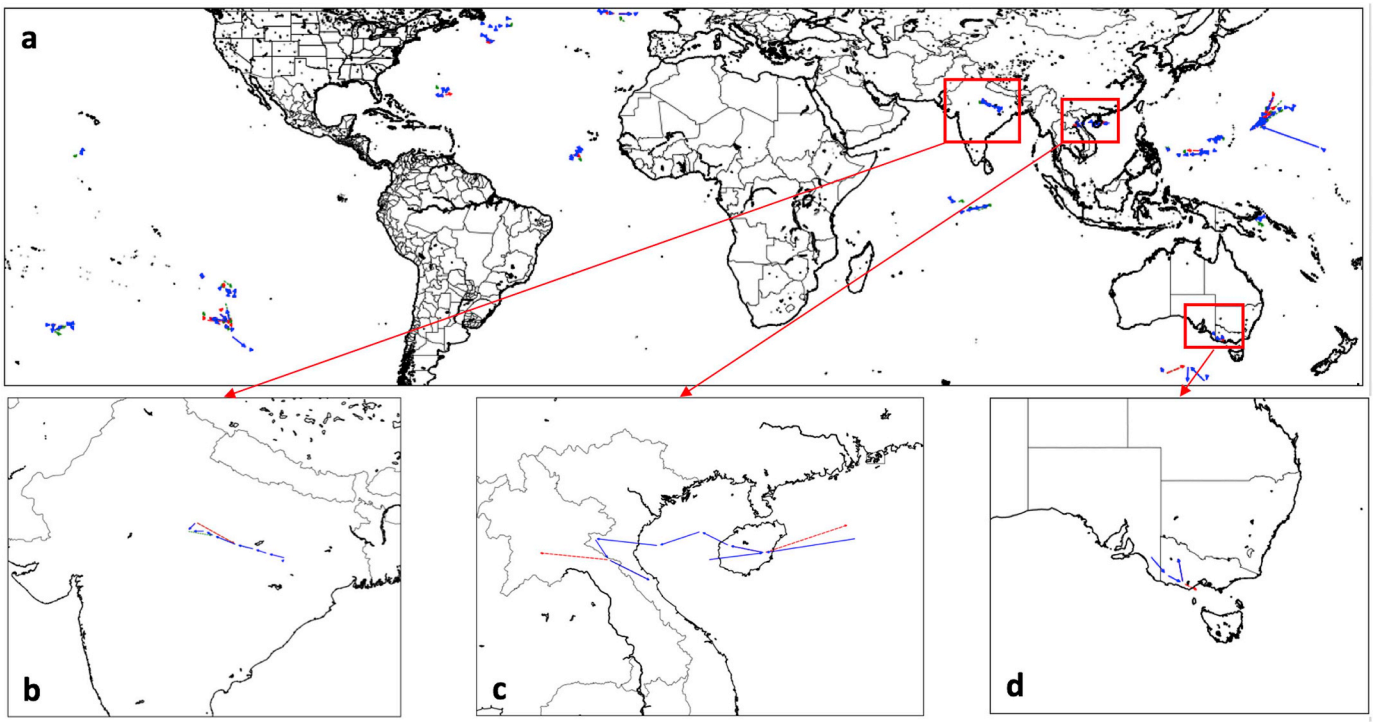

Fig. 20. 3D precipitation test result. Figure a: global detected precipitation tracks; Figure b, c, and d: zoomed in examples. Blue arrow: continuation; green arrow: merge; red arrow: split. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

6. Conclusion

A framework is proposed to automatically identify and track meteorological phenomena to represent their evolvement. The framework introduces the identification and tracking approaches to detect events and analyzes the evolvements and dynamic movements of the detected events. The feature identification approach was derived from previous research (Yu and Yang, 2017), whereas the tracking approach considers the variable size of dust feature and adopts a second check for additional temporal linkages. Detected tracks are optimized through merging tracks that are possibly sequential over time. Tracks are also simplified for a better visualization.

In the future the capability of this framework to operate using machine learning and knowledge reasoning algorithms to analyze seasonal variation, annual cycle, and inter-annual cycle of dust events is a high priority. In addition, this framework will be used on dust observation datasets to monitor dust transport and provide information on early warning of severe dust events. The framework of identifying and tracking natural phenomena can also be utilized for modelers as a toolkit to automatically process their real-time forecast result, including storm cells, ocean eddies, and cyclones. As high resolution atmospheric and climate models produce big data, cloud computing provides increasing value in high efficiency and scalability to the data management and analytical process (Yang et al., 2017).

Acknowledgements

This research is supported by NSF I/UCRC (IIP-1338925) and OAC (1835507). We thank the anonymous reviewers for their insightful comments and reviews. Dr. George Taylor edited the paper. Ms. Qian Liu provided the one-day TRMM data and its description.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cageo.2018.10.003.

References

Adrienko, N., Adrienko, G., 2011. Spatial generalization and aggregation of massive movement data. IEEE Trans. Visual. Comput. Graph. 17 (2), 205–219.
Ammar, K., El-Metwally, M., Almazroui, M., Wahab, M.A., 2014. A climatological analysis of Saharan cyclones. Clim. Dynam. 43 (1–2), 483–501.
Basart, S., Pérez, C., Nickovic, S., Cuevas, E., Baldasano, J., 2012. Development and evaluation of the BSC-DREAM8b dust regional model over Northern Africa, the Mediterranean and the Middle East. Tellus B 64 (1), 18539.
Borbély-Kiss, I., Kiss, A.Z., Koltay, E., Szabo, G., Bozo, L., 2004. Saharan dust episodes in Hungarian aerosol: elemental signatures and transport trajectories. J. Aerosol Sci. 35 (10), 1205–1224.
Choi, J., Olivera, F., Socolofsky, S.A., 2009. Storm identification and tracking algorithm for modeling of rainfall fields using 1-h NEXRAD rainfall data in Texas. J. Hydrol. Eng. 14 (7), 721–730.
Dixon, M., Wiener, G., 1993. TITAN: thunderstorm identification, tracking, analysis, and nowcasting-A radar-based methodology. J. Atmos. Ocean. Technol. 10 (6), 785–797.
Duchi, R., Cristofanelli, P., Landi, T.C., Arduini, J., Bonafe, U., Bourcier, L., Busetto, M., Calzolari, F., Marinoni, A., Putero, D., Bonasoni, P., 2016. Long-term (2002–2012) investigation of Saharan dust transport events at Mt. Cimone GAW global station, Italy (2165 m asl). Elem. Sci. Anth. 4.
El-Askary, H., Kafatos, M., Hegazy, M., 2002. June. Environmental monitoring of dust storms over the Nile Delta, Egypt using MODIS satellite data. In: Proceedings of Third International Symposium Remote Sensing of Urban Areas, Istanbul, Turkey, June 11 - 13, 2002. İTÜ, Istanbul, Turkey, pp. 452.
Han, L., Fu, S., Zhao, L., Zheng, Y., Wang, H., Lin, Y., 2009. 3D convective storm identification, tracking, and forecasting—an enhanced TITAN algorithm. J. Atmos. Ocean. Technol. 26 (4), 719–732.
India Meteorological Department (Ministry of Earth Sciences), 2017. Rainfall Statistics of India – 2016. Report NO. ESSO/IMD/HS/R. F. REPORT/01(2017)/23. Hydromet Division, India Meteorological Department, Lodi Road, New Delhi, India, pp. 17 (total pages: 105).
Israelevich, P.L., Ganor, E., Levin, Z., Joseph, J.H., 2003. Annual variations of physical properties of desert dust over Israel. J. Geophys. Res.: Atmosphere 108 (D13).
Johnson, J.T., MacKeen, P.L., Witt, A., Mitchell, E.D.W., Stumpf, G.J., Eilts, M.D., Thomas, K.W., 1998. The storm cell identification and tracking algorithm: an enhanced WSR-88D algorithm. Weather Forecast. 13 (2), 263–276.
Lakshmanan, V., Hondl, K., Rabin, R., 2009. An efficient, general-purpose technique for identifying storm cells in geospatial images. J. Atmos. Ocean. Technol. 26 (3), 523–537.
Lorensen, W.E., Cline, H.E., 1987. Marching cubes: a high resolution 3D surface construction algorithm. In: ACM SIGGRAPH '87 Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques. ACM, New York, NY, pp. 163–169.
Moulin, C., Lambert, C.E., Dayan, U., Masson, V., Ramonet, M., Bousquet, P., Legrand, M., Balkanski, Y.J., Guelle, W., Marticorena, B., Bergametti, G., 1998. Satellite

climatology of African dust transport in the Mediterranean atmosphere. J. Geophys. Res.: Atmosphere 103 (D11), 13137–13144.
Moy, K., Li, W., Tran, H.P., Simonis, V., Story, E., Brandon, C., Furst, J., Raicu, D., Kim, H., 2015. Computational methods for tracking, quantitative assessment, and visualization of C. elegans locomotory behavior. PloS One 10 (12), e0145870.
NASA Earth Observatory, 2014a. Cape Verde under Dust [online]. NASA Earth Observatory, NASA Goddard Space Flight Center, Greenbelt, MD, USA Available from: http://earthobservatory.nasa.gov/IOTD/view.php?id=83260, Accessed date: 27 October 2016.
NASA Earth Observatory, 2014b. Oman Dust Event [online]. NASA earth Observatory, NASA Goddard Space Flight Center, Greenbelt, MD, USA Available from: http://earthobservatory.nasa.gov/IOTD/view.php?id=83869, Accessed date: 27 October 2016.
Nickovic, S., Kallos, G., Papadopoulos, A., Kakaliagou, O., 2001. A model for prediction of desert dust cycle in the atmosphere. J. Geophys. Res.: Atmosphere 106 (D16), 18113–18129.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12 (Oct), 2825–2830.
Pérez, C., Nickovic, S., Baldasano, J.M., Sicard, M., Rocadenbosch, F., Cachorro, V.E., 2006a. A long Saharan dust event over the western Mediterranean: Lidar, Sun photometer observations, and regional dust modeling. J. Geophys. Res.: Atmosphere 111 (D15).
Pérez, C., Nickovic, S., Pejanovic, G., Baldasano, J.M., Oezsoy, E., 2006b. Interactive dust-radiation modeling: a step to improve weather forecasts. J. Geophys. Res.: Atmosphere 111 (D16).
Prospero, J.M., Lamb, P.J., 2003. African droughts and dust transport to the Caribbean: climate change implications. Science 302 (5647), 1024–1027.
Pultar, E., Cova, T.J., Yuan, M., Goodchild, M.F., 2010. EDGIS: a dynamic GIS based on space time points. Int. J. Geogr. Inf. Sci. 24 (3), 329–346.
Rezazadeh, M., Irannejad, P., Shao, Y., 2013. Climatology of the Middle East dust events. Aeolian Res. 10, 103–109.
Samtaney, R., Silver, D., Zabusky, N., Cao, J., 1994. Visualizing features and tracking their evolution. Computer 27 (7), 20–27.
Stein, A.F., Draxler, R.R., Rolph, G.D., Stunder, B.J., Cohen, M.D., Ngan, F., 2015. NOAA's HYSPLIT atmospheric transport and dispersion modeling system. Bull. Am. Meteorol. Soc. 96 (12), 2059–2077.
Van Der Walt, S., Colbert, S.C., Varoquaux, G., 2011. The NumPy array: a structure for efficient numerical computation. Comput. Sci. Eng. 13 (2), 22–30.
Van Rossum, G., Drake Jr., F.L., 1995. Python Tutorial. Centrum voor Wiskunde en Informatica, Amsterdam, The Netherlands.
Varga, G., Újvári, G., Kovács, J., 2014. Spatiotemporal patterns of Saharan dust outbreaks in the Mediterranean basin. Aeolian Res. 15, 151–160.
Wilks, D.S., 2006. Statistical Methods in the Atmospheric Sciences, second ed. Academic press, Cambridge, Massachusetts, pp. 627.
Worboys, M., 2005. Event-oriented approaches to geographic phenomena. Int. J. Geogr. Inf. Sci. 19 (1), 1–28.
World Meteorological Organization (WMO), 2013. Establishing a WMO Sand and Dust Storm Warning Advisory and Assessment System Regional Node for West Asia: Current Capabilities and Needs. WMO, Geneva, Switzerland.
Yang, C., Wu, H., Huang, Q., Li, Z., Li, J., 2011. Using spatial principles to optimize distributed computing for enabling the physical science discoveries. Proc. Natl. Acad. Sci. Unit. States Am. 108 (14), 5498–5503.
Yang, C., Huang, Q., Li, Z., Liu, K., Hu, F., 2017. Big Data and cloud computing: innovation opportunities and challenges. Int. J. Digital Earth 10 (1), 13–53.
Yuan, M., 2001. Representing complex geographic phenomena in GIS. Cartogr. Geogr. Inf. Sci. 28 (2), 83–96.
Yu, M., Yang, C., 2017. A 3D multi-threshold, region-growing algorithm for identifying dust storm features from model simulations. Int. J. Geogr. Inf. Sci. 31 (5), 939–961.
Zucker, S.W., 1976. Region growing: childhood and adolescence. Comput. Graph. Image Process. 5 (3), 382–399.