Analysis of convective cell evolution with split and merge events using a graph-based methodology

Jenna Ritvanen[1,2], Martin Aregger[3], Dmitri Moisseev[1,2], Urs Germann[4], Alessandro Hering[4], and Seppo Pulkkinen[1]

[1]Finnish Meteorological Institute, Helsinki, Finland
[2]Institute for Atmospheric and Earth System Research, Faculty of Science, University of Helsinki, Helsinki, Finland
[3]Institute of Geography, University of Bern, Bern, Switzerland
[4]MeteoSwiss, Locarno-Monti, Switzerland

Correspondence: Jenna Ritvanen (jenna.ritvanen@fmi.fi)

Abstract.

Convective storms are associated with several hazards, including heavy rainfall, hail, and lightning, which pose severe risks to society. While the nowcasting (i.e., short-term forecasting from 5 minutes to 6 hours) of storm locations has been extensively studied, nowcasting storm development remains a challenge. Nowcasting rapid, non-linear convective storm evolution requires finding connections between observations and storm evolution and representing them in the nowcasting model. Convective cell identification and tracking algorithms are commonly used for nowcasting and analysis of convective storms. This analysis is complicated by the splits and merges that occur in the cell tracks, either due to the physical processes or data quality issues. Consequently, the splits and merges are often excluded from the analysis. Here, we present a methodology for analyzing cell development around time of interest that explicitly includes the splits and merges in the analysis. The time of interest can be the time when the nowcast is created or the occurrence of some fingerprint of meteorological processes, for example, Z_{dr} columns. We represent the cell tracks as directed graphs where we select event nodes to represent the times of interest. For each event node, a subgraph of related cells from both the past and future of the event node is selected. We propose rules for selecting the subgraphs with the aim of retaining the available information in the subgraph at each time step. Once selected, the cell features in the subgraphs are aggregated into time series for analysis. We demonstrate the methodology through case studies of convective storms with Z_{dr} column features signalling updrafts and apply it to analyze split and merge events using three years of warm-season (MJJAS) operational radar data from the Swiss national weather radar network, with a focus on the total rainfall amount produced by the cells. Splits and merges occur in 7.2% of all identified cells, and are more frequent in cells with larger vertically integrated liquid (17.9%) or containing Z_{dr} columns (11.7%). Typically, cell mergers are associated with growth in total rainfall and cell area, and cell splits are associated with decrease in total rainfall.

1 Introduction

Convective storms are associated with several meteorological phenomena, such as heavy rainfall, lightning, hail, and strong winds, that pose severe hazards to society. Because convective storms tend to evolve rapidly, producing accurate and timely

hazard forecasts and warnings, such as flash flood warnings, requires short-term forecasting with lead times ranging from 5 minutes to 6 hours, i.e., nowcasting (World Meteorological Organization, 2017). The nowcasting of convective storms in operations is a challenging task. Although the development stages of convective storms are relatively well understood through modelling and observational case studies (Brooks et al., 2018), acquiring and representing the knowledge of the storm evolution from observations for use in real-time operational nowcasting models remains a challenge. Furthermore, the rainfall amounts produced by the storms are expected to increase due to climate change, emphasizing the need for accurate short-term forecasting and warning systems (Rädler et al., 2019; Taszarek et al., 2021; Utriainen et al., 2025).

Weather radar measurements are particularly well suited for analyzing and nowcasting convective storms because of their high spatial and temporal resolution (e.g., 100 m to 1 km and 1 to 5 mins), the wide spatial coverage provided by national radar networks in many countries, and their ability to estimate surface rainfall more effectively than satellites (Li et al., 2024; De Luca et al., 2025). In addition to estimating rainfall, radar measurements can provide other information relevant to the evolution of the storms. In recent years, several *fingerprints* of convective storm processes have been identified in polarimetric radar measurements. For example, differential reflectivity (Z_{dr}) columns have been shown to be proxies for updrafts (Kumjian et al., 2014). These fingerprints can precede the intensification of the storms or the onset of hail or heavy rainfall, and thus may have potential for improving the nowcasts (Picca et al., 2010; Kumjian et al., 2014; Snyder et al., 2015; Kingfield and Picca, 2018; Van Den Broeke, 2017; Wilson and Broeke, 2022).

Representing the convective storms as objects, i.e., convective cells, provides a convenient approach for associating the temporal evolution, different fingerprints of the convective storm processes, as well as storm-related hazards into a single conceptual model. These phenomena often manifest at scales significantly larger than a single radar pixel or through different modalities, such as binary detection of lightning versus continuous measurement of rainfall intensity. Consequently, linking these phenomena with grid-based approaches, which treat radar pixels individually, is challenging. However, handling these different properties together becomes simpler when the convective storms are represented as objects combining all radar pixels within the object (Rossi, 2015).

Convective cells are usually identified with contour-based methods, where the cells are defined as contiguous areas exceeding a threshold aimed to differentiate between stratiform and convective rainfall (e.g., 35–40 dBZ in radar reflectivity), with possibly some post-processing applied to further refine the extent of the identified cells (see e.g., Steiner et al., 1995; Lakshmanan et al., 2003; Hering et al., 2004; Muñoz et al., 2018). Depending on the targeted phenomena, other identification criteria may also be used, such as image texture (Guyot et al., 2023). Once identified, cells at consecutive time steps are matched to each other based on some algorithm-specific criteria, e.g., similarity or distance between the cells, to form cell tracks that describe the temporal evolution of the cells. A variety of algorithms have been developed for cell identification and tracking from weather radar observations (e.g., Dixon and Wiener, 1993; Johnson et al., 1998; Handwerker, 2002; Lang, 2002; Hering et al., 2004; Kyznarová and Novák, 2009; Muñoz et al., 2018; Skinner et al., 2018; Hu et al., 2019; Zan et al., 2019; Hou and Wang, 2017; Raut et al., 2021; Shi et al., 2024). The cell tracks can be used to study convective storm evolution and climatology in radar data or simulations (e.g., Kyznarová and Novák, 2009; Liu and Li, 2016; Hu et al., 2019; Tuftedal et al., 2024; Oue et al., 2022; Feldmann et al., 2021); study convective initiation (Merk and Zinner, 2013); nowcast future locations of the cells (Dixon

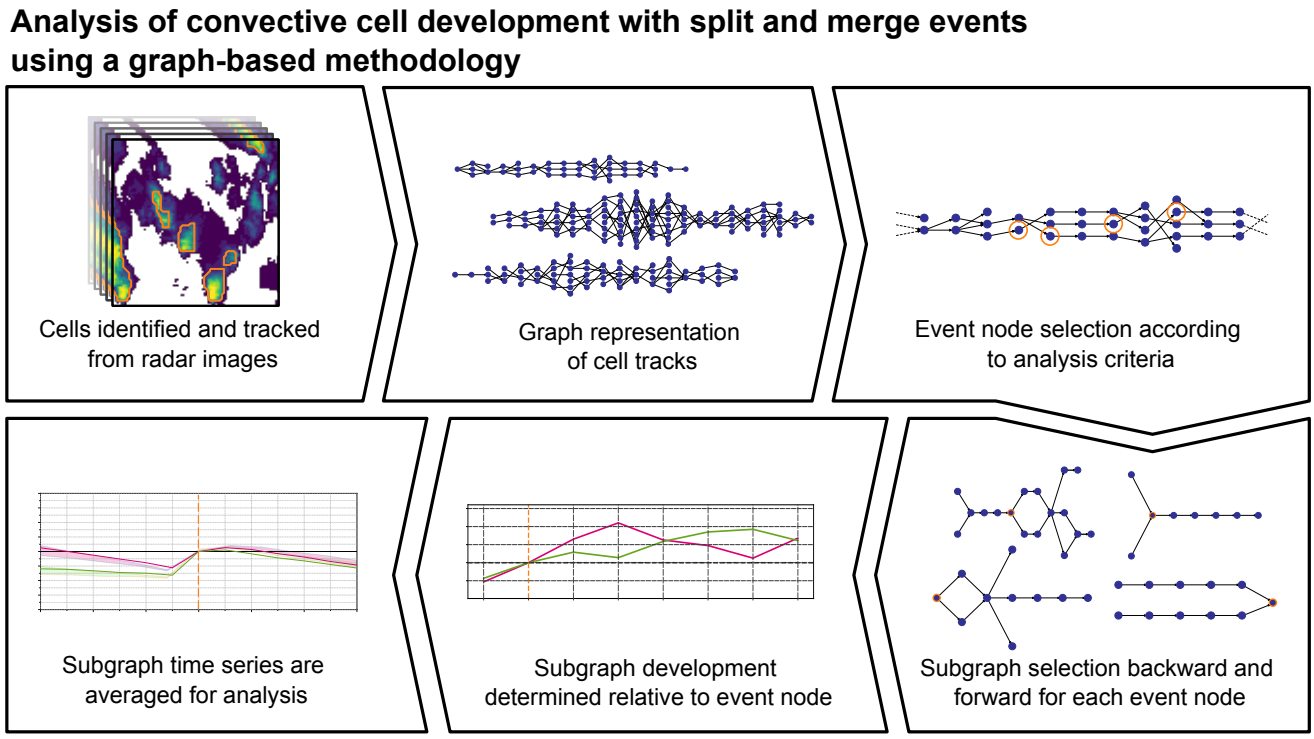

Figure 1. Overview of the presented analysis methodology.

and Wiener, 1993; Rossi et al., 2015); estimate and nowcast hazards related to the cells (Rossi et al., 2012; Tervo et al., 2019; Heinselman et al., 2024); or evaluate the forecast skill of grid-based nowcasting systems (Ritvanen et al., 2025b). Historically, most attention has focused on nowcasting the location of convective cells with extrapolation methods, whereas nowcasting their evolution has remained a challenge.

A major challenge in the analysis of convective cell evolution through cell tracking is the occurrence of splits and merges in the tracks when a cell has multiple parent cells or child cells instead of a single cell. Splits and merges can be caused by two separate factors. First, splitting and merging are genuine meteorological processes observed in convective storms, driven by various forcing mechanisms (Bluestein et al., 1990; Westcott, 1984). Additionally, horizontal motion can cause multiple identified cells to converge into one or, similarly, one cell to multiple, especially if the cells are defined to contain areas larger than only the convective core of the storm. We refer to these as *true* merges and splits. Second, merges and splits can result from inconsistencies in the identification algorithm or data artefacts (e.g., strong attenuation), which may cause a cell to be artificially identified as multiple cells, causing a split or merge to be mistakenly identified in the tracks, or from overly permissive matching criteria in the cell tracking algorithm. Even though the impact of these artificial splits and merges can be mitigated by data quality control and fine-tuning of the cell identification and tracking algorithms (Lamer et al., 2023), true splits and merges are inherent features of storm evolution and excluding them from analysis can remove useful information.

Furthermore, the separation between true and artificial splits is not always obvious; in some situations, the evolution of the cells can be so chaotic that even experienced meteorologists may find it difficult to visually identify the splits or merges.

Most cell tracking algorithms provide information on splits and merges, but these events lead to complex track structures that complicate analysis of the evolution. A common approach when analyzing the tracks is to determine a "most representative" path through a track, typically based on minimizing variation in some attribute, such as radar reflectivity factor inside the cell, across the cell track (Liu and Li, 2016; Cheng et al., 2024; Tseng et al., 2025; Bournas and Baltas, 2024; Ritvanen et al., 2025b). This approach assumes that a "good" cell track should preserve storm-related attributes (Lakshmanan and Smith, 2010), and thus minimizing variability in the attributes yields the track that best describes the storm. In this approach, during a split, the original cell track is continued in the "most representative" cell among the multiple potential child cells, while other cells initiate new tracks. Similarly, in a merge, the track of the parent cell most similar to the new cell continues, and other tracks are terminated. Another commonly used approach is to simply exclude any tracks containing splits or merges from the analysis (Lamer et al., 2023; Tuftedal et al., 2024).

The "most representative" path approach is useful for analysing the cell tracks for long time periods or full cell life cycles. In such cases, favouring long, stable cell tracks is most likely to produce good results. However, when we want to analyze the evolution of the cells in a limited time window related to some event, or to use the analysis for nowcasting purposes, such as to predict the immediate development of the cells, this approach is less suitable. In forecasting, the future state of the cells is not known, so it cannot be used for determining the "most representative" path. Rather, we want to gather all available information in a short time window surrounding the analysis time that could contain information useful for predicting the development of cells. Furthermore, the importance of the splits and merges to the analysis depends on the application. For example, in the dataset analyzed in this study (see Section 2.1), the fraction of all identified cells with splits or merges, i.e., cells with multiple parent or child cells, is 7.2%; however, 11.7% of the identified cells that are associated with a Z_{dr} column feature and 17.9% of cells where the maximum vertically integrated liquid (VIL) inside the cell is at least 20 $\mathrm{kg\,m^{-2}}$ have splits or merges. Thus, especially if the analysis is focused on these cells, it is sensible to include the split and merge branches of the cell tracks in the analysis.

A logical approach to explicitly include the splits and merges in the analysis is to represent the cell tracks as spatio-temporal graphs (Liu et al., 2016; Hou and Wang, 2017). In this approach, each cell is represented as a node, and the temporal connections between cells at consecutive time steps as edges in the graph. Split cells are nodes with multiple out-going edges and merge cells are nodes with multiple incoming edges. In this approach, defining the "most representative" path is not necessary and all branches of the tracks can be accounted for together. Spatio-temporal graphs have been used in other fields for geospatial data analysis (e.g., Liu et al., 2019; Wang et al., 2020; Xue et al., 2019b). However, compared to many other datasets, cell tracks present challenges due to their (usually) large amount and the variability of the tracks, ranging from short-lived isolated cells to large, long-lived organized convective systems. The variability of the cell track graphs makes it challenging to define isomorphisms (transformations) between the cell track graphs that are often used in quantitative graph analysis (e.g., Yan et al., 2016). Previous studies applying the cell track graph approach have largely focused on quantifying split and merge statistics,

e.g., their counts or spatial distribution, or on full storm lifecycle without emphasizing the splits or merges (Liu et al., 2016; Liu and Li, 2016; Yin et al., 2022; Xue et al., 2019a).

However, to analyze and predict the development of cells in a short time window surrounding some specific event (such as the occurrence of a Z_{dr} column feature or the nowcast creation time), we need to be able to extract the parts of the cell track graphs that are relevant for the analysis; the question of how to do this systematically has not been addressed by previous studies. In this study, we develop a methodology that extracts the cells that are related to any instantaneous event from the cell track graph, enabling quantitative analysis of cell development in relation to the event. The proposed methodology enables analysis of the cell development in the presence of splits or merges in the cell track. The workflow is illustrated in Figure 1, and the terminology used in the study is listed in Table 1. First, cell tracks are represented as directed graphs. Then we select event nodes that are of interest to the analysis from the cell track graphs. For each event node, we select a subgraph of related cells from both the past and future of the event node. We propose the selection rules for the subgraphs with the aim of retaining the available information in the subgraph at each time step. Once selected, these subgraphs are aggregated into time series for each event node, which can then be analyzed to assess the impact of the event node on convective cell development.

We demonstrate the methodology through case studies, as well as by applying it to a dataset consisting of split and merge events using three years of warm-season (May-September) operational radar products from the Swiss weather radar network. The cells in the dataset were identified for the purpose of analyzing factors impacting convective rainfall. We present statistics on the splits and merges in the dataset and discuss their significance to the analysis of the cell tracks. Then we analyze the impact of the splits and merges to the total rainfall in the cells and the cell area using the methodology.

The rest of this article is structured as follows. Section 2 describes the methodology in detail, as well as the data used in the analysis. Section 3 presents two case studies demonstrating the methodology, and Section 4 discusses the results of applying the methodology to the dataset of split and merge events. Finally, Section 5 summarizes the article.

2 Methodology and data

2.1 Radar data

In this study, we use weather radar data from the Swiss national weather radar network. Mainly, two different radar products are used: firstly, the convective cells are identified from vertically integrated liquid (VIL; Greene and Clark, 1972; Hu et al., 2019; Tuftedal et al., 2024) and, secondly, the rainfall amount inside the cells is determined from the radar-only quantitative precipitation estimation (QPE) product (Germann et al., 2006, 2022). Both radar products are produced operationally by MeteoSwiss with a temporal resolution of 5 minutes and a spatial resolution of 1 kilometre as a composite product of 5 C-band radars on a Cartesian grid. The size of the products is 710×640 kilometres, and the product domain is demonstrated in Fig. 3. The dataset used in the analysis presented in Section 4 is extracted from May to September from the years 2021-2023.

Table 1. Description of terms used in the study.

Term	Explanation
Cell	Contiguous area identified from a radar image using a cell identification algorithm; represents an observation of a convective storm
Cell track	Continuous sequence of temporally connected cells that represents the evolution of a convective storm produced by a cell tracking algorithm
Cell track graph	Graph representation of a cell track where each cell is a node and temporal connections between the cells are the edges between the nodes
Event node	Cell that is of interest to the analysis; time step at which the event node exists is denoted t_0
Event subgraph	Subgraph of nodes selected from a cell track graph that contains cells that might influence the event node or that the event node might influence in a short time window
Predecessor cell	Cell that is temporally connected to the current cell (possibly through some other cell) and occurs before it
Parent cell	Cell that occurs at the previous time step and is temporally connected to the current cell
Successor cell	Cell that is temporally connected to the current cell (possibly through some other cell) and occurs after it
Child cell	Cell that occurs at the next time step and is temporally connected to the current cell
Split cell	Cell that is temporally connected to multiple cells at the next time step (i.e., has multiple child cells) but to only one cell at the previous time step (i.e., has only one parent cell)
Merged cell	Cell that is temporally connected to multiple cells at the previous time step (i.e., has multiple parent cells) but to only one cell at the next time step (i.e., has only one child cell)
Merge-split cell	Cell that is temporally connected to multiple cells at the previous time step (i.e., has multiple parent cells) and to multiple cells at the next time step (i.e., has multiple child cells)
New cell	Cell that has no parent cells but at least one child cell
Terminal cell	Cell that has no child cells but at least one parent cell

The VIL, in units of kilograms per square meter, is calculated by integrating radar reflectivity observations in a vertical column (Greene and Clark, 1972):

$$\mathrm{VIL} = 3.44 \times 10^{-6} \int_{h_{\mathrm{bottom}}}^{h_{\mathrm{top}}} Z^{0.57} dh \tag{1}$$

where h is the height in meters, h_{bottom} and h_{top} are the echo base and top heights, respectively, and Z is the radar reflectivity in linear units of millimetres to the sixth power per cubic meter. The minimum non-zero value of the VIL product is $0.5\ \mathrm{kg\,m^{-2}}$, and the product is stored in an 8-bit format with a resolution of $0.5\ \mathrm{kg\,m^{-2}}$.

The rainfall product is created from radar reflectivity observations using the $Z - R$ relation $Z = 316R^{1.5}$ where the radar reflectivity Z is in linear units of millimetres to the sixth power per cubic meter, and the rainfall rate R is in units of millimetres per hour (Germann et al., 2006; Joss et al., 1998). The data are further processed to remove ground clutter and non-meteorological echoes, correct for visibility and vertical profile of reflectivity, and correct for bias compared to rain gauge measurements (Germann et al., 2006), before being stored in an 8-bit format. Furthermore, to avoid overestimation in the presence of hail, rain rates are saturated at approximately $120\ \mathrm{mm\,h^{-1}}$ (approximately 56 dBZ). From the rainfall product, we determine the total rainfall inside the cells as the volume rain rate (RVR; in units of cubic meters per hour) that is the integrated rain rate over the cell area at a given time step. The definition of volume rain rate is similar to what was used, for example, by Rosenfeld (1987); Hu et al. (2019); Feng et al. (2018) and Ritvanen et al. (2025b).

Additionally, in the case studies in Section 3, we use the Z_{dr} column height (Z_{dr}C) determined from the radar observations to demonstrate the impact of splits and merges for the continuity of Z_{dr}C signals in the convective cells. Z_{dr} columns are vertically contiguous areas of high Z_{dr} (in this case, $Z_{\mathrm{dr}} \geq 1$ dB) above the environmental $0\,^{\circ}\mathrm{C}$ level that can be used as a proxy for updrafts in convective storms (e.g., Illingworth et al., 1987; Kumjian et al., 2014; Picca et al., 2010). For more details about the Z_{dr}C data, refer to Snyder et al. (2015); Aregger et al. (2025) and Appendix B.

2.2 Cell track graph construction

The first step of the presented analysis methodology involves identifying and tracking the convective cells within the data. Once the cells are tracked, the cell tracks are constructed into graphs. In the graph construction, each distinct cell track is presented as a directed simple acyclic graph $\mathcal{G} = (V, E)$ where V represents the nodes (cells) belonging to the graph and E represents the edges (temporal connections) between the nodes produced by the tracking algorithm. More specifically, the nodes are defined as a set $V = \{(t, id, A)\}$, where t is the time step where the cell was identified, id is an identifier that uniquely identifies the cell at time step t, and A is a set of attributes associated with the cell (e.g., the volume rain rate or the cell area). The edges are defined as a set $E \subseteq \{(v_1, v_2) = ((t_1, id_1, A_1), (t_2, id_2, A_2)) \mid v_1, v_2 \in V \text{and } t_1 < t_2\}$, i.e., the connections between two nodes v_1 and v_2 in the cell track, defined in an order so that v_1 occurs before v_2. Furthermore, for each node v, we can construct the set of parent nodes (mathematically known as *in-neighbourhood*) $N^-(v) = \{u \mid (u, v) \in E\}$, and the set of child nodes (*out-neighbourhood*) $N^+(v) = \{u \mid (v, u) \in E\}$. Using these definitions, we can define:

- a *merge* cell as a node with more than one parent but at most one child: $|N^-(v)| > 1$ and $|N^+(v)| \leq 1$,

- a *split* cell as a node with more than one child but at most one parent: $|N^-(v)| \leq 1$ and $|N^+(v)| > 1$,
- a *merge-split* cell, formed from multiple merging cells and splitting into multiple cells at the next time step, as a node with both more than one parent and more than one child: $|N^-(v)| > 1$ and $|N^+(v)| > 1$
- a *new* (born) cell as a node with no parents but at least one child: $|N^-(v)| = 0$ and $|N^+(v)| > 0$, and
- a *terminal* (dying) cell as a node with at least one parent but no child cells: $|N^-(v)| > 0$ and $|N^+(v)| = 0$.

Each cell track graph consists of all identified cells connected to each other through some other cell in the track, either in the past or future, allowing the cell track graphs to cover arbitrarily many consecutive time steps. Because of this, a single cell can only be part of one cell track graph. However, multiple cell track graphs originating from different cells can co-exist at the same time steps.

Note that the methodology described here is independent of the identification and tracking algorithms, provided that the tracking algorithm supplies information on splits and merges occurring in the cell tracks. Additionally, any potential spatial connections between the cells at one time step, such as, spatial clustering or multiscale cell identification (e.g., Lakshmanan et al., 2003; Hou and Wang, 2017), are not considered in this study and are left for future development. If algorithms that produce spatial connections, such as clusters, are used, the clusters or a single level of the multiscale cells could replace the convective cells in the graphs. As with any cell-based analysis, the cell identification and tracking algorithms impact the results obtained from the analysis, so the algorithms and objects used in the graphs should be selected so that they describe the phenomena studied as well as possible.

In this study, the cells were defined for the purpose of analyzing factors impacting convective rainfall in operational weather radar products. The cells are identified from the VIL product using a constant threshold of 1.0 $\mathrm{kg\,m^{-2}}$, with a minimum detected cell size of 10 $\mathrm{km^2}$, and a tracking algorithm based on the algorithms used by Rossi et al. (2012, 2015) and Tervo et al. (2019) is employed. Appendix A describes the cell identification and tracking algorithms.

2.3 Subgraph selection

After the cell track graphs are constructed, the subsequent steps involve selecting the event nodes of interest and extracting the subgraph for each event node using a breadth-first approach (Cormen, 2009). The event nodes should be selected to represent the instantaneous events that are analyzed. In the results presented later in this study, the event nodes include all merge, split, and merge-split cells. Depending on the goals of the study, other criteria for selecting the event nodes could involve, for example, peak in rainfall or the occurrence of certain polarimetric features. For each selected event node, we denote time step of the event node as t_0, the time steps backwards from the event node as $t_{-1}, t_{-2}, \ldots$, and the time steps forwards after the event node as $t_1, t_2, \ldots$.

For each event node v, the subgraph is constructed in two stages: backward and forward selection. In backward selection, we select predecessor cells that are connected to the event node, and in forward selection, successor cells that are connected to the event node. An example of the selection is shown in Figure 2. The selection aims to retain the information at each time

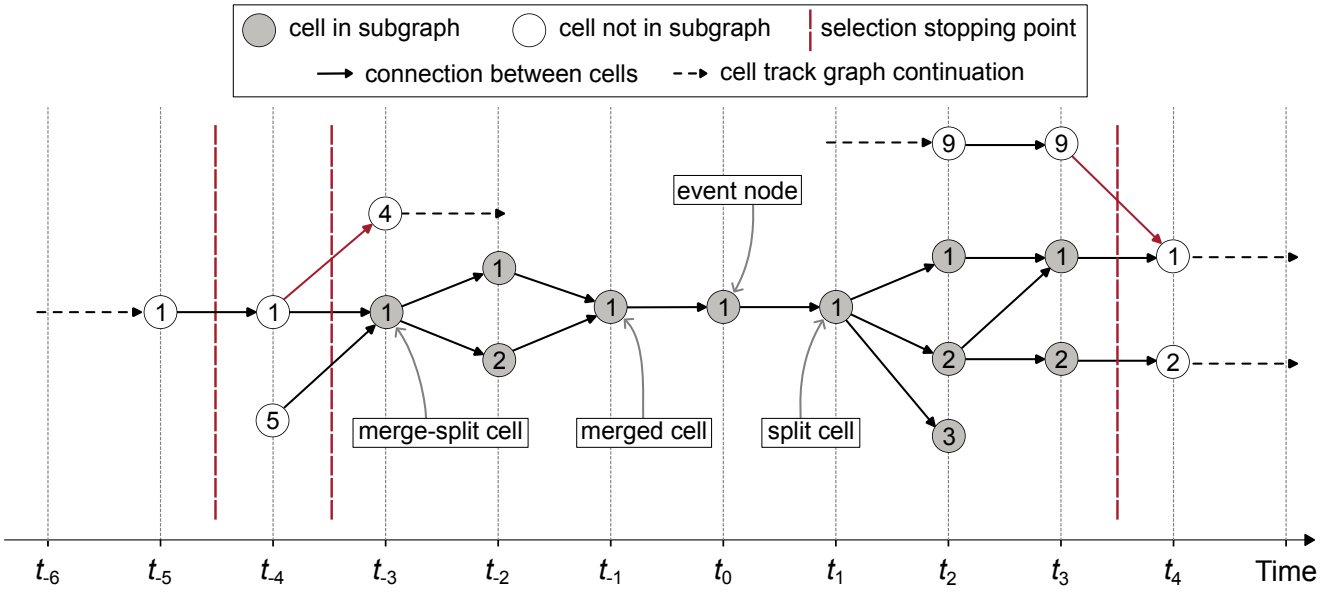

Figure 2. Example of cell subgraph selection from the cell track graph. For the event node labelled 1 at time t_0 the subgraph selection stopping points are presented with dashed vertical lines and cells accepted into the subgraph are coloured as grey. The arrows in the graph represent the temporal connections between the cells.

step in the subgraph; if information is lost or added at some time step, the selection is stopped. Information here is understood to mean all the observed cell attributes, such as the rainfall in the cells, the cell area, or the presence of Z_{dr} column features, and their temporal development. In backward selection, information could be lost through splits where some child cells are not in the current subgraph, or added in new cells. In forward selection, new information could be added by a merge where some parent cells are not in the current subgraph. The selection is done so that at each time step either all cells connected to the event node are included in the subgraph, or none of them. The cells belonging to the backward subgraph are denoted as $S_m^-(v)$ and the forwards subgraph as $S_n^+(v)$, where m and n are the maximum number of time steps traversed backwards and forwards, respectively.

Algorithm 1 describes the selection of the cells $S_m^-(v)$ in the subgraph backward from the event node. To select the cells, a backward traversal from the event node is conducted. At each time step $t_{-1}, \ldots, t_{-n}$, the parents of all cells that are part of the current subgraph at the previous time step are inspected. In the backward subgraph, two special cases are considered: 1) if the cell splits and 2) if the cell is a new cell.

For splitting cells, the cell is accepted in the subgraph only if all of its child cells are already in the subgraph; otherwise, the cell is not accepted, and the subgraph selection is stopped at the previously inspected time step. For example, in Fig. 2, at t_{-3} cell number 1 is a split cell with both of its child cells (1 and 2 at time step t_{-2}) in the subgraph, and thus accepted. However, at t_{-4} cell 1 is a split cell for which one child cell (4 at t_{-3}) is not a predecessor of the event node, and therefore, cell 1 at

t_{-4} is not accepted into the subgraph. Since information is discarded at time step t_{-4}, the subgraph selection is stopped at the previous time step t_{-3}.

In the case of new cells, the cell is accepted in the subgraph, but since the new cell contains information that does not exist at the next inspected time step, the subgraph selection is stopped after the time step of the new cell. This is demonstrated in Fig. 2, where at t_{-4}, cell 5 is a new cell that would cause the subgraph selection to be stopped after time step t_{-4}. Note that in the demonstration in Fig. 2, cell 5 at t_{-4} would not be included due to the split occurring in cell 1, but if that split did not occur, cell 5 would be included, and the subgraph selection would be stopped after t_{-4}.

Similarly, Algorithm 2 describes the forward subgraph $S_n^+(v)$ selection. At each time step $t_1, \ldots, t_m$, the children of all cells that are part of the subgraph at the previous time step are inspected. For the forward subgraph selection, the special cases are the merge cells, which are accepted only if all of their parent cells are included in the subgraph at the previous time step. For example, in Fig. 2, at t_3, cell 1 is a merge whose parent cells are all included in the subgraph, and it is thus accepted. However, for the merge cell 1 at t_4, one parent (cell 9 at t_3) is not a successor of the event node, and hence cell 1 at t_4 would introduce new information to the subgraph and it is not accepted. Not accepting cell 1 at t_4 removed some information at t_4, so the entire time step t_4 must be disregarded, and the subgraph selection is stopped after t_3. Note that in the forward subgraph selection terminal (dying) cells do not stop the subgraph selection (as opposed to new cells in the backwards subgraph selection), as they do not introduce new information into the subgraph.

After both the backward and forward subgraphs have been constructed, the full subgraph for the event node v can be defined as a graph $\mathcal{T}_{n,m}(v) = (S_{n,m}(v), C_{n,m}(v))$ where the nodes are defined as $S_{n,m}(v) = S_n^-(v) \cup \{v\} \cup S_m^+(v)$ and the edges $C_{n,m}(v) = \{(x, y) \in E \mid x, y \in S_{n,m}(v)\}$, where E are the edges in the cell track graph.

The algorithms presented here can be applied to tracks in real-time, i.e., without information on the future development of the tracks. Using incomplete cell tracks can impact the composition of the cell track graphs, as the cell tracks could become connected at a later time. However, the only limitation for subgraph selection and subsequent analysis would be a restriction on how many time steps forward the subgraphs can be selected (if those time steps do not exist in the data). Because the forward selection of the cells in the subgraph at one time t_i does not depend on the next time step t_{i+1}, the forward selection could be done iteratively as new time steps are observed without changing the selections at previous time steps.

2.4 Subgraph development analysis

After subgraph construction, each event node v is represented by a subgraph with possibly multiple cells at the time steps $t_{-n}, \ldots, t_{-1}, t_1, \ldots, t_m$ (note that at t_0, the only cell is the event node v). The subgraphs are aggregated at each time step to produce a single time series centered at the time of each event node.

The method of aggregation over cells at one time step depends on the attribute studied, and should be selected so that the resulting variable is sensible. The volume rain rate and area of the cells are used in this study; these variables are summed over the cells at each time step to obtain estimates for the total volume rain rate and area of the subgraph. For variables where summing is not appropriate, such as mean or maximum rain rate, values can be aggregated by averaging or taking the maximum or minimum. The average can also be weighted by some attribute, such as cell area.

Algorithm 1 Backward subgraph $S_n^-(v)$ selection for event node v.

$S_n^-(v) \leftarrow \varnothing$
$\mathcal{C}_{-1} \leftarrow N^-(v)$
for $i = -1$ to $-n$ **do**
  `stop_at_this_level` $\leftarrow$ **false**
  `stop_after_this_level` $\leftarrow$ **false**
  $\mathcal{C}_{i-1} \leftarrow \varnothing$
  for all $c_k \in \mathcal{C}_i$ **do**
    `included` $\leftarrow$ true
    if $|N^+(c_k)| > 1$ **and not** $N^+(c_k) \subset S_m^-(v)$ **then** {Cell splits and all child cells are not included in the subgraph}
      `stop_at_this_level` $\leftarrow$ **true**
      `included` $\leftarrow$ **false**
    end if
    if $|N^-(c_k)| = 0$ **then** {Cell is a new cell}
      `stop_after_this_level` $\leftarrow$ **true**
    end if
    if `included` **then** {Add current cell to subgraph and its parents to cells investigated at next iteration}
      $S_n^-(v) \leftarrow S_n^-(v) \cup \{c_k\}$
      $\mathcal{C}_{i-1} \leftarrow \mathcal{C}_{i-1} \cup N^-(c_k)$
    end if
  end for
  if `stop_at_this_level` **then** {Remove all cells at this time step from the subgraph and do not traverse further}
    $S_n^-(v) \leftarrow S_n^-(v) \setminus \mathcal{C}_i$
    break
  end if
  if `stop_after_this_level` **then** {Do not traverse further}
    break
  end if
end for

For any aggregated variable x_i, where i is the time step for which the variable is calculated, the relative change compared to time step t_0 is defined as

$$Qx_i = \frac{\Delta x_i}{x_0} = \frac{x_i - x_0}{x_0} \tag{2}$$

Note that with this definition, the relative change at t_0, Qx_0, is always zero, but the variable x_0 is not. In fact, Eq. 2 is not defined if $x_0 = 0$, and the relative change Qx_i can be misleading if x_0 is very small. For variables where very small or zero

Algorithm 2 Forward subgraph $S_m^+(v)$ selection for event node v.

$S_m^+(v) \leftarrow \varnothing$
$\mathcal{C}_1 \leftarrow N^+(v)$
for $i = 1$ to m **do**
  `stop_at_this_level` $\leftarrow$ **false**
  $\mathcal{C}_{i+1} \leftarrow \varnothing$
  for all $c_k \in \mathcal{C}_i$ **do**
    `included` $\leftarrow$ true
    if $|N^-(c_k)| > 1$ **and not** $N^-(c_k) \subset S_m^+(v)$ **then** {Cell is a merge and all parent cells are not included in the subgraph}
      `stop_at_this_level` $\leftarrow$ **true**
      `included` $\leftarrow$ **false**
    end if
    if `included` **then** {Add current cell to subgraph and its children to cells investigated at next iteration}
      $S_m^+(v) \leftarrow S_m^+(v) \cup \{c_k\}$
      $\mathcal{C}_{i+1} \leftarrow \mathcal{C}_{i+1} \cup N^+(c_k)$
    end if
  end for
  if `stop_at_this_level` **then** {Remove all cells at this time step from the subgraph and do not traverse further}
    $S_m^+(v) \leftarrow S_m^+(v) \setminus \mathcal{C}_i$
    break
  end if
end for

values are possible, it is recommended to either apply some transformation or redefinition of the variable or also study the absolute difference Δx_i.

Finally, information on subgraph development can be obtained from the values of x_i or Q_i. Note that the number of subgraphs at each time step $t_{-n}, \ldots, t_m$ is likely different, with the maximum number of samples existing at t_0, because every subgraph exists at least at time t_0 (event node time step) but not necessarily at the previous or next time steps. Therefore, results should be interpreted with care; the value Qx_i describes the relative change only in the samples of subgraphs for which cells could be selected from t_0 to t_i.

3 Case studies

We demonstrate the methodology using two case studies of rainfall events with merge-split cells, one from June 7, 2022, and another from July 13, 2021. An overview of the rainfall fields at the case study times are shown in Figure 3.

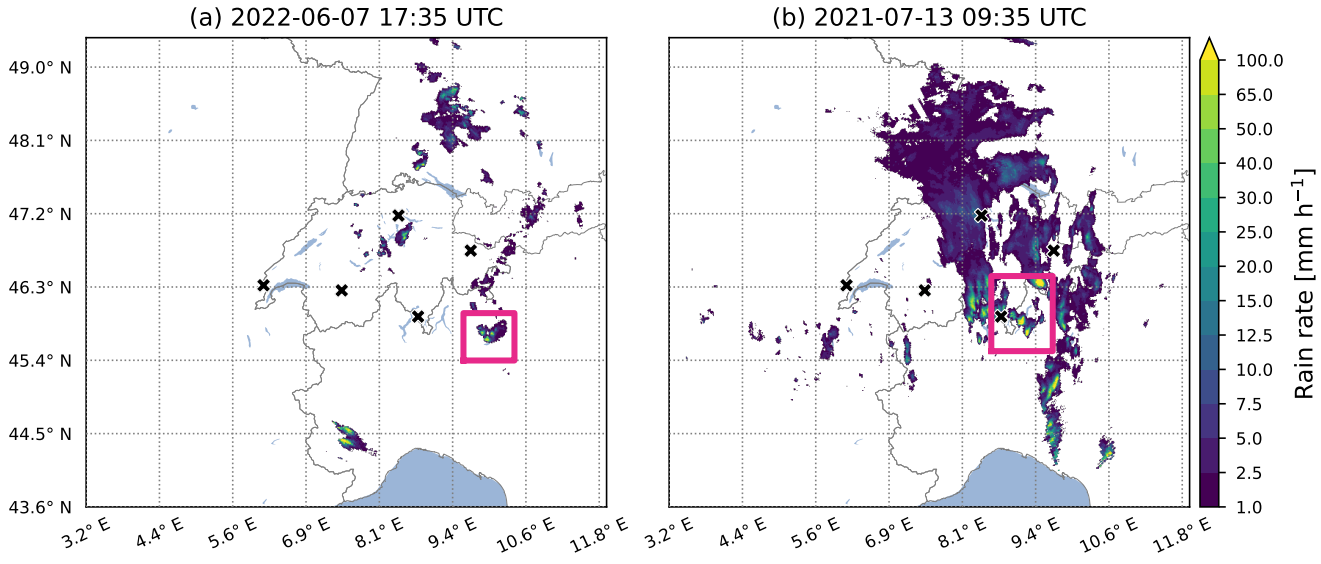

Figure 3. Overview of the rainfall field for the case studies at (a) June 7th 2022 17:35 UTC and (b) July 13th 2021 09:35 UTC. The magenta boxes indicate the area shown in (a) Fig. 4 and (b) Fig. 5. The crosses indicate radar locations, and the distance between the grid lines is 100 km.

The first case study subgraph on June 7, 2022, shown in Figure 4, occurs as a part of a cell track existing between 16:45 and 19:50 UTC, consisting of an isolated convective system with at most three cells identified at any time. The merge-split event node, for which the subgraph is constructed, occurs at 17:35 UTC when the two cells identified at 17:30 merge into one cell. This cell then splits into two cells at 17:40. After this, the cell labelled 2 continues to grow in both cell area and mean rain rate. The other cell, labelled 1, begins decaying and splits into two more cells at 17:50, which are no longer identified after 18:00. The opposing development of the two branches of the subgraph after the event node, one growing and one decaying, is captured in the time series graphs shown in Figure 6b-d. Additionally, Figure 6f shows the impact of the decaying subgraph branch on the total cell area: although the total cell area is consistently larger than at t_0, the relative change in total cell area decreases after 17:45 due to the decaying branch, while the relative increase in the volume rain rate, dominated by the growing subgraph branch, is much higher after 17:45.

The benefit of accounting for splits and merges in the analysis of even such a simple system is best observed when studying the Z_{dr}C signature in the cells (see Fig. 4). The Z_{dr} column first appears at 17:35 (in the event node) but is observed in the area of the cell that later splits into the decaying branch (labelled as 1) of the subgraph. In this decaying branch, the Z_{dr} column is observed only once more at 17:40. However, in the other branch, the Z_{dr} column feature is observed at all time steps between 17:40 and 18:05. Therefore, if the cell track were cut at the split occurring after 17:35 by defining the "most representative" path through the track (bold lines in Fig. 6b-e and h-l), the observed Z_{dr}C time series (Fig. 6f) for the growing branch of the subgraph (labelled 2) would not be complete, and the first time step at which the Z_{dr} column is detected would be delayed.

The second case study, shown in Figure 5, involves a merge-split event node from 09:35 UTC on July 13, 2021. This case study occurs on a day with significant convective activity in the domain. The case study event node is a part of a cell track that occurs between 08:25 and 10:40 UTC, with several other cell tracks in the surrounding area. As seen in Figure 5, the merge-split event consists of two cells, originally initiated independently, merging at 09:35 and then immediately splitting into two cells at 09:40. After this split, both branches of the subgraph split again after 15 minutes. Figure 6l shows that before the merge-split, both the total volume rain rate and area were increasing, mainly driven by growth in the cell labelled as 1. Peak values in total volume rain rate and area were obtained at 5-10 minutes after the event node, after which both begin decreasing as the cells decay.

Similar to the first case study, the benefits of considering splits and merges in the analysis are particularly evident in the Z_{dr}C (Fig. 5). If the "most representative" branch were selected at each split or merge event cell for analysis, and other branches considered either new or terminated, the Z_{dr}C signal, which now appears continuously through the whole subgraph (Fig. 6k), would be scattered across multiple tracks, and the full extent of the signal could easily be lost in the analysis. This would happen, in this case study, regardless of the method used to determine the "most representative" path through the cell track.

Even though the cell behaviour, including splits and merges and the exact time at which they are observed, clearly depends on the input data used, as well as the identification and tracking algorithms and their parameters, in this case study the splits and merges cannot be removed completely by algorithm design. For example, visually, the cell labelled 1 at 09:15 (Fig. 5) splits into two cells. Here the split is observed at 09:40, while some identification algorithm may have observed it earlier; however, it can be regarded as a *true* split. Thus, if we are interested in the development of the cell labelled 1 at 09:15 or the cells resulting from the split in a short time window, analysing the whole track together, including the split, provides the most information.

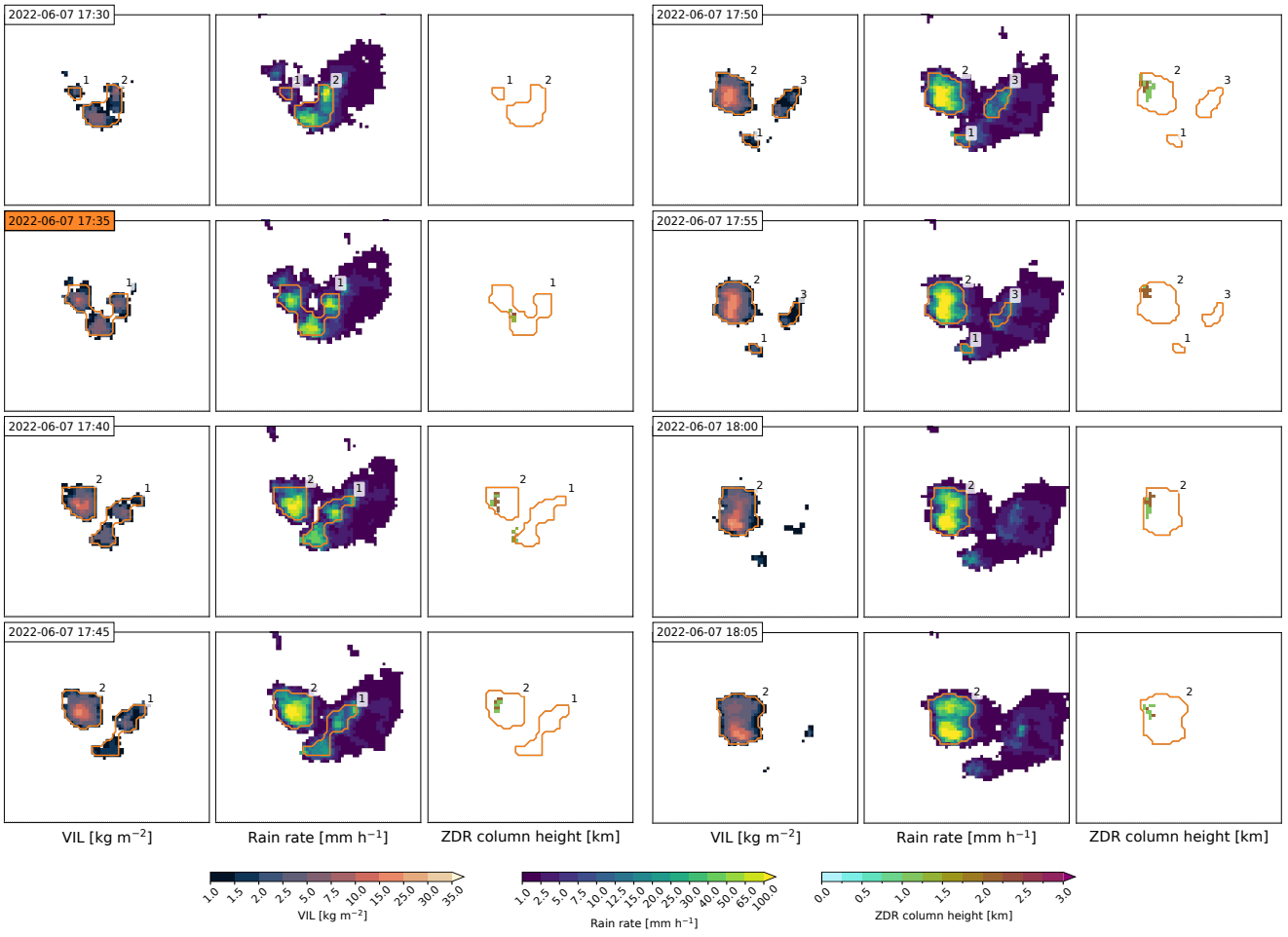

Figure 4. Time series of the case study on June 7th 2022 17:30 – 18:05 UTC. The t_0 time step (orange color background) is 17:35 UTC. For each time step, the left panel shows the vertically integrated liquid (VIL) the middle panel the rain rate, and the right panel the Z_{dr} column height fields. Cells outlined with orange are part of the case study subgraph, with the labels corresponding to the numbers in Fig. 6a-f. The panels show an area of approximately 70×65 km.

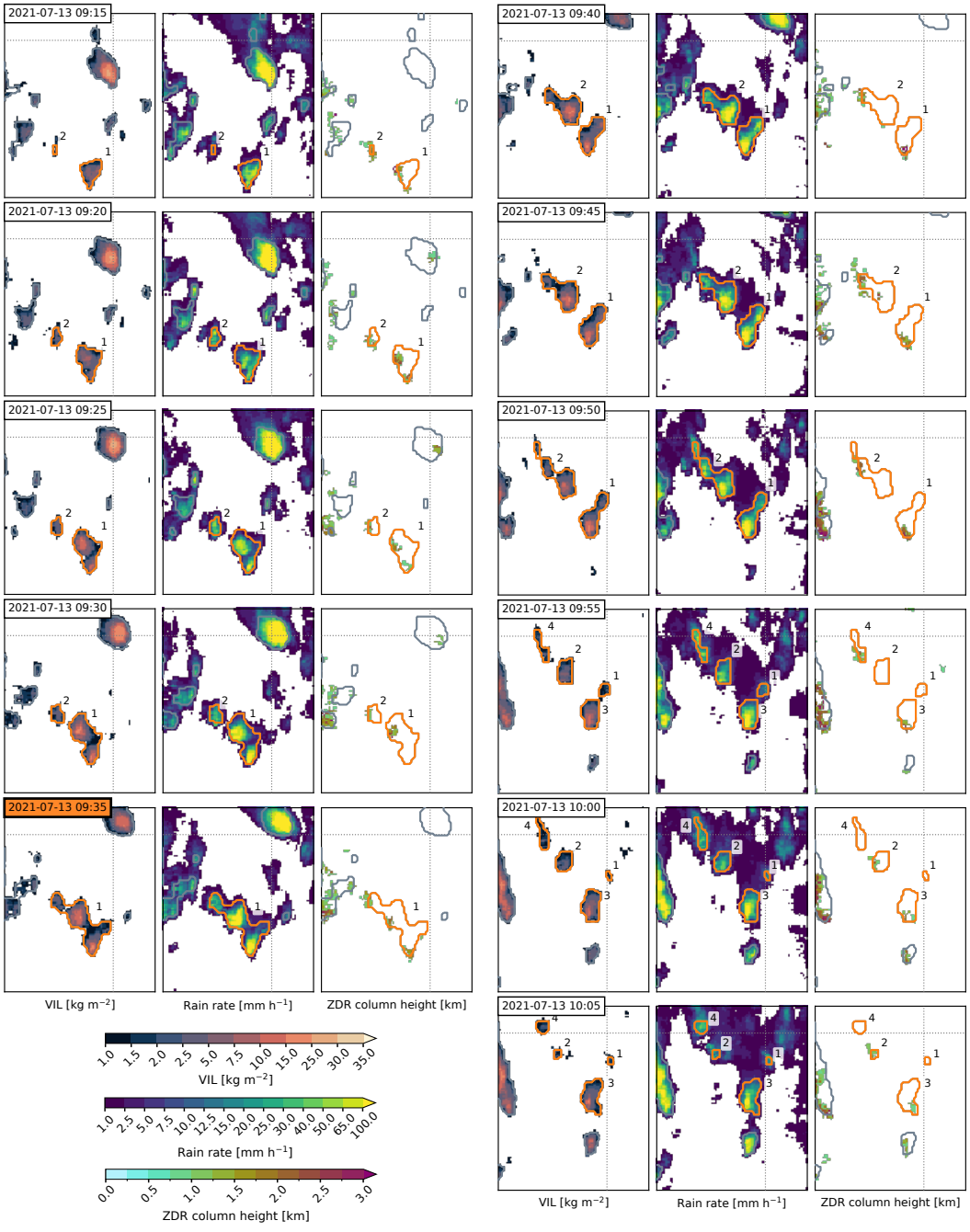

Figure 5. Time series of the case study on July 13th 2021 09:15 – 10:05 UTC. The t_0 time step (orange color background) is 09:35 UTC. For each time step, the left panel shows the vertically integrated liquid (VIL), the middle panel the rain rate, and the right panel the Z_{dr} column height fields. Cells outlined with orange are part of the case study subgraph, with the labels corresponding to the numbers in Fig. 6g-l. Cells outlined with grey are cells identified in the area that are not part of the case study subgraph. The panels show an area of approximately 85×100 km.

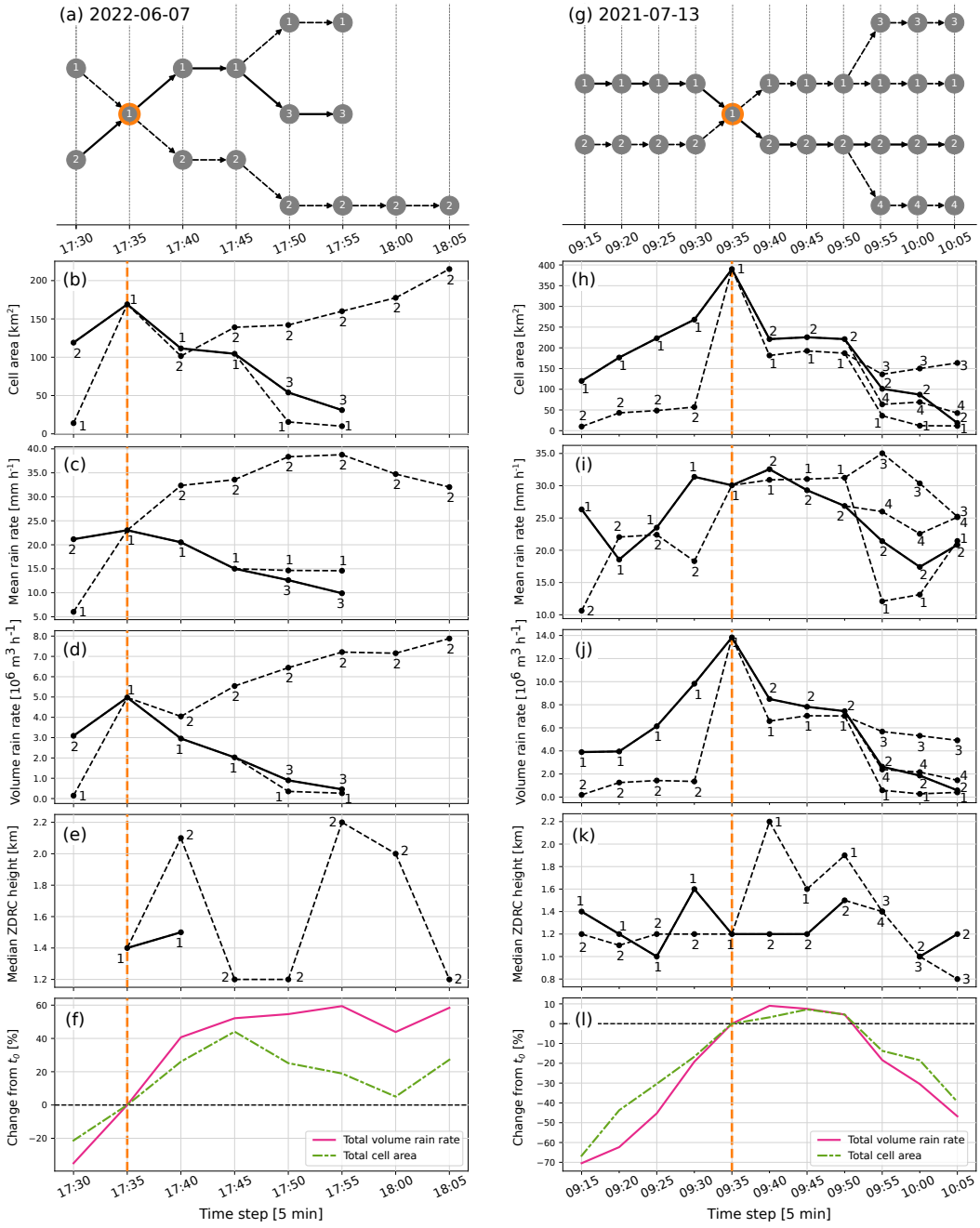

Figure 6. For the two case studies, (a, g) graph diagrams of the case study subgraphs, and time series of (b, h) cell area, (c, i) cell mean rain rate, (e, j) volume rain rate, (e, k) median Z_{dr} column height, and (f, l) relative change from t_0 time step in total volume rain rate and cell area. In (a) and (g), the orange-outlined node is the merge-split event node that the subgraph is selected for. In the time series (b-e) and (h-k), the orange dashed vertical line shows the t_0 time step, and the bold lines indicate the "most representative" cell track that would be obtained by continuing the track of the cell with the largest area at each split or merge. Note that some cells have no values for median Z_{dr} column height which causes parts of the cell tracks to be missing in panels (e) and (k).

4 Analysis of split and merge events

Next we apply the presented methodology to analyze the split and merge events in the 3-year dataset. To extract the dataset, we first identified and tracked convective cells in the radar data from May to September during 2021 to 2023. Cell track graphs were constructed for all cell tracks that lasted at least 10 minutes (i.e., two time steps). Details of the cell identification and tracking algorithms used in the study are provided in Appendix A. From this dataset, cells splitting at the next time step (split events), cells formed by a merge (merge events), and cells formed by a merge and splitting at the next time step (merge-split events) were selected as the event nodes, and a subgraph was extracted for each event node. The following section describes the split, merge, and merge-split events, and then investigates the convective cell development related to these events.

Note that, as with any cell-based analysis, the statistics and results presented are dependent on the selected cell identification and tracking methods and their parameters, as well as the data to which the algorithms are applied (e.g., climatology or data processing algorithms). Thus, the aim is not to present universal results about splits and merges, but rather to discuss factors that should be considered when analyzing the cell tracks and how these appear in the dataset. The cell identification and tracking methods used in this study were selected with the aim of analyzing factors contributing to convective rainfall in operational weather radar products; for example, cells are defined at a low VIL threshold to include more rainfall in the cells. This will impact, for example, at which stage the split or merge is identified and can thus impact how they appear in the results. The statistics presented in this section are valid for cell events at single timesteps and do not describe entire cell lifecycles.

4.1 Statistics of splits and merges in the cell dataset

When considering whether to include splits and merges in the analysis of convective cell development compared to selecting only a "most representative" path through the cell tracks, three questions should be considered. First: how many cells have splits and merges; second: among the cells involved in a split or merge events, is there a clear "most representative" cell to be selected, and third: is the impact of the other cells to the development negligible? In this section, we aim to examine these questions by analyzing the number of cells involved in splits or merges and their areas. Because the definition of a "most representative" cell or path through the track depends on the selected algorithm, here we simply use the cell area to characterize the impact of the cells in splits and merges, with the assumption that larger cells would have a greater impact to the event or development than smaller cells. However, this simplified analysis might not be directly comparable with algorithms that define the "most representative" path in the cell track using information from outside the cells involved in splits or merges (such as global features of the cell track; see, e.g., Liu et al., 2016).

In the dataset, the fraction of split or merge cells is 7.2% (53,059 out of 735,163) from all identified cells, with splits and merges comprising 3.0% (21,951) and 3.2% (23,391) of the dataset, respectively, and merge-split cells 1.0% (7,717). However, the fraction of splits and merges increases when looking at more intense cells. In cells with maximum VIL inside the cell at least 20 $\mathrm{kg\,m^{-2}}$, the fraction of splits or merges is 17.9% (6,770 out of 37,786). Furthermore, in cells associated with a Z_{dr}C feature, the fraction of splits or merges is 11.7% (21,364 out of 181,835). Thus, the impact of splits and merges to the analysis depends on the selection of the cells that the analysis focuses on.

One factor describing the complexity of a merge or split is the number of cells involved in the event. Figure 7a shows the number of cells involved in split or merge events. Approximately 90 % of split and merge events involve only two cells, and approximately 1 % of events include four or more cells. That is, most split events consist of one cell splitting into two and most merge events are two cells merging into one. Conversely, in merge-split events, only 70 % of the events involve two cells merging into one that then splits into two cells. In the remaining events, either the merge or split, or both, involve more than two cells. This indicates that merge-split events tend to be more complex than events with only a split or a merge, making it more difficult to define a "most representative" parent or child cell.

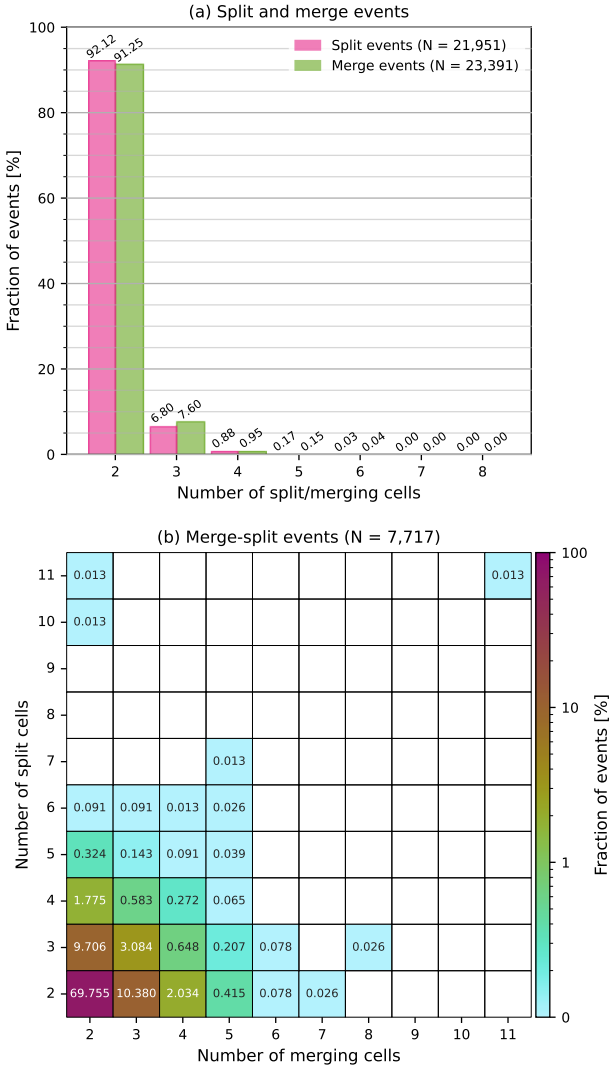

Figure 7. Number of cells participating in (a) split and merge events and (b) merge-split events. In (b) the x-axis shows the number of cells merging and the y-axis the number of cells resulting from the split in the merge-split event.

Figure 8 shows the histograms of cell areas for split, merge, and merge-split events, and for the cells involved in the splits and merges in the events compared to all identified cells. Note that for the cells involved in the splits or merges, the histograms only include the cells that are part of the subgraph of the split or merge event node. The split and merge event cells (Fig. 8a and b) tend to be significantly larger compared to all cells. These event cells are either cells that will split into multiple cells or cells formed by a merge of multiple cells. Because the split or merge event is captured at a random stage of the split or merge process, the event cells can be thought to consist of multiple cells identified as one, which makes them naturally larger than identified isolated cells. However, for the cells involved in split, merge, or merge-split events, the histograms of cell areas are similar compared to all cells, with slightly lower peaks at small areas, indicating that the split or merging cells tend to be only slightly larger than other cells.

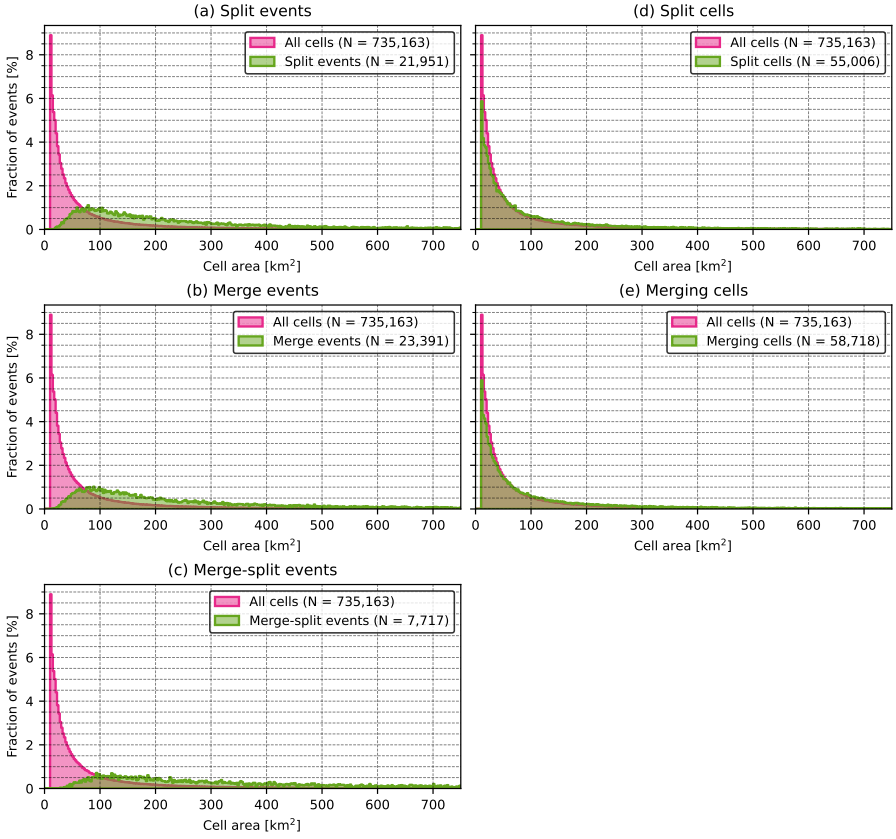

Figure 8. Histograms of cell area in (a) split events, (b) merge events, (c) merge-split events, (d) the cells merging in merge and merge-split events, and (e) split cells in split and merge-split events. The histogram is shown for the cells (a) before splitting in split events and (b) after merging in merge events. In merge-split events (c), the cell is the cell formed by the merge that has not yet split. Note that for the cells involved in the splits (d) or merges (e), the histograms only include the cells that are part of the subgraph of split, merge, or merge-split event.

If selecting the "most representative" cell in a split or merge event based on the cell area, one of the cells should be clearly larger and thus have a greater impact than the other cells. Figure 9 shows the ratio of the smallest cell area to the largest among the cells resulting from the split in split events and the cells merging in merge events. When the largest cell involved is larger than 100 km^2, the area of the smallest cell is likely less than 30 % of the largest cell's area, and the larger the largest cell, the smaller the ratio of the cell areas. For these split and merge events, the largest cell could be selected as the "most representative" cell likely without much loss of information. However, for events where the largest cell is smaller than 100 km^2, there is no clearly preferred size of the smallest cell relative to the largest, and selecting the "most representative" cell from cells of similar sizes is more likely to result in information loss in the "most representative" track. Additionally, approximately 40 % of events fall in the category of areas smaller than 100 km^2, and the number of events decreases in the higher area categories. This is also indicated in Figure 8d and 8e, where most cells are smaller than 100 km^2. Thus, in a large portion of split and merge events, the definition of the "most representative" track is questionable when looking at the immediate development of the cells.

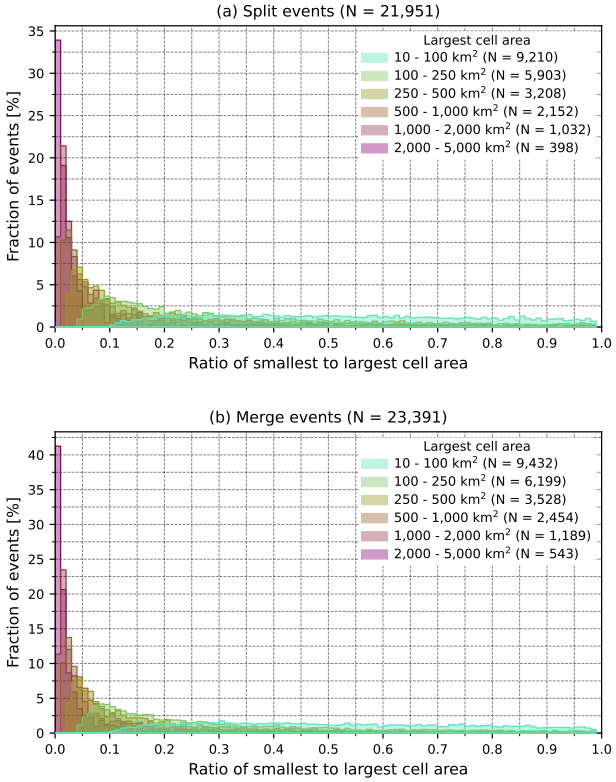

Figure 9. The area ratio of smallest to largest participating cell in (a) split and (b) merge events.

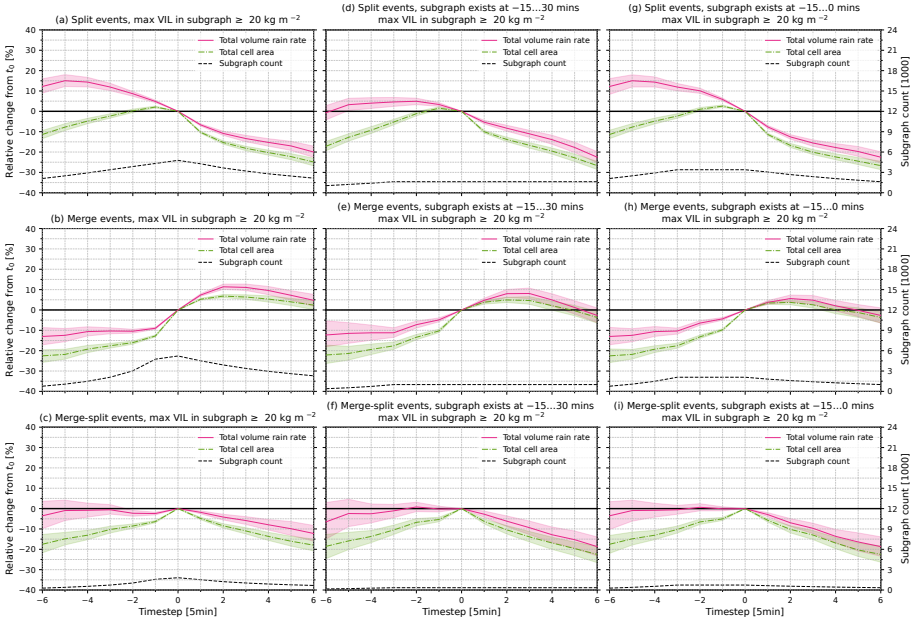

Figure 10. Mean relative change in total cell volume rain rate and total cell before and after (a, d, g) split, (b, e, h) merged, and (c, f, i) merge-split events. Panels (a-c) are calculated over all subgraphs in the dataset where the maximum VIL inside any cell is $\geq 20\ \mathrm{kg\,m^{-2}}$, while (d-f) are calculated over subgraphs that exist at least from -15 to 30 minutes, and (g-i) for subgraphs that exist at least from -15 minutes to t_0. The shaded areas indicate the 95 % confidence intervals of the mean.

4.2 Cell development surrounding splits and merges

Finally, we examine the development of the convective cells related to the events. Figure 10a-c illustrates the mean relative change in total volume rain rate and cell area within the -30 to 30 minute window surrounding the event. To emphasize the most severe convective cells, Figure 10 presents subgraphs where at least one cell has a maximum VIL value greater or equal to 20 $\mathrm{kg\,m^{-2}}$. For comparison, the same figure calculated over all subgraphs is shown in Figure A1.

Split events (Fig. 10a) appear to occur following a decrease in total volume rain rate, but an increase in total cell area. This suggests that splits occur during a decaying stage of cell development, where rainfall diffuses within the cells, resulting in smaller volume rain rates but larger cells. After the split, both the total volume rain rate and cell area decrease sharply. However, a part of the decrease occurring between t_0 and t_1 can be attributed to the area lost between the cells during the split.

Merge events (Fig. 10b), in contrast to split events, seem to occur after a sharp increase in both total volume rain rate and cell area between t_{-1} and t_0. This may be at least partly due to horizontal expansion of the identified cell into the areas between the merging cells. After the merge event, there is some short-lived increase in total volume rain rate and area, but on average, both begin decreasing slowly after 10 to 15 minutes.

In merge-split events (Fig. 10c), the trends in relative change before the event are similar to merge events and afterwards to split events; however, the rate of increase or decrease is lower. This suggests that, even though the merge-split events might

involve more cells than split and merge events, the development is still similar, and separate treatment for merge-split events might not be necessary.

When analyzing cell development in the events, it is important to consider the time window used in subgraph selection, as it can impact the results. As an example, the mean relative change is shown for two other samples: in Figure 10d-f, the relative change is averaged over subgraphs that exist at least between -15 and $+30$ minutes, and in Figure 10g-i, for subgraphs that exist between -15 minutes and t_0. This corresponds, for example, to a situation where a forecast model is trained on all samples that exist from -15 to $+30$ minutes, but during inference, the model is applied to all subgraphs that exist between -15 minutes and t_0 (as at inference, it is unknown if the subgraph exists beyond t_0). In this case, the model might, on average, slightly overestimate growth after merge events and underestimate decay after splits. This effect is more clearly observed when examining all events without filtering based on the maximum VIL in Figure A1.

Note that the existence of the subgraph is not equivalent to the existence of the cell track continuing from the event node: a subgraph existing at a certain time step means that we were able to select the subgraph at that time step, and, conversely, if a subgraph does not exist at some time step, it means that we were not able to select it. This could be either due to the subgraph selection being stopped previously, or because there are no cells in the cell track at that time step (relative to the event node for which the subgraph is selected). Thus, we should be careful when interpreting the results as they apply only to the subgraphs existing at the studied time steps, and conclusions cannot be drawn from the length of the time interval during which a subgraph exists.

5 Conclusions and summary

Cell identification and tracking algorithms are commonly used for analyzing the lifecycle of convective cells and nowcasting their future development based on weather radar data. However, splits and merges in the tracks complicate the use of these algorithms, especially for quantitative analysis of large datasets. To simplify the analysis, a common approach to handle the splits and merges is to define "most representative" tracks. However, in rapidly evolving convective systems, splits and merges of the cells are unavoidable, and ignoring them removes valuable information from the analysis. Previous studies have mainly focused on evolution of the cell tracks over their full life cycles, and not on how to address the splits and merges when analyzing the cell development in short time windows.

The aim of this study was to develop a methodology for quantitative analysis of convective cell development before and after selected events, allowing for the inclusion of splits and merges in the analysis, suitable for analysis across large datasets. In this methodology, cell tracks, created by identifying and tracking convective cells in radar data, are represented as directed graphs. *Event nodes*, whose development we are interested in, are then selected from the cell tracks. After this, for each event node, a subgraph of predecessor cells that are connected to the event node and successors cells the event node is connected to is selected according to the algorithms presented in the study, aiming to retain the available information in the subgraph at each time step. Once selected, the cell features in the subgraphs are aggregated over each time step to produce a time series.

We demonstrated the methodology by applying it to a dataset of cell tracks identified from operational weather radar products from the Swiss national radar network, using data from May to September over three years. The cell identification and tracking algorithms were applied to study total rainfall and its development in convective cells. To demonstrate the impact of splits and merges on cell development, all split, merge and merge-split cells were selected as the event nodes and subgraphs were constructed for them. We demonstrated the methodology with case studies, and examined through statistical analyses of the split and merge events why, in this dataset, including the splits and merges provides critical information for the analysis of cell development. It was documented that splits and merges occurred in 7.2% of all identified cells, and they were more frequent in cells with maximum VIL $\geq 20.0\,\mathrm{kg\,m^{-2}}$ (17.9%) or containing Z_{dr}C features (11.7%). Commonly used approach of determining the "most representative" cell in a split or merge is not well defined for small cells ($< 100\ \mathrm{km}^2$). In such cases we found no clear preference for a "most representative" cell among the cells involved in the split or merge. In cases where at least one of the cells is larger than $100\ \mathrm{km}^2$, the other cells tend to be significantly smaller. When looking at the cell development and how it is affected by merges and splits, we found that cell merges were associated with growth in total rainfall and cell area, while cell splits were associated with decrease in total rainfall.

The benefits of the proposed methodology are its versatility and scalability over large datasets. It can be applied to cell tracks obtained through any identification and tracking algorithms, allowing for use in various applications, and it can be applied to quantitatively analyze large datasets. However, even though explicitly including the splits and merges in the analysis can partially mitigate uncertainties caused by inconsistent cell identification and tracking, the results will, as with any cell-based analysis, depend on the cell identification and tracking algorithms used. Furthermore, the time window over which the subgraphs are analyzed impacts the available sample and results, influencing the conclusions drawn. The methodology cannot be applied to study full cell life cycles because the subgraph selection can be stopped even though the cell track would continue; rather, the focus is on cell development within a pre-determined short time window surrounding events of interest.

The proposed methodology can be further developed for more advanced analysis of cell development. While currently cell features in the subgraphs are aggregated into a single time series, they could be analyzed using algorithms specialized for graphs. For example, the past and future parts of the subgraphs can be separate graphs and could be, for example, clustered using graph clustering algorithms (Bunke, 2000; Günter and Bunke, 2002; Bunke et al., 2003) or graph neural networks (e.g. Xia et al., 2025) to match and group similar subgraphs. This would also allow for determining the similarity of two subgraphs, which could be used, for example, in nearest-neighbour or analogue-based nowcasting (Foresti et al., 2015; Shehu and Haberlandt, 2022). However, defining mappings between the subgraphs (Yunwen Xu et al., 2013), for example, to estimate similarity between to subgraphs is not straightforward due to the different sizes of the subgraphs and the lack of shared nodes. Thus, any mappings or definitions of similarity depend on which properties of the subgraphs are relevant for the application and how similarity between two cells is measured.

The proposed methodology can be used in various cell tracking applications. It can be applied, similarly to this study, to produce climatological information on cell development under certain conditions. Additionally, the constructed subgraphs can be expanded to consider also spatial connections between the cells, such as clustering or multiscale cell identification (e.g. Hou and Wang, 2017) to allow for the analysis of larger convective systems. The methodology could also be used to evaluate

how well grid-based nowcasting models reproduce the development of convective cells, similarly as in Ritvanen et al. (2025b), while accounting for both true splits and merges in the cell tracks as well as artificial splits and merges caused by blurring in the nowcasts. In addition to cell development, the subgraphs could be used to analyze the severity of cells related to hazards and to create nowcasting models to predict future hazard levels and produce warnings (e.g. Rossi et al., 2012; Tervo et al., 2019). Beyond convective cells, the methodology can be applied also to any dataset that can be represented as spatio-temporal graphs. The methodology presented in this study offers opportunities to significantly advance the knowledge of convective storm development and other systems with complex dynamics.

Appendix A: Convective cell identification and tracking algorithm

The cell identification and tracking algorithms applied in this study are simplified versions of what was applied by Rossi et al. (2012, 2015) and Tervo et al. (2019). The main changes in the algorithms are identifying the cells from vertically integrated liquid fields (VIL) instead of radar reflectivity, and not applying a spatial clustering to the cells.

For cell identification, the VIL field is first pre-processed by transforming the field into binary field using a threshold of 1.0 $\mathrm{kg\,m^{-2}}$. To reduce noise in the fields, morphological opening and closing operations with kernel sizes of 3×3 kilometres are applied to the fields. After this, the cells are identified by finding closed contours in the resulting binary field. All cells smaller than 10 $\mathrm{km^2}$ (i.e., 10 pixels) are discarded. The VIL threshold of 1.0 $\mathrm{kg\,m^{-2}}$ was selected as it is the lowest value discernable from background values (0.5 $\mathrm{kg\,m^{-2}}$) in the VIL product; in previous studies where cells have been identified from VIL fields thresholds as low as 0.01 $\mathrm{kg\,m^{-2}}$ have been used (Hu et al., 2019; Lamer et al., 2023; Tuftedal et al., 2024).

In cell tracking, at each time step first a motion field is determined between the previous time step t_{-1} and current time step t. The motion field is determined from the rainfall field, as that is expected to be more continuous than the VIL field and thus provide motion fields of better quality. The cells existing at time step t_{-1} are then advected to t using the mean velocity inside the cells as the velocity estimate, similarly as in Muñoz et al. (2018). After this, the overlapping area between each advected cell from t_{-1} with each cell at the current time step t is calculated. Any cell pair where the overlapping area is at least 10 % of the minimum of the two cells' area is considered a connection, and the cell at time step t is marked as a child cell of the cell from time step t_{-1}, and, vice versa, the cell at time step t_{-1} is marked as the parent cell of the cell at time step t.

Appendix B: Quality control of Z_{dr} column data used in case studies

The Z_{dr} column height (Z_{dr}C) fields used in the case studies in Section 3 were initially calculated by the algorithm presented by (Snyder et al., 2015). For more specific details regarding the implementation, for example, the pre-processing of Z_{dr} measurements or the environmental zero isotherm used in the calculation, we refer the reader to Aregger et al. (2025).

After the initial Z_{dr}C field, i.e. the maximum height of $Z_{\mathrm{dr}} \geq 1$ dB above environmental zero isotherm, is calculated, we apply additional quality control to the field to reduce noise caused by high Z_{dr} values most likely not related to convection. The requirements for a Z_{dr}C value in a grid point to be accepted are

1. MaxEcho $\geq 30\ \mathrm{dBZ}$
2. $800\ \mathrm{m} \leq Z_{\mathrm{dr}}\mathrm{C} \leq 3000\ \mathrm{m}$
3. $Z_{\mathrm{dr}}\mathrm{C} \leq \mathrm{ETML}_{20}$
4. Area of $Z_{\mathrm{dr}}\mathrm{C} \geq 5\ \mathrm{km}^2$

where MaxEcho is the maximum reflectivity in the vertical column in the grid point and ETML_{20} is the 20 dBZ echo top height above the environmental zero isotherm. The MaxEcho and echo top values are taken from the operational radar products produced by MeteoSwiss, and the environmental zero isotherm is the same as used in the $Z_{\mathrm{dr}}\mathrm{C}$ calculation.

Compared to the quality filtering applied in Aregger et al. (2025), our filtering is more cautious to account for the lower Z_{dr} threshold (1 dB vs. 2 dB) and the different definition of the convective cells used in the study. From the quality conditions, the MaxEcho condition (1) follows Aregger et al. (2025). The minimum and maximum height conditions (2) were set to remove $Z_{\mathrm{dr}}\mathrm{C}$ features most likely not related to updrafts. The echo top condition (3) was similarly used to remove areas where the high Z_{dr} occurs outside convective storms, e.g., caused by pristine ice crystals. Finally, the condition for the contiguous area of the $Z_{\mathrm{dr}}\mathrm{C}$ was set to remove spurious features and noise.

When assigning the $Z_{\mathrm{dr}}\mathrm{C}$ values to the case studies, we assign to each cell all $Z_{\mathrm{dr}}\mathrm{C}$ areas that the cell overlaps or touches. If a single $Z_{\mathrm{dr}}\mathrm{C}$ area touches or overlaps with multiple cells, it is assigned to the cell with whom it has the largest overlapping area.

Appendix C: Convective cell development in all split, merge, and merge-split events

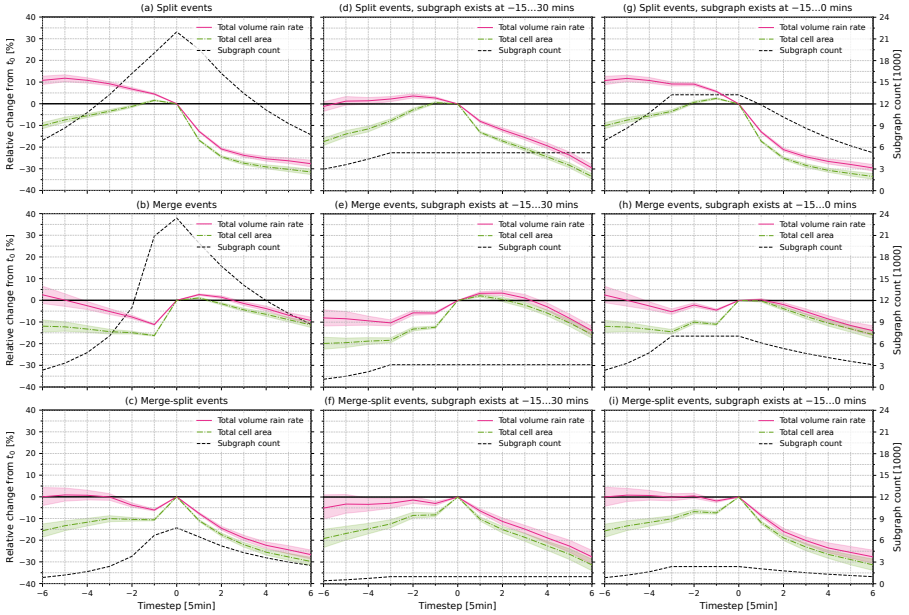

Figure A1. Same as Figure 10, but calculated without selecting events based on maximum VIL. Mean relative change in total cell volume rain rate and total cell before and after (a, d, g) split, (b, e, h) merged, and (c, f, i) merge-split events. Panels (a-c) are calculated over all subgraphs in the dataset, while (d-f) are calculated over subgraphs that exist at least from -15 to 30 minutes, and (g-i) for subgraphs that exist at least from -15 minutes to t_0. The shaded areas indicate the 95 % confidence intervals of the mean.

Code and data availability. The source code used to produce the results in this manuscript is available online at https://doi.org/10.5281/zenodo.17609981 (Ritvanen, 2025). The identified cells and their attribute data, as well as the cell tracks and subgraphs, and numerical versions of the result figures are available online at https://doi.org/10.57707/fmi-b2share.c857ccb10eb547d2a21384cc37ddaf7b (Ritvanen et al., 2025a). The original radar data used to identify and track the cells are not published with this manuscript. MeteoSwiss has decided to make its data publicly available under open-data. Implementation is currently underway. For more information refer to https://www.meteoswiss.admin.ch/services-and-publications/service/open-data.html.

Author contributions. JR, DM and SP conceptualized the study. JR developed the methodology with input from DM and SP, implemented the methodology, processed the data, performed the analysis and wrote the manuscript with input from all authors. MA produced the Z_{dr} column dataset. UG and AH provided the Swiss radar data. All authors have accepted the final version of the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Disclaimer. TEXT

Acknowledgements. The scientific colour maps described in Crameri et al. (2020) and Crameri (2023) were used in this study to prevent exclusion of readers with colour vision deficiencies. The colour maps are available at Crameri (2023).

Financial support. JR was funded by a grant from the Vilho, Yrjö and Kalle Väisälä Fund of the Finnish Academy of Science and Letters; SP was supported by the Research Council of Finland project PINCAST (grant no. 341964); and MA was supported by the Swiss Science Foundation project no. CRSII5 201792 J.K.

References

Aregger, M., Martius, O., Germann, U., and Hering, A.: Differential Reflectivity Columns and Hail: Linking C-band Radar-Based Estimated Column Characteristics to Crowdsourced Hail Observations in Switzerland, Quarterly Journal of the Royal Meteorological Society, 151, e5003, https://doi.org/10.1002/qj.5003, 2025.

Bluestein, H. B., McCaul, E. W., Byrd, G. P., Walko, R. L., and Davies-Jones, R.: An Observational Study of Splitting Convective Clouds, Monthly Weather Review, 118, 1359–1370, https://doi.org/10.1175/1520-0493(1990)118<1359:AOSOSC>2.0.CO;2, 1990.

Bournas, A. and Baltas, E.: Development of a Storm-Tracking Algorithm for the Analysis of Radar Rainfall Patterns in Athens, Greece, Water, 16, 2905, https://doi.org/10.3390/w16202905, 2024.

Brooks, H. E., Doswell III, C. A., Zhang, X., Chernokulsky, A. M. A., Tochimoto, E., Hanstrum, B., de Lima Nascimento, E., Sills, D. M. L., Antonescu, B., and Barrett, B.: A Century of Progress in Severe Convective Storm Research and Forecasting, Meteorological Monographs, 59, 18.1–18.41, https://doi.org/10/ggwphp, 2018.

Bunke, H.: Recent Developments in Graph Matching, in: Proceedings 15th International Conference on Pattern Recognition. ICPR-2000, vol. 2, pp. 117–124 vol.2, ISSN 1051-4651, https://doi.org/10.1109/ICPR.2000.906030, 2000.

Bunke, H., Foggia, P., Guidobaldi, C., and Vento, M.: Graph Clustering Using the Weighted Minimum Common Supergraph, in: Graph Based Representations in Pattern Recognition, edited by Hancock, E. and Vento, M., pp. 235–246, Springer, Berlin, Heidelberg, ISBN 978-3-540-45028-3, https://doi.org/10.1007/3-540-45028-9_21, 2003.

Cheng, Y.-S., Wang, L.-P., Scovell, R. W., and Wright, D.: Exploring the Use of 3D Radar Measurements in Predicting the Evolution of Single-Core Convective Cells, Atmospheric Research, 304, 107 380, https://doi.org/10.1016/j.atmosres.2024.107380, 2024.

Cormen, T. H., ed.: Introduction to Algorithms, MIT Press, Cambridge, Mass., 3. ed edn., ISBN 978-0-262-03384-8 978-0-262-53305-8, 2009.

Crameri, F.: Scientific Colour Maps (8.0.1), Zenodo, https://doi.org/10.5281/zenodo.5501399, 2023.

Crameri, F., Shephard, G. E., and Heron, P. J.: The Misuse of Colour in Science Communication, Nature Communications, 11, 1–10, https://doi.org/10/ghg5rd, 2020.

De Luca, D. L., Napolitano, Francesco, Kim, Dongkyun, Onof, Christian, Biondi, Daniela, Wang, Li-Pen, Russo, Fabio, Ridolfi, Elena, Moccia, Benedetta, and Marconi, F.: Rainfall Nowcasting Models: State of the Art and Possible Future Perspectives, Hydrological Sciences Journal, pp. 1–20, https://doi.org/10.1080/02626667.2025.2490780, 2025.

Dixon, M. and Wiener, G.: TITAN: Thunderstorm Identification, Tracking, Analysis, and Nowcasting—A Radar-based Methodology, Journal of Atmospheric and Oceanic Technology, 10, 785–797, https://doi.org/10/dc5g2t, 1993.

Feldmann, M., Germann, U., Gabella, M., and Berne, A.: A Characterisation of Alpine Mesocyclone Occurrence, Weather and Climate Dynamics, 2, 1225–1244, https://doi.org/10.5194/wcd-2-1225-2021, 2021.

Feng, Z., Leung, L. R., Houze Jr., R. A., Hagos, S., Hardin, J., Yang, Q., Han, B., and Fan, J.: Structure and Evolution of Mesoscale Convective Systems: Sensitivity to Cloud Microphysics in Convection-Permitting Simulations Over the United States, Journal of Advances in Modeling Earth Systems, 10, 1470–1494, https://doi.org/10.1029/2018MS001305, 2018.

Foresti, L., Panziera, L., Mandapaka, P. V., Germann, U., and Seed, A.: Retrieval of Analogue Radar Images for Ensemble Nowcasting of Orographic Rainfall, Meteorological Applications, 22, 141–155, https://doi.org/10.1002/met.1416, 2015.

Germann, U., Galli, G., Boscacci, M., and Bolliger, M.: Radar Precipitation Measurement in a Mountainous Region, Quarterly Journal of the Royal Meteorological Society, 132, 1669–1692, https://doi.org/10.1256/qj.05.190, 2006.

Germann, U., Boscacci, M., Clementi, L., Gabella, M., Hering, A., Sartori, M., Sideris, I. V., and Calpini, B.: Weather Radar in Complex Orography, Remote Sensing, 14, 503, https://doi.org/10.3390/rs14030503, 2022.

Greene, D. R. and Clark, R. A.: Vertically Integrated Liquid Water—A New Analysis Tool, Monthly Weather Review, 100, 548–552, https://doi.org/10.1175/1520-0493(1972)100<0548:VILWNA>2.3.CO;2, 1972.

Günter, S. and Bunke, H.: Self-Organizing Map for Clustering in the Graph Domain, Pattern Recognition Letters, 23, 405–417, https://doi.org/10.1016/S0167-8655(01)00173-8, 2002.

Guyot, A., Brook, J. P., Protat, A., Turner, K., Soderholm, J., McCarthy, N. F., and McGowan, H.: Segmentation of Polarimetric Radar Imagery Using Statistical Texture, Atmospheric Measurement Techniques, 16, 4571–4588, https://doi.org/10.5194/amt-16-4571-2023, 2023.

Handwerker, J.: Cell Tracking with TRACE3D—a New Algorithm, Atmospheric Research, 61, 15–34, https://doi.org/10/cvzsp7, 2002.

Heinselman, P. L., Burke, P. C., Wicker, L. J., Clark, A. J., Kain, J. S., Gao, J., Yussouf, N., Jones, T. A., Skinner, P. S., Potvin, C. K., Wilson, K. A., Gallo, B. T., Flora, M. L., Martin, J., Creager, G., Knopfmeier, K. H., Wang, Y., Matilla, B. C., Dowell, D. C., Mansell, E. R., Roberts, B., Hoogewind, K. A., Stratman, D. R., Guerra, J., Reinhart, A. E., Kerr, C. A., and Miller, W.: Warn-on-Forecast System: From Vision to Reality, Weather and Forecasting, 39, 75–95, https://doi.org/10.1175/WAF-D-23-0147.1, 2024.

Hering, A. M., Morel, C., Galli, G., Senesi, S., Ambrosetti, P., and Boscacci, M.: Nowcasting Thunderstorms in the Alpine Region Using a Radar Based Adaptive Thresholding Scheme, in: Proc. Third European Conf. on Radar Meteorology, pp. 206–211, ERAD, Visby, Sweden, 2004.

Hou, J. and Wang, P.: Storm Tracking via Tree Structure Representation of Radar Data, Journal of Atmospheric and Oceanic Technology, 34, 729–747, https://doi.org/10/f93gwb, 2017.

Hu, J., Rosenfeld, D., Zrnic, D., Williams, E., Zhang, P., Snyder, J. C., Ryzhkov, A., Hashimshoni, E., Zhang, R., and Weitz, R.: Tracking and Characterization of Convective Cells through Their Maturation into Stratiform Storm Elements Using Polarimetric Radar and Lightning Detection, Atmospheric Research, 226, 192–207, https://doi.org/10/ggk6g7, 2019.

Illingworth, A. J., Goddard, J. W. F., and Cherry, S. M.: Polarization Radar Studies of Precipitation Development in Convective Storms, Quarterly Journal of the Royal Meteorological Society, 113, 469–489, https://doi.org/10.1002/qj.49711347604, 1987.

Johnson, J. T., MacKeen, P. L., Witt, A., Mitchell, E. D. W., Stumpf, G. J., Eilts, M. D., and Thomas, K. W.: The Storm Cell Identification and Tracking Algorithm: An Enhanced WSR-88D Algorithm, Weather and Forecasting, 13, 263–276, https://doi.org/10/fb54r3, 1998.

Joss, J., Schädler, B., Galli, G., Cavalli, R., Boscacci, M., Held, E., Bruna, G. D., Kappenberger, G., Nespor, V., and Spiess, R.: Operational Use of Radar for Precipitation Measurements in Switzerland, vdf Hochschulverlag AG, ETH Zurich, Switzerland, 1998.

Kingfield, D. M. and Picca, J. C.: Development of an Operational Convective Nowcasting Algorithm Using Raindrop Size Sorting Information from Polarimetric Radar Data, Weather and Forecasting, 33, 1477–1495, https://doi.org/10.1175/WAF-D-18-0025.1, 2018.

Kumjian, M. R., Khain, A. P., Benmoshe, N., Ilotoviz, E., Ryzhkov, A. V., and Phillips, V. T. J.: The Anatomy and Physics of ZDR Columns: Investigating a Polarimetric Radar Signature with a Spectral Bin Microphysical Model, Journal of Applied Meteorology and Climatology, 53, 1820–1843, https://doi.org/10.1175/JAMC-D-13-0354.1, 2014.

Kyznarová, H. and Novák, P.: CELLTRACK — Convective Cell Tracking Algorithm and Its Use for Deriving Life Cycle Characteristics, Atmospheric Research, 93, 317–327, https://doi.org/10/chxgvk, 2009.

Lakshmanan, V. and Smith, T.: An Objective Method of Evaluating and Devising Storm-Tracking Algorithms, Weather and Forecasting, 25, 701–709, https://doi.org/10.1175/2009WAF2222330.1, 2010.

Lakshmanan, V., Rabin, R., and DeBrunner, V.: Multiscale Storm Identification and Forecast, Atmospheric Research, 67–68, 367–380, https://doi.org/10/cpf7f7, 2003.

Lamer, K., Kollias, P., Luke, E. P., Treserras, B. P., Oue, M., and Dolan, B.: Multisensor Agile Adaptive Sampling (MAAS): A Methodology to Collect Radar Observations of Convective Cell Life Cycle, Journal of Atmospheric and Oceanic Technology, 40, 1509–1522, https://doi.org/10.1175/JTECH-D-23-0043.1, 2023.

Lang, P.: KONRAD, Umweltwissenschaften und Schadstoff-Forschung, 14, 212–212, https://doi.org/10/b38x5t, 2002.

Li, J., Zheng, J., Li, B., Min, M., Liu, Y., Liu, C.-Y., Li, Z., Menzel, W. P., Schmit, T. J., Cintineo, J. L., Lindstrom, S., Bachmeier, S., Xue, Y., Ma, Y., Di, D., and Lin, H.: Quantitative Applications of Weather Satellite Data for Nowcasting: Progress and Challenges, Journal of Meteorological Research, 38, 399–413, https://doi.org/10.1007/s13351-024-3138-6, 2024.

Liu, J., Xue, C., Dong, Q., Wu, C., and Xu, Y.: A Process-Oriented Spatiotemporal Clustering Method for Complex Trajectories of Dynamic Geographic Phenomena, IEEE Access, 7, 155 951–155 964, https://doi.org/10.1109/ACCESS.2019.2949049, 2019.

Liu, W. and Li, X.: Life Cycle Characteristics of Warm-Season Severe Thunderstorms in Central United States from 2010 to 2014, Climate, 4, 45, https://doi.org/10.3390/cli4030045, 2016.

Liu, W., Li, X., and Rahn, D. A.: Storm Event Representation and Analysis Based on a Directed Spatiotemporal Graph Model, International Journal of Geographical Information Science, 30, 948–969, https://doi.org/10.1080/13658816.2015.1081910, 2016.

Merk, D. and Zinner, T.: Detection of Convective Initiation Using Meteosat SEVIRI: Implementation in and Verification with the Tracking and Nowcasting Algorithm Cb-TRAM, Atmospheric Measurement Techniques, 6, 1903–1918, https://doi.org/10/gb8sd5, 2013.

Muñoz, C., Wang, L.-P., and Willems, P.: Enhanced Object-Based Tracking Algorithm for Convective Rain Storms and Cells, Atmospheric Research, 201, 144–158, https://doi.org/10/gcvgxr, 2018.

Oue, M., Saleeby, S. M., Marinescu, P. J., Kollias, P., and van den Heever, S. C.: Optimizing Radar Scan Strategies for Tracking Isolated Deep Convection Using Observing System Simulation Experiments, Atmospheric Measurement Techniques, 15, 4931–4950, https://doi.org/10.5194/amt-15-4931-2022, 2022.

Picca, J. C., Kumjian, M., and Ryzhkov, A. V.: ZDR Columns as a Predictive Tool for Hail Growth and Storm Evolution, in: 25th Conf. on Severe Local Storms, p. 11.3, American Meteorological Society, Denver, CO, USA, 2010.

Rädler, A. T., Groenemeijer, P. H., Faust, E., Sausen, R., and Púčik, T.: Frequency of Severe Thunderstorms across Europe Expected to Increase in the 21st Century Due to Rising Instability, npj Climate and Atmospheric Science, 2, 1–5, https://doi.org/10/ggvkxt, 2019.

Raut, B. A., Jackson, R., Picel, M., Collis, S. M., Bergemann, M., and Jakob, C.: An Adaptive Tracking Algorithm for Convection in Simulated and Remote Sensing Data, Journal of Applied Meteorology and Climatology, 60, 513–526, https://doi.org/10/gh7m8x, 2021.

Ritvanen, J.: fmidev/convective-cell-graph-analysis: Graph-based Analysis of Convective Cell Development, Zenodo [code], https://doi.org/10.5281/zenodo.17540363, 2025.

Ritvanen, J., Aregger, M., Moisseev, D., Germann, U., Hering, A., and Pulkkinen, S.: Data for Manuscript "Analysis of Convective Cell Development with Split and Merge Events Using a Graph-Based Methodology" by Ritvanen et al., b2share [data set], https://doi.org/10.57707/fmi-b2share.c857ccb10eb547d2a21384cc37ddaf7b, 2025a.

Ritvanen, J., Pulkkinen, S., Moisseev, D., and Nerini, D.: Cell-Tracking-Based Framework for Assessing Nowcasting Model Skill in Reproducing Growth and Decay of Convective Rainfall, Geoscientific Model Development, 18, 1851–1878, https://doi.org/10.5194/gmd-18-1851-2025, 2025b.

Rosenfeld, D.: Objective Method for Analysis and Tracking of Convective Cells as Seen by Radar, Journal of Atmospheric and Oceanic Technology, 4, 422–434, https://doi.org/10.1175/1520-0426(1987)004<0422:OMFAAT>2.0.CO;2, 1987.

Rossi, P. J.: Object-Oriented Analysis and Nowcasting of Convective Storms in Finland, Doctoral thesis, Aalto University, Helsinki, Finland, ISBN 978-952-60-6441-3, 2015.

Rossi, P. J., Hasu, V., Halmevaara, K., Mäkelä, A., Koistinen, J., and Pohjola, H.: Real-Time Hazard Approximation of Long-Lasting Convective Storms Using Emergency Data, Journal of Atmospheric and Oceanic Technology, 30, 538–555, https://doi.org/10/f4rxhk, 2012.

Rossi, P. J., Chandrasekar, V., Hasu, V., and Moisseev, D.: Kalman Filtering–Based Probabilistic Nowcasting of Object-Oriented Tracked Convective Storms, Journal of Atmospheric and Oceanic Technology, 32, 461–477, https://doi.org/10/f652r2, 2015.

Shehu, B. and Haberlandt, U.: Improving Radar-Based Rainfall Nowcasting by a Nearest-Neighbour Approach – Part 1: Storm Characteristics, Hydrology and Earth System Sciences, 26, 1631–1658, https://doi.org/10.5194/hess-26-1631-2022, 2022.

Shi, Z., Wen, Y., and He, J.: A Clustering-Based Method for Identifying and Tracking Squall Lines, Atmospheric Measurement Techniques, 17, 4121–4135, https://doi.org/10.5194/amt-17-4121-2024, 2024.

Skinner, P. S., Wheatley, D. M., Knopfmeier, K. H., Reinhart, A. E., Choate, J. J., Jones, T. A., Creager, G. J., Dowell, D. C., Alexander, C. R., Ladwig, T. T., Wicker, L. J., Heinselman, P. L., Minnis, P., and Palikonda, R.: Object-Based Verification of a Prototype Warn-on-Forecast System, Weather and Forecasting, 33, 1225–1250, https://doi.org/10.1175/WAF-D-18-0020.1, 2018.

Snyder, J. C., Ryzhkov, A. V., Kumjian, M. R., Khain, A. P., and Picca, J.: A ZDR Column Detection Algorithm to Examine Convective Storm Updrafts, Weather and Forecasting, 30, 1819–1844, https://doi.org/10.1175/WAF-D-15-0068.1, 2015.

Steiner, M., Houze, R. A., and Yuter, S. E.: Climatological Characterization of Three-Dimensional Storm Structure from Operational Radar and Rain Gauge Data, Journal of Applied Meteorology, 34, 1978–2007, https://doi.org/10/d96zc8, 1995.

Taszarek, M., Allen, J. T., Brooks, H. E., Pilguj, N., and Czernecki, B.: Differing Trends in United States and European Severe Thunderstorm Environments in a Warming Climate, Bulletin of the American Meteorological Society, 102, E296–E322, https://doi.org/10.1175/BAMS-D-20-0004.1, 2021.

Tervo, R., Karjalainen, J., and Jung, A.: Short-Term Prediction of Electricity Outages Caused by Convective Storms, IEEE Transactions on Geoscience and Remote Sensing, pp. 1–9, https://doi.org/10/gf6hz6, 2019.

Tseng, C.-Y., Wang, L.-P., and Onof, C.: Modelling Convective Cell Life Cycles with a Copula-Based Approach, Hydrology and Earth System Sciences, 29, 1–25, https://doi.org/10.5194/hess-29-1-2025, 2025.

Tuftedal, K. S., Treserras, B. P., Oue, M., and Kollias, P.: Shallow- and Deep-Convection Characteristics in the Greater Houston, Texas, Area Using Cell Tracking Methodology, Atmospheric Chemistry and Physics, 24, 5637–5657, https://doi.org/10.5194/acp-24-5637-2024, 2024.

Utriainen, L., Virman, M., Laakso, A., Ritvanen, J., Jylhä, K., and Merikanto, J.: Less Frequent but More Intense Summertime Precipitation in Finland: Results from a Convection-Permitting Climate Model, Boreal Environment Research, 30, 93–109, https://doi.org/10.60910/BER2025.XV04-3N48, 2025.

Van Den Broeke, M. S.: Polarimetric Radar Metrics Related to Tornado Life Cycles and Intensity in Supercell Storms, Monthly Weather Review, 145, 3671–3686, https://doi.org/10.1175/MWR-D-16-0453.1, 2017.

Wang, H., Du, Y., Yi, J., Wang, N., and Liang, F.: Mining Evolution Patterns from Complex Trajectory Structures—A Case Study of Mesoscale Eddies in the South China Sea, ISPRS International Journal of Geo-Information, 9, 441, https://doi.org/10.3390/ijgi9070441, 2020.

Westcott, N.: A Historical Perspective on Cloud Mergers, Bulletin of the American Meteorological Society, 65, 219–226, https://doi.org/10.1175/1520-0477(1984)065<0219:AHPOCM>2.0.CO;2, 1984.

Wilson, M. B. and Broeke, M. S. V. D.: Using the Supercell Polarimetric Observation Research Kit (SPORK) to Examine a Large Sample of Pretornadic and Nontornadic Supercells, E-Journal of Severe Storms Meteorology, 17, 1–38, https://doi.org/10.55599/ejssm.v17i2.85, 2022.

World Meteorological Organization: Guidelines for Nowcasting Techniques, Tech. Rep. WMO-No.1198, World Meteorological Organization, Geneva, Switzerland, ISBN 978-92-63-11198-2, 2017.

Xia, Y., Chen, J., Li, X., and Gao, J.: DeepNM: Incremental Graph Matching Based on Sinkhorn Similarity, IEEE Transactions on Knowledge and Data Engineering, 37, 5141–5157, https://doi.org/10.1109/TKDE.2025.3583059, 2025.

Xue, C., Liu, J., Yang, G., and Wu, C.: A Process-Oriented Method for Tracking Rainstorms with a Time-Series of Raster Datasets, Applied Sciences, 9, 2468, https://doi.org/10.3390/app9122468, 2019a.

Xue, C., Wu, C., Liu, J., and Su, F.: A Novel Process-Oriented Graph Storage for Dynamic Geographic Phenomena, ISPRS International Journal of Geo-Information, 8, 100, https://doi.org/10.3390/ijgi8020100, 2019b.

Yan, J., Yin, X.-C., Lin, W., Deng, C., Zha, H., and Yang, X.: A Short Survey of Recent Advances in Graph Matching, in: Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval, ICMR '16, pp. 167–174, Association for Computing Machinery, New York, NY, USA, ISBN 978-1-4503-4359-6, https://doi.org/10.1145/2911996.2912035, 2016.

Yin, J., Pan, Z., Rosenfeld, D., Mao, F., Zang, L., Zhu, Y., Hu, J., Chen, J., and Gong, J.: Full-Tracking Algorithm for Convective Thunderstorm System From Initiation to Complete Dissipation, Journal of Geophysical Research: Atmospheres, 127, e2022JD037 601, https://doi.org/10.1029/2022JD037601, 2022.

Yunwen Xu, Salapaka, S. M., and Beck, C. L.: A Distance Metric between Directed Weighted Graphs, in: 52nd IEEE Conference on Decision and Control, pp. 6359–6364, IEEE, Firenze, ISBN 978-1-4673-5717-3 978-1-4673-5714-2 978-1-4799-1381-7, https://doi.org/10.1109/CDC.2013.6760895, 2013.

Zan, B., Yu, Y., Li, J., Zhao, G., Zhang, T., and Ge, J.: Solving the Storm Split-Merge Problem—A Combined Storm Identification, Tracking Algorithm, Atmospheric Research, 218, 335–346, https://doi.org/10/gf52zh, 2019.