


Article

# A Graph-Based Spatiotemporal Data Framework for 4D Natural Phenomena Representation and Quantification—An Example of Dust Events

Manzhu Yu 

Department of Geography, Pennsylvania State University, State College, PA 16801, USA; mqy5198@psu.edu

Received: 26 December 2019; Accepted: 20 February 2020; Published: 22 February 2020



**Abstract:** Natural phenomena are intrinsically spatiotemporal and often highly dynamic. The increasing availability of simulation and observation datasets has provided us a great opportunity to better capture and understand the complexity and dynamics of natural phenomena. Challenges are posed by the formalization of the representation of such phenomena in terms of their non-rigid boundaries and the quantification of event dynamics over space and time. The objectives of this research are to (1) conceptually represent the natural phenomenon as an event, and (2) quantify the dynamic movements and evolutions of events using a graph-based approach. This proposed data framework is applied to a dust simulation dataset to represent the 4D dynamic dust events. Dust events are identified, and movements are tracked to reconstruct dust events in the Northern Africa region from December 2013 to November 2014. Quantified dynamics of different dust events are demonstrated and verified to be in alignment with observations.

**Keywords:** spatiotemporal representation; dust event; 4D; graph-based; dynamics quantification

## 1. Introduction

Natural phenomena are intrinsically spatiotemporal and often highly dynamic [1]. The increasing availability of simulation and observation datasets has enabled scientists and researchers to better capture and understand the complexity and dynamics of natural phenomena. With higher spatiotemporal resolution and dimensionality, a more complete view of natural phenomena can be obtained, such as detailed information of the vertical and temporal dimensions (4D: latitude, longitude, altitude, and time). GIScience methodologies and techniques have been developed to represent and examine the dynamics in geographic changes through quantitative and qualitative ways, especially assisting us in understanding “where and when natural phenomena happen”, “how long a natural phenomenon lasts”, or “what the common transport pathway is for a natural phenomenon.”

Existing spatiotemporal data models [2–4] have the capability of representing various changes in 3D (latitude, longitude, and time), such as creation, alteration, destruction, reincarnation, split, merge, and reallocation. However, unlike those phenomena occurring on the Earth’s surface, such as flooding [3], land use and land cover change [5], or deforestation [6], most geographic phenomena exist in a 4D context (latitude, longitude, altitude, and time). Geographic objects in 3D space interact more frequently than the ones in 2D space. Therefore, it is important to represent, analyze, and model the dynamics of natural phenomena in a 4D environment. In this research, volumetric components, as well as the handling of complex object interactions, are proposed in the spatiotemporal representations.

One of the challenges is to incorporate the fuzzy or non-rigid boundary of natural phenomena into the spatiotemporal framework. The intrinsic properties of the non-rigid natural phenomena include dynamics and deformation and can be constantly changing their internal structure while transport in space and time [7]. Therefore, representing these internal structure change needs to be included in

the spatiotemporal framework as well. In the object-based image analysis field, in order to describe natural variability, models need to be capable of expressing the fuzzy boundaries and being adaptable to unforeseeable changes [8]. In addition, hierarchical relationships at different levels of abstraction in space and time need to be further explored [9]. Inspired by the above researches, we have integrated the hierarchy theory in our non-rigid object representation to enhance the event representation.

In addition, existing spatiotemporal data frameworks have been largely focused on spatial reasoning or temporal reasoning separately, whilst spatiotemporal quantitative analysis can assist the better understanding and future predictions of natural phenomena events. Through an object-oriented method, McIntosh and Yuan [10] quantified the characterization and similarity of geographic events regarding the static and dynamic characteristics and relationships of objects. Spatiotemporal statistical methods that quantify the changes or dynamics have also been developed, including spatiotemporal scan statistics [11], spatial time series [12], and SpatioTemporal Autoregressive Regression [13]. In addition, Bothwell and Yuan [14] quantified the movements and directions in continuous space and time based on the principles of fluid kinematics. However, a full exploration of quantifying the event evolution in a holistic data framework is still missing. Although quantifying the attributes of an event can assist with better understanding the events as a whole, it is more critical to examine the intermediate and internal dynamics of event evolution over space and time. Therefore, in this research, we propose an event-based measurement to quantify the dynamics, not just the attributes, of natural phenomena evolutions.

In this paper, our objectives are to (1) conceptually represent a natural phenomenon as an event using a graph-based approach, and (2) quantify the evolutions and dynamics of the events. Works relating to spatiotemporal representation and quantification are reviewed in Section 2. The proposed data framework is described in Section 3 and applied to represent and quantify the constant movements of dust events as a case study in Section 4. Section 5 provides a discussion on the potential applications and limitations of the proposed framework. And finally, Section 6 concludes this research and proposes potential future work.

## 2. Related Works

### 2.1. Spatiotemporal Representations

The integration of time into the spatial data models can be traced back to the early 1980s, where time-stamped layers are integrated into relational databases [15,16]. Over the years, there has been a great amount of research focusing on designing data models that can represent and characterize complex dynamical phenomena using simulation and observation data, including object-oriented modeling [17,18], domain-based modeling [19], event-based modeling [20,21], and graph-based modeling [22,23]. A more complete review on spatiotemporal data models can be found in Siabato et al. [24]. As concluded by Siabato et al. [24], a majority of data models focus on object-oriented approaches.

As a general theory to represent the complexity of geo-information, Goodchild et al. [25] proposed the concepts of geo-atoms, geo-objects, and geo-dipoles. The geo-atoms, the 'atomic form' of location and properties, can form geo-objects. Geo-objects are dynamic based on three conditions: (i) possible movement; (ii) possibly changing geometry; and (iii) homogeneous or heterogeneous/evolving internal structure. Geo-dipoles represent the interactions based on the locations of geo-atoms, such as the direction, the distance, the interaction, and the flow. Since natural phenomena generally do not have crisp boundaries, they are broadly represented as fuzzy spatial objects in spatiotemporal databases [26,27], thus providing a flexible query and reasoning on imperfect spatial and temporal information [28]. Object-oriented database models were also developed to store, manage, query, and index fuzzy spatial objects with information imperfectness [29]. Topological relationships, such as overlapping and intersectional, between fuzzy spatial objects were also proposed to handle simple and complex fuzzy regions [30]. Moving beyond object interactions, Worboys [31] proposed an upgraded

representation from object and interactions to an event-oriented view of the world, as it can represent the object–event relationships as well as the event–event relationships. As asserted in Yuan and Hornsby [32], event-based approaches “focus on the dynamic happening as a whole, and not just the time of the event.” In this research, we pursue an event-based spatiotemporal representation and consider natural phenomena occur in the form of an event.

Graph-based strategies have been proposed for event-based spatiotemporal representation. Renolen [22] proposed the History Graph Model to visualize the evolution of geographic information resulting from their changes by events over time using Petri Nets. Del Mondo et al. [23] represented the 2D land-use with evolving entities using a graph-based model and formalized the constraints for spatiotemporal databases as well as the semantic constraints on the filiation relations, including expansion, contraction, split, separation, merge, and annexation. Based on the evolution graph of land use and land cover change, Guttler et al. [33] quantified the temporal variations of different types of natural, semi-natural and agricultural areas in the South of France. As summarized by Siabato et al. [24], one advantage of modeling in graphs is its flexibility of integrating different query languages and semantic constraints on the modeling strategy. Therefore, in this research, we propose a graph-based representation for natural phenomena movement tracking with a formalized quantification approach considering both spatial and non-spatial dynamics over time.

## 2.2. Quantification of Event Dynamics

Spatiotemporal data frameworks reviewed above are generally focused on qualitative representation, but they lack the capability of producing numerical understanding of the spatiotemporal dynamics of natural phenomena. Various methods have been developed to quantify spatiotemporal dynamics, as reviewed by Long and Nelson [34] and An et al. [35], to better analyze the dynamics for (point-based) individual movement data or (polygon-based/volume-based) natural phenomena. For example, Kulldorff [36] and Kulldorff et al. [37] developed a space-time clustering method based on scan statistics to produce clusters with specific spatial and temporal ranges that are statistically significant. This spatiotemporal scan statistics method is suitable for clustering individual movement data that are irregularly distributed in space and time. For polygon-based movement data, Robertson et al. [38] measured events for the changes of size and direction in moving polygons. Size changes are quantified based on area overlaps, while directional changes are quantified by the directional distribution degree and the directional rate of spread.

The quantification of changes or dynamics have been integrated into spatiotemporal representations. For example, Bothwell and Yuan [14] identified kinematic flows and objects to represent geographic processes using velocity as the fundamental unit. The kinematic approach is applied to identify outliers in global temperature for further investigation. McIntosh and Yuan [10] quantified the static attributes of geographic objects using elongation, orientation, and distribution; as well as quantifying the dynamic attributes of geographic relations inside events using growth, granularity of change, and relative movements. Utilizing these quantification metrics, rainfall events can be assessed for similarity, to retrieve most similar events based on spatiotemporal queries. The boundary of a natural phenomenon is not generally based on a single criterion; thus, in our approach, we adopt a hierarchical approach to determine the spatiotemporal dynamics of natural phenomena in the analysis of multiple intensity granularity.

## 3. A spatiotemporal Data Framework for 4D Natural Phenomena Representation and Quantification

This research proposes a spatiotemporal data framework to represent and quantify the dynamics of natural phenomena as events. An event is defined as “something that happens at a particular time and place” by Allan et al. [39], which is broadly accepted in the literature. There are three major entities inside the data framework for events (Figure 1): (1) *ST-object* that represents an object that evolves in space and time; (2) *ST-relation* that represents seven types of relationships between *ST-objects* between consecutive timestamps; (3) *ST-event* that represents the lifecycle of an occurrence

of natural phenomena from the start till the end, possibly involving multiple *ST-objects* and the *ST-relations* associated with these *ST-objects*. In addition, to integrate the hierarchical theory into the object representation, *ST-objects* are represented in different levels of intensity (e.g., temperature for an urban heat island, dust concentration for a dust storm, and base reflection for a thunderstorm), where a higher level of intensity occupies a smaller spatial coverage, i.e., one or multiple *ST-objects* with a higher level of intensity aggregate into an *ST-object* with a lower level of intensity. These intensity values are determined based on the understanding of a particular type of natural phenomena, and the number of different levels can be flexible. A particular *ST-event* is then represented as a graph, consisting of *ST-objects* as nodes and *ST-relations* as edges. A graph-based quantification method is further proposed to numerically understand the dynamics of changes between timestamps and the whole event. Details are introduced in the following subsections.

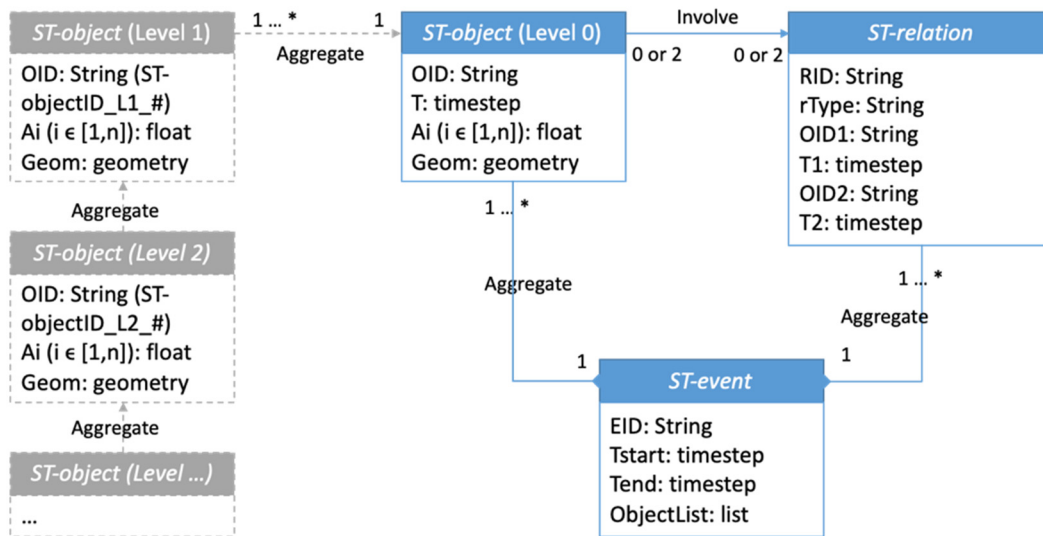


Figure 1. Overall conceptual data framework.

### 3.1. ST-Object

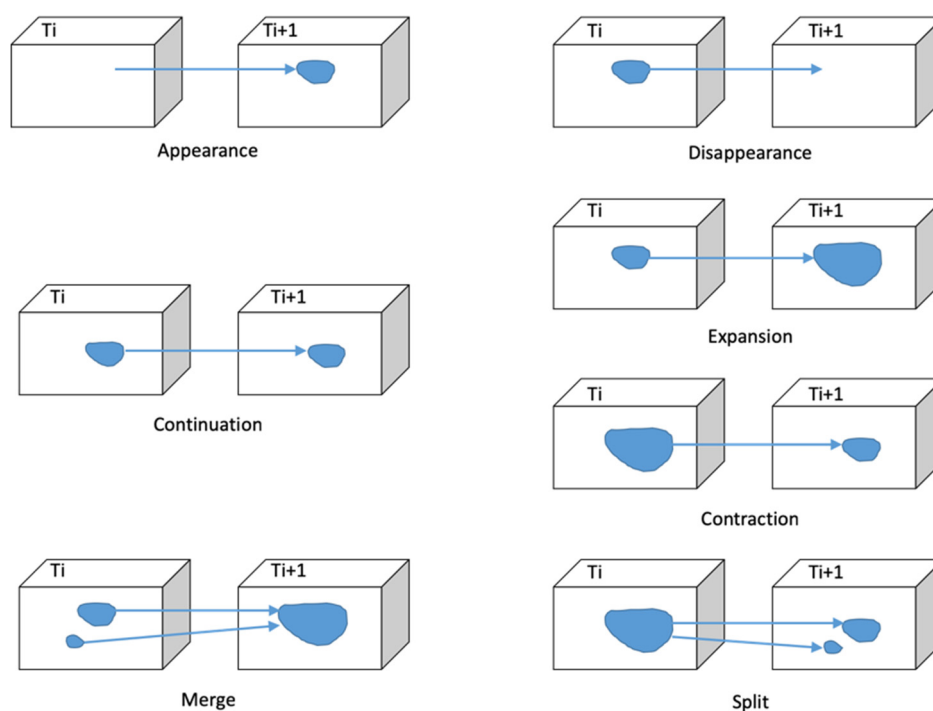
A spatiotemporal object (*ST-object*) represents a moving object with possibly changing shape and evolving internal structure, i.e., its geometry with spatial attributes (shape and location) can change in different timestamps. Each geo-object is represented by its identity, time, thematic attributes, and geometry at time:  $Id(O), t, A_1(id,t), \dots, A_n(id,t), Geom(id,t)$  respectively. In keeping with the rapid increase of 4D Big Data, such as meteorological satellite retrieval and forecasting of results, we embed the  $Geom(id,t)$  of an *ST-object* in a 4D context. For each *ST-object*, the thematic attributes and geometry are evolving over time. The geometries of geo-objects are usually indeterminate for natural phenomena, as there is no single threshold appropriate for the identification of all geographical objects for a natural phenomenon. Thus, we embed the indeterminacy of boundary in the data framework using a hierarchical approach. By modeling the natural phenomenon at several hierarchies, internal structures of the natural phenomenon of interest can show a distinct pattern at a given hierarchy. This approach can also represent the hierarchical relationships of nested objects, e.g., a high temperature zone is contained inside a low temperature zone.

### 3.2. ST-Relation

A spatiotemporal relation (*ST-relation*) connects *ST-object* at consequent timestamps, i.e., appearance, disappearance, expansion, contraction, continuation, splitting, and merging (Figure 2). A continuation relation,  $r_{continue}$ , represents when object  $o$  at time  $t_i$  continues at time  $t_{i+1}$  and its geometry remains constant, where the *ST-object*'s identifier does not change. Expansion ( $r_{expand}$ ) and contraction ( $r_{contract}$ ) are two other types of relations when the *ST-object* continues to exist, but its

geometry grows or decays. Splitting ( $r_{split}$ ) and merging ( $r_{merge}$ ) are two relations when the *ST-objects* cease to exist through splitting and merging. Appearance and disappearance are two relations that involve an *ST-object* from only one particular timestamp, which represents the beginning or the end of the particular *ST-object*. A specific *ST-relation* ( $R$ ) is represented by the involved objects' identity, time, and type of relation:  $Id(o_1), t_i, r, Id(o_2), t_{i+1}$ ; where  $r \in [r_{continue}, r_{expand}, r_{contract}, r_{split}, r_{merge}]$ ,  $i \in [1, n]$ . All attributes serve as the primary key of this table. Given objects  $O_1(id_1, t_1, \bar{a}_1, g_1)$ ,  $O_2(id_2, t_2, \bar{a}_2, g_2)$ , and  $O_3(id_3, t_3, \bar{a}_3, g_3)$ , the spatiotemporal relations can be defined as follows:

- **Expansion** (Geometric Growth):  $O_2(id_2, t_2, \bar{a}_2, g_2)$  continues existing and is geometrically growing from  $O_1(id_1, t_1, \bar{a}_1, g_1)$ , where  $id_1 = id_2$ , and  $t_1 < t_2$ .
- **Contraction** (Geometric Decay):  $O_2(id_2, t_2, \bar{a}_2, g_2)$  continues existing and is geometrically decaying from  $O_1(id_1, t_1, \bar{a}_1, g_1)$ , where  $id_1 = id_2$ , and  $t_1 < t_2$ .
- **Continuation** (Geometry remains constant):  $O_2(id_2, t_2, \bar{a}_2, g_2)$  continues existing and its geometry remains constant from  $O_1(id_1, t_1, \bar{a}_1, g_1)$ , where  $id_1 = id_2$ ,  $g_1 = g_2$ , and  $t_1 < t_2$ .
- **Splitting**:  $O_1(id_1, t_1, \bar{a}_1, g_1)$  splits into  $O_2(id_2, t_2, \bar{a}_2, g_2)$ , and  $O_3(id_3, t_2, \bar{a}_3, g_3)$ , where  $id_1 \neq id_2 \neq id_3$ . Object  $O_1$  ceases to exist after the split and objects  $O_2$  and  $O_3$  never exist before. Then the graph will store the associated entries,  $(id_1, t_1, r_s, id_2, t_2)$  and  $(id_1, t_1, r_s, id_3, t_2)$ .
- **Merging**:  $O_1(id_1, t_1, \bar{a}_1, g_1)$  and  $O_2(id_2, t_1, \bar{a}_2, g_2)$  merge into  $O_3(id_3, t_2, \bar{a}_3, g_3)$ , where  $id_1 \neq id_2 \neq id_3$ . The merged objects  $O_1$  and  $O_2$  cease to exist and the new object  $O_3$  is formed by the merge and did not exist before. Then the relation table will store the associated entries,  $(id_1, t_1, r_m, id_3, t_2)$  and  $(id_2, t_1, r_m, id_3, t_2)$ .
- **Appearance**:  $O_1(id_1, t_2, \bar{a}_1, g_1)$  is not related to any object at  $t_1$ ;
- **Disappearance**:  $O_1(id_1, t_1, \bar{a}_1, g_1)$  is not related to any object at  $t_2$ .



**Figure 2.** ST-relations between objects at  $t_i$  and  $t_{i+1}$ .

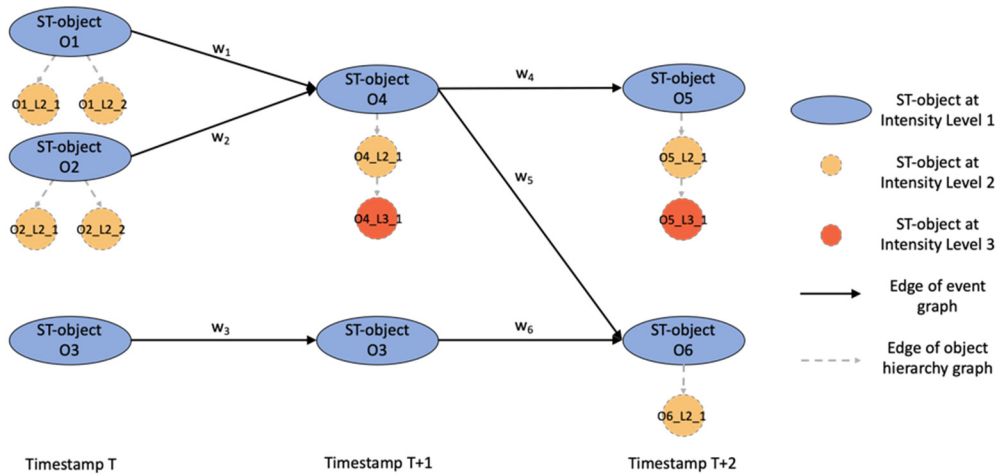
### 3.3. ST-Event

A spatiotemporal event (*ST-event*) is represented as a single occurrence for a definite time series which can involve one or more *ST-objects* that have associated *ST-relations*. A specific event  $E$  is represented by its identity, starting time, ending time, and the corresponding list of *ST-objects* and

*ST-relations*:  $EID, t_i, t_j, \{O_m, \dots, O_n\}, \{R_k, \dots, R_l\}$ , where  $t_i \leq t_j$ . AN *ST-event* is represented as a graph linking the objects of timestamp  $t$  with the objects of timestamp  $t+1$ . Each node of the graph represents a specific object, and each edge represents the overlapping interactions between objects from consequent timestamps. Each edge corresponds with its own weight, which represents the degree of spatial overlap between two objects (Equation (1)).

$$w_{i,j} = \frac{2 * overlappingVolume(O_i, O_j)}{Volume(O_i) + Volume(O_j)} \quad (1)$$

where  $O_i$  represents an object of timestamp  $t$  and  $O_j$  represents an object of timestamp  $t+1$ . Therefore, for two objects that have continuation relation, their weight is exactly 1. Objects that have expansion, contraction, merge, or split relations have a weight larger than 0 and smaller than 1. For appearance and disappearance relations, there is no need to calculate the weights, since one of the objects (i.e.,  $O_i$  and  $O_j$ ) does not exist. Each graph represents an *ST-event*, and the number of levels is the number of timestamps that this *ST-event* exists. For each *ST-object*, the object hierarchy is stored as a sub-graph, and can be integrated with the overall time directed graph of the associated *ST-event* (Figure 3).



**Figure 3.** Spatiotemporal (ST)-event and object hierarchy represented in an event graph.

### 3.4. Graph-Based Measurements of ST-Event Evolution Dynamics

In order to quantify the dynamics based on the *ST-event* graphs, we propose to measure the variations inside a *graph* to quantitatively represent how an *ST-event* evolves in time regarding its spatial and non-spatial characteristics. Guttler et al. [33] proposed the quantification of the variations considering the 2D area change of the land cover, and the Euclidean distances between the changing land cover areas. Similar to Guttler et al. [33], we represent a graph as  $G$ , the set of objects covered by  $G$  at the timestamp  $t$  as  $G_t$ , and the weight of the link between object  $O_i$  and object  $O_j$  from consecutive timestamps as  $w_{i,j}$ . In our approach, however, we enhance this quantification by considering the spatial distances as well as the change of 3D volume and intensity level of the natural phenomena.

The Variation (*Var*) of a graph between two consecutive timestamps is calculated with the Equation (2).

$$Var(G_t, G_{t+1}) = \sum_i^{O_i \in G_t} \frac{Volume(O_i)}{Volume(G_t)} \cdot \frac{\sum_j^{O_j \in G_{t+1}} w_{i,j} \cdot (\alpha \cdot dist(O_i, O_j) + \beta \cdot diff(O_i, O_j))}{\sum_j^{O_j \in G_{t+1}} w_{i,j}} \quad (2)$$

There are two parts inside the *Var* formula, where the first part represents the importance of the object  $O_i$  compared to the entire set of objects at timestamp  $t$ . A higher object importance will influence more on the total variation value. The second part investigates the changes between an object at timestamp  $t$

and all objects at timestamp  $t+1$  that have linkages to it. This change is calculated as a weighted sum of the 3D Euclidian distances  $dist(O_i, O_j)$  and intensity differences  $diff(O_i, O_j)$  between the object  $O_i$  and  $O_j$ . The weight  $w_{i,j}$  measures the spatial overlap o between  $O_i$  and  $O_j$ . The intensity differences  $diff(O_i, O_j)$  are calculated based on the absolute differences between the intensity-weighted volumes of the object  $O_i$  and  $O_j$ , represented in Equation (3).

$$diff(O_i, O_j) = abs \left( \frac{\sum_k^{L_k \in O_i} Volume(L_k) \cdot Intensity(L_k)}{\sum_k^{L_k \in O_i} Volume(L_k)} - \frac{\sum_m^{L_m \in O_j} Volume(L_m) \cdot Intensity(L_m)}{\sum_m^{L_m \in O_j} Volume(L_m)} \right) \quad (3)$$

where the intensity levels of the object  $O_i$  (at timestamp  $t$ ) and the object  $O_j$  (at timestamp  $t+1$ ) are represented as  $L_k$  and  $L_m$ , and different intensity levels are taken into consideration for the overall volume calculation. This volume calculation strategy ensures that a small volume with a higher level of intensity can still have a high intensity-weighted volume value, and that it can still result in a high difference value if the intensity changes significantly while the volume does not. In addition, this intensity-weighted volume calculation naturally reflects and quantifies the internal structure of the objects.

The 3D distances  $dist(O_i, O_j)$  and intensity differences  $diff(O_i, O_j)$  are weighted by  $\alpha$  and  $\beta$  respectively. The interdependence of these two variables needs to be calculated based on the univariate analysis of variance (ANOVA). If the interdependence is statistically significant, then  $\alpha$  and  $\beta$  need to be adjusted based on the coefficient of different variables. If not, then  $\alpha$  and  $\beta$  are assigned equal weights.

The Global Variation (*GlobalVar*) for a graph is calculated to summarize the variations from each pair of consecutive timestamps. The summarized *GlobalVar* evaluates the attributes of an event as a whole. The *GlobalVar* value of a particular graph reflects the extent of evolvments of the event within the spatiotemporal range. A high *GlobalVar* value indicates that there might be significant temporal evolutions occurred within the event's lifecycle.

#### 4. Dust Event Representation and Quantification

In order to evaluate the applicability of the spatiotemporal data framework for representing and quantifying the dynamics of natural phenomena, dust events in Northern Africa, the Middle East, and the Mediterranean were used to illustrate the instantiation of the proposed conceptual data framework and obtain information of dust events' spatiotemporal evolution and transport characteristics. Dust events are extracted from 4D simulation dataset using object identification and tracking algorithms. The overall procedure, illustrated in Figure 4, includes the following: (1) identifying static dust storm objects; (2) extracting the interactions of objects over consequent timestamps; (3) reconstructing dust storm events (ST-events); and (4) quantifying the variations of dust storm events.

##### 4.1. Data and Implementation

Dust events are detected from the 4D simulation data generated by the operational dust model BSC-DREAM8bv2.0 [40] maintained at the Barcelona Supercomputing Center. The spatial extent covers North Africa, the Middle East, and the Mediterranean (0.7 S–64.3 N, 25.7 W–59.3 E) with the spatial resolution of ~30 km. The vertical dimension covers 24 layers from the Earth surface to 15 km. The temporal extent covers 12 months, from December 2013 to November 2014, with a temporal resolution of one hour. The simulated variable used in this research is the dust concentration with a unit of  $\mu\text{g}/\text{m}^3$ , which reflects the intensity of dust particles in the atmosphere (Figure 5).

The implementation of dust event representation and quantification (detailed in Sections 4.2–4.5) is conducted in Python. The identified dust objects, relations, and events are stored in graph-based data structure using NetworkX (<https://networkx.github.io/>), which is a Python library for the creation and manipulation of graphs. The total size of the 4D dust simulation data is 40.0 GB. The identified dust objects are recorded and indexed by their unique  $Id(O)$  and their associated attributes and geometries are stored in separate auxiliary files, which totals at 1.8 GB. The reconstructed dust events are stored in

a graph-based data structure, which totals at 4.7 MB. The experiments were conducted on an 8-core Intel i7-9700k with 16.0 GB RAM. The execution time for each step takes 24 min 32 s for identifying dust objects, 13 min 12 s for tracking the linkages of objects over time, and 6 s for reconstructing dust events.

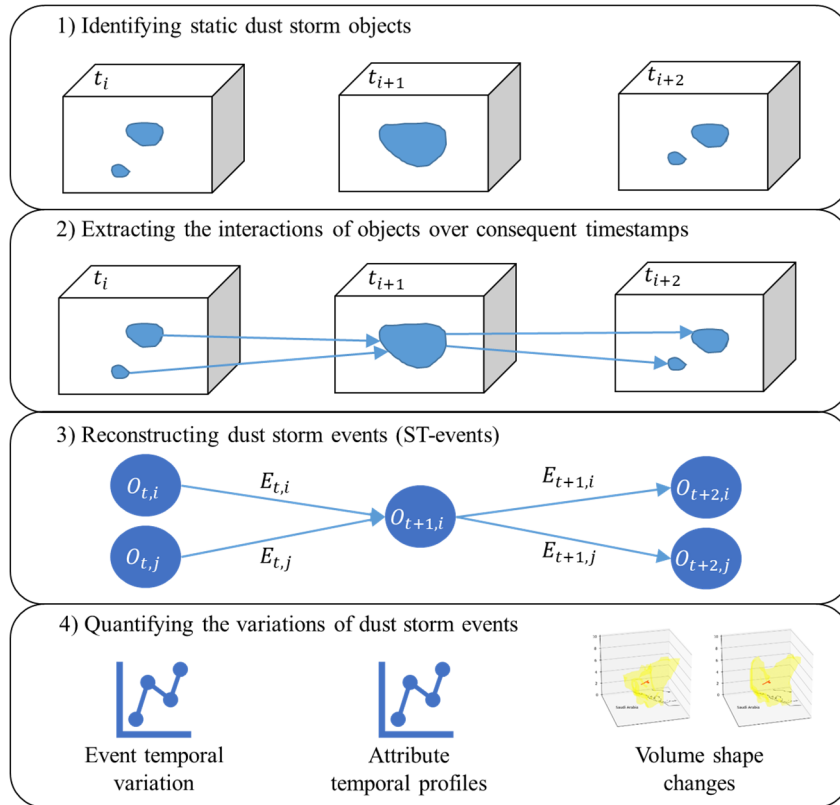


Figure 4. Overall procedure of dust event representation and quantification.

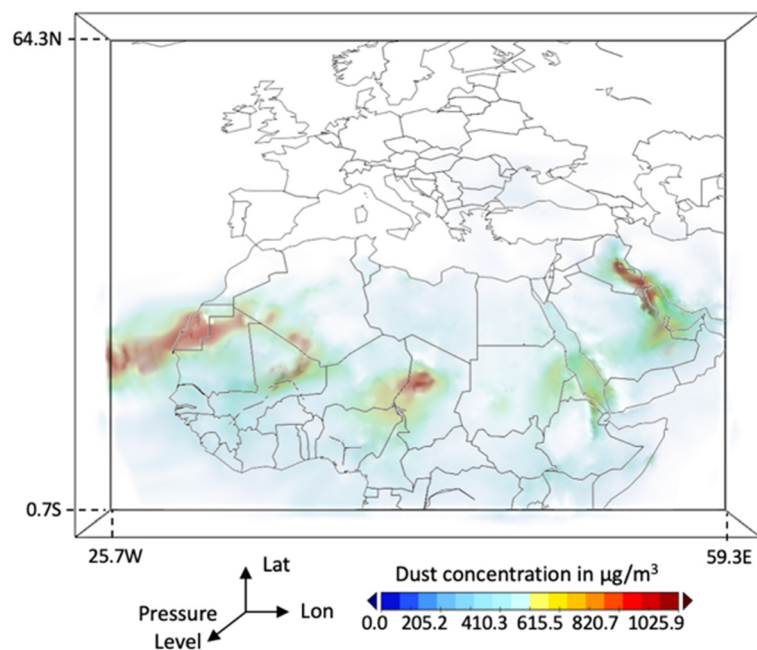


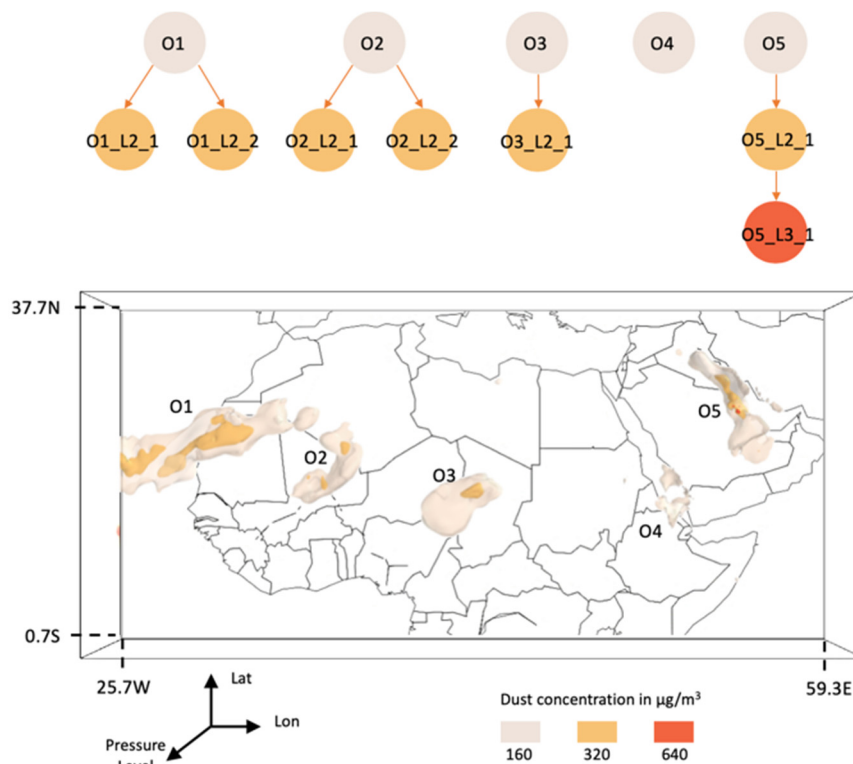
Figure 5. Dust concentration at a certain timestamp.



#### 4.2. Identifying Dust Objects

Dust objects at each timestamp are identified using a region-grow based algorithm [41], in which a dust object is defined as a “contiguous volume with dust concentration value larger than a threshold ( $D_{th}$ ), while its volume larger than a threshold ( $V_{th}$ ).” In this experiment, the dust concentration threshold is a set of 160, 320, 640  $\mu\text{g}/\text{m}^3$  (as intensity level 1, 2, and 3) and the volume threshold is 10 voxels. The multi-threshold of dust concentration is referred by the legend of dust concentration maps of model results from Barcelona Supercomputing Center (i.e., 20, 40, 80, 160, 320, 640, 1280, 2650  $\mu\text{g}/\text{m}^3$ ) (<https://ess.bsc.es/bsc-dust-daily-forecast>), and is selected based on the health impact indications of the surface dust concentration to vulnerable populations as negligible, marginal, and severe dust intensity [42]. Although the multi-thresholding approach used here and detailed in Yu and Yang [41] is designed particularly for dust object identification, it is flexible and can be easily adapted to the identification of other atmospheric objects.

Figure 6 demonstrates the identified five objects (O1~O5) at Intensity Level 1, and the objects contained inside them. For example, Object O1\_L2\_1 is identified at Intensity Level 2 and is contained inside Object O1; and Object O5\_L3\_1 is identified at Intensity Level 3 and is contained inside Object O5\_L2\_1, which is inside Object O5. The relationships—*Contains*—between objects identified at different intensity levels are computed through the identification algorithm [36].



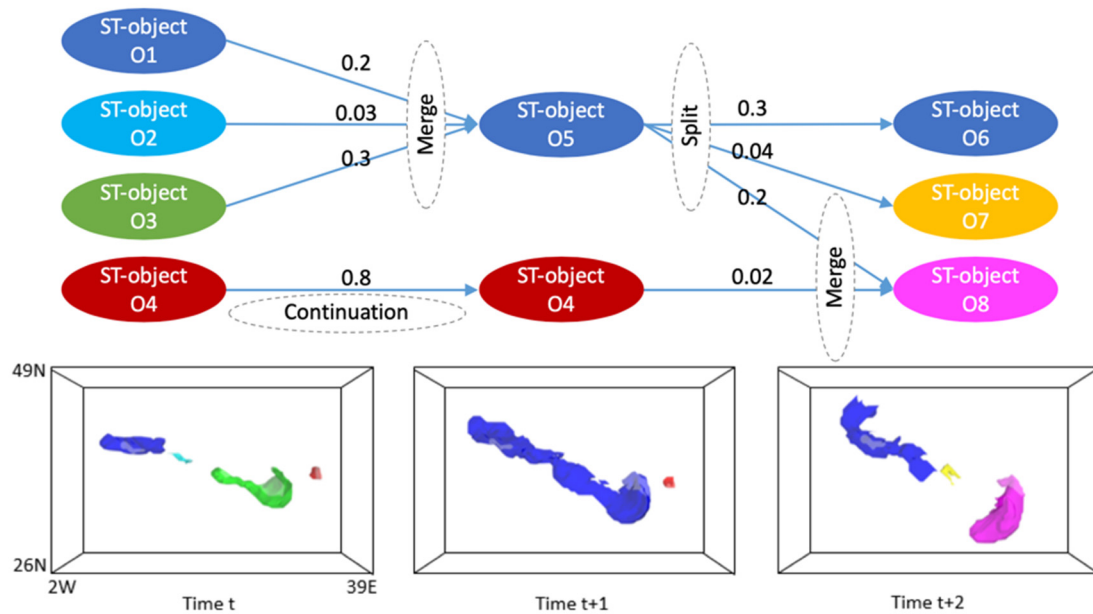
**Figure 6.** 3D visualization of dust concentration with different levels of hierarchy.

#### 4.3. Tracking the Linkages of Objects over Consequent Timestamps

The tracking algorithm is an overlap-based method, detailed in Yu et al. [43]. The identified dust objects from consecutive timestamps are compared regarding partial spatial overlaps to identify the potential linkages, and a best match is found from all possible combinations of objects from each pair of timestamps.

Based on the proposed framework, *ST-relations* are automatically identified and recorded from the tracking results. An appearance is considered as the start of a specific *ST-object*, and a disappearance is considered as the end of the object. In between, an *ST-object* is temporally linked by continuation,

expansion, or contraction, which are recognized as *ST-relations* within the same *ST-object*. Other linkages, including merging and splitting, are recognized as *ST-relations* between different *ST-objects*. Figure 7 demonstrates an example of three consecutive timestamps, where *ST-objects* O1, O2, and O3 merge into *ST-object* O5, and then O5 splits into *ST-objects* O6, O7, and O8; and *ST-object* O4 continues to exist and merged with *ST-object* O5 into *ST-object* O8.



**Figure 7.** Example of dust storm object continuing, merging, and splitting at consecutive timestamps.

#### 4.4. Reconstructing Dust Events

Dust events are then reconstructed by linking *ST-objects* that have *ST-relations* among each other. Each *ST-event* is represented by a graph connecting *ST-objects* with *ST-relations*, demonstrating how this event evolves and moves in space and time. The spatial and non-spatial attributes with each *ST-object* at a particular timestamp are associated with the nodes of the graph, and the weights representing spatial overlap are associated with the edges of the graph. Sample attributes are maintained in the graphs as attributes of the nodes, including centroid (i.e., the center of mass for a dust object), volume (i.e., the amount of cubic space inside of a dust object), and average concentration (i.e., the average dust concentration of all cubic space inside of a dust object). Object hierarchy subgraphs associated with individual *ST-objects* are integrated into the event graph.

In this experiment, there are 1796 *ST-events*/graphs reconstructed for the entire spatiotemporal range of the experiment data (Table 1). The total number of nodes (*ST-objects*) per graph ranges from 11 to 1959, with a mean value of 47. The total number of edges (*ST-relations*) per graph ranges from 10 to 1980, with a mean value of 46. For each event, the temporal variation between each time pair and the global variation are calculated. For variance calculation,  $\alpha$  and  $\beta$  are determined as independent or statistically different based on the univariate ANOVA test; thus  $\alpha$  and  $\beta$  are both set to be 1. The *GlobalVar* per graph ranges from 3.47 to 14,216.18, with a mean value of 363.75. The *ST-events* with high *GlobalVar* values might reflect significant evolutions within the event's lifecycle regarding the temporal changes in volume, intensity, and distance transported. However, an *ST-event* with a high *GlobalVar* value does not necessarily indicate a significant evolution, as the event might be long-lasting so that the accumulated *GlobalVar* is higher. An *ST-event* with a low *GlobalVar* value reflects that the event might be short-lived and has stable variations in volume, intensity, and distance transported.

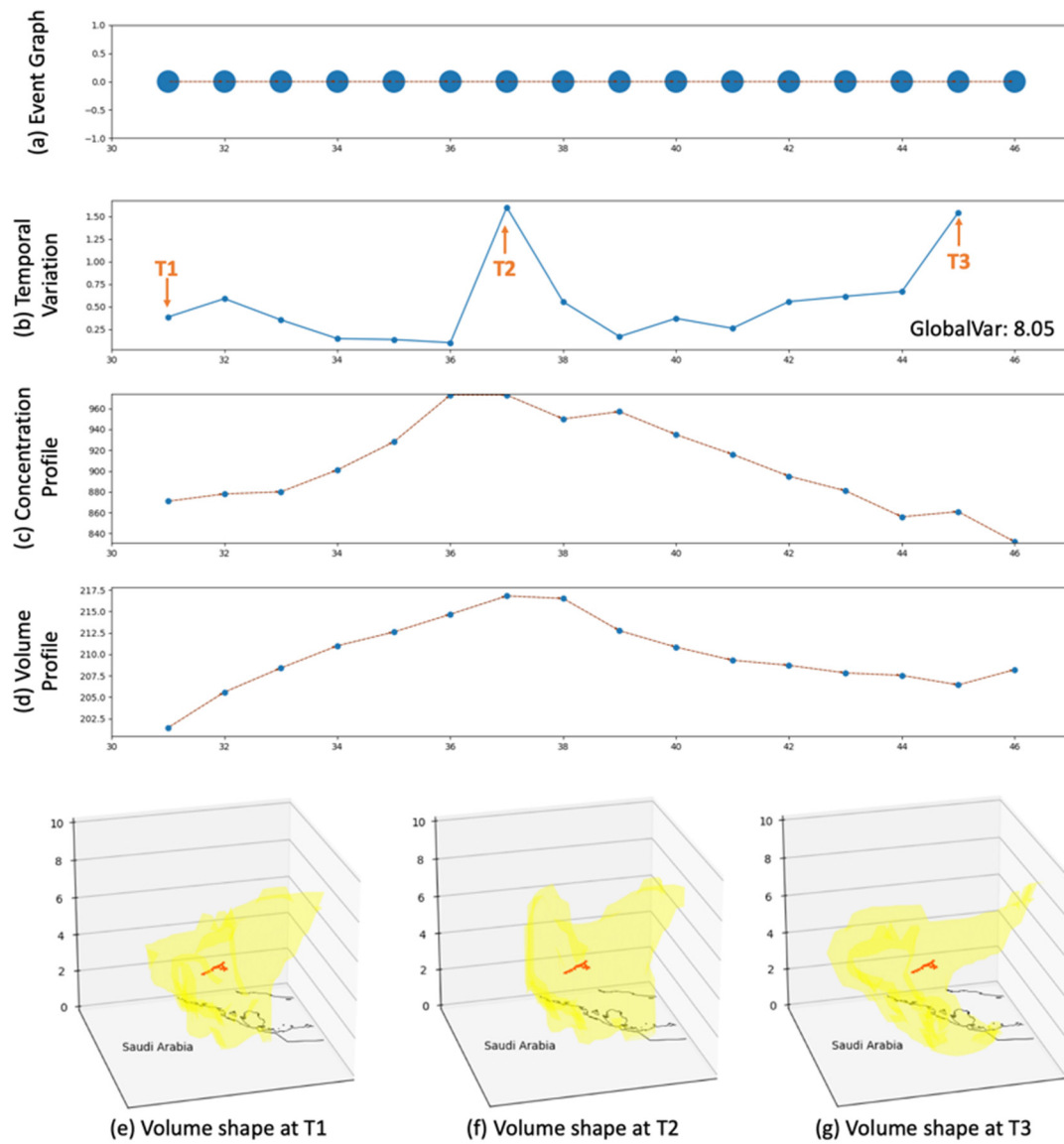
**Table 1.** Global graph statistics.

	Event Duration (in hrs)	Number of Nodes	Number of Edges	GlobalVar
<b>Min</b>	8	11	10	3.47
<b>Mean</b>	34	47	46	363.75
<b>Max</b>	844	1959	1980	14,216.18

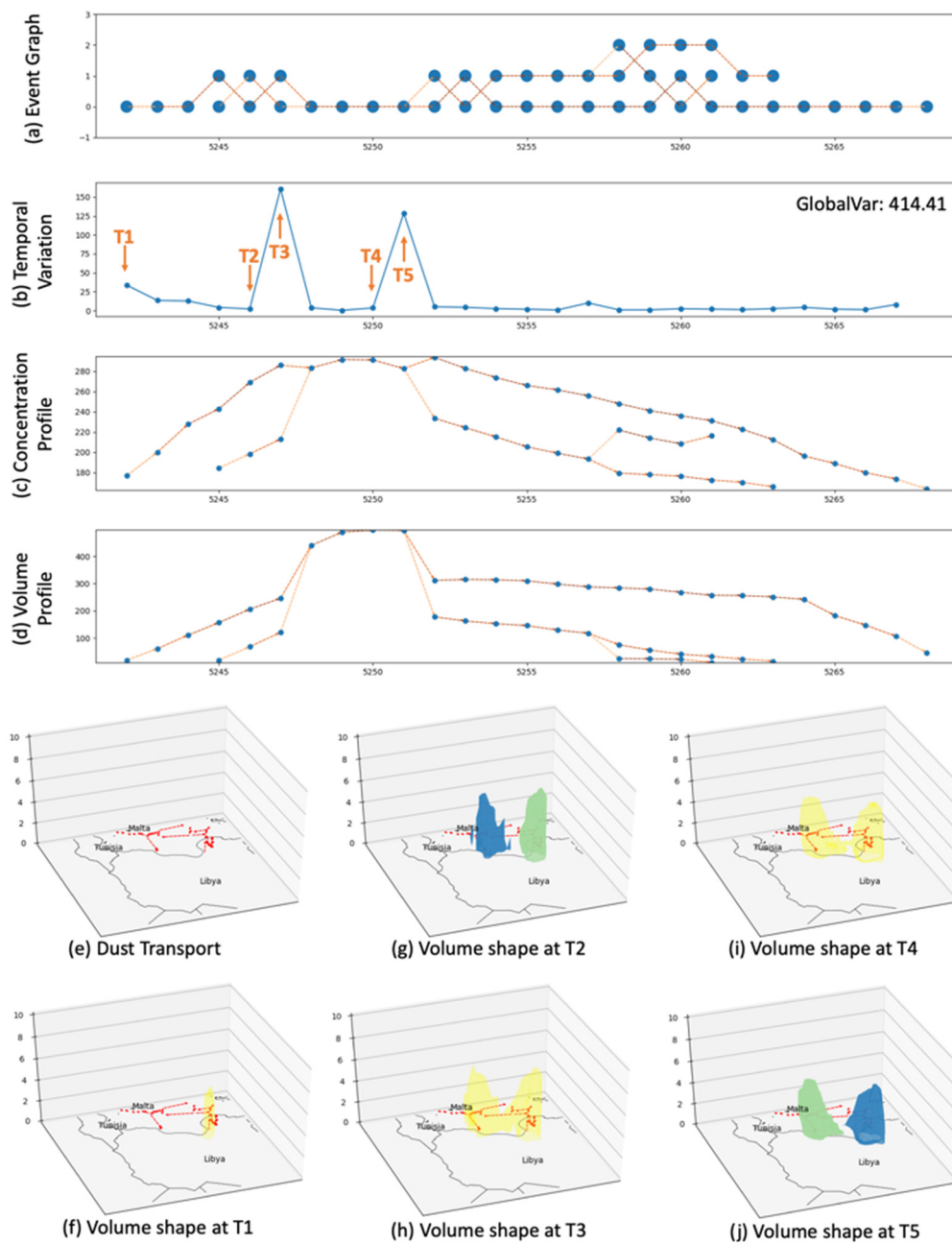
#### 4.5. Graph-Based Measurements of Dust Event Evolution Dynamics

In order to better illustrate the graph structure and content, we select three event graphs representing different levels of *GlobalVar* (low, medium, and high). The first event graph with a relatively low *GlobalVar* represents part of the Southern Arabian Peninsula dust event in July 2014 (<https://earthobservatory.nasa.gov/images/85370/persistent-dust-storms-on-the-southern-arabian-peninsula>). Within the dust event, dust with a high concentration transported from the interior of Sudan to the Red Sea, carried by the northwest winds to the southeast. As indicated by the event graph (Figure 8a), this *ST-event* consists of an *ST-object* with continuing types of *ST-relation*—continuation, expansion, and contraction. The temporal variation graph (Figure 8b) shows that this event evolves slowly in time with low variation values and a relatively low *GlobalVar* value. Both the dust concentration and the volume profile graphs (Figure 8c,d) describe the growing and decaying stages of this event. The dust transport (red trajectories in Figure 8e–g) shows that this event travels from northwest of Saudi Arabia towards central Saudi Arabia while the dust plume covered both Saudi Arabia and United Arab Emirates.

The second event graph represents the Libya dust event in early September 2014, where a cold and windy front moving across the Libya desert caused a haboob (<https://earthobservatory.nasa.gov/images/84014/dust-plume-over-the-red-sea>). A haboob is a type of intense dust event usually caused by strong winds of thunderstorm outflow. As indicated by the event graph (Figure 9a), this *ST-event* consists of seven *ST-objects* with different types of *ST-relations*—continuation, expansion, contraction, splitting, and merging. The temporal variation graph (Figure 9b) shows that this event has a relatively high *GlobalVar* value contributed majorly by two disturbingly high variation values at T3 and T5. The temporal profiles of dust concentration and volume (Figure 9c,d) reveal the evolution in time that two *ST-objects* merged into one, then split again and decayed gradually. Figure 8e demonstrated the overall dust transport pathway for the September 2014 Libya dust event, where dust originating from the Libyan desert (Figure 9f) was transported westward to Tunisia while picking up dust from other dust sources. The highlighted times are instances where the dust event evolves in a significant way, thus illustrating the changes in dust object volume and shape to further investigate at those times (Figure 9g–j). High variation values happen when dust objects both merge and split almost equally, which is shown in Figure 9c,d.



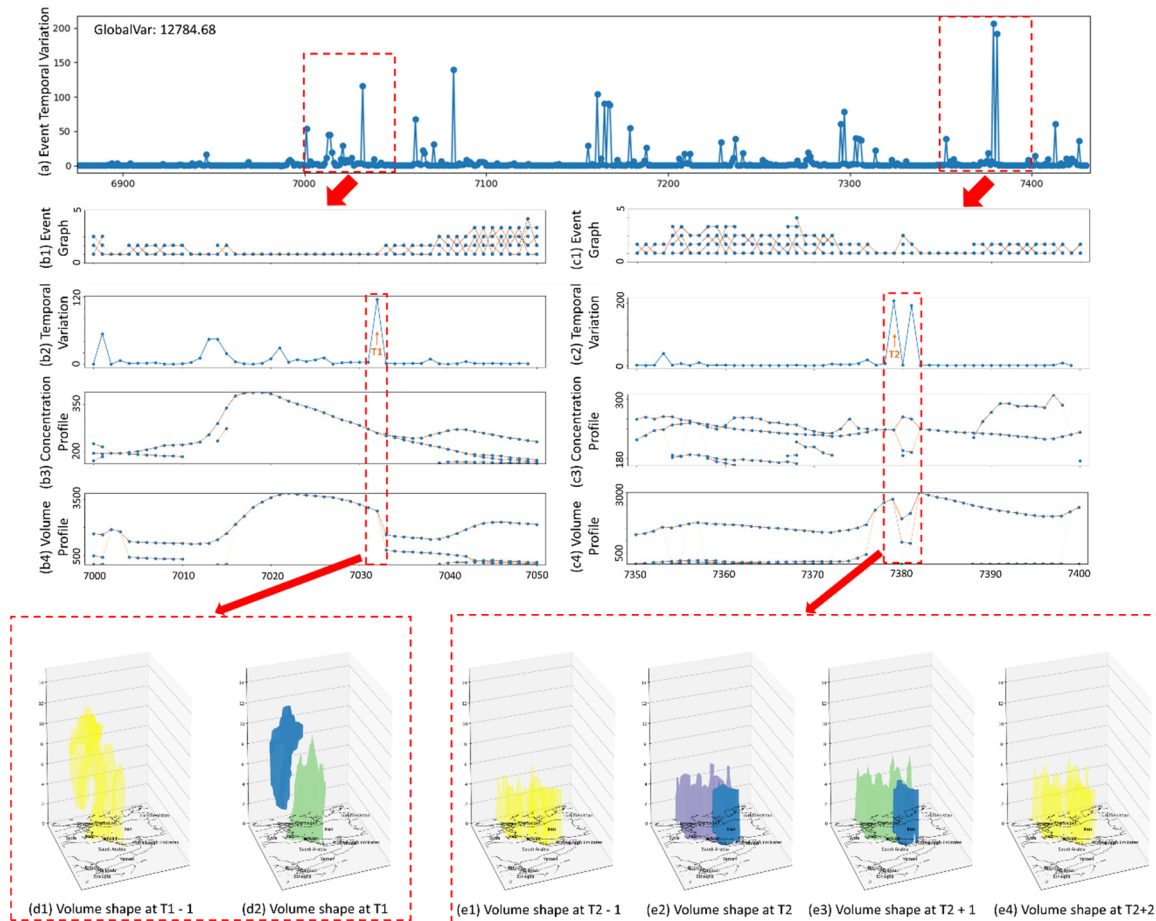
**Figure 8.** Event graph for the Southern Arabian Peninsula dust event in late July 2014. In Figure a, c, and d, each arrow is color-coded by the weight of the edge.



**Figure 9.** Event graph and temporal variation of the September 2014 Libya dust event.

The third event graph represents the Middle East dust event in September 2014, where a long-term dust storm with high dust concentration and volume. It combined the characteristics of a haboob and a shamal. A haboob is usually a short-lived but dramatic dust event appearing in a wall of dust sweeping across the ground, while a shamal is a pattern of persistent northwesterly winds in the Middle East region. When dust events are associated with shamal, they can persist over a wide spatial and temporal range. This specific event is not recorded in NASA Earth Observatory, but a similar one was recorded in the same time of 2015 (<https://earthobservatory.nasa.gov/images/86539/dust-marches-across-iraq-and-iran>). The event represented in Figure 10 demonstrates that this event last for 23 days and has the one of the highest *GlobalVar* values (within the experiment data) of 12784.68. The characteristics of a haboob in this event can be reflected in Figure 10(b1–b4), where dust has a relatively low concentration but a large volume. One of the first peaks in variation is highlighted as T1, where a high wall of dust splits into two objects with almost even volume. Similarly, from the highlighted time T2, it is observed that

the high variations occur when large volumes of dust split or merge. The characteristic of a shamal in this event is reflected by the overall movement of the dust transport from the northwest to the Persian Gulf, suggesting late-summer shamal winds. In addition, the long-lived dust event is sustained by the long-lasting shamal wind bringing up and maintaining dust particles in the atmosphere.



**Figure 10.** Event graph and temporal variation of the September 2014 dust event marching across Iraq and Iran. Figures d1, d2, e1–e4 highlight the volume and shape changes of the dust event and omit the event trajectory for the purpose of decluttered visualization.

## 5. Discussions

The proposed framework allows the description of multiple *ST-objects* aggregating into an *ST-event* when these *ST-objects* have *ST-relations* at any time. However, the framework is limited in that an *ST-object* cannot be associated with several *ST-events* simultaneously. The reason the proposed framework is designed in this way is that multiple *ST-objects* (happening at the same time and having *ST-relations* at any stage during their lifecycles) are considered to represent the state of an event at a particular point in time. *ST-events* can happen simultaneously, but these *ST-events* do not have *ST-relations* at any stage during their lifecycles.

The proposed framework represents a data-driven analytical framework for understanding the evolution of natural phenomena, instead of a model based on physical laws. The representation and quantification of natural phenomena over space and time can be used in various ways to assist the understanding of natural phenomena, thus integrating or complementing physical models with data-driven analytics. The proposed framework can exploit the spatiotemporal dependencies and dynamics of natural phenomena, including the geometry, context, and spread trends. Based on these spatiotemporal dependencies and dynamics produced by the proposed framework, long-term spatiotemporal statistics and machine learning models can be built to better capture spatiotemporal

patterns of a certain phenomenon. Integrating the data-driven analysis and physical models can potentially provide more accurate simulation and forecasting of the natural phenomena.

The proposed framework is applicable with minor adjustments to other weather events, including thunderstorms, jet streams, and ocean eddies. One of the examples applying the identification and tracking methods to satellite precipitation observations to generate rainfall events was reported in a previous article [43], and it is expected that temporal variations of these rainfall events can be further calculated using the proposed framework in this study.

## 6. Conclusions and Future Work

This research proposes a spatiotemporal data framework to represent the characteristics and dynamics of natural phenomena in an event-oriented way. The spatiotemporal data framework consists of three major entities: *ST-object*, *ST-relation*, and *ST-event*. Considering the non-rigid boundaries of natural phenomena, *ST-objects* are represented in a hierarchical way, where lower intensity *ST-objects* contains higher intensity ones, and intensity values can be flexible. AN *ST-event* is represented as a graph, where *ST-objects* and *ST-relations* that are associated with this event are nodes and edges. Based on the graph representation, the measurements of temporal variations and the global variation are proposed to quantify the dynamics within the event's lifecycle and the characteristic of the entire event. This graph-based quantification takes into account the following factors: spatial overlaps, object importance, distance change, and volume-intensity change; thus, it can represent a relatively objective measurement of the event variation. Such a holistic spatiotemporal data framework with event representation and evolution quantification can assist in the better understanding of the spatiotemporal dynamics of natural phenomena.

The proposed spatiotemporal data framework is applied to the one-year dust simulation dataset from the operational model maintained at the Barcelona Supercomputing Center. Dust objects are identified at each timestamp and the movements of dust events are tracked through consecutive timestamps. Dust events are then reconstructed and represented as event graphs. Different dust events along with the event graph, temporal variations, temporal profiles of dust volume and concentration, and the transport pathways are demonstrated and verified with the archived dust events at NASA Earth Observatory.

One of the future research directions is to integrate heterogenous databases for a multi-modal database to further assist the understanding of the relationship between natural phenomena with other physical and social factors, such as the relationship between dust events and disease outbreaks [44], and the relationship between Sahara dust events and hurricanes [45]. For example, to understand whether dust transport is a contributing factor to the outbreaks of infectious disease, we can establish a multi-modal database system that integrates a graph database for dust transport, a relational database for disease outbreak records, and a non-relational database for real-time environmental sensors for aerosols. As environmental big data evolves with real-time analytical demands, such a multi-modal database will be of significant benefit to scientific researchers and the industry.

The example framework presented in this research utilizes dust event identification and tracking as an example and leverages a standard desktop computing environment to handle the sample data. Another future direction resides in utilizing advanced computing capabilities, such as Graphics Processing Unit (GPU) computing, cloud computing, and big data management systems, to accelerate the process of identifying and tracking the evolution of natural phenomena and enable the real-time analytics when integrating the increasingly available environmental big data. Such event representation and evolution quantification framework can further be integrated into early warning systems to provide dynamic and customized early warning messages to vulnerable populations with specific needs.

**Acknowledgments:** The author thanks Matt Rice for proofreading the manuscript, and the anonymous reviewers for comments that greatly improved the manuscript.

**Conflicts of Interest:** The author declare no conflict of interest.

## References

1. Yang, C.; Goodchild, M.; Huang, Q.; Nebert, D.; Raskin, R.; Xu, Y.; Bambacus, M.; Fay, D. Spatial cloud computing: How can the geospatial sciences use and help shape cloud computing? *Int. J. Digit. Earth* **2011**, *4*, 305–329. [[CrossRef](#)]
2. Yuan, M. Representing complex geographic phenomena in GIS. *Cartogr. Geogr. Inf. Sci.* **2001**, *28*, 83–96. [[CrossRef](#)]
3. Hubbard, S.; Hornsby, K.S. Modeling alternative sequences of events in dynamic geographic domains. *Trans. GIS* **2011**, *15*, 557–575. [[CrossRef](#)]
4. Yi, J.; Du, Y.; Liang, F.; Zhou, C.; Wu, D.; Mo, Y. A representation framework for studying spatiotemporal changes and interactions of dynamic geographic phenomena. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 1010–1027. [[CrossRef](#)]
5. Arenas, H.; Harbelot, B.; Cruz, C. LC3: A Spatiotemporal Data Model to Study Qualified Land Cover Changes. In *Land Use and Land Cover Semantics*; Ahlqvist, O., Varanka, D., Fritz, S., Janowicz, K., Eds.; CRC Press: Boca Raton, FL, USA, 2018; pp. 228–259.
6. Lu, M.; Pebesma, E.; Sanchez, A.; Verbesselt, J. Spatio-temporal change detection from multidimensional arrays: Detecting deforestation from MODIS time series. *ISPRS J. Photogramm. Remote Sens.* **2016**, *117*, 227–236. [[CrossRef](#)]
7. Shen, D.; Takara, K.; Liu, Y. 3D GIS Modeling of Soft Geo-Objects: Taking Rainfall, Overland Flow, and Soil Erosion as an Example. *Geospat. Technol. Environ. Soc. Appl.* **2016**. [[CrossRef](#)]
8. Hofmann, P. A Fuzzy Belief-Desire-Intention Model for Agent-Based Image Analysis. *Mod. Fuzzy Control Syst. Appl.* **2017**, 281–296.
9. Chen, J.; Cohn, A.G.; Liu, D.; Wang, S.; Ouyang, J.; Yu, Q. A survey of qualitative spatial representations. *Knowl. Eng. Rev.* **2015**, *30*, 106–136. [[CrossRef](#)]
10. McIntosh, J.; Yuan, M. Assessing similarity of geographic processes and events. *Trans. GIS* **2005**, *9*, 223–245. [[CrossRef](#)]
11. Campelo, C.E.; Bennett, B. Representing and reasoning about changing spatial extensions of geographic features. In Proceedings of the International Conference on Spatial Information Theory, Scarborough, UK, 2–6 September 2013; Tenbrink, T., Stell, J., Galton, A., Wood, Z., Eds.; Springer: Cham, Switzerland, 2013; pp. 33–52.
12. Erhardt, T.M.; Czado, C.; Schepsmeier, U. R-vine models for spatial time series with an application to daily mean temperature. *Biometrics* **2015**, *71*, 323–332. [[CrossRef](#)]
13. Anselin, L. *Spatial Econometrics: Methods and Models*; Springer: Dordrecht, The Netherlands, 2013.
14. Bothwell, J.; Yuan, M. Apply concepts of fluid kinematics to represent continuous space–time fields in temporal GIS. *Ann. GIS* **2010**, *16*, 27–41. [[CrossRef](#)]
15. Armstrong, M.P. Temporality in spatial databases. In Proceedings of the GIS/LIS'88: Accessing the World, American Society for Photogrammetry and Remote Sensing, Falls Church, VA, USA, 30 November–2 December 1988; pp. 880–889.
16. Langran, G.; Chrisman, N.R. A framework for temporal geographic information. *Cartographica* **1988**, *25*, 1–14. [[CrossRef](#)]
17. Worboys, M.F.; Hearnshaw, H.M.; Maguire, D.J. Object-oriented data modelling for spatial databases. *Int. J. Geogr. Inf. Syst.* **1990**, *4*, 369–383. [[CrossRef](#)]
18. Khaddaj, S.; Adamu, A.; Morad, M. Construction of an integrated object oriented system for temporal GIS. *Am. J. Appl. Sci.* **2005**, *2*, 1584–1594. [[CrossRef](#)]
19. Yuan, M. Use of a Three-Domain Representation to Enhance GIS Support for Complex Spatiotemporal Queries. *Trans. GIS* **1999**, *3*, 137–159. [[CrossRef](#)]
20. Peuquet, D.J.; Duan, N. An event-based spatiotemporal data model (ESTDM) for temporal analysis of geographical data. *Int. J. Geogr. Inf. Syst.* **1995**, *9*, 7–24. [[CrossRef](#)]
21. Galton, A.; Worboys, M. Processes and events in dynamic geo-networks. In Proceedings of the International Conference on GeoSpatial Semantics, Mexico City, Mexico, 29–30 November 2005; Springer: Berlin/Heidelberg, Germany, 2005.
22. Renolen, A. History graphs: Conceptual modeling of spatio-temporal data. *GIS Front. Bus. Sci.* **1996**, *2*, 46.



23. Del Mondo, G.; Rodríguez, M.A.; Claramunt, C.; Bravo, L.; Thibaud, R. Modeling consistency of spatio-temporal graphs. *Data Knowl. Eng.* **2013**, *84*, 59–80. [[CrossRef](#)]
24. Siabato, W.; Claramunt, C.; Ilarri, S.; Manso-Callejo, M.Á. A Survey of Modelling Trends in Temporal GIS. *ACM Comput. Surv. (CSUR)* **2018**, *51*, 30. [[CrossRef](#)]
25. Goodchild, M.F.; Yuan, M.; Cova, T.J. Towards a general theory of geographic representation in GIS. *Int. J. Geogr. Inf. Sci.* **2007**, *21*, 239–260. [[CrossRef](#)]
26. Schneider, M. Spatial Plateau Algebra for implementing fuzzy spatial objects in databases and GIS: Spatial plateau data types and operations. *Appl. Soft Comput.* **2014**, *16*, 148–170. [[CrossRef](#)]
27. Cheng, H. Modeling and querying fuzzy spatiotemporal objects. *J. Intell. Fuzzy Syst.* **2016**, *31*, 2851–2858. [[CrossRef](#)]
28. de Caluwe, R.; De Tré, G.; Bordogna, G. (Eds.) *Spatio-Temporal Databases: Flexible Querying and Reasoning*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013.
29. Shukla, P.K.; Darbari, M.; Singh, V.K.; Tripathi, S.P. A survey of fuzzy techniques in object oriented databases. *Int. J. Sci. Eng. Res.* **2011**, *2*, 1–11.
30. Carniel, A.C.; Schneider, M.; Ciferri, R.R.; de Aguiar Ciferri, C.D. Modeling fuzzy topological predicates for fuzzy regions. In Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Dallas, TX, USA, 4 November 2014; pp. 529–532.
31. Worboys, M. Event-oriented approaches to geographic phenomena. *Int. J. Geogr. Inf. Sci.* **2005**, *19*, 1–28. [[CrossRef](#)]
32. Yuan, M.; Hornsby, K.S. *Computation and Visualization for Understanding Dynamics in Geographic Domains: A Research Agenda*; CRC Press: Boca Raton, FL, USA, 2007.
33. Guttler, F.; Ienco, D.; Nin, J.; Teisseire, M.; Poncelet, P. A graph-based approach to detect spatiotemporal dynamics in satellite image time series. *ISPRS J. Photogram. Remote Sens.* **2017**, *130*, 92–107. [[CrossRef](#)]
34. Long, J.A.; Nelson, T.A. A review of quantitative methods for movement data. *Int. J. Geogr. Inf. Sci.* **2013**, *27*, 292–318. [[CrossRef](#)]
35. An, L.; Tsou, M.H.; Crook, S.E.; Chun, Y.; Spitzberg, B.; Gawron, J.M.; Gupta, D.K. Space–time analysis: Concepts, quantitative methods, and future directions. *Ann. Assoc. Am. Geograph.* **2015**, *105*, 891–914. [[CrossRef](#)]
36. Kulldorff, M. Prospective time periodic geographical disease surveillance using a scan statistic. *J. R. Stat. Soc. Ser. A (Stat. Soc.)* **2001**, *164*, 61–72. [[CrossRef](#)]
37. Kulldorff, M.; Heffernan, R.; Hartman, J.; Assunção, R.; Mostashari, F. A space–time permutation scan statistic for disease outbreak detection. *PLoS Med.* **2005**, *2*, e59. [[CrossRef](#)]
38. Robertson, C.; Nelson, T.A.; Boots, B.; Wulder, M.A. STAMP: Spatial–temporal analysis of moving polygons. *J. Geogr. Syst.* **2007**, *9*, 207–227. [[CrossRef](#)]
39. Allan, J.; Carbonell, J.G.; Doddington, G.; Yamron, J.; Yang, Y. Topic detection and tracking pilot study final report. In Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop, Lansdowne, VA, USA, 8–11 February 1998.
40. Basart, S.; Pérez, C.; Nickovic, S.; Cuevas, E.; Baldasano, J. Development and evaluation of the BSC-DREAM8b dust regional model over Northern Africa, the Mediterranean and the Middle East. *Tellus B Chem. Phys. Meteorol.* **2012**, *64*, 18539. [[CrossRef](#)]
41. Yu, M.; Yang, C. A 3D multi-threshold, region-growing algorithm for identifying dust storm features from model simulations. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 939–961. [[CrossRef](#)]
42. De Longueville, F.; Ozer, P.; Doumbia, S.; Henry, S. Desert dust impacts on human health: An alarming worldwide reality and a need for studies in West Africa. *Int. J. Biometeorol.* **2013**, *57*, 1–19. [[CrossRef](#)] [[PubMed](#)]
43. Yu, M.; Yang, C.; Jin, B. A framework for natural phenomena movement tracking—Using 4D dust simulation as an example. *Comput. Geosci.* **2018**, *121*, 53–66. [[CrossRef](#)]
44. Griffin, D.W.; Kellogg, C.A. Dust storms and their impact on ocean and human health: Dust in Earth’s atmosphere. *EcoHealth* **2004**, *1*, 284–295. [[CrossRef](#)]
45. Evan, A.T.; Dunion, J.; Foley, J.A.; Heidinger, A.K.; Velden, C.S. New evidence for a relationship between Atlantic tropical cyclone activity and African dust outbreaks. *Geophys. Res. Lett.* **2006**, *33*. [[CrossRef](#)]

