Topology-based Analysis for Multimodal Atmospheric Data of Volcano Eruptions

Alexander Kuhn, Wito Engelke, Markus Flatken, Hans-Christian Hege, Ingrid Hotz

Abstract Many scientific applications deal with data from a multitude of different sources, e.g., measurements, imaging and simulations. Each source provides an additional perspective on the phenomenon of interest, but also comes with specific limitations, e.g. regarding accuracy, spatial and temporal availability. Effectively combining and analyzing such *multimodal* and *partially incomplete* data *of limited accuracy* in an *integrated* way is challenging. In this work, we outline an approach for an integrated analysis and visualization of the atmospheric impact of volcano eruptions. The data sets comprise observation and imaging data from satellites as well as results from numerical particle simulations. To analyze the clouds from the volcano eruption in the spatiotemporal domain we apply topological methods. Extremal structures reveal structures in the data that support clustering and comparison. We further discuss the robustness of those methods with respect to different properties of the data and different parameter setups. Finally we outline open challenges for the effective integrated visualization using topological methods.

1 Introduction

The analysis of atmospheric gas and particle traces, such as of SO_2 or ash particles, is of fundamental importance for a better understanding of global atmospheric processes. Specifically volcano eruption events can produce massive amounts of such 'tracers' over a short time period. These substances can have severe global impact and may trigger complex atmospheric interactions. To better understand those processes an increasing number of modalities (including measurements and simulations) is utilized. One major challenge is to extract and combine the essential information, which is spread over various, mostly (w.r.t. space and time) sparse data sources. This requires a careful integration of information from each modality, considering its valuableness and limitations. For instance, sparse but particularly reliable measurement data can be used to calibrate the simulations producing denser

1

data, while simulation data provide a means to interpolate measurement data and to fill spatial and temporal gaps. The increased amount of information available today provides more complete representations of the physical phenomena, but increases the complexity of the analysis; hence a main objective are effective methods for data filtering, information reduction and abstraction. In this work we address those challenges by applying topology-based methods. Using the example of ash clouds we show that topological data analysis can serve as an effective tool to address essential analysis tasks. We focus on the extraction of extremal graphs as means for feature-oriented data reduction and as basis for visual comparison of the data from different sources. Spatio-temporal clustering in the space-time domain is used for a visual representation of the evolution of aforementioned atmospheric events. Hence, our main contributions are:

- Integration of spatially and/or temporally sparse data into a space-time domain
- Topological analysis and visualization of features in the common domain
- Topology-based spatio-temporal segmentation of features in the space-time domain, and fused visualization thereof.

The specific application considered is the analysis of pollutant clouds that emerged from volcanic eruptions. We are using data provided for the 2014 IEEE Visualization Contest [**?**]. The work is a follow-up of a contest contribution [**?**] that focuses on topological aspects and deepens them.

2 Related Work

Climate research is a data-intensive field. Depending on the specific application the data comes from various sources, e.g. large-scale simulations or observations from satellites. Accordingly, understanding and analyzing the data plays an important role. While visualization is used on an everyday basis, it is mostly limited to simple methods such as diagrams, statistical plots, color plots in a geographical context. More advanced visualization tools are often not well known [**?**]. A particularly important topic is cloud evolution, which requires identification and tracking of clouds. Many methods have been proposed for this purpose, but they are often very complicated and only suitable for a specific application. For example Kober et al. follow a multi-scale approach based on image pixel displacement to track thunderstorm clouds from satellite data [**?**]. Extraction of contours and isosurfaces as well as tracking their change over time is part of multiple other approaches. Gambheer et al. track clouds by considering the overlap of sub-level sets across time steps [**?**]. Isosurface cloud-tracking in VR environments was introduced by Griffith et al. [**?**]. An example for cumulus cloud tracking is based on region growing followed by a space time segmentation [**?**].

Many of the ideas used in these papers fit well into the framework of *topological* feature identification and tracking. A perspective that seems not yet to be present in meteorology and climate research. For instance, topological methods can be very powerful regarding abstraction, simplification and comparison of features as clouds.

This has been shown in a recent paper, where cloud systems are tracked by employing topological segmentation for the identification of clouds and an optical flow method for sub-scale motion tracking [**?**]. Topological data analysis plays an increasing role in the field of visualization with many different applications. The full topological information of a scalar field is given by its Morse decomposition or topological graph. There are different strategies available to extract the topological graph from sampled scalar fields. The most common approaches for its computation go back either to Edelsbrunner [**?**] (using a piecewise linear interpolation) or to Forman [**?**] (who proposed a discrete Morse theory). Nowadays, efficient algorithms exist implementing these approaches [**?, ?, ?, ?**]. In many applications it is not necessary to compute the complete topological graph. A reduced structure keeping track only of changes in the number of components is the contour tree [**?**]. A structure that frequently represents features of interest is the extremal graph, which focuses on maxima and ridge lines [**?**]. Overviews on topological approaches in the context of vector fields are presented by Laramee et al. [**?**] and Pobitzer et al. [**?**]. Also tracking of topological structures over time has been considered. The proposed methods can roughly be categorized according to the geometric and topological criteria they are using to establish correspondences between features at successive time points. Geometric methods considering overlap of regions have been used for the tracking of burning cells [**?, ?**]. An example for topological tracking is mapping of critical points across time steps defining a feature flow field [**?**]. Other approaches are based on Jacobi sets [**?**].

3 Description of the Application and Data Sources

The aim is to analyze the atmospheric impact of volcanic eruptions covered by several measurement modalities and simulations during the time from 30.05.2011 until 07.09.2011. In the focus of the presented analysis are three dominant volcanic events that have been recorded during this period: The eruptions of the volcanos Puyehue-Cordón Caulle (Ranco Province, Chile), Nabro (Red Sea Region, Eritrea), and Grimsvötn (South-East Iceland). The eruptions can be characterized by considering atmospheric tracers, like ash particles and SO_2 gas concentrations. The tracers form 'plumes', whose spatio-temporal evolution can be observed in all modalities. An overview of the temporal availability of the modalities is shown in Figure 1.

All data sources are inherently time-dependent and have a common geographic reference grid (longitude, latitude, and optionally height). The presented methods address the following questions that have been raised by the domain scientists [**?**]:

1. What are suitable methods to combine and relate the given modalities into a common reference space (integrated domain)? See Section 4.1.
2. How to reduce the amount of data to focus on the phenomena of interest (i.e., events related to volcano eruptions)? See Section 4.2.
3. How to integrate sparse data sources, derive compact feature-oriented visualizations, and how to associate this information to specific events? See Section 5.
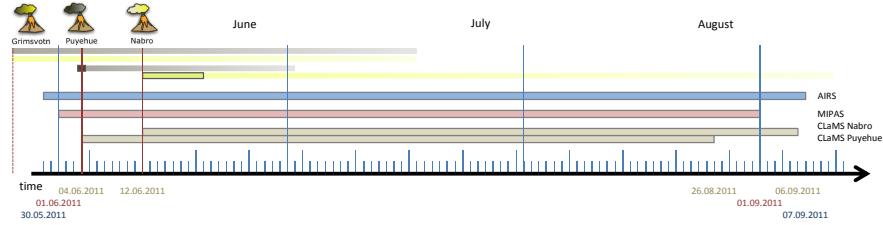
Fig. 1 Overview over the temporal availability of the various data: The three eruption events and the time span of the respective measurements or simulations are displayed. The yellow and gray bars indicate the 'lifetime' of the ejected sulfur, respectively ash particles.
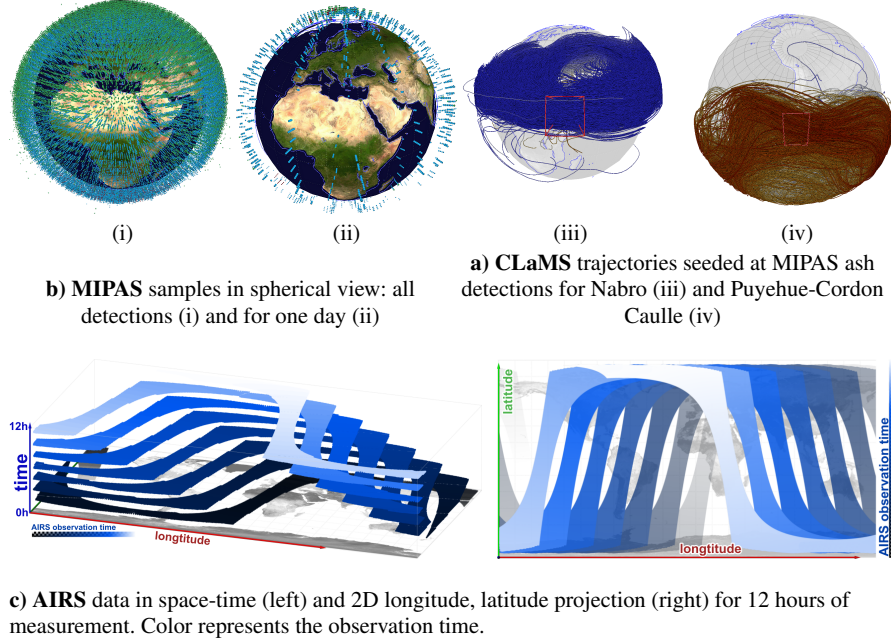
3.1 Input Data Sources

The first type of data was obtained using the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) [?]. MIPAS measurements provide vertical sampling profiles at altitudes between 5-70 km, with approximately 14 orbits per day. The measuring technique is highly sensitive towards aerosol tracers and offers a good vertical resolution. The data is stored as single trajectory, containing the sampling points (longitude, latitude, altitude, and time) and values for different events (clear sky, ice detection, ash, and sulfate aerosol detection) [?]. Note that each event is represented as a binary value that indicates whether a predefined threshold has been exceeded. All sampling points ($\sim$ 1.3 Million points at 48 MB, starting from 01.06.2011 until 01.09.2011) of the MIPAS data set are shown in Figure 2 a).

The second data source are simulated trajectories using the Chemical Lagrangian Model of the Stratosphere (CLaMS) developed at the Institute for Energy and Climate Research, RWTH Aachen Univ. [?, ?]. This is an hierarchical model to describe the global chemical transport and contains a selected subset of pre-integrated trajectories, seeded at MIPAS detections. These are sulfur detections on the northern hemisphere for Grimsvötn and Nabro ($\sim$ 55.000 trajectories at 1GB) and MIPAS ash detections on the southern hemisphere for Puyehue-Cordón Caulle ($\sim$ 5.800 trajectories at 62MB). The input boundary conditions for the simulations are based on ERA interim data [?]. CLaMS trajectories are characterized by an high spatial and temporal resolution and carry information about physical scalars (e.g., pressure, temperature, potential vorticity). However, since numerical simulation can only approximate the real world phenomenon, the reliability of the trajectories decreases with their temporal distance to the seeding point. The trajectories are shown in Figure 2 b).

The third data source are measurements from the Atmospheric Infrared Sounder (AIRS), acquired by the NASA Aqua satellite. It measures thermal emissions in the atmosphere [?]. The satellite scans horizontal cross-sections of the atmosphere at very high resolutions and performs 14.5 orbits per day. Individual scanning samples are organized on a high-resolution quad-strip, which has been cut into 200 segments describing 12 hours of measurement with $\sim$1.4 Million quads at 95 MB each (in

total 19 GB). Every segment provides almost global coverage (i.e., temporal delay produces gaps between neighboring strips) and provides index information about SO_2 and ash concentrations [**?**]. Note that the index summarizes atmospheric information of a vertical column, hence height information is lost during acquisition. The original data acquired at a 12 hour interval is shown at Figure 2 c).

(i)   (ii)   (iii)   (iv)

a) CLaMS trajectories seeded at MIPAS ash detections for Nabro (iii) and Puyehue-Cordon Caulle (iv)

b) MIPAS samples in spherical view: all detections (i) and for one day (ii)

c) AIRS data in space-time (left) and 2D longitude, latitude projection (right) for 12 hours of measurement. Color represents the observation time.

Fig. 2 Overview of the available data sets: a) MIPAS satellite measurements, b) CLaMS simulated trajectories, and c) AIRS satellite measurements.

4 Analysis of the Common Reference Domain

To combine and analyze the different data sources we proceed as follows: The basis is a projection of all individual data sets into one *common space-time domain*; second, *extremal structures* for both major measurement sources are extracted for visual comparison; finally, a *space-time segmentation* method is applied for a *combined visualization* integrating standard methods as isosurfaces, trajectory rendering for a detailed visual analysis.

4.1 Construction of the Common Reference Domain

In a first step, we sample the given data sets (Section 3.1) into a common reference domain. It is constructed as discrete regularly-sampled 4D domain with the dimen-

sions longitude, latitude, altitude, and time. For visualization purposes, we mainly refer to the 3D subspace consisting of longitude, latitude, and time. Thereby, the measurement data builds the core of the new data set. Each source adds specific information to this domain according to the characteristic of the respective modality, e.g., SO_2 or ash concentrations, number of detection events. To increase the spatial and temporal coverage we interpolate the data on basis of the simulated CLaMS trajectories. In detail we introduce two filtering procedures:

1. **Gaussian filtering of the raw data:** The goal of this step is to assign a small volume to the point measurements and to the one-dimensional trajectories. This is achieved using a Gaussian convolution kernel, which adds an *isotropic footprint* with decreasing intensity to the samples in space *and/or* time. The size of the kernel is chosen very small in the order of a couple of gird cells. After this filtering step the field is still sparse and does not cover the entire domain.

2. **Spatio-temporal interpolation using trajectories:** The CLaMS particle simulation provides the necessary information for a realistic interpolation of the sparsely filled domain. It is assumed that the detected particles roughly follow these trajectories. This leads to a convolution of the measurements along the CLaMS trajectories with a *non-linear* and *anisotropic* spatio-temporal footprint. Technically, we apply an advection of the measurements along the trajectories. The length of the chosen trajectory segment τ and the decay of the signal along the trajectory are parameters of the method. In the following we assume a linear decay to zero within an interval $[+\tau, -\tau]$. For small values of τ the accuracy of the trajectories is high enough to obtain a reliable approximation of the particle distribution.

If not mentioned differently, we use a sampling resolution of $720 \times 360 \times 800$ cells for the reference domain. Detailed views for sub-spaces of higher resolution or dimension may be extracted and analyzed using the same methodology. In the following the data specific projections are described in more detail.

MIPAS Data Processing: The MIPAS samples are point samples that only carry a flag indicating a detection. The first step to integrate the MIPAS data is to count the number of detections per grid cell. Next, we apply at first the isotropic Gaussian filter and then the anisotropic interpolation filter. For the Figures a kernel size of 5 cells and $\sigma = 0.4$ is used. A temporal reasonable range of τ is in the order of 48 hours (see [**?**, **?**] for details). This result of the MIPAS integration is a scalar density field describing the spatial and temporal distribution of the particles.

AIRS Data Processing: AIRS data has a high spatial but low temporal resolution due to the orbiting of the measuring satellite. There is no altitude information attached to these measurements. The temporal gaps between two measurements are in the order of 12 hours. A Gaussian filter in temporal direction provides a simple but not very accurate solution to bridge the gaps. A more realistic result is obtained

by interpolating the data using the CLaMS data as described above.

CLaMS Data Processing: For the particle simulation trajectories are seeded at relevant MIPAS detections during the Puyehue-Cordón Caulle and Nabro eruptions. The detections are pre-filtered by domain experts before the simulation [**?**]. Thus, the density of the trajectories depends on the number of MIPAS samples and they are only available at irregularly distributed locations. Its reliability is decreasing with the simulation time. Therefor we use the CLaMS data mainly as basis for the interpolation of the satellite observation-data via advection.

To speed up the advection procedure along the trajectories, we generate an auxiliary time-dependent vector field from the CLaMS trajectories. Therefor, we compute a weighted average of the velocity information within each cell. The velocity is given as the tangent directions of the trajectories passing through that cell. The weight accounts for the decreasing reliability along the trajectories. Here we use a linear decay along the trajectory. Note that the resulting field is not an approximation of the meteorological wind field used for the simulation, since the simulation also accounts for chemical reactions in the atmosphere [**?**, **?**].

4.2 Topological Analysis in the Common Reference Domain

Integration of all data into the common space-time domain yields fields that characterize the spatio-temporal distribution of ash and sulfur. In our framework, topological techniques are applied to analyze and visualize these fields. The benefit of topological analysis is an abstraction that yields explicit geometrical structures. These are graphs, segmented volumes, or surfaces characterizing the underlying field. The applied methods provide also access to topological simplification and filtering. The topological features can be used to link and compare the given modalities against each other, e.g., reference ash against SO_2, or advected MIPAS against AIRS. Further they can be associated to observable phenomena, like specific volcano eruptions, at an abstract feature-based level. In the following we explore possibilities to characterize physical features, as eruption plumes, using *extremal structures* in the respective field. For an algorithmic outline to derive *approximate* topological structures similar to topological spines [**?**, **?**] and segmentations of a scalar field $s(\mathbf{x})$, see Algorithm 1.

Extremum Graph Extraction An approximate extrema graph [**?**] is used to describe the spatial structure of the particle distribution for every time slice. Extrema graphs represent the target phenomena, in our case ash and SO_2 plumes, as spatially embedded graphs. They connect saddle points with extremal points (see Alg. 1) according to the approximate Morse-Smale complex of the function s. Note that in this step a persistence based filtering is not necessary since the construction of the common reference domain already comprises a smoothing and down-sampling step. Instead we define a minimum threshold h_{min} filtering out parts of the graph below this value. The graph is used in three ways: Visualization – The graph describes the spatial connectivity of extremal events in each time slice. It visually captures advec-

Data: Regular scalar field $s(\mathbf{x})$
Result: Segmentation by hierarchical labels, topological graph
1) Init union-find data structure $UF(G_j)$,
where each group G_j has a label j and stores a set of grid nodes
2) Sort all sample points, based on its value into sorted list L
(ascending: maxima, descending: minima features)
while *L contains grid nodes g_i with values larger than threshold h_{min}* **do**
    classify current grid node g_i:
    if *g_i is isolated extrema* **then**
        add new group to UF
    else if *g_i is adjacent to a single group G_j* **then**
        add g_i to G_j
    else if *g_i is adjacent to multiple groups (saddle point)* **then**
        process node g_i: construct graph edge (see Sec. 4.2) or
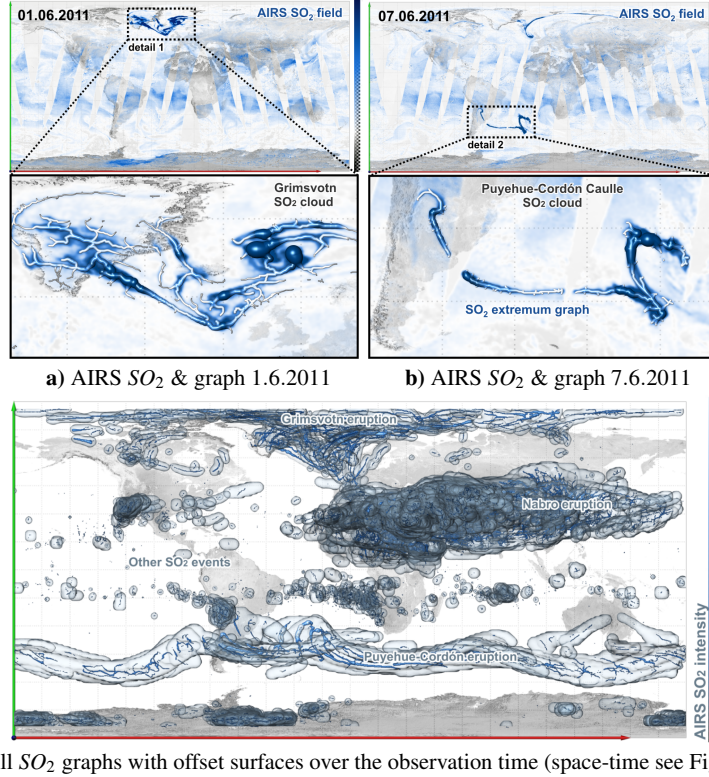   merge two groups (see Sec. 4.2)
end

Algorithm 1: Approximated topological feature algorithm

tion patterns that are eminent in temporal snapshots in AIRS or advected MIPAS data (see Figure 6). It also serves as basis for visual comparison of the fields from the different modalities as illustrated in Figure 3. Filtering – The graphs are further used as input for filtering and for measurements. This allows to focus on specific topological events. It further supports feature-based filtering with respect to size or the location of the plume. Space-time segmentation – The extremal graph lives by construction in single time slices. To obtain a connected structure in space-time the graph is further used as input for spatio-temporal clustering, see below.

Topology-based Space-Time Segmentation To capture the temporal evolution of extremal structures we use a topological method to segment regions within the common reference domain. This segmentation results in labels for extremal events that can be associated to individual physical phenomena (see Figure 4 b). Two adjacent groups are merged, if the difference of their maximum values is lower than a user-defined persistence threshold p_{min}. Technically this corresponds to a watershed algorithm [**?**] with a persistence-based hierarchical merging step [**?**]. We can obtain one connected cluster for dominant events choosing a maximum persistence value for $p_{min} = max(s(\mathbf{x}))$. The results of this segmentation are shown in Figure 4.

4.3 Topology-based Analysis and Visualization

The extraction of topological features, such as graphs or segmentation surfaces (Sec. 4.2), allows characterizing and grouping extremal features of the underlying scalar fields. Extreme graphs capture the spatial structure, e.g., longitudinal and lateral extents, as well as spatial connectivity of extremal regions. To enhance the information content of the graph, we map the local values of the scalar field to its thickness and color (see Figure 5 a)). Extremal structures can further be emphasized by visualizing corresponding iso-surfaces in the underlying advection fields (see Sec. 3.1).

a) AIRS SO_2 & graph 1.6.2011   **b)** AIRS SO_2 & graph 7.6.2011

d) All SO_2 graphs with offset surfaces over the observation time (space-time see Figure 4).
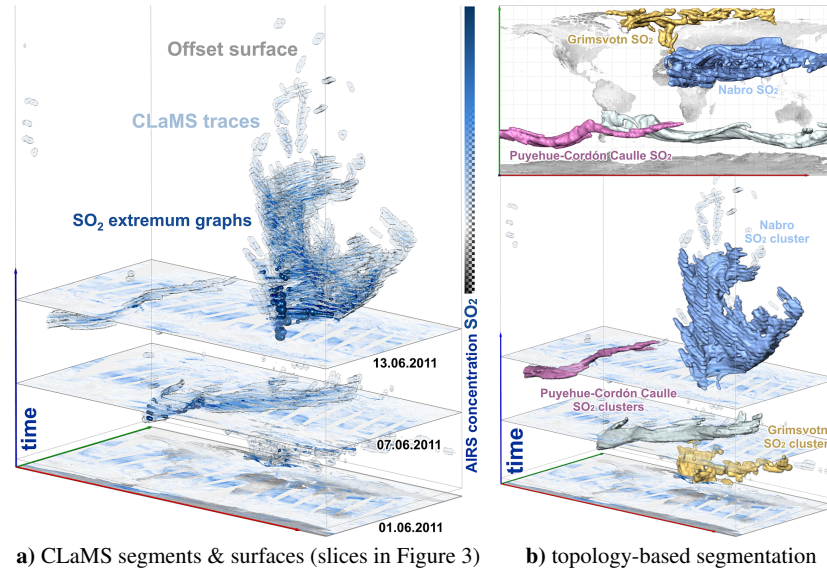
Fig. 3 Combining AIRS and CLaMS data: AIRS and CLaMS data describe the spatio-temporal structure of occurring SO_2 events and characterize the evolution of the SO_2 plumes: The SO_2 cloud created by the Puyehue-Cordón Caulle eruption moves rapidly eastwards, following the major jet streams on the southern hemisphere. In contrast, the Nabro event remains 'trapped' in a larger vortex structure and is distributed across central Asia, while the Grimsvötn SO_2 cloud circulates towards the north pole. In combination, the space-time view of multiple modalities conveys location and time of strongest value concentrations and their distribution during the plume development.

5 Results

We use the topology-based methods from Section 4.2 to visualize characteristics of the data in the common reference domain. Figure 5 shows a visualization of the integrated data in a space-time frame. The combination of filtering, interpolation and topology allows to extract a set of distinct extremal objects that can be associated to the major events of interest. The respective parameter setup allows to derive different levels of detail. Thus, not only the three dominant volcano eruptions but also a set of small-scale events are captured, which exhibit a much lower particle density.

To get a better impression of the movement of the particles we further augment this visualization with CLaMS trajectory segments, illustrated in Figure 5 b). The display of events associated to specific eruptions allows to compare ash concentrations based on AIRS satellite and MIPAS advected measurements. Mapping ad-

a) CLaMS segments & surfaces (slices in Figure 3) **b)** topology-based segmentation
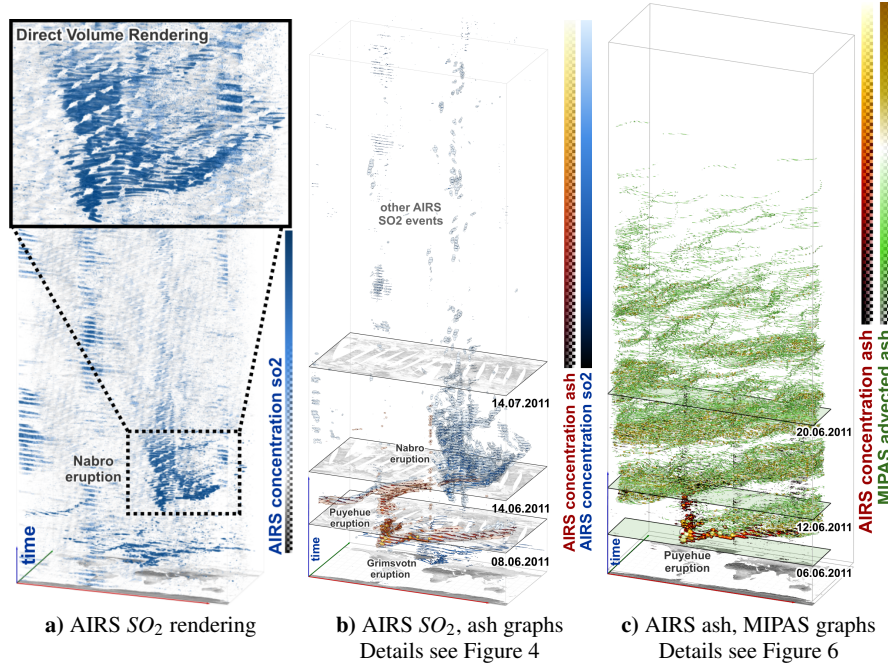
Fig. 4 Combining AIRS and CLaMS data: The space-time segmentation of AIRS SO_2 graph fields conveys location and time of largest features and their distribution during the plume development. The segmentation determines cluster volumes (filtered such that only volumes larger than 10.000 cells remain) that can be used to classify individual events.

ditional scalar information to the geometry of the topological graphs emphasize regions of specifically high scalar values. This is a clear advantage compared to slice-based color plots or direct volume visualizations. The spatial connectivity of the extremum graphs serves as a suitable description of the spatial structures of the corresponding plumes. Figure 5 c) shows a visualization of ash events form AIRS (red) and advected MIPAS (green) data. The higher sensitivity of the MIPAS data in space-time, compared to AIRS ash detection is clearly visible.

Detailed visualizations of single time slices, see Figure 6, illustrate the different characteristics of both measurement modalities. While AIRS data captures very detailed transport structures of the ash cloud, the MIPAS-CLaMS combination results in a blurred reconstruction and misses thin cloud structures. The concentration derived from the binary MIPAS signal is only an approximation of the true values. Due to its higher sensitivity, advected MIPAS better illustrates the spread of the ash cloud during the time of observation over the southern earth hemisphere.

Figure 3 illustrates the benefits of the high spatial resolutions of the AIRS measurements. Depending on the filtering scales, it is possible to detect also small-scale volcanic eruptions. An example is the Lokon-Epung event in Indonesia around July 2011 and multiple SO_2 and ash emissions associated to the Etna Volcano in Sicily. These smaller events are often not covered by the MIPAS data or are represented only by a very small number of detections. There are also multiple additional SO_2 events that are typically associated to mining or shipping trails.

a) AIRS SO_2 rendering

b) AIRS SO_2, ash graphs
Details see Figure 4

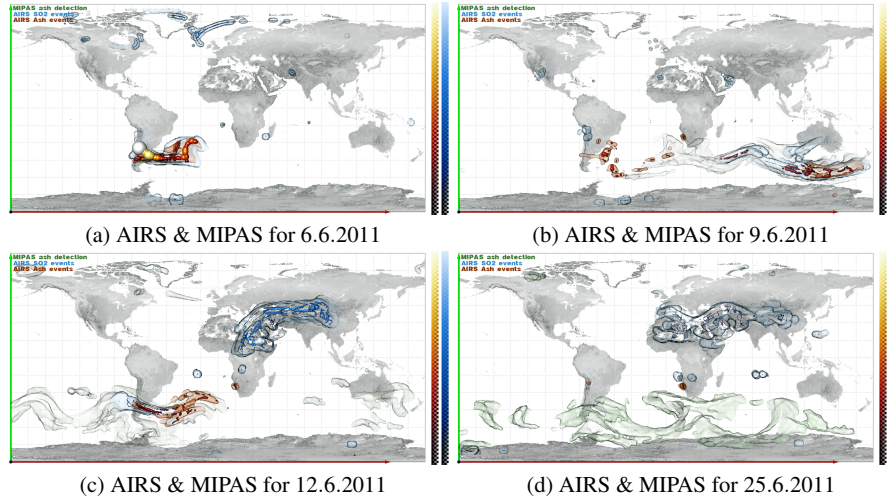**c)** AIRS ash, MIPAS graphs
Details see Figure 6

Fig. 5 Space-time AIRS, CLaMS, MIPAS: AIRS data is used to analyze SO_2 and ash concentration during the eruption events in space and time. In subfigures b) and c) the extremal graphs are displayed for each time slice. They are colored and sized by AIRS field values to emphasize strong production events. Transparent surfaces enclose the extremal graphs connecting them in temporal direction. The extend of these surfaces in time direction (z-axis) characterizes the life time of a plum. In c) the circulations of the tracers in the southern hemisphere can be observed nicely as visual stripes.

5.1 Limitations

Topological analysis tools have been shown to be valuable for reduction of complex input data to a smaller and visually assessable set of extremal structures. However, transferring sparsely sampled data from different modalities into a common reference domain still requires multiple filtering and interpolation steps (see Sec. 4.1). These use implicit assumptions, heuristic parametrizations and approximations of the real world phenomena, which are not covered by the data. Critical applications, as flight route planning after volcanic eruptions, require a more profound assessment of the impact of those techniques w.r.t. uncertainty and sensitivity in the final results. We assume that by using higher-order interpolation methods (e.g., for CLaMS and AIRS data) and samplings of the integrated domain, approximations can get more accurate, while some fundamental sources of uncertainty will remain (e.g., reliability of the CLaMS simulation, AIRS sensitivity).

a) AIRS ash & MIPAS advection field for 6.6. and 12.6.2011   **b)** AIRS, MIPAS graphs
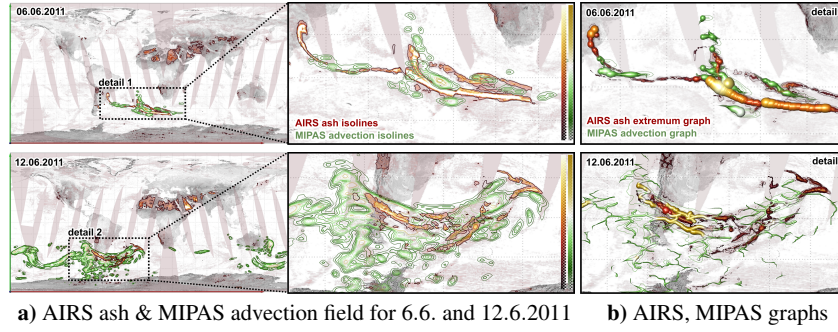
Fig. 6 Combining AIRS, MIPAS and CLaMS: Comparison of the AIRS ash data and the advected MIPAS ash field for the Puyehue-Cordon Caulle eruption. a) shows the reference fields for both modalities (AIRS, advected MIPAS). The detail views reveal the strengths and weaknesses of both information sources: AIRS captures fine spatial advection patterns, but suffers a limited detection accuracy. MIPAS advection fields are more sensitive, but do not convey information of the strength of the detection and sometimes miss samples across thin cloud structures.

(a) AIRS & MIPAS for 6.6.2011   (b) AIRS & MIPAS for 9.6.2011

(c) AIRS & MIPAS for 12.6.2011   (d) AIRS & MIPAS for 25.6.2011

Fig. 7 Integrated Visualization: This image shows snapshots of an integrated visualization of AIRS SO_2 (blue) and AIRS ash (red) with its corresponding graphs (scaled and colored by AIRS value) and MIPAS advection surfaces (green transparent surfaces). Each images captures the situation of one day and highlights the different characteristics of the Puyehue-Cordon Caulle and the Nabro eruptions, especially with respect to ash production (animation in accompanying video).

6 Conclusion

In summary, we have described a basic procedure to construct a common data domain and integrated data from a variety of input modalities, including measurements (MIPAS, AIRS) and simulations (CLaMS), to analyze volcanic eruptions. We have showed that topological processing provides suitable toold to characterize extremal features of such integrated data. The availability of explicit feature descriptors (i.e., topological graphs, segmentations) allows for improved filtering, compact visualization, feature-based comparison and visual analysis. The presence of topology features allows utilizing a broad set of tools known from the topology-related literature (e.g., persistence-based filtering, topology-driven rendering, high-dimensional visualization). As topological methods have been shown to scale in settings with large amounts of data (e.g., [**?**, **?**]) and in high-dimensional settings (e.g., [**?**]), similar analysis procedures should be able to cover future high-resolution input modalities. For future work we plan to analyze the impact of filtering methods in further detail. Of interest are also feature-based statistics (e.g., life-time and size distributions) and automated quantitative analytics of the data (automated event detection and classification).