Geosci. Model Dev. Discuss., 4, 3013–3045, 2011
www.geosci-model-dev-discuss.net/4/3013/2011/
doi:10.5194/gmdd-4-3013-2011

Geoscientific Model Development Discussions

Detection, tracking and event localization of interesting features in 4-D atmospheric data

S. Limbach[1,2], **E. Schömer**[1], **and H. Wernli**[2]

[1]Institute for Computer Science, Johannes-Gutenberg University, Mainz, Germany
[2]Institute for Atmosphere and Climate Science, ETH, Zurich, Switzerland

Received: 1 November 2011 – Accepted: 2 November 2011 – Published: 17 November 2011

Correspondence to: S. Limbach (limbach@uni-mainz.de)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

We introduce a novel algorithm for the efficient detection and tracking of interesting features in spatial-temporal atmospheric data, as well as for the precise localization of the occurring genesis, lysis, merging and splitting events. The algorithm is based on the well-known region growing segmentation method. We extended the basic idea towards the analysis of the complete 4-D dataset, identifying segments representing the spatial features and their development over time. Each segment consists of one set of distinct 3-D features per time step. The algorithm keeps track of the successors of each 3-D feature, constructing the so-called event graph of each segment. The precise localization of the splitting events is based on a search for all grid points inside the initial 3-D feature which have a similar distance to all successive 3-D features of the next time step. The merging event is localized analogously considering inverted direction of time. We tested the implementation on a four-dimensional field of wind speed data from European Centre for Medium-Range Weather Forecasts (ECMWF) analyses and computed a climatology of upper-tropospheric jet streams and their events. We compare our results with a previous climatology, investigate the statistical distribution of the merging and splitting events, and illustrate the meteorological significance of the jet splitting events with a case study. A brief outlook is given on additional potential applications of the 4-D data segmentation technique.

1 Introduction

Albeit highly variable, the atmospheric flow can be described by characteristic and frequently recurring flow features, like for instance tropopause-level jet streams, and surface cyclones and anticyclones. These flow features are particularly important since they are typically associated with certain weather conditions (e.g., sunny and dry weather with subtropical anticyclones; stormy weather and intense precipitation with extratropical cyclones) and with specific dynamical processes (e.g., cyclogenesis

GMDD
4, 3013–3045, 2011

in the left exit region of upper-level jets). Therefore, several algorithms have been developed during the last decades to objectively and efficiently identify atmospheric flow features from large datasets. These algorithms allow, for instance, to produce so-called synoptic climatologies of specific features and, if applied to homogeneous climate data sets, to quantify potential trends in the frequency of these features.

Early examples of such algorithms are the cyclone identification techniques by Lambert (1988), Murray and Simmonds (1991), and König et al. (1993). The earliest of these approaches considers cyclones as point objects and provides climatological density maps of these features. They are typically identified as local extrema of a particular field (e.g., minimum sea level pressure). In addition, the later techniques also considered the temporal coherency of the features, which led to the computation of feature tracks, and feature genesis and lysis points. For a concise review on cyclone identification and tracking methods, the reader is referred to Ulbrich et al. (2009). The algorithm introduced by Wernli and Schwierz (2006) considers cyclones explicitly as finite-size two-dimensional features (instead of point objects). Similar approaches have been used to identify, for instance, upper-tropospheric jet streams (Koch et al., 2006) and upper-tropospheric cut-off cyclones (Wernli and Sprenger, 2007) as two-dimensional features. The objective identification of these features was based upon either the topology of the two-dimensional field (e.g., considering the outermost closed contour surrounding a local extremum) or a simple threshold (e.g., considering the region where a field exceeds a certain value). So far, all these two-dimensional feature identification algorithms have been applied to individual time steps of climatological atmospheric data (e.g., every 6 h if using recent reanalysis data sets) and additional tracking algorithms have been used to meaningfully connect the identified features in time. As a consequence, spatial and temporal dimensions of atmospheric data have been treated very differently when using existing feature identification and tracking algorithms. Recently, Schiemann et al. (2009) and Manney et al. (2011) introduced schemes that focus on the jet axis (i.e., the location in meridional vertical cross-sections where the horizontal wind is maximum) and thereby avoided the vertical averaging of

GMDD
4, 3013–3045, 2011

the wind field as performed by Koch et al. (2006). The specific objectives of this study was, respectively, the investigation of the seasonal variability of the westerly jet near the Tibetan plateau (Schiemann et al., 2009) and of the complex three-dimensional trace gas distribution near the tropopause (Manney et al., 2011).

In this study we introduce a novel approach to the identification and tracking of atmospheric flow features, based upon the fundamental concept of region growing segmentation. Several variations of region growing segmentation algorithms exist (Zucker, 1976). Region growing has its origin in image processing, where it has a wide range of applications. Siegesmund (2006) applied a region growing method for the segmentation of ozone holes from time-series of two-dimensional ozone data. Inspired by his application of region growing techniques in atmospheric sciences, we created and implemented an algorithm for the full four-dimensional segmentation of atmospheric phenomena. Our algorithm differs in three major points from existing region-growing based feature detection and tracking algorithms, as found for example in Fonseca and Ii (1996); Reinders (2001); Derrien and Gléau (2007):

- Our algorithm works on time series of three-dimensional input data. An efficient implementation and capable data structures are required to process the increased amount of data efficiently.
- Instead of searching a single segment, our implementation performs a complete segmentation of the data set in order to detect all existing features of a certain category.
- The algorithm detects the occurrence of genesis, lysis, merging, and splitting events and localizes them on a per-grid-point basis.

We implemented the algorithm as part of a novel software tool for the analysis and segmentation of atmospheric data (Limbach et al., 2009). As a first application, we used the algorithm for the segmentation of ozone holes in the Antarctic. Next, we computed climatologies of upper-tropospheric jet streams and their events, as documented

GMDD
4, 3013–3045, 2011

in this study. The data basis of the jet stream segmentation were operational meteorological analyses from the European Center for Medium-Range Weather Forecasts (ECMWF).

In the upcoming section, the fundamental terms and objects of our algorithm are defined. In Sect. 3, the different steps of the complete segmentation algorithm are described in detail. While these two sections cover the general ideas and mechanisms of our segmentation algorithm, Sect. 4 deals with the setup and the results of a concrete application of the algorithm for the identification of jet streams. The section starts with an example case study of a Rossby wave breaking and an associated jet stream merging event. Next, it covers the computation and analysis of a two-years climatology of jet streams and their merging and splitting events. The last section provides a summary of the results and a short outlook on future developments and applications of the novel algorithm.

2 Foundations

2.1 Input

Region growing segmentation algorithms, both in their traditional and in the novel, extended form presented here, require a discretized, sampled signal as input data. In our applications, the input data consists of a series of discretized 3-D data sets representing the state of one or more atmospheric parameters at fixed time instants. The resolution of the three spatial dimensions (longitude, latitude and pressure) and of the time dimension may vary with respect to the concrete application and the available data sources. The underlying continuous domain of our data, however, remains the same:

Definition 1 (Data domain)**.** *We consider the atmospheric data we are interested in to be defined on the domain* $\Omega := [-180, 180) \times [-90, 90] \times \mathbb{R} \times \mathbb{R} \subset \mathbb{R}^4$. *The first two components of the domain represent the geographic longitude and latitude in degrees,*

GMDD
4, 3013–3045, 2011

respectively. The third component represents the vertical dimension, either Cartesian height (in m) or more commonly pressure (in hPa), and the last component represents time.

In their idealized, continuous form, the atmospheric parameters we are interested in can be seen as the mapping $p : \Omega \rightarrow V$. The co-domain V of this mapping depends on the actual features we want to track. Most of the time we are interested in n real-valued atmospheric variables, so our co-domain has the form $V \subseteq \mathbb{R}^n$. If we are, for example, interested in the segmentation of jet streams, a value $x \in V$ could represent the horizontal wind speed as a single scalar, or it could represent a horizontal wind vector $(u, v) \in \mathbb{R}^2$. Some more complex atmospheric structures may require the combination of several other, distinct measures.

In our practical application the input data is a sampled set of discretized values of the continuous atmospheric data lying on a point lattice within the data domain Ω. The exact form of a single sample depends on the concrete application and the objects we want to identify and track.

Definition 2 (Input data)**.** *Our input data consists of the set of samples* $X := \{x_{i,j,k,t} \mid i = 1, \ldots, i_{\max};\ j = 1, \ldots, j_{\max};\ k = 1, \ldots, k_{\max};\ t = 1, \ldots, t_{\max}\}$*. Indices* i, j, k *specify the spatial position of the sample on the point lattice, index* t *indicates the time step. The set* X_t *denotes all samples of a single time step* t*.*

Although we impose no constraints on the actual form of the point lattice, for the sake of reasonable results the indices of the samples should reflect the sample's adjacencies, that is, neighbouring samples should only differ by ±1 in one index. So far we worked with data on regular longitude/latitude grids (represented by indices i and j, respectively), with varying pressure given as a hybrid combination of the layer of the data set (index k) and the surface pressure (depending on i, j and t). In meteorological terms this corresponds to hybrid $p - \sigma$ coordinates. When using such a non-isotropic lattice, one has to take care as soon as additional information is derived from the results of the segmentation, such as the size of a segment or its centre of mass. The handling

GMDD
4, 3013–3045, 2011

of the indices at the poles and at the −180°/180° longitudinal transition requires some particular attention as well.

2.2 Output

The goal of a segmentation algorithm is to partition the set of input data into connected subsets of samples, where each subset ideally corresponds to the exact location of the phenomenon one wants to identify. In our case, since our set of input samples X is four-dimensional, the resulting *segments* will be four-dimensional objects as well. The construction of these 4-D-segments is accomplished in two steps:

1. We iterate over all time steps of the input data and partition the three-dimensional set of samples X_t into 3-D-subsets of samples corresponding ideally to exactly one instance of the atmospheric phenomenon we want to track at the given time step. We call these subsets three-dimensional *features* (see Fig. 1). This step is called the *feature detection* step.

2. We *track* and group the features of different time steps, such that we obtain information about the development of the atmospheric phenomena over time. This step is called *feature tracking*, and the resulting sets of connected three-dimensional features are our final four-dimensional *segments*.

More formally, we define the three-dimensional features as follows:

Definition 3 (Features)**.** *The pairwise disjoint sets of connected samples representing one occurrence of the detected atmospheric phenomenon at a single time step are called* features*. We denote the ith feature of time step t as $F_{t,i} \subseteq X_t$. F_t is the set of all features at the time step t. F is the set of all features at all time steps.*

Our algorithm outputs the information obtained during feature tracking in form of a so-called *event graph*. An event graph, like any graph, consists of a set of *nodes* connected by *edges*. In our case, the set of nodes corresponds to the set of all detected

GMDD
4, 3013–3045, 2011

3-D-features. If a connection between two 3-D-features of two consecutive time steps was detected within the feature tracking step, this connection is represented as an edge in our graph. The formal definition of the event graph is as follows:

Definition 4 (Event graph)**.** *The graph* $G := (F, E)$*, where* F *is the set of all detected features and* $E \subseteq \{(a,b) \mid a \in F_t;\ b \in F_{t+1};\ t = 1, \ldots, t_{\max} - 1\}$ *is the set of all edges representing a direct connection between features of two consecutive time steps, is called* event graph*.*

Note that there are several ways to define what "direct connection between features of two consecutive time steps" means. In many basic applications of our segmentation algorithm it is sufficient to define such a connection as a spatial overlapping of the respective features. More involved techniques, however, are often required if the features to track are small and fast moving.

Our final four-dimensional segments, each of which represents one atmospheric phenomenon and its development over time, are already contained in the event graph G as the connected sets of 3-D-features (see schematic example in Fig. 2). Such distinct sets of connected nodes of a graph, together with the connecting edges, are called *connected components*.

Definition 5 (Segment)**.** *Let* $G_S = (S, E_S)$ *denote a connected component of* G*. Then* S *is called a* segment*, representing all features associated with* one *atmospheric phenomenon as it develops over time and* G_S *is called the* event graph *of* S *containing all edges* E_S *between connected features of* S*.*

From the connectivity information provided by each event graph G_S, we can derive information about the occurring genesis, lysis, merging, and splitting events of each segment. The events can be detected through an inspection of the connecting edges E_S of G_S in the following way:

– A *genesis event* is detected if a node in the event graph has no incoming edges. For example, there is a genesis event at the node depicted by the green ellipse in Fig. 2.

GMDD
4, 3013–3045, 2011

– A *lysis event* is registered at nodes without outgoing edges. See the red ellipse in Fig. 2 for an example.

– If a node has more than one incoming edge, we register a *merging event*. This is the case at the node with the red incoming edges in Fig. 2.

– A *splitting event* exists at nodes with multiple outgoing edges. All nodes with green outgoing edges in Fig. 2 are associated with a splitting event.

Since features at the first time step have no incoming edges, we cannot tell genesis events, defined the way described above, apart from situations where a phenomenon simply enters the sampled data domain. In order to avoid spurious results, we exclude the first time step from the formal definition of our genesis events. We have an analog situation regarding lysis events at the last time step and we therefore exclude them as well. Taking this into account, we obtain the following, formal definitions:

Definition 6 (Genesis event)**.** *There is a* genesis event *of segment* S *at time step* $t+1$, $t \in \mathbb{N}$, *iff* $\exists j : \forall i : F_{t+1,j} \in S \wedge (F_{t,i}, F_{t+1,j}) \notin E_S$.

Definition 7 (Lysis event)**.** *There is a* lysis event *of segment* S *at time step* t, $t < t_{\max} \in \mathbb{N}$, *iff* $\exists i : \forall j : F_{t,i} \in S \wedge (F_{t,i}, F_{t+1,j}) \notin E_S$.

Definition 8 (Merging event)**.** *There is a* merging event *of segment* S *between time steps* t *and* $t+1$, *iff* $\exists i : \exists j : \exists k : i \neq j \wedge (F_{t,i}, F_{t+1,k}) \in E_S \wedge (F_{t,j}, F_{t+1,k}) \in E_S$.

Definition 9 (Splitting event)**.** *There is a* splitting event *of segment* S *between time steps* t *and* $t+1$, *iff* $\exists i : \exists j : \exists k : j \neq k \wedge (F_{t,i}, F_{t+1,j}) \in E_S \wedge (F_{t,i}, F_{t+1,k}) \in E_S$.

Note that for some applications it could be reasonable to replace genesis events at the first time step and lysis events at the last time step by so-called *entry* and *exit* events, respectively.

Our algorithm is capable of estimating the locations of the occurring events not only on a per-feature but on a per-sample basis. In case of genesis and lysis events, all

GMDD
4, 3013–3045, 2011

samples of each single involved feature are associated with the respective event. The detection of the locations of merging and splitting events is more involved, see Sect. 3.3 for a detailed description. In all cases, we denote the result of these attributions as follows.

Definition 10 (Event locations)**.** *The localization of the occurring events is represented by the set* $T \subset X \times \{\text{“genesis”}, \text{“lysis”}, \text{“merging”}, \text{“splitting”}\}$*. This set contains all involved samples together with an annotation indicating the event types that occur at the respective positions of the samples on the point lattice.*

2.3 Region growing segmentation

Many traditional region growing methods begin their segmentation at a given starting point (or seed point) $x_0 := x_{i_0,j_0,k_0,t_0}$. This is the first sample of the resulting segment S. The “growing” of the region is accomplished by looking at all direct neighbors of x_0 (eight samples $x_1, \ldots, x_8$ in the four-dimensional case). The algorithm decides which of these neighbouring samples are to be added to S based on a given homogeneity predicate. Next, all new neighbors of the extended segment S are checked as well and the process is iterated until there are no new samples to be added to S. The predicate used for testing new candidate samples is called the *homogeneity criterion*.

Definition 11 (Homogeneity criterion)**.** *The* homogeneity criterion *is a predicate* $p : \mathcal{P}(X) \to \{\text{TRUE}, \text{FALSE}\}$ *which decides whether a set of samples* $S \subseteq X$ *belongs to the same segment (*$p(S) = \text{TRUE}$*), or not (*$p(S) = \text{FALSE}$*).*

The concrete formulation of such a predicate p strongly depends on the type of sampled input data available, and on the actual objects to be detected. In general, p could involve any global characteristics of each set of samples, for example the mean sampled value. As long as the objects to be detected allow it, and with respect to the algorithm's performance, the decision can be restricted to characteristics of a local neighborhood of the candidate sample to be added. A typical restricted homogeneity criterion depending only on two adjacent samples can be defined as follows:

GMDD
4, 3013–3045, 2011

Definition 12 (Restricted homogeneity criterion)**.** *The* restricted homogeneity criterion $h : X \times X \rightarrow \{\text{TRUE}, \text{FALSE}\}$ *decides whether a given pair of neighboring samples* $x := x_{i,j,k,t} \in X$ *and* $x' := x_{i+i',j+j',k+k',t+t'} \in X$ *with* $i', j', k', t' \in \{-1, 0, 1\}; |i'| + |j'| + |k'| + |t'| = 1$ *belongs to the same segment, or not.* h *has to be commutative.*

A common approach for such a predicate is to define a threshold on the difference between the sampled values of x and the adjacent sample x'.

Several other common modifications to the region growing approach as described above exist. For example, the input data could initially be partitioned into cells containing a fixed number of samples, and the algorithm tests and merges whole cells instead of single samples. Another common modification is the selection of more than one sample or cell as seed points at the beginning of the region growing process. See, for example, Zucker (1976); Jain et al. (1995); Tönnies (2005) for several variations of region growing and merging algorithms.

3 Extended region growing segmentation

In this section, we describe the details of our extended region growing segmentation algorithm for the detection and tracking of interesting structures in atmospheric data. In contrast to many traditional region growing methods, which require one or more distinct seed points at the beginning, our algorithm performs a *complete* segmentation of the set of samples X. This means that we analyze the whole dataset in order to find all disjunct four-dimensional segments that can be associated with the atmospheric phenomenon we are looking for. Regarding the large, four-dimensional sets of data we want to run our algorithm on, the manual selection of one or more single seed points is no longer applicable. An automatic search of the whole dataset for potential seed points followed by traditional region growing is, depending on the size of the data set, often not efficient enough. Because of this, we designed our new method to only require one sequential iteration over the input samples. Due to efficiency considerations, we

GMDD
4, 3013–3045, 2011

decided to use the restricted homogeneity criterion h during this one-pass iteration. The restricted criterion also ensures that the sequence in which samples are added to a growing feature is irrelevant. For compensation of the missing selection of a seed point and the restrictions on the homogeneity criterion, we introduce two additional predicates: The *local selection criterion* and the *global selection criterion*.

Definition 13 (Local selection criterion)**.** *The* local selection criterion $l : X \rightarrow \{\text{TRUE}, \text{FALSE}\}$ *decides whether or not a single sample belongs to a potential three-dimensional feature, based on any of its local characteristics.*

This criterion can be applied to discard unsuitable samples right from the start. Depending on the objects to be detected, this local criterion may be sufficient for a full classification of whether a sample belongs to a feature or not. As it is the case for the homogeneity criterion, a concrete formulation should only depend on local properties of the respective sample.

Definition 14 (Global selection criterion)**.** *The* global selection criterion $g : \mathcal{P}(X) \rightarrow \{\text{TRUE}, \text{FALSE}\}$ *decides whether or not to keep a candidate three-dimensional feature based on any of its global characteristics.*

The global selection criterion allows the formulation of global requirements of any feature. For example, one could test whether a feature covers sampled values below or above a given threshold, or whether the candidate feature contains a minimum total number of samples.

While the formulation of the homogeneity criterion implicitly determines the form of all potential connected features, the two selection criteria work as additional filters at different stages of the algorithm. The main task to be fulfilled before the algorithm can be applied to different types of atmospheric phenomena is to find adequate and applicable predicates h, l and g.

Now that the input and output of the algorithm, as well as all required predicates have been defined, we are ready to describe the steps of the algorithm in detail. Algorithm 1

GMDD
4, 3013–3045, 2011

shows the general outline of the segmentation algorithm. In the following sections, we will investigate the details of all functions mentioned in this algorithm outline.

3.1 Feature detection

During the feature detection phase, denoted by the *feature_detection*(X_t, h, l, g) function call in the previous algorithm outline, all sets of samples which fulfill the local selection criterion and which are connected by means of the homogeneity criterion are constructed. On the way to the final set of features, the algorithm creates *candidate* features which are later either kept, discarded or merged with other candidate features.

The algorithm iterates sequentially over all samples of X_t, starting at $x_{1,1,1,t}$ and traversing the remaining samples by increasing the first three indices lexicographically. During the iteration, the algorithm keeps track of any association of a sample with any candidate feature. As soon as a sample $x := x_{i,j,k,t}$ fulfills the local selection criterion l, we look at all of the up to three already visited adjacent samples $x_{i-1,j,k,t}$, $x_{i,j-1,k,t}$, and $x_{i,j,k-1,t}$. From these samples we select those that fulfill the homogeneity criterion h. If no such samples exist, we associate x with a new candidate feature. If one or more neighbour samples are already associated with a candidate feature and fulfill the homogeneity criterion, we know that x is a connection between these features. In this case we merge all of them into one single feature and associate x with it. By using appropriate data structures in our implementation, we assure that all references to the candidate features are updated accordingly with constant computational costs. When the iteration ends, we have constructed a set of candidate features, which we finally check using the global selection criterion for features, g. All remaining candidate features are output as real features of the given time step.

3.2 Feature tracking

The next important task of our algorithm is the tracking of features. In the outline, the feature tracking step is represented by the *feature_tracking*(F_{t-1}, F_t) function call. The

GMDD
4, 3013–3045, 2011

goal of feature tracking is to identify the relations between features of the previous time step F_{t-1} and features of the current time step F_t by grouping features which belong to the same instance of an atmospheric phenomenon. There are many possible approaches to achieve this goal, depending primarily on the type of object to track. Although the feature tracking is specified as a separate step in the algorithm outline, it is often advisable to perform a lot of pre-processing already during the feature detection phase.

In our current implementation, we primarily check for spatial overlapping. In order to do this efficiently, we extend the data structure representing candidate features by a list of adjacent features of the next time step. We populate this list during the feature detection phase as soon as we finish the association of a sample $x_{i,j,k,t}$ with a candidate feature. If the sample at the same location of the previous time step, $x_{i,j,k,t-1}$, is associated with a candidate feature, we update its list of successive features accordingly. In the feature tracking step, it remains to iterate over those lists creating edges for all connections between real features.

In addition, we search for non-overlapping features of similar size within a limited range and connect them as well. The size and position of each feature are attributes we also compute on-the-fly during the feature detection phase, based on the single sizes and positions of all samples we add to each candidate feature.

3.3 Event localization

As soon as all connections between the three-dimensional features are established, we are ready to localize the occurring events on a per-grid-point basis. The localization of the genesis and lysis events (indicated by the *find_genesis_event_samples* and *find_lysis_event_samples* function calls) is straightforward: We associate all samples of features without a preceding feature with a genesis event, and all samples of features without a successive feature with a lysis event. We exclude genesis events at the first time step and lysis events at the last time step because we have no information about the history or continuation of these features.

GMDD
4, 3013–3045, 2011

The basic idea behind the localization of the merging and splitting events (*find_merging_event_samples* and *find_splitting_event_samples*) is a search for positions on the point lattice which have the same distance, measured as number of steps on the lattice, to all preceding or successive features. To find those points, we let the features "grow" and search for the positions where they touch. We now describe the method for localizing a merging event in detail. The localization of a splitting event is done analogously.

Assume there is a merging of features $F_{t-1,1},...,F_{t-1,m}$ into feature $F_{t,1}$, which is identified by the edges of the event graph. At first, we initialize an empty set N in which we will insert elements of the form $(i,j,k,n) \in \mathbb{N}^4$. We use these tuples in order to relate certain positions on the point lattice of our samples, specified by means of indices i,j,k, to different numerical tags n. Initially, for each existing pair of spatially overlapping samples $(p,q) := (p_{i,j,k,t-1}, q_{i,j,k,t})$ with $p \in F_{t-1,g}$ and $q \in F_{t,1}$, we add the tuple (i,j,k,g) to N. In other words, we associate all overlapping sample positions with the number of the respective feature from the previous time step, see the top-left panel of Fig. 3 for a two-dimensional example of this process. Next, we iteratively let these tagged regions grow by tagging all untagged positions adjacent to these regions with the respective numbers. This growing is limited to positions of $F_{t,1}$, see the top-right panel of Fig. 3. If in any growing step a single position is to be tagged by two or more different numbers, or if an already tagged position is to be additionally tagged by a different number, we found a position where multiple growing regions touch. We add a special event tag to N, replacing all other tags at this position, which indicates a merging. This special tag is excluded from any future growing steps. The iteration ends as soon as all positions corresponding to $F_{t,1}$ are tagged, see the bottom-left image of Fig. 3.

In a final growing phase, we let the regions marked by the special event tag grow for a fixed small number of steps. As before, this region growing is limited to the positions covered by samples of $F_{t,1}$. The idea behind this second growing phase is to transform the original thin border from the first growing phase into an area with

GMDD
4, 3013–3045, 2011
4-D feature detection, tracking and event localization
S. Limbach et al.

a certain, controllable amount of fuzzyness. The bottom-right picture of Fig. 3 shows the final result of an example, two-dimensional event localization. A real example of the three-dimensional merging localization from our jet stream segmentation is depicted in Fig. 5.

4 A climatology of upper-tropospheric jet streams and their events

We implemented and tested the new segmentation algorithm for different types of atmospheric phenomena. In this section we present results from one of these applications, which was to compute a two-years climatology of three-dimensional upper-tropospheric jet streams and their merging and splitting events. Wind fields were taken from the operational ECMWF analyses for the years 2007 and 2008, available every 6 h on 60 vertical levels interpolated to a regular longitude/latitude grid with a resolution of 1 degree. Before presenting the climatological results, we start with a brief case study of an interesting episode of Rossby wave breaking and an associated jet stream merging event over the North Atlantic. The aim of this case study is to illustrate the relevance of such events for atmospheric dynamics.

4.1 A Rossby wave breaking event over the North Atlantic

A prominent chain of events including rapid cyclogenesis, a prominent warm conveyor belt, the formation of a blocking anticyclone, a subsequent Rossby wave breaking, and eventually a jet stream merging event occurred during the time period 20–23 January 2007 over the North Atlantic. Isentropic charts of potential vorticity (PV) on 315 K at 12:00 UTC 20 January reveal a prominent trough with stratospheric PV (i.e., more than 2 pvu) over Eastern Canada and an equally prominent ridge with tropospheric PV (i.e., less than 2 pvu) downstream, over the Western North Atlantic (Fig. 4a). Rapid surface cyclogenesis occurred during the previous day beneath the upper-level trough leading to a mature cyclone with a core pressure of less than 970 hPa situated over the Gulf of St. Lawrence. Trajectory calculations (not shown) indicate that the rapid cyclone

GMDD
4, 3013–3045, 2011

evolution was associated with a prominent warm conveyor belt (Browning, 1990; Wernli and Davies, 1997), which ascends from the cyclone's warm sector almost to the 310-K isentrope and enlarges the upper level ridge during the following day (Fig. 4b). In parallel, the trough over the Western North Atlantic elongates into a filamentary "PV streamer" (Appenzeller and Davies, 1992) and a downstream trough evolves to the west of Europe. In between, the upper-level ridge develops into a persistent atmospheric blocking. At this time intense jet streams are present on 315 K (black contours) along both flanks of the PV streamer and the downstream trough. On the 350-K isentrope, jets exist over the US east coast, over Central Europe, and over Northern Africa. Whereas the first two of these jets are partially aligned with the jet systems on 315 K, the African jet is shallower and only present on the higher isentrope.

During the following 30 h the blocking becomes more prominent (reaching a maximum sea level pressure larger than 1045 hPa), the downstream trough protrudes to the Iberian Peninsula where it triggers the formation of a Mediterranean cyclone (not shown), and the anticyclonically curved jet stream to the north of the blocking on 315 K intensifies (Fig. 4c). Six hours later, at 00:00 UTC 23 January (Fig. 4d), the downstream trough reaches into the Western Mediterranean and its associated jet stream on 315 K becomes vertically aligned with the northern extension of the African jet on 350 K. A similar event has been described by Martius et al. (2010) (their Fig. 5) who emphasized the importance of such a jet merging event for a kinetic energy transfer from the extratropical to the subtropical waveguide. Recently, Martius and Wernli (2011) corroborated the relevance of these extratropical wave breaking events (and associated jet mergings) for the intensification of the subtropical jet over Africa.

Figure 5 shows the three-dimensional structure of the jet streams associated with the jet stream merging event over Gibraltar between 18:00 UTC 22 January and 00:00 UTC 23 January, as identified with the new segmentation algorithm (see the beginning of the next subsection for more details). The previously discussed jet streams to the north of the blocking and over Europe, and the elongated subtropical jet reaching from Northern Africa to the Western North Pacific are clearly visible. The red region in the second

GMDD
4, 3013–3045, 2011

panel highlights the merging of the extratropical and subtropical jets, as discussed above. This brief case study illustrates that jet merging events can be associated with prominent events of extratropical Rossby wave breaking and an associated momentum transfer from a midlatitude to a subtropical jet stream. According to climatologies of Rossby wave breakings (e.g., Wernli and Sprenger, 2007) such events are most likely to occur over the Eastern North Atlantic/Western Mediterranean and Eastern North Pacific/Western North America. It will be therefore interesting to consider the climatological occurrence of jet streams and their merging (and splitting) events in the following subsections.

4.2 Frequency of jets, jet merging, and jet splitting

For the detection of upper tropospheric jet streams, we choose the local selection criterion to be a height-dependent threshold on the wind speed. We accept all samples below 100 hPa (i.e., grid points with pressure larger than 100 hPa) with a horizontal wind speed exceeding 40 m s^{-1}. This criterion is motivated by the wind speed threshold criterion used by Koch et al. (2006), who considered jet streams as two-dimensional features of the vertically integrated wind speed between 100 and 400 hPa. Strong winds in the stratosphere, that is at levels above 100 hPa, are excluded to focus on jet streams in the upper troposphere. Compared to Koch et al. (2006) we use a 10 m s^{-1} larger threshold, due to the extension of considering jet streams as fully 3-D features instead of 2-D features of the vertically averaged wind speed. We do not state any explicit homogeneity criterion or global feature selection criterion. This has the effect that all adjacent samples fulfilling the local selection criterion are assigned to the same jet stream and none of our detected three-dimensional features is discarded.

With the results of the segmentation, we are able to compile a climatology of the frequency of jet streams and their genesis, lysis, merging, and splitting events. Figure 6 depicts the spatial distribution of all detected jet streams during the two-year time period. In order to obtain a two-dimensional plot, we unified all jet stream samples of each time step at each horizontal position. For this, we calculated the jet stream frequency

GMDD
4, 3013–3045, 2011

at each horizontal position as the ratio between the number of time steps at which a jet stream was detected at any observed vertical level (i.e., levels below 100 hPa), and the total number of time steps.

Our results agree favorably with the previous climatology by Koch et al. (2006), despite the different time periods (2007–2008 vs. 1979–1993). The overall frequency maximum is found over Japan (with values exceeding 90 %) and secondary maxima are located over Newfoundland and Libya/Egypt. The general spiral-like shape of the region with high jet stream frequencies is also well reproduced, corroborating the reliability of the new approach.

In addition, with the aid of our new method it is straightforward to compute climatological frequency distributions of merging and splitting events, as shown in Fig. 7. These patterns are obviously very different from the overall jet stream frequency distribution. Clear maxima of both mergings and splittings occur in the Western Northern Hemisphere, in particular over North America. Secondary maxima are found to the north of the Tibetan Plateau and over North Africa. This indicates on the one hand that the very high jet stream frequency over the Western North Pacific is associated with very stable jets that experience comparatively little merging and splitting events. On the other hand the much higher frequency of these events in regions mentioned above is qualitatively consistent with a high frequency of Rossby wave breaking (Wernli and Sprenger, 2007). Future studies will be required to better understand the global linkage between wave breaking and jet stream merging and splitting events.

4.3 Lifetime and stability of jet segments

The climatology indicates that jet merging and splitting is more frequent in one half of the Northern Hemisphere (from approximately 120° W to 60° E) than in the other half. We call this region the “North Atlantic region”, in contrast to the “North Pacific region” where the identified jet stream segments appear to be more persistent. In order to get statistical evidence of the differences in the stability of these segments in the two semi-hemispheres, we further investigate the size and lifetime of these jet segments.

GMDD
4, 3013–3045, 2011

As a direct consequence of the way we perform our feature tracking (see Sect. 3.2), some of our 4-D segments may consist of multiple major, separate jet stream phenomena that merge or split at some point in time. Here we are interested in obtaining the statistical distribution of the time span between consecutive merging and splitting events associated with the identified 4-D segments. This will allow differentiating between, for instance, very stable, long-lived and maybe quasi-stationary jet stream segments, and highly transient segments that break apart and re-merge at frequent intervals. In order to achieve this goal, it is useful to introduce so-called *sub-segments*.

A sub-segment is a subset of a 4-D segment without any major merging and splitting event. The decision whether or not a connection between two 3-D features of a segment is considered a connection between nodes of the same sub-segment is taken as follows: if any one feature of the segment has only one successor, the connection between the features is registered as a sub-segment continuation. For any feature $F_{t,i}$ with more than one successor, we pick the successor $F_{t+1,j}$ whose size is closest to the size of $F_{t,i}$. If now, looking backward in time, out of all of $F_{t+1,j}$'s predecessors $F_{t,i}$ is the one with the closest size to $F_{t+1,j}$, we register a valid sub-segment continuation between these two features. Out of these sub-segment continuations we are able to construct the complete sub-segments of each 4-D segment.

We additionally compute the average center of mass, the size and the lifetime of each sub-segment and use this information to attribute each one to either the North Atlantic or the North Pacific region. Figure 8 shows the statistical results of this processing. There are many relatively small and short-living segments in both regions. Large segments are less frequent. Very large segments, which contain about 15 000 grid points or more, have a lifetime of only up to 400 h in the North Atlantic. In contrast, segments of this size (or larger) all have a lifetime of about 400–4000 h in the North Pacific. Clearly, these huge and long-lived segments dominate the jet stream pattern in the North Pacific area and contribute essentially to the maximum jet stream frequency in this region. This finding is also consistent with the previously discussed results on the frequency of jet stream splitting and merging events.

GMDD
4, 3013–3045, 2011

5 Conclusions

We presented a new region growing segmentation method with per-grid-point localization of merging and splitting events, together with an analysis of upper-tropospheric jet streams in the Northern Hemisphere and their stability in different regions. So far we have been testing our implementation on two types of atmospheric phenomena: ozone holes and upper-tropospheric jet streams. In the future we plan to apply our method also to other atmospheric flow phenomena. Currently, we are working on the identification of cyclones. In contrast to existing methods working with 2-D structures (e.g., Raible et al. (2008) and studies mentioned in the introduction) we plan to detect and track the cyclones as full 3-D objects. For the detection of three-dimensional cyclones it may be helpful to perform an analysis of the topology of the underlying 3-D vector field, similar to the methods of Mahrous et al. (2004). The tracking of smaller and faster moving features, as well as the localization of the merging and splitting events for non-overlapping features are additional issues which have to be addressed. Tracking techniques based on statistical methods or techniques from fields such as image or signal processing (for example optical flow estimation) could be refined and adapted to cope with the tracking of smaller features. Another approach could involve the adaption of the region growing technique: By letting the features identified at a given time step grow with respect to their expected direction of movement, one ends up with a set of possible positions of the features at the following time step. If we analyze any occurring overlappings of these grown regions, either with each other or with the features of the next time step, we could possibly extract information about mergings, splittings and continuations of such rapidly moving features as well.

So far our experience with the new segmentation method in terms of computational costs is very positive. The analysis of the 2-yr wind field dataset used in this study was performed in about five hours on a standard Linux PC. This indicates that the method is also capable of analyzing very large datasets, for instance output of multi-model long-term climate simulations and/or from ensemble simulations. This will offer novel pathways for an in-depth analysis of flow features in very large climate datasets.

GMDD
4, 3013–3045, 2011

Acknowledgements. We would like to thank Patrick Jöckel from the DLR (Deutsches Zentrum für Luft- und Raumfahrt) for his help at the initial stage of this work, and for his introduction to and helpful suggestions on the topic of ozone hole detection. This project has been funded by the *Center for Computational Sciences in Mainz*, http://www.csm.uni-mainz.de.

References

Appenzeller, C. and Davies, H. C.: Structure of stratospheric intrusions into the troposphere, Nature, 358, 570–572, 1992. 3029

Browning, K. A.: Organization of clouds and precipitation in extratropical cyclones, in: Extratropical Cyclones, edited by: Newton, C. and Holopainen, E., vol. The Erik H. Palmen Memorial Volume, Amer. Meteor. Soc., Boston, pp. 129–153, 1990. 3029

Derrien, M. and Gléau, H. L.: Temporal-differencing and region-growing techniques to improve twilight low cloud detection from SEVIRI data, in: Proceedings of Joint 2007 EUMETSAT Meteorological Satellite Conference and the 15th Satellite Meteorology & Oceanography Conference of the American Meteorological Society, 2007, Amsterdam, 24–28 September 2007, available at: http://www.eumetsat.int/home/Main/AboutEUMETSAT/Publications/ConferenceandWorkshopProceedings/2007/SP_1232700283028?l=en, last access: 16 November 2011. 3016

Fonseca, L. M. G. and Ii, F. M.: Satellite imagery segmentation: a region growing approach, in: VIII Brazilian Symposium on Remote Sensing, pp. 677–680, Salvador, Bahia, 14–19 April 1996. 3016

Jain, R., Kasturi, R., and Schunck, B. G.: Machine Vision, McGraw-Hill, New York, 1995. 3023

Koch, P., Wernli, H., and Davies, H.: An event-based jet-stream climatology and typology, Int. J. Climatol., 26, 283–301, 2006. 3015, 3016, 3030, 3031

König, W., Sausen, R., and Sielmann, F.: Objective identification of cyclones in GCM simulations, J. Climate, 6, 2217–2231, 1993. 3015

Lambert, S. J.: A cyclone climatology of the Canadian Climate Centre general circulation model, J. Climate, 1, 109–115, 1988. 3015

Limbach, S., Marto, M., Jöckel, P., Schömer, E., and Wernli, H.: Die Analyse atmospärischer

GMDD

4, 3013–3045, 2011

Strömungen, Natur & Geist – Das Forschungsmagazin der Johannes Gutenberg-Universität Mainz, 2, 22–25, 2009. 3016

Mahrous, K., Bennett, J., Scheuermann, G., Hamann, B., and Joy, K.: Topological segmentation of three-dimensional vector fields, IEEE T. Vis. Comput. Gr., 10(2), 198–205, 2004. 3033

Manney, G. L., Hegglin, M. I., Daffer, W. H., Santee, M. L., Ray, E. A., Pawson, S., Schwartz, M. J., Boone, C. D., Froidevaux, L., Livesey, N. J., Read, W. G., and Walker, K. A.: Jet characterization in the upper troposphere/lower stratosphere (UTLS): applications to climatology and transport studies, Atmos. Chem. Phys., 11, 6115–6137, doi:10.5194/acp-11-6115-2011, 2011. 3015, 3016

Martius, O. and Wernli, H.: A trajectory based investigation of physical and dynamical processes that govern the temporal evolution of the subtropical jet streams over Africa, J. Atmos. Sci., submitted, 2011. 3029

Martius, O., Schwierz, C., and Davies, H. C.: Tropopause-level waveguides, J. Atmos. Sci., 67, 866–879, 2010. 3029

Murray, R. J. and Simmonds, I.: A numerical scheme for tracking cyclone centres from digital data. Part I: development and operation of the scheme., Aust. Meteorol. Mag., 39, 155–166, 1991. 3015

Raible, C. C., Della-Marta, P. M., Schwierz, C., Wernli, H., and Blender, R.: Northern hemisphere extratropical cyclones: a comparison of detection and tracking methods and different reanalyses, Mon. Weather Rev., 136, 880–897, 2008. 3033

Reinders, F.: Feature-based visualization of time-dependent data, Ph. D. thesis, University of Delft, Delft, 2001. 3016

Schiemann, R., Lüthi, D., and Schär, C.: Seasonality and interannual variability of the westerly jet in the Tibetan Plateau region, J. Climate, 22, 2940–2957, 2009. 3015, 3016

Siegesmund, M.: Visuell gestützte Mustererkennung zur Identifikation und Klassifikation von Ozonlöchern, Master's thesis, Universität Rostock, Rostock, 2006. 3016

Tönnies, K.: Grundlagen der Bildverarbeitung, Pearson Studium, Munich, 2005. 3023

Ulbrich, U., Leckebusch, G. C., and Pinto, J. G.: Extra-tropical cyclones in the present and future climate: a review, Theor. Appl. Climatol., 96, 117–131, 2009. 3015

Wernli, H. and Davies, H. C.: A Lagrangian-based analysis of extratropical cyclones, I, the method and some applications, Q. J. Roy. Meteor. Soc., 123, 467–489, 1997. 3029

Wernli, H. and Schwierz, C.: Surface cyclones in the ERA-40 dataset (1958–2001). Part I: novel identification method and global climatology, J. Atmos. Sci., 63, 2486–2507, 2006. 3015

GMDD

4, 3013–3045, 2011

Wernli, H. and Sprenger, M.: Identification and ERA15 climatology of potential vorticity streamers and cut-offs near the extratropical tropopause, J. Atmos. Sci., 64, 1569–1586, 2007. 3015, 3030, 3031

Zucker, S. W.: Region growing: childhood and adolescence, Comput. Vision. Graph., 5, 382–399, 1976. 3016, 3023

GMDD

4, 3013–3045, 2011

Algorithm 1

Input: The set of sampled atmospheric data X, the homogeneity criterion h, the local selection criterion l and the global selection criterion g. **Output:** The event graphs G_{S_i} containing all segments $S_1, \ldots, S_n$ corresponding to the detected atmospheric phenomena together with the precise event locations T.

$F := \emptyset$ ▷ The set of all features.
$E := \emptyset$ ▷ The set of all connecting edges.
$T := \emptyset$ ▷ The set of all event locations.
for $t := 1, \ldots, t_{\max}$ **do**
  $F_t := feature\_detection(X_t, h, l, g)$
  $F := F \cup F_t$
  if $t > 1$ **then**
    $E := E \cup feature\_tracking(F_{t-1}, F_t)$
    $T := T \cup find\_lysis\_event\_samples(F_{t-1})$
    $T := T \cup find\_splitting\_event\_samples(F_{t-1})$
    $T := T \cup find\_genesis\_event\_samples(F_t)$
    $T := T \cup find\_merging\_event\_samples(F_t)$
  end if
end for
$segment\_id := 1$
for each connected component (S, E_S) of (F, E) **do**
  $S_{segment\_id} := S$
  $E_{S_{segment\_id}} := E_S$
  $G_{S_{segment\_id}} := (S_{segment\_id}, E_{S_{segment\_id}})$
  $segment\_id++$
end for
return $G_{S_1}, \ldots, G_{S_{segment\_id-1}}$ and T

GMDD
4, 3013–3045, 2011

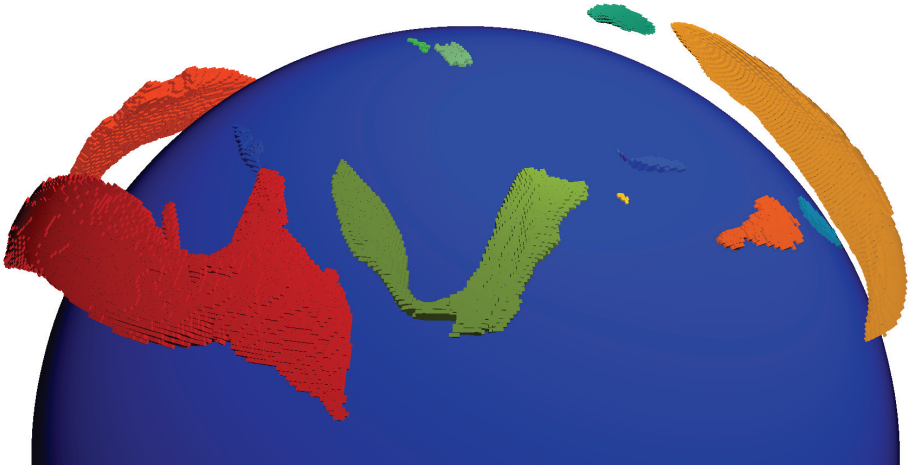

Fig. 1. Example illustration showing all three-dimensional features at one time step of a segmentation of jet streams in the Northern Hemisphere (perspective projection, vertically exaggerated). Different colors indicate different features.

GMDD
4, 3013–3045, 2011

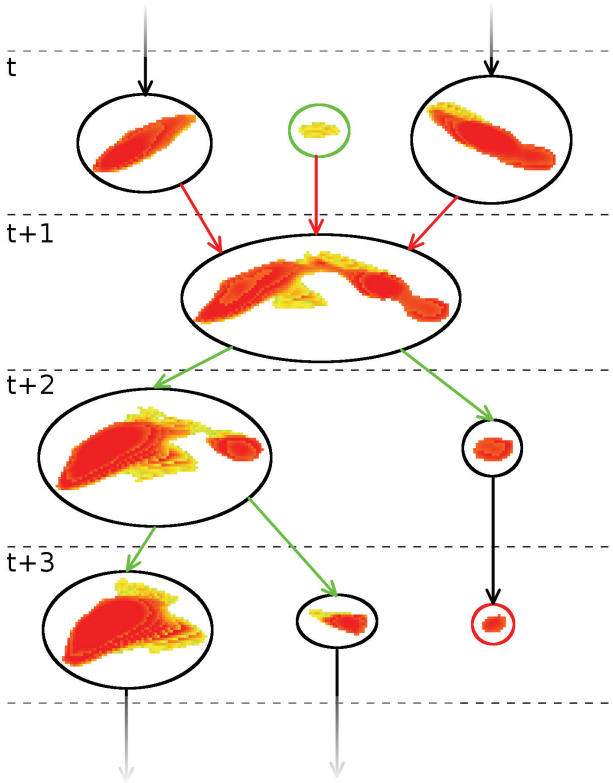

Fig. 2. Part of an example event graph of a single 4-D-segment. Nodes are depicted by ellipses including images of the corresponding 3-D-features – here 3-D-visualizations of jet streams. Connecting edges are depicted by arrows, dashed lines indicate the border between features of different time steps. The green ellipse indicates a genesis event, and the red ellipse a lysis event.

GMDD
4, 3013–3045, 2011

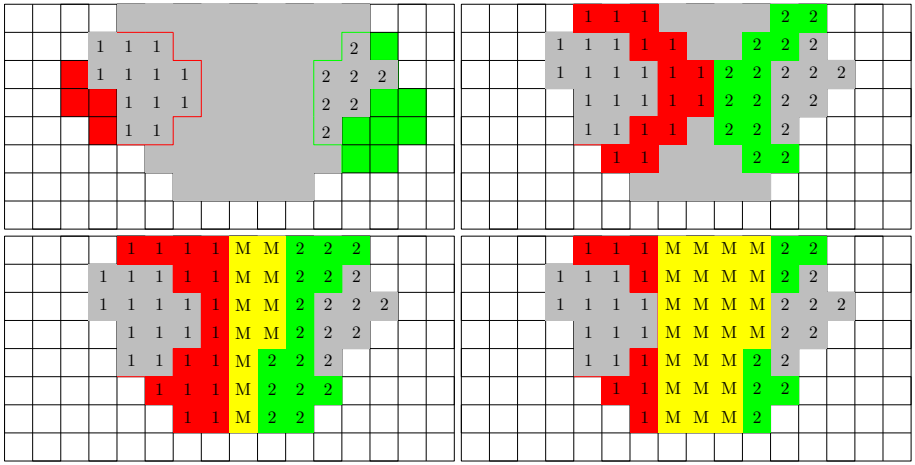

Fig. 3. Two-dimensional example of the applied steps for localizing a merging event. In the top-left picture, gray grid points indicate the location of the merged feature at time step t. Red and green grid points indicate the initial location of the separated features at the previous time step $t-1$. Grid points with the tag "1" mark the positions where the red and gray features overlap. Grid points with the tag "2" indicate positions where the green and gray features overlap. The pictures to the right and below show the second step of the first growing phase and the final tagging, respectively. Newly tagged regions are depicted in red and green, the positions where both regions touch are indicated by the tag "M" on yellow background. The picture at the bottom-right shows the final state at the end of the second growing phase.

GMDD
4, 3013–3045, 2011

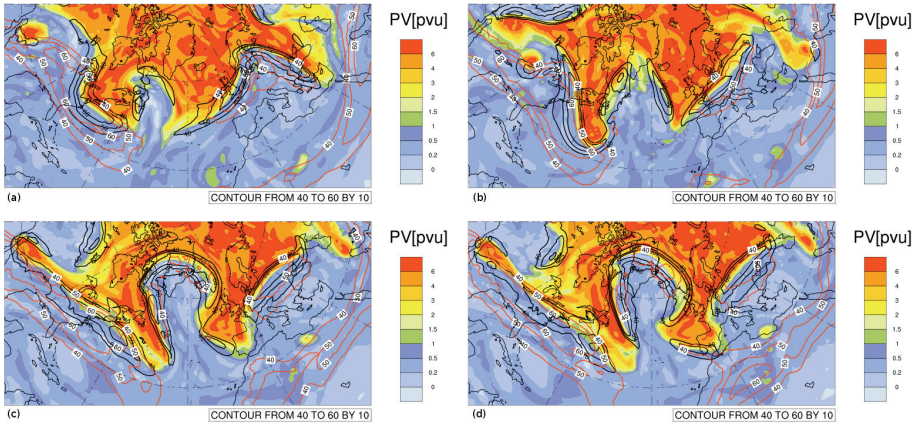

Fig. 4. Isentropic potential vorticity on 315 K (colors, in pvu) and wind speed on 315 K (black contours for 40, 50, and 60 m s^{-1}) and on 350 K (orange contours for 40, 50, and 60 m s^{-1}) at **(a)** 12:00 UTC 20 January 2007, **(b)** 12:00 UTC 21 January 2007, **(c)** 18:00 UTC 22 January 2007, and **(d)** 00:00 UTC 23 January 2007. The 2-pvu contour denotes the dynamical tropopause.

GMDD
4, 3013–3045, 2011

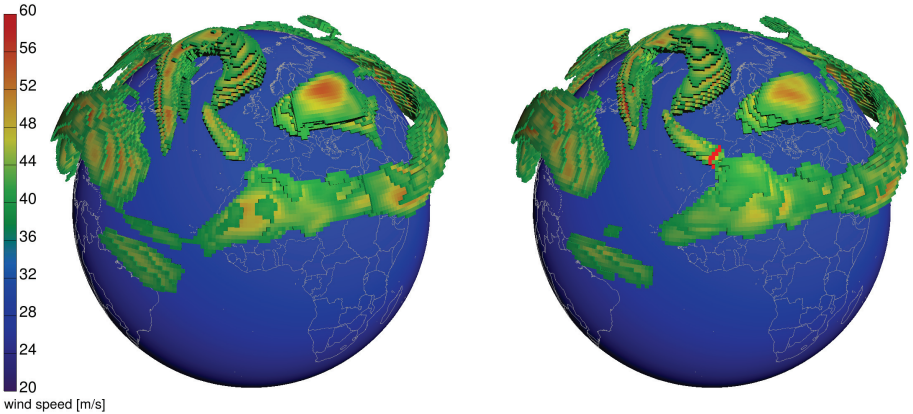

Fig. 5. Two successive time steps of a jet stream segmentation at 18:00 UTC 22 January 2007 (left) and 00:00 UTC 23 January 2007 (right). The red highlighted region in the panel on the right indicates the location of a merging event. At the previous time step (see panel on the left) the two jet stream features were still separated. All non-event samples are shaded according to the wind speeds at their respective positions.

GMDD
4, 3013–3045, 2011

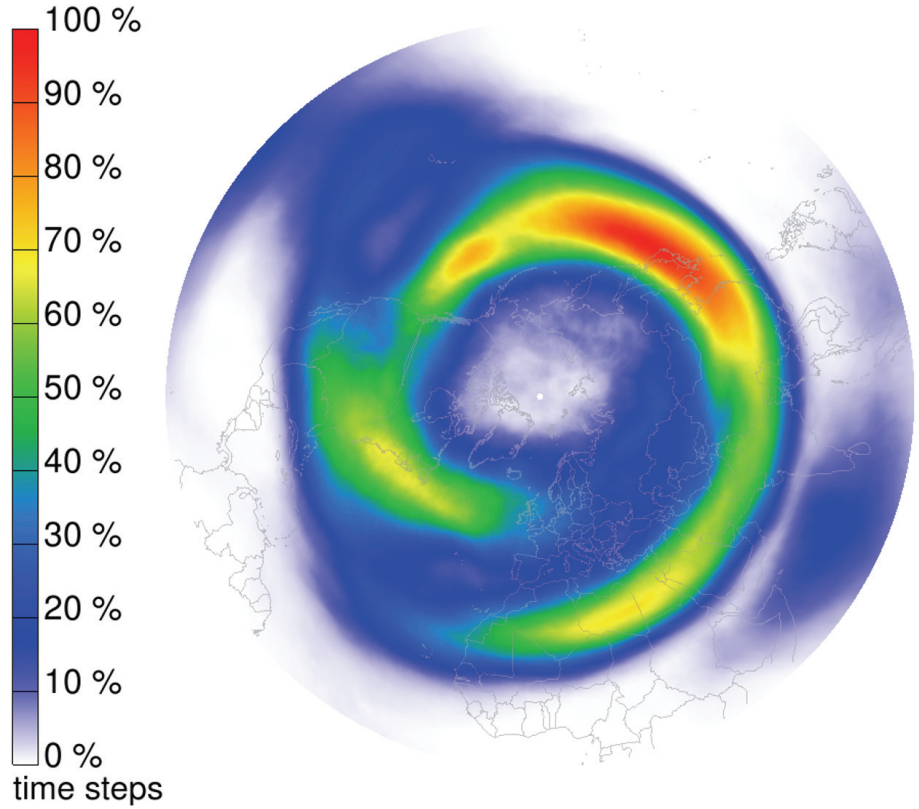

Fig. 6. Two-year climatology of the spatial distribution of jets identified with the new segmentation method. Values indicate the frequency of jet stream occurrences (in %).

GMDD
4, 3013–3045, 2011

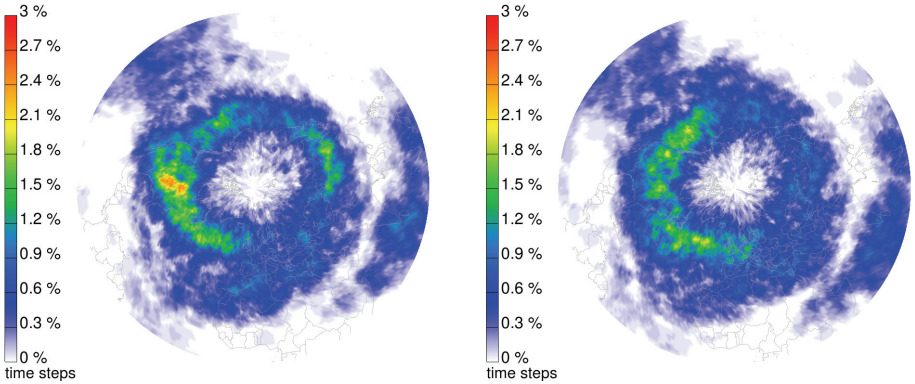

Fig. 7. Two-year climatology of the spatial distribution of (left) jet stream merging and (right) jet stream splitting events. Values indicate the event frequencies (in %).

GMDD
4, 3013–3045, 2011

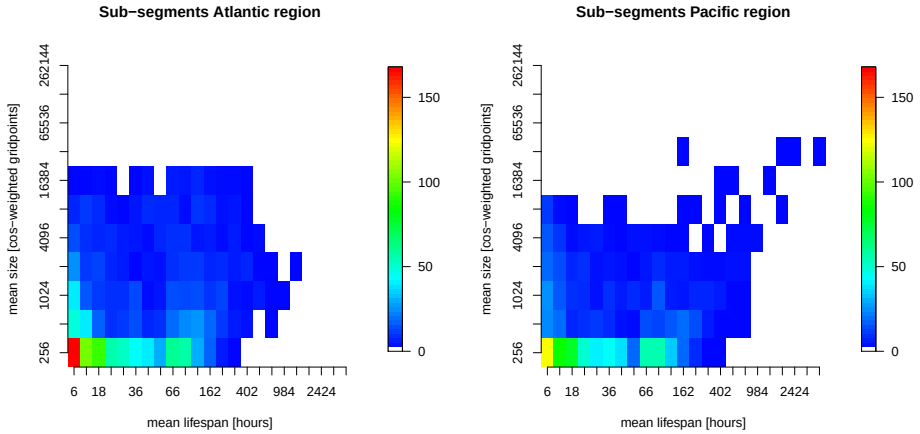

Fig. 8. Histograms relating the size to the lifetime of jet stream sub-segments in the North Atlantic (left) and North Pacific (right) region, respectively.

GMDD
4, 3013–3045, 2011