Spatiotemporal Detection and Uncertainty Visualization of Atmospheric Blocking Events

Mingzhe Li ⓘD, Peer Nowack ⓘD, and Bei Wang ⓘD

Abstract— Atmospheric blocking events are quasi-stationary high-pressure systems that disrupt the typical paths of polar and subtropical air currents, often producing prolonged extreme weather events such as summer heat waves or winter cold spells. Despite their critical role in shaping mid-latitude weather, accurately modeling and analyzing blocking events in long meteorological records remains a significant challenge. To address this challenge, we present an uncertainty visualization framework for detecting and characterizing atmospheric blocking events. First, we introduce a geometry-based detection and tracking method, evaluated on both pre-industrial climate model simulations (UKESM) and reanalysis data (ERA5), which represent historical Earth observations assimilated from satellite and station measurements onto regular numerical grids using weather models. Second, we propose a suite of uncertainty-aware summaries: contour boxplots that capture representative boundaries and their variability, frequency heatmaps that encode occurrences, and 3D temporal stacks that situate these patterns in time. Third, we demonstrate our framework in a case study of the 2003 European heatwave, mapping the spatiotemporal occurrences of blocking events using these summaries. Collectively, these uncertainty visualizations reveal where blocking events are most likely to occur and how their spatial footprints evolve over time. We envision our framework as a valuable tool for climate scientists and meteorologists: by analyzing how blocking frequency, duration, and intensity vary across regions and climate scenarios, it supports both the study of historical blocking events and the assessment of scenario-dependent climate risks associated with changes in extreme weather linked to blocking.

Index Terms—Uncertainty visualization, atmospheric blocking events, feature detection, feature tracking, weather science, climate science, meteorology, visualization applications.

◆

---

1 INTRODUCTION

Atmospheric blocking events are large-scale, persistent weather patterns that occur in the mid-latitudes. Their defining feature is a quasi-stationary high-pressure system that lingers over a geographical region for several days (typically five or more), disrupting the usual paths of the polar jet streams [26, 36]. Blocking events are frequently associated with extreme weather, most notably summer heatwaves and winter cold spells [20, 55]. By altering the surrounding atmospheric circulation, they are also linked to upstream and downstream extremes, such as the devastating floods in Libya and central Greece in September 2023, which arose from an omega blocking event over Central Europe [37]. Analyzing trends in blocking event occurrences, both across continents and in climate change simulations, is therefore of critical importance to the scientific community and society at large.

However, detecting and characterizing blocking events in long atmospheric time series remains a challenging task. They are characterized by quasi-stationary high-pressure anomalies that remain over geographical regions for typically five or more days. While this may appear straightforward to define and detect at first glance, individual blocking events exhibit highly diverse characteristics [20], leading to ambiguity in their definition and detection. Blocking indices commonly define blocking in terms of different anomalies and gradients of key dynamical variables such as potential vorticity, geopotential height, potential temperature, and sea-level pressure. Consequently, the use of different blocking indices yields substantial discrepancies in both the frequency and spatial distribution of blocking events [46].

A further challenge lies in the representation of blocking in state-of-the-art climate models, which are essential for estimating future scenario-dependent changes in climate and atmospheric dynamics. Despite notable advances, including increases in spatial resolution [42],

climate models still exhibit systematic low biases in blocking frequency compared to observations [56].

In summary, tracking progress in the modeling and analysis of scenario-dependent future changes in blocking event frequency and characteristics is critical for understanding potential shifts in extreme weather events. This requires robust methods for the automated detection of blocking events in decadal- to centennial-scale climate simulations. Ideally, such methods should also capture detailed event characteristics, including the size of the high-pressure system, its horizontal and vertical structure, its persistence, and its seasonal and regional variability in likelihood of occurrence.

To address these challenges, Thomas et al. [46] recently developed expert-labeled ground truth datasets for atmospheric summer blocking events over Europe, based on approximately 41 years of ERA5 reanalysis observations [16] and a 101-year climate model simulation [43]. Using this dataset, they demonstrated that a machine learning approach based on self-organizing maps outperforms major blocking indices. However, their method does not provide detailed information on the spatiotemporal characteristics of blocking events, which are essential for long-term analyses of climatological changes in blocking.

Building on the work of Thomas et al. [46], we aim to advance methodologies for characterizing how blocking frequency, duration, and intensity vary across regions and climate scenarios, thereby enabling a more comprehensive assessment of the risks and spatiotemporal characteristics of prolonged extreme events. To that end, we propose an uncertainty visualization framework for detecting and characterizing variations in blocking events across multiple spatiotemporal scales. Our contributions are as follows:

- **Geometry-based detection and tracking.** We present a geometry-based pipeline that identifies and tracks high-pressure systems from preprocessed geopotential-height fields, producing results that are both interpretable and reproducible.
- **Improved detection quality.** Our framework achieves higher detection accuracy and F1-scores on both reanalysis (ERA5) and climate-model (UKESM) datasets compared to two baseline blocking indices, namely SOM-BI [46] and DG83 [8], when evaluated against expert-labeled ground truth.
- **Event-centric representation.** Unlike SOM-BI, which provides pattern labels rather than discrete events, our method explicitly detects and tracks individual blocking events, returning their spatial footprints (boundaries, areas, lifetimes, and trajectories) as the primary

- *Mingzhe Li is with the University of Notre Dame; this work was completed while at the University of Utah. E-mail: mli33@nd.edu*
- *Peer Nowack is with the Karlsruhe Institute of Technology. E-mail: peer.nowack@kit.edu*
- *Bei Wang is with the University of Utah. E-mail: beiwang@sci.utah.edu*

representation and backbone for subsequent analysis.

- **Uncertainty-aware summaries.** Building on the event representation, we introduce uncertainty-aware visualizations—including contour boxplots, frequency heatmaps, and 3D temporal stacks—that succinctly convey the typical structure, spread, location, and seasonal evolution of blocking events. To the best of our knowledge, these event-level uncertainty summaries represent a novel contribution to the study of blocking and complement existing blocking indices.

2 RELATED WORK

Atmospheric blocking indices. Atmospheric blocking is strongly associated with extreme weather events in the mid-latitudes. The primary feature of each blocking event is a quasi-stationary high-pressure system, which can be observed in surface pressure variations but typically even more clearly in other atmospheric dynamical variables used to characterize large-scale weather systems, such as mid-tropospheric geopotential height anomalies. Given the nature of high-pressure systems and the associated clear-sky conditions, blocking regionally leads to anomalously warm conditions during summer and cold conditions during brisk winter days. In addition, blocking events affect the general background atmospheric flow in their vicinity, leading to anomalous heat transport patterns. As a result, the hottest areas associated with blocking-related heatwaves are often found outside the main high-pressure system. Due to their effect on large-scale atmospheric flow, blocking can also lead to extreme rainfall elsewhere. For example, when an omega blocking event over Central Europe trapped a low-pressure storm over central Greece, it led to days of extreme rainfall and subsequent catastrophic flooding in September 2023 [37].

To better understand blocking and to investigate the influence of, e.g., changes in blocking frequencies and duration under climate change, meteorologists have developed a variety of approaches to automatically detect blocking in long atmospheric dynamics time series. Since manually labeling events seemed prohibitive, these methods mostly refer to several major blocking indices (BIs) that have been suggested to objectively identify events [8, 23, 31, 46, 47]. Notably, while previous comparisons of BIs show similar climatologies (i.e., maps of long-term average blocking frequencies across a spatial region), studies also found significant discrepancies. These have naturally led to questions concerning how to objectively measure skill in blocking detection, also depending on meteorological season [3, 7, 32, 46]. Therefore, Thomas et al. [46] introduced a new expert-labeled ground-truth dataset for summer blocking events over Europe. In their comparison of three widely used BIs, the index developed by Dole and Gordon (which we refer to as DG83) [8] performed the best, and they additionally demonstrated that a different approach using machine learning—based on self-organizing maps [21] to categorize weather states—led to further performance gains in blocking detection. They referred to their index as SOM-BI. In the following, we will use the ground-truth dataset developed in [46], taking the DG83 and SOM-BI methods as our baselines.

Other frequently used methods to study the climatology and characteristics of blocking include K-means clustering analyses to study weather regimes [6, 10, 27, 45, 49, 50]. In addition, there have been recent deep-learning efforts for detecting and analyzing blockings, including a two-stage CNN approach that detects and localizes blocking patterns in global climate model output [29] with a BI as ground truth, and a UNet-based model that aims to reconstruct summer blocking frequencies over the Last Millennium from tree-ring-based temperature fields [19]. A complementary line of research focuses on forecasting the blocking maintenance using deep learning models. For example, Zhang et al. [60] employed transfer learning to train a network that predicts whether a nascent Atlantic block will persist several days after its onset, using the DG83 index as ground truth. This forecasting objective differs from approaches that analyze full temporal sequences to detect and summarize complete blocking episodes.

Tracking atmospheric features. Many atmospheric features can be detected and analyzed via feature extraction and tracking. Beyond high-pressure systems contributing to blocking events, recent work has tracked cyclones and pressure-perturbations using topology-aware

segmentation, with robustness analyses across thresholds and resolutions [54, 58, 59]. Related pipelines track low-level cloud systems by extracting cloud objects/systems via thresholding and assembling trajectories across time [12, 24, 44], and track atmospheric rivers by identifying elongated moisture plumes and following them across successive time steps [35, 39]. Earlier efforts also track deep-convective clouds directly from satellite images using topological and computer-vision techniques [9, 11]. Across these lines, common ingredients of tracking atmospheric features are region/feature extraction, matching features on consecutive times, treatment of merges/splits, and explicit lifetime/size criteria. In this spirit, our blocking detection method adopts a simple region-overlap criterion to enforce quasi-stationarity of high-pressure systems while remaining easy to reproduce and compare.

Uncertainty visualizations for atmospheric events. Ensemble visualization methods can summarize how atmospheric events vary across members, times, or models. Wang et al. [51] classified these methods by data types, visual encodings, and analysis tasks, noting that ensembles inherently express uncertainty through multiple realizations of the same quantity; see other uncertainty-visualization surveys [18, 30, 33].

In meteorology and climatology, uncertainty visualizations have long supported ensemble analysis: Ensemble-Vis provides interactive overviews and statistical summaries to expose probabilistic characteristics in forecasting and climate applications [34]; time-varying weather ensembles have been examined for parameter sensitivity and resolution effects [4]; point/curve-based uncertainties have been shown with graduated glyphs and ribbons (Noodles) [40]; and uncertainty-aware visual analysis of atmospheric rivers with topological tools summarizes variability and occurrence across ensembles [22]. For feature-centric event summaries, where boundaries or tracks are the primary objects, contour boxplots compute order statistics on ensembles of contours to show a representative boundary (median) and central/outer envelopes; the approach extends to streamlines, pathlines, and curve boxplots, and has been used for hurricane tracks and weather-forecast ensembles [28, 53]. These techniques collectively support concise summaries of atmospheric events by revealing typical geometry, spread, and outliers, and they complement frequency heatmaps introduced in Sec. 5.

3 TECHNICAL BACKGROUND: CONTOUR BOXPLOT

Given an ensemble of contours, Whitaker et al. [53] introduced contour boxplots to give a statistically grounded, shape-aware summary. A contour boxplot encompasses a representative boundary, central spread, and outliers, analogous to a classic boxplot but tailored to contours.

We first review classic boxplot, and its usage in terms of summarization. A boxplot is a compact summary for a set of numbers: it shows the median along with range indicators for different percentiles to convey the spread of the set. A key step of constructing a boxplot is choosing a sensible center (i.e., the median). One way is to define the median by how often a value lies between pairs of other values, which we call the *band depth* of a value in a set. For 1D data, the band depth of a number is the fraction of all pairwise intervals that contain it; the number with the highest depth is the median. For example: for a set $\{2, 4, 5, 7, 12\}$, there are 10 pair intervals. The number 5 lies inside 4 of them, namely $\{2, 7\}$, $\{2, 12\}$, $\{4, 7\}$, and $\{4, 12\}$. Meanwhile, 4 and 7 lie inside 3 pairs, and 2 and 12 do not lie inside any pair. Thus, the number 5 has the greatest depth and is the median of the set.

A contour boxplot extends the boxplot idea from numbers to contours. Given an ensemble of contours, we assign each contour a band-depth score: for every pair of contours, take the region between them (their *band*); a contour's depth is the fraction of these pairwise bands that contain it. Contours that sit *between* many other pairs are considered to be more central; those that rarely fall between others are considered outliers. The contour boxplot encodes the distribution of contour shapes by ordering contours by their *band-depth*: the median contour is the one with the highest depth (most central); a central envelope (e.g., 50%) shows where the middle portion of contours typically lies; and an outer envelope (e.g., 100%) shows the full spread across all contours, with optional outlier flagged for very low-depth cases. In this way, a single figure communicates the typical position of the contour and the variability around it directly in the shape of contours.
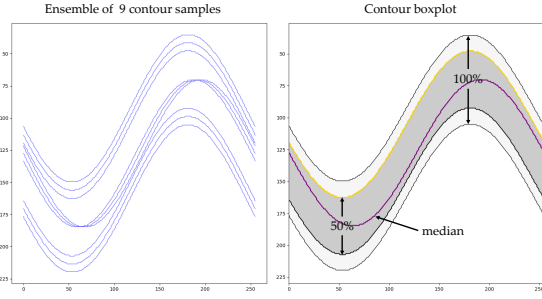
Fig. 1: An example of contour boxplot. Left: an ensemble of nine contours generated by sine waves shifted horizontally and vertically. Right: a contour boxplot showing the median contour (purple), the 50% central envelope (the inner band in gray), and the 100% envelope (the outer band in light grey).

We illustrate the concept with a simple ensemble of contours in Fig. 1. The left panel shows contours generated from sine waves shifted horizontally and vertically, while the right panel summarizes the ensemble with a contour boxplot. The median curve in purple (i.e., the most central contour by band depth) captures the central tendency. The inner gray band denotes the 50% central envelope, enclosing the five most central contours, while the outer band marks the 100% envelope, capturing the full spread of all nine contours and highlighting divergence at peaks and valleys. In short, the median provides a representative shape of the ensemble, while the envelopes convey its variability.

4 DATASETS AND PREPROCESSING

In this section, we describe the data preparation steps undertaken prior to the detection and characterization of blocking events.

4.1 Datasets

We utilize two datasets in our experiments. The first is **ERA5**, the fifth-generation global reanalysis from the European Centre for Medium-Range Weather Forecasts (ECMWF) [16], which serves as a proxy for observational data. ERA5 provides hourly estimates of a wide range of atmospheric state variables, including geopotential, sea-level pressure, and wind speed, at an original spatial resolution of $0.25° \times 0.25°$. For our analysis, we downsample the data to $1° \times 1°$ resolution by averaging over each 4×4 grid cell block.

The second dataset is derived from a pre-industrial climate simulation using the UK Earth System Model (**UKESM**) [43], conducted as part of the Coupled Model Intercomparison Project Phase 6 (CMIP6). The UKESM output (version UKESM1-0-LL) has a spatial resolution of $1.25° \times 1.875°$, corresponding to a global grid of 144×192 cells.

Typical atmospheric blocking events persist for several days; accordingly, we track high-pressure systems over multiple days. To improve robustness against small atmospheric perturbations, we compute *daily averages* for both datasets. For ERA5, we use data from 1979 to 2019, while for UKESM we draw from an arbitrarily defined timeline[1] in the pre-industrial simulations spanning 1960 to 2060, under constant climate forcing conditions.

Ground-truth datasets. Thomas et al. [46] created expert-labeled datasets of blocking events in Europe for both ERA5 and UKESM. In this paper, we treat these as ground-truth datasets (GTDs). These datasets provide binary labels indicating the presence or absence of blocking events during the June–July–August (JJA) period in a region of Europe ($30 - 75°$N, $10°$W$-40°$E), following the IPCC AR6 definitions [17]. The GTDs are constructed following the definition that blocking events persist for at least five consecutive days. The ERA5 GTD spans May 28 to September 4 for each year from 1979 to 2019, whereas the UKESM GTD covers May 27 to September 4 for the years 1960 to 2060, with UKESM employing a standardized 30-day calendar for each month.

---

[1]In the simulation, timeline is not real time but a simulation time index.

4.2 Data Preprocessing

Geopotential height. For both the UKESM and ERA5 datasets, we use the geopotential height at 500 hPa (Z_{500}) as a standard variable to characterize atmospheric circulation states. Geopotential height is obtained by dividing the geopotential field from ERA5 by the standard gravitational constant (≈ 9.80665, m/s^2). We focus on the 500 hPa level, a widely used diagnostic for blocking events [32, 46], as it lies sufficiently above the Earth's surface to minimize the influence of orographic effects and boundary layer noise, thereby highlighting large-scale circulation patterns.

Geopotential height anomaly. To better characterize weather anomalies relative to the expected atmospheric background at any given time of year, we analyze anomalies of geopotential height following previous work [15, 32, 46]. We first compute the *long-term daily mean* (LTDM) of the Z_{500} field at each grid point. Specifically, for each location $(\text{latitude}_i, \text{longitude}_i)$, we calculate the mean value of Z_{500} for each calendar day across the full dataset period—101 years for UKESM and 41 years for ERA5—referred to as the *daily mean*. This procedure yields a 365-day seasonal cycle at each location.

To smooth the LTDM fields, we approximate this seasonal cycle using the first six Fourier harmonics, thereby reducing noise caused by the limited sample size (e.g., 41 years in ERA5). The result is a smooth climatological background field that we subtract from each daily value to obtain daily Z_{500} anomalies (see Fig. 2 for an example).

Formally, we define the resulting *daily anomaly field* as $Z'_{500} = Z_{500} - \widehat{Z}_{500}^{\text{LTDM}}$, where $\widehat{Z}_{500}^{\text{LTDM}}$ is the smoothed LTDM field. Additionally, we linearly *detrend* the anomaly fields, following [46], to remove the influence of background warming in ERA5 and residual model drift in UKESM due to imperfect model spin-up to pre-industrial conditions.
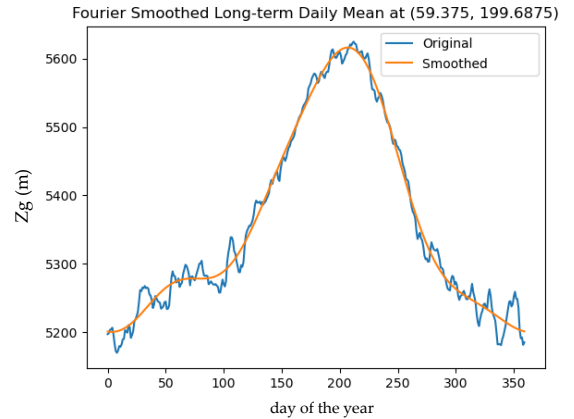

Fig. 2: ERA5 dataset: the long-term daily mean (blue) of the geopotential height (Zg, unit: meter or m) at 59.375°N, 160.3125°W and its smoothed curve (yellow) by keeping the first six Fourier harmonics.

Normalization of geopotential height anomaly. To make daily anomalies comparable across regions and seasons, we rescale them so that large values are judged relative to the local seasonal variability rather than against a single fixed threshold [32]. Starting from the anomaly field Z'_{500} defined above, we compute at each grid cell a *long-term daily standard deviation*, denoted $\widehat{\sigma}_L$, from the raw Z_{500} field: for each calendar day, we take the standard deviation across all years and then smooth the resulting 365-day cycle using the first six Fourier harmonics, consistent with the LTDM procedure.

We then normalize as $Z_{500}^{\text{norm}} = \frac{Z'_{500}}{\max(100, \widehat{\sigma}_L)}$, where the 100 m floor prevents artificially large ratios in regions or seasons with very small background variability (where division by a near-zero standard deviation would otherwise amplify noise and yield spurious exceedances). After this rescaling, values of Z_{500}^{norm} indicate how unusual a day is for that location and time of year. A candidate high-pressure system is then defined as any contiguous region where $Z_{500}^{\text{norm}} \geq \lambda$, allowing a single threshold λ to be applied consistently across regions and seasons.

5 METHOD

This section outlines our pipeline. We begin by tracking connected high-pressure systems using a geometry-based criterion (Sec. 5.1). Next, we identify blocking events as systems that persist for at least five consecutive days (Sec. 5.2), with parameter tuning described in Sec. 6. Finally, we introduce our main contribution: an uncertainty visualization framework for characterizing the spatiotemporal patterns of blocking events (Sec. 5.3).

5.1 Tracking High-Pressure Systems

Extracting high-pressure systems. Mid-latitude blocking is typically associated with quasi-stationary high-pressure systems in the normalized 500 hPa geopotential-height anomaly field Z_{500}^{norm}. For each day t (restricted to the European domain), we form the superlevel set $S_\lambda(t) = \{ x \mid Z_{500}^{\mathrm{norm}}(x,t) \geq \lambda \}$. A *superlevel-set component* is defined as a contiguous region within this set, representing the footprint of a candidate high-pressure system on that day.

Geometry-based tracking. Let $A \subseteq S_\lambda(t)$ and $B \subseteq S_\lambda(t+1)$ be components on consecutive days. We measure their (latitude-weighted) area overlap by $\Omega(A, B) = \sum_{x \in A \cap B} w(x)$, where $w(x) = \cos(\mathrm{lat}(x))$. We connect A to B if $\Omega(A, B) \geq C$ for some threshold C. Processing days sequentially yields trajectories that capture quasi-stationary evolution while permitting modest day-to-day displacements. Default values of parameters are reported in Sec. 6.

Ambiguity in system correspondences. Because systems may appear, disappear, merge, or split, correspondences need not be unique. We retain all trajectories satisfying $\Omega \geq C$; these plausible continuations are carried forward to the persistence test below.

5.2 Detecting Blocking Events

Persistence test and labels. Following Thomas et al. [46], a *blocking event* is present when a tracked system persists for at least five consecutive days. Let $\mathcal{T}$ be the set of trajectories constructed above, and let $D(\tau)$ denote the set of calendar days covered by a trajectory $\tau \in \mathcal{T}$. We label a day t as *positive* if

$$\exists \tau \in \mathcal{T} \text{ such that } \big|D(\tau)\big| \geq 5 \quad \text{and} \quad t \in D(\tau).$$

That is, *all* days covered by any trajectory whose duration is at least five consecutive days receive a positive label. Equivalently, whenever a trajectory contains any 5-day contiguous span, every day on that trajectory is marked positive; e.g., a system lasting July 1–8 renders July 1–8 positive.

Robustness to merges and splits. Because high-pressure systems may appear, disappear, merge, or split, multiple trajectories can interact. We define a blocking event as any trajectory that remains continuous over a five-day window, without enforcing unique correspondences between systems. This approach preserves robustness to minor boundary changes, while the overlap threshold C enforces day-to-day quasi-stationarity. All the positively labeled components constitute our ensemble of *blocking footprints*; their boundaries are the *blocking boundaries* used in Sec. 5.3.

5.3 Uncertainty Visualization

While the tracking of high-pressure systems is relatively standard, relying on a region-overlap criterion, the detection of blocking events requires careful parameter tuning (as detailed in Sec. 6). Our primary contribution is an uncertainty visualization framework for characterizing the spatiotemporal patterns of blocking events.

Blocking footprints can vary across days, months, and years, even on the same calendar date. To convey their typical locations and variations over time, we summarize ensembles of blocking events using three uncertainty-aware visual encodings:

- **Contour boxplots** that capture blocking boundary variations;
- **Frequency heatmaps** that encode occurrences;
- **3D temporal stacks** that situate these patterns in time.

To capture the frequency, duration, and intensity of blocking events across different spatiotemporal scales, we construct ensembles of varying types for each dataset. In the ERA5 dataset spanning 41 years, for example, a fixed calendar day (e.g., July 4) defines a *daily ensemble* that aggregates all blocking events detected on that day across the 41-year period. Likewise, a fixed calendar month (e.g., July) defines a *monthly ensemble* that collects all blocking events detected throughout that month across the 41 years. Similarly, a summer season (e.g., June, July, and August) gives rise to a *seasonal ensemble* that contains all blocking events detected throughout the season in the 41-year period.

Contour boxplots. Given an ensemble of blocking events, we collect all blocking boundaries into a single set of contours $\mathcal{G} = \{\gamma_1, \ldots, \gamma_n\}$. We then compute:

1. **Pairwise bands.** For each contour γ_i, iterate over all unordered pairs (γ_j, γ_k) for all $j < k$, and form the *band* between γ_j and γ_k, that is, the region lying between the two blocking boundaries within the domain.
2. **Mismatch matrix.** For each band, measure the fraction of γ_i that lies outside that band, and store this information in a mismatch matrix M (with elements valued within $[0, 1]$).
3. **Relaxed depth with ε.** We treat γ_i as *inside* a band if its mismatch for that pair is $\leq \varepsilon$. The *relaxed band-depth score* $D_\varepsilon(\gamma_i)$ is defined as the fraction of all pairs for which γ_i is labeled as *inside*. We then sweep over a small grid of ε values and select the smallest ε that stabilizes the ranking of the most central contours and produces a well-defined central region, as determined by the values of D_ε.
4. **Contour boxplot.** The following three elements constitute the contour boxplot we display:

- The **median** is the contour with the largest relaxed band-depth D_ε.
- The **50% envelope** is the union-minus-intersection of the deepest half of contours quantified by D_ε.
- The **100% envelope** is the union-minus-intersection of all contours.

See [53] for more details on implementing the contour boxplot, including tuning the parameter ε.

Frequency heatmaps. For a given ensemble of blocking events, we construct a frequency heatmap by counting, at each grid cell, the number of days in which that cell lies within a detected blocking footprint. The resulting heatmap encodes the frequency of blocking events across the ensemble, highlighting where blocking events are most likely to occur within a daily or monthly ensemble.

The frequency map highlights where blocking events repeatedly occur and how concentrated they are, while the contour boxplot summarizes the *representative* boundary (median) and the *spread* of boundaries (50% and 100% envelopes). Taken together, the median provides a representative footprint, the 50% envelope identifies the common core, the 100% envelope captures the full spatial extent observed in the ensemble, and the frequency heatmap indicates the likelihood of blocking events.

As a concrete example, Fig. 3 summarizes the daily ensemble for June 5 in the ERA5 dataset (1979–2019). In 23 of the 41 years, June 5 was classified as a (positive) blocking day. The frequency heatmap shows, across those 23 years, how often each location fell within a detected blocking footprint. The left panel overlays the 23 extracted block boundaries, while the right panel summarizes them with a contour boxplot. In this boxplot, the median contour is the most central member of the ensemble and serves as the representative. The 50% contour encloses the central half of all boundaries, while the 100% contour traces the union of all boundaries. For the June-5 daily ensemble, the 100% contour delineates the full spatial extent of detected blocking events, with its southern edge marking the farthest south any June-5 blocking event has reached.

3D temporal stacks. As shown in Fig. 3, a contour boxplot combined with a frequency heatmap summarizes the variability of blocking events for a single calendar day across 41 years, but it does not capture how their shapes and occurrences evolve throughout the season. To explicitly incorporate temporal information, we construct two complementary 3D temporal stacks along the time dimension for a given ensemble: the *3D*
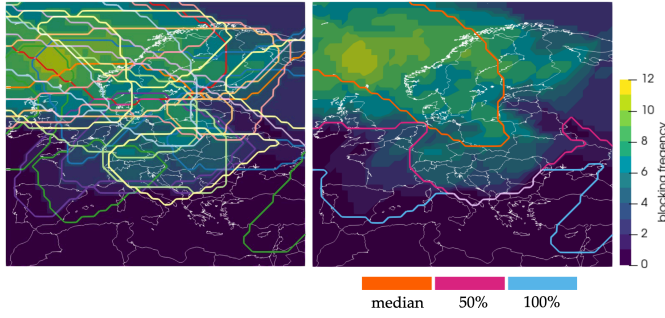
4

Fig. 3: ERA5 dataset (1979–2019), June 5 daily ensemble. Left: the boundaries of 23 ensemble members, which are blocking events detected across 41 years. Right: the contour boxplot of the ensemble, including the median blocking boundary (orange), 50-percentile central envelope (meganta), and 100-percentile envelope (blue). The background frequency heatmap highlights blocking frequency.

median stack and the *3D frequency stack*.

To construct the 3D median stack, we extract the median contour from each day's contour boxplot (the representative boundary) and place it in a 3D scene, with latitude and longitude on the horizontal axes and calendar time on the vertical axis. Stacking these daily medians produces a filament-like sketch of how a representative footprint shifts and persists over time. To reduce visual clutter, we restrict the stack to median contours without envelopes.

To construct the 3D frequency stack, we compute a frequency heatmap for each day and stack them along the same time axis to form a 3D volume, where higher values indicate dates and locations with more frequent blocking events.

Fig. 4: 3D temporal stacks of a seasonal ensemble (ERA5 dataset, May 28–Sep 4, 1979–2019). Left: a 3D median stack, where daily median contours are stacked by date along the z-axis (earlier dates at the bottom). Right: a 3D frequency stack, constructed by stacking daily frequency heatmaps into a 3D volume along the same z-axis, with color encoding frequency. The red box highlights a mid-June interval during which the median contours cluster over the west of Europe. Together, these 3D temporal stacks summarize the evolution and spatial persistence of blocking events throughout the season.

The two 3D temporal stacks provide a spatiotemporal overview that complements the 2D contour plots and frequency heatmaps. In the 3D median stack, clusters of vertically aligned median contours indicate multi-day concentrations of blocking events. In the 3D frequency stack, bright regions highlight periods and locations with frequent blocking events. These 3D temporal stacks are illustrated in Fig. 4.

6 IMPLEMENTATION

Our implementation is available at `https://github.com/tdavislab/atmos-blocking-uncertainty`.

Configurations. All experiments are conducted on a laptop with a 12th Gen Intel(R) Core(TM) i9-12900H 2.50 GHz CPU with 32 GB memory. Data processing, tracking, evaluation, and visualizations are

all implemented in Python. We use the *ParaView 5.13.3* [1] for the frequency heatmap and volume rendering of the 3D frequency stack.

Pipeline overview. According to Sec. 5, we first track high-pressure systems directly in the daily normalized 500 hPa geopotential-height anomaly field, $Z_{500}^{norm}(x, t)$. For each JJA day t, we threshold at level λ and take the connected components of the superlevel set $\{Z_{500}^{norm} \geq \lambda\}$ as system candidates. To track systems across time, we employ a region-overlap strategy by counting intersecting pixels weighted by their latitudes $w = \cos(\text{lat}(x))$, where $\text{lat}(x)$ is the latitude of the pixel x. Pairs of systems whose weighted overlap exceeds a threshold C are connected and thus producing time-ordered trajectories as days are processed sequentially. A day is labeled *positive* if it belongs to any trajectory lasting at least five consecutive days.

Parameter tuning. We tune two hyperparameters: the superlevel-set threshold λ and the overlap threshold C. Following DG83 [8], which sets $\lambda = 1.5$ for normalized Z_{500}, we conduct an exhaustive grid search with $\lambda \in \{1.0, 1.1, \ldots, 2.0\}$ and $C \in \{5, 6, \ldots, 40\}$. For each dataset, the first half of the years is used as a tuning set for five-fold cross-validation: the years are partitioned into five folds, each parameter pair (λ, C) is scored by training on four folds and validating on the held-out fold, and the pair maximizing the mean F1-score across folds is selected. The chosen (λ, C) is then fixed for all subsequent analyses. Because our detector is a transparent threshold-and-tracking procedure (i.e., a white-box index), we follow common practice in the blocking-index literature and report results on the full dataset rather than restricting to a separate test set [46]. The optimal set of parameters for the two datasets is

- ERA5: $\lambda = 1.2$, $C = 31$
- UKESM: $\lambda = 1.0$, $C = 31$

7 EXPERIMENTAL RESULTS

To demonstrate the utility of our framework, we first evaluate its performance in detecting blocking events, compared against two baseline methods: SOM-BI [46] and DG83 [8, 32]. The baselines are described in Sec. 7.1, and the evaluation results are presented in Sec. 7.2. We then provide a case study of the 2003 European heatwave in Sec. 7.3, showcasing our uncertainty visualizations. Finally, in Sec. 7.4, we summarize the spatiotemporal patterns of blocking events and visually analyze their variations.

7.1 Baseline Detection Methods

Both DG83 [8, 32] and SOM-BI [46] detect blocking events from persistent anomalies in the 500 hPa geopotential height field (Z_{500}).

DG83. In the original work of Dole and Gordon [8], Z_{500} anomalies are computed as departures from a long-term seasonal mean and rescaled by a latitude factor $(\sin 45° / \sin \phi)$ (ϕ: latitude) to account for the meridional variation of planetary vorticity. A grid point is considered *blocked* (i.e., part of a blocking footprint) if the scaled anomaly exceeds a fixed threshold (originally ≥ 100 m) for a sustained duration (originally ≥ 10 days).

For global intercomparison, Pinheiro et al. [32] modified DG83 by replacing the fixed 100 m threshold with a spatiotemporally varying threshold of 1.5σ (with a floor of 100 m), where σ is the smoothed long-term standard deviation of geopotential height at each grid point. Blocking events are matched across days when they overlap in at least one pixel and satisfy a duration constraint of ≥ 5 days; we adopt this modified DG83 by Pinheiro et al. as our baseline for comparison.

SOM-BI. SOM-BI [46] is a state-of-the-art blocking index based on geopotential height. It trains a self-organizing map (SOM) on daily Z_{500} anomaly maps over Europe to learn a compact set of representative circulation patterns, referred to as SOM nodes. During training, each day is assigned to its best-matched node, and the matched node together with its neighbors on the SOM grid are updated toward that day's field using a decaying learning rate and neighborhood size. This process is iterated until the node patterns converge, after which each day is mapped to its best-matched node.

SOM-BI then applies a 5-day sliding window over the dataset, treating each sequence of best-matched nodes as a candidate *blocking signature*. Using ground-truth labels of blocked 5-day periods, all distinct signatures are scored, and the subset that maximizes the F1-score is retained. A day is labeled to be *blocked* (i.e., exhibiting blocking event) if any 5-day window containing it matches one of the retained signatures.

Comparison with our approach. Both DG83 and our detection method apply a superlevel-set threshold to the normalized Z_{500} anomaly field, whereas SOM-BI employs an unsupervised pattern-classification framework based on a self-organizing map [8, 46]. The key difference is that our method explicitly enforces spatial stationarity: components from consecutive days must overlap by an area greater than C, ensuring that only slowly evolving, spatially coherent high-pressure systems are linked into a trajectory and counted toward the multi-day lifetime requirement. By contrast, DG83 requires only minimal day-to-day overlap (as little as a single pixel) between footprints. As a result, rapidly translating highs that do not constitute a blocking event may still satisfy DG83's duration rule, whereas our area threshold C enforces quasi-stationarity.

To maintain precision, DG83 typically relies on a relatively high anomaly threshold (e.g., 1.5σ), which helps reduce false positives but also makes the method less sensitive to moderate events. In contrast, our overlap criterion tolerates modest spatial displacements while still enforcing quasi-stationarity, thereby allowing the use of a lower λ without inflating noise.

On the other hand, SOM-BI differs in spirit: it first learns the representative of daily Z_{500} anomaly patterns and then labels days via the best-performing 5-day pattern sequences against the ground truth [46]. This can be robust to synoptic variability, but the classification hinges on SOM design and training labels. Besides, SOM-BI yields only binary labels rather than explicit boundaries of high-pressure systems. In contrast, our tracking directly produces contiguous footprints each day, which we use consistently for case studies involving contour boxplots and frequency heatmaps.

7.2 Blocking Event Detection Evaluation

Evaluation with ground-truth labels. We assess all blocking event detection methods against the ground-truth (GTD) labels. For DG83 and our detection method, scores are computed from our own implementation. For SOM-BI, we report the published results of Thomas et al. [46], obtained on the same datasets and timespan.

We evaluate day by day against the ground-truth labels using *accuracy*, *precision*, *recall*, and *F1*-score. A "positive" label means *blocked* and a "negative" label means *not blocked*. A day is a true positive (TP) if both the output label and GTD are positive; a true negative (TN) if both are negative; a false positive (FP) if the tool labels it positive but GTD is negative; and a false negative (FN) if the tool labels it negative but GTD is positive.

Detection summary. Standard confusion metrics are computed over the evaluation period, including accuracy, precision, recall and F1-score. These metrics are reported in Tab. 1 for ERA5 (reanalysis) and UKESM (simulation) datasets across the three detectors (DG83, SOM-BI, and ours). For **ERA5**, all methods achieve comparable accuracy and precision, but our method attains the highest recall and therefore the top F1-score. For **UKESM**, our method performs consistently strongest overall: its precision is on par with DG83, while recall is substantially higher, yielding a higher F1-score.

This result reflects our design choices: unlike DG83, which relies on a relatively high anomaly threshold to safeguard precision, our method enforces quasi-stationarity through a region-overlap constraint, filtering out fast-moving high-pressure systems before they can contribute to false positives. This constraint enables the use of a more permissive anomaly threshold without sacrificing precision (ERA5: $\lambda = 1.2$ vs. 1.5 in DG83; UKESM: $\lambda = 1.0$ vs. 1.5), while recovering additional modest events and improving recall.

Relative to SOM-BI, we achieve similar or higher F1-score on both datasets and, moreover, produce day-by-day contiguous footprints of the high-pressure systems, which feed directly into our downstream

analysis. In comparison, SOM-BI yields a binary time series and typically relies on composite of additional heatmaps (such as Z_{500}, vertically integrated potential vorticity) for spatial context.

Disagreement with DG83. Because both DG83 and our method apply thresholds to the normalized Z_{500} field, we examine cases where their GTD-based labels disagree (Tab. 2). With a more lenient normalized anomaly threshold λ, our method correctly identifies 161 additional blocked days in ERA5 and 440 in UKESM that DG83 misses, at the cost of introducing more false positives (161 in ERA5 and 340 in UKESM) that DG83 correctly marks as unblocked. This reflects the standard precision-recall trade-off: lowering λ increases sensitivity (recall) but can also admit marginal cases.

Two design choices temper this loss of precision in our framework: the region-overlap requirement C and trajectory building, which filter out fast-moving highs and favor quasi-stationary systems. Consistent with this, DG83 flags only 14 (ERA5) and 109 (UKESM) blocked days that our method does not, which are likely intensified systems that failed our stationarity and lifetime criteria. In contrast, our method correctly screens out 100 (ERA5) and 305 (UKESM) days as unblocked that DG83 labels as blocked.

Overall, relative to DG83, our approach shifts the balance toward higher recall for modest or slightly drifting events, while maintaining competitive precision through the overlap threshold C.

Prevalence and bias. While our precision remains comparable to the baselines, the number of days labeled as blocked by our detector is slightly higher than the GTD base rate in ERA5 (37.3% vs. 33.4%), indicating a mild positive bias toward detection. In UKESM, our prevalence closely matches GTD (29.4% vs. 29.0%), whereas DG83 under-detects (25.9%). Taken together, these results suggest our method trades a small increase in prevalence for meaningful gains in recall, with precision preserved by the explicit overlap-based stationarity check.

Potential improvement. Among the three blocking event detection tools, only our method reaches a considerable difference between the precision and recall on the ERA5 dataset, indicating bias towards over-detection than under-detection. This is because our parameter tuning process prioritizes maximizing the F1-score on the validation data without much consideration about the difference between precision and recall. We may improve the parameter tuning process to find the parameter setting that minimizes the difference between precision and recall to better balance the ratio of detected blocked days.

Out-of-season agreement. Although λ and C are tuned using only JJA (summer) cases, we also test generalization to the rest of the year. To mitigate seasonal amplitude differences, we pre-process Z_{500} anomalies using the global/seasonal normalization introduced by Pinheiro et al. [32], and we compare our monthly outputs against DG83 implemented with the same normalization; see Sec. 4.2 for the normalization details. Tab. 3 reports monthly precision, recall, and F1-score for this pairwise agreement.

For **ERA5**, within JJA (the tuning season), the F1-score agreement with DG83 is 0.885 (June), 0.855 (July), and 0.806 (August). Outside JJA, the lowest monthly F1-score is 0.777 (November), with the next-lowest 0.840 (January). All values are comparable to the in-season range, indicating broadly similar alignment across months, with November showing the largest divergence.

For **UKESM**, a similar pattern holds: in JJA the F1-scores are 0.756 (June), 0.812 (July), and 0.802 (August); outside JJA, the lowest F1-score occurs in October (0.710), followed by May (0.780). Across all the months aside from October, there are no pronounced spikes or dips in F1-score. Overall, the absence of strong seasonal variations in monthly F1-score suggests that, with the same normalization applied, our method maintains a generally consistent level of agreement with DG83 beyond the season used for tuning.

7.3 Case Study: 2003 European Heatwave

The 2003 European heatwave is a canonical benchmark for blocking event detection: from late July into mid-August, a persistent, quasi-stationary anticyclonic circulation settled over Western Europe, displacing the North Atlantic jet poleward, suppressing precipitation, and

Table 1: Detection statistics across three detection methods, with respect to the ground-truth labels for both the ERA5 and the UKESM datasets during the JJA period. The best performance for each metric is in bold. In the ground-truth labels, 33.4% days are blocked for the ERA5 dataset (1979-2019) and 29.0% days are blocked for the UKESM dataset (1960-2060), respectively.

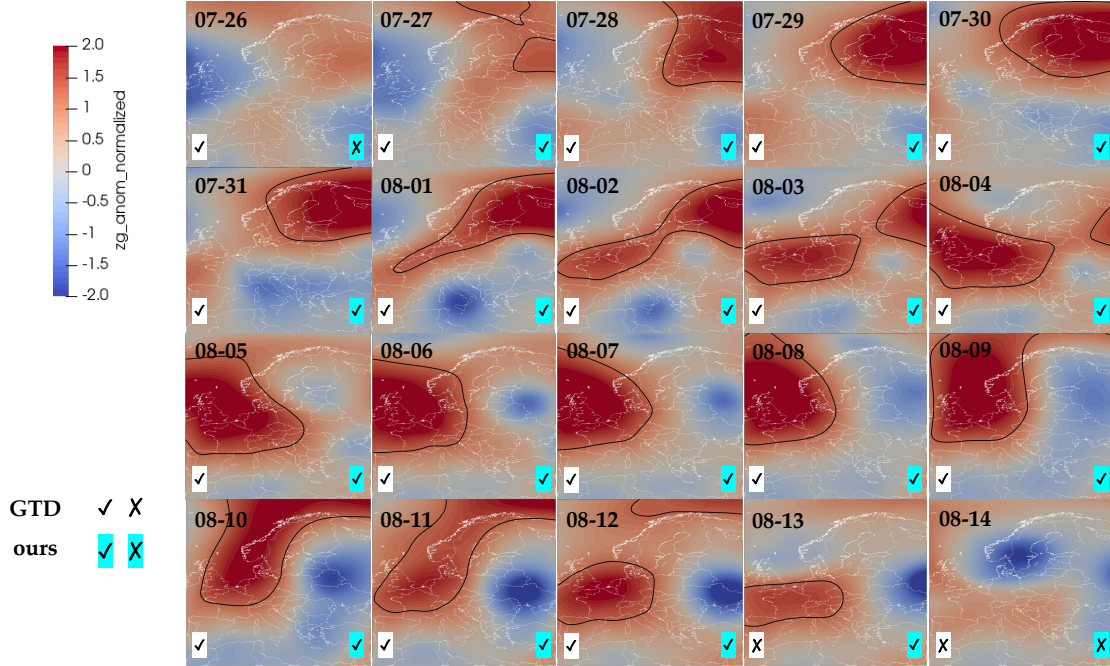| Methods	ERA5					UKESM				
	Days blocked	Accuracy	Precision	Recall	F1-score	Days blocked	Accuracy	Precision	Recall	F1-score
DG83	32.3%	0.82	0.72	0.72	0.72	25.9%	0.87	0.80	0.72	0.75
SOM-BI	34.6%	0.82	**0.73**	0.75	0.74	29.6%	0.83	0.70	0.72	0.71
ours	37.3%	**0.84**	0.72	**0.83**	**0.77**	29.4%	**0.90**	**0.81**	**0.83**	**0.82**

Fig. 5: ERA5 dataset: the frequency heatmap of normalized Z_{500} anomalies over Europe for July 26 - August 14, 2003. White lines show national borders and coastlines. The black contour (isovalue=1.2) delineates the high-pressure boundary as defined by our detection method. Bottom annotations in each map report our detection and the ground-truth labels (GTD) for each date; tick = blocked, cross = not blocked.

Table 2: The disagreement in detection results between DG83 and our method for data within the JJA period.

Dataset	GTD labels	only ours correct	only DG83 correct	both correct	both incorrect
ERA5	Blocked	161	14	935	213
	Not blocked	100	161	2204	312
UKESM	Blocked	440	109	1939	361
	Not blocked	305	340	6124	280

Table 3: The agreement of our detection results with respect to the DG83 detection results for different months.

Month	ERA5			UKESM		
	Precision	Recall	F1-score	Precision	Recall	F1-score
01	0.741	0.968	0.840	0.675	0.936	0.784
02	0.746	0.970	0.843	0.686	0.943	0.794
03	0.804	0.963	0.876	0.690	0.927	0.791
04	0.800	0.931	0.860	0.705	0.954	0.811
05	0.773	0.932	0.845	0.693	0.893	0.780
06	0.823	0.958	0.885	0.695	0.828	0.756
07	0.807	0.911	0.855	0.779	0.849	0.812
08	0.740	0.886	0.806	0.754	0.856	0.802
09	0.797	0.896	0.844	0.700	0.845	0.766
10	0.777	0.959	0.858	0.587	0.897	0.710
11	0.669	0.925	0.777	0.660	0.956	0.781
12	0.778	0.977	0.866	0.689	0.948	0.798

favoring clear-sky subsidence and extreme near-surface warming [5]. According to the record, the most intense phase occurred around August 6-12 as the anticyclonic block centered near France, with Z_{500} anomalies and stagnant flow maintaining accumulated heat over the region [14, 25, 48]. The summer-mean temperature anomalies exceeded 3°C over broad areas and locally reached several standard deviations above the 1961–1990 baseline [13, 41]. The societal toll was enormous: more than 70,000 people died across Europe, with France particularly affected [38]. As a case study, this heatwave provides a well-documented sequence of onset, maintenance, and decay of a high-impact blocking event against which to evaluate both detection skill and attribution of sustained heat extremes.

Case-study detection. Fig. 5 shows normalized Z_{500} anomalies from the ERA5 observations, with a black contour (isovalue 1.2) marking the diagnosed high-pressure domain from July 26 to August 14, 2003. From July 27 through August 13, every day is detected as "blocked" by our method. The detected high-pressure system evolves from an initial peak over northeastern Europe on July 27–28 to a westward-expanded ridge that covers France, Germany, the U.K., and the adjacent North Atlantic by August 4–12. The system then weakens on August 13 and dissipates on August 14.

This day-by-day evolution aligns with independent analyses for the 2003 European heatwave, which document exceptionally large positive Z_{500} anomalies over western Europe during early-mid August and a subsequent decline thereafter [5, 48]. Importantly, the signal we highlight with contours (anomalously high Z_{500}) captures the same large, quasi-stationary high-pressure pattern that defines a blocking event in practice. When this positive anomaly spreads over a wide area and lasts for several days, winds weaken and the air mass stagnates, allowing heat to build. In the 2003 heatwave case, our uninterrupted run of blocked days aligns with the observed intensification and lifetime of extreme heat over France and neighboring regions in early-mid August [5, 48].

Evaluation. We compare our blocking event detections with the ground truth labels. Only two dates disagree: July 26 and August 13. On July 26, our method finds no grid cells above the scalar threshold, whereas the GTD marks the day as blocked; on August 13, a pronounced high-pressure anomaly remains over western Europe, yet the GTD is marked as unblocked. Given the judgment involved in expert ground-truth labeling, minor discrepancies are unsurprising. For example, on July 26 (Fig. 5, top-left), the normalized anomaly falls just below our threshold,

7

yet a nascent positive signal is evident and subsequently intensifies into the high-pressure system observed the following day.
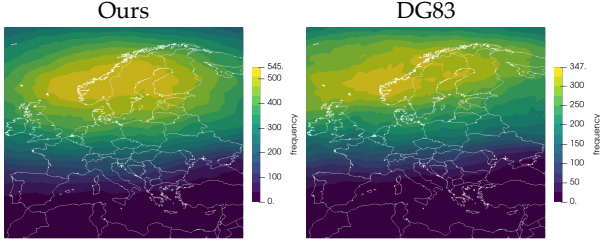

Fig. 6: ERA5 dataset (May 28-Sep 4, 1979-2019). The spatial distribution of blocking event frequency over Europe.

7.4 Uncertainty Analysis and Visualization

Our detection method not only identifies days with blocking events but also depicts the boundary of the high-pressure systems associated with them. Based on our detection results, we present an extensive analysis of the spatiotemporal distribution of blocking events in Europe. We focus our analysis on the ERA5 dataset since it reflects the real-world observations, but there is no specific limitation in extending our visualization to UKESM or other datasets.

Overall spatial distribution. Spatial patterns of blocking event frequency are fundamental to blocking climatology and model evaluation, as emphasized by prior studies [32, 46]. Motivated by this, we built a frequency heatmap by running our detection method on every day in the analysis period of an ensemble. At each grid point, we compute the frequency of days detected as blocked from the normalized Z_{500} anomaly field using the same threshold and domain as in the case studies. The resulting frequency heatmap highlights coherent regional organization, offering a compact summary of where blocking events preferentially occur.

The frequency heatmap showing the spatial distribution of blocking events is presented in Fig. 6. Overall, the spatial blocking frequency is distributed according to the latitude. Higher latitude regions, especially Scandinavia and the North Atlantic, have a high frequency of blocking events. As the latitude decreases, the blocking frequency decreases. This aligns with the overall observation for the frequency heatmaps from the DG83 in the right of Fig. 6. However, we notice that the overall frequency from our detection method is higher than that of DG83. This is expected because our method uses a lower scalar threshold than DG83 when detecting blocking events, which typically obtains larger areas for high-pressure systems.

Contour boxplot for daily blocking distribution. In addition to the overall summary, we are interested in the variation of the spatial distribution on different days of the year, which may follow different patterns due to seasonal variations. To better analyze such variation, for each specific day across 41 years in the ERA5 dataset, we compute the contour boxplot and frequency heatmap for blocking events.

Fig. 7 presents a series of daily contour boxplots for June 9-20, computed with a number of daily ensembles. For each calendar day, the panel pairs a frequency heatmap with a contour boxplot of detected blocking boundaries, both computed across the 41-year dataset. Note that the contributing years can differ from day to day, since not every year exhibits a blocking event on every date.

We observe day-to-day shifts of the median contour in Fig. 7. From June 9–11, it remains at higher latitudes, mainly over Scandinavia, the North Atlantic, and the Arctic. On June 12, it moves southward, spanning western, central, and parts of eastern Europe. During June 13–17, it lingers over western Europe, especially the U.K. and France. It then shifts eastward on June 18–20, centering over France and Germany.

The accompanying frequency heatmaps show a related evolution in blocking frequency. On June 9–13, the frequency pattern is diffuse, with no clear signature of the June 12 southward shift. On June 14–17, higher frequencies concentrate over western Europe, consistent with the median contour's placement. On June 18–20, the frequency increases over central and eastern Europe, tracking the eastward move of the median contour.

Taken together, these visualizations suggest a tentative mid-June progression in this multi-decadal archive: early-period blocking spread across mid–high latitudes, a mid-month focus over western Europe, and a late-period increase over central and eastern Europe. Given the roughly four decades of data and varying year-to-year participation for each date, we interpret these as indicative patterns rather than definitive climatological features.

Spatiotemporal summaries with 3D temporal stacks. Beyond daily 2D visualizations (contour boxplots and frequency heatmaps), we summarize blocking behavior using two complementary 3D temporal stacks (Fig. 4). Time is mapped to the vertical axis (May 28-Sep 4). On the left, the 3D median stack displays daily median contours (blocking boundaries) from the contour boxplots stacked along the time axis; the orange strands represent the daily median contours. On the right, the 3D frequency stack combines daily frequency heatmaps into a single 3D volume, where intensity encodes how often blocking events occur at each location over time.

These 3D temporal stacks reveal the broad space-time organization of blocking events. For example, the red box on the left of Fig. 4 highlights June 13-20, when the median contours cluster over western Europe, consistent with the daily panels in Fig. 7. Such aggregations provide a compact, qualitative view of when and where blocking events concentrate during the season.

Read across days, the median stack illustrates shifts in the central tendency of blocking footprints rather than the continuous motion of a single event. Vertically coherent segments in the median stack therefore indicate multi-day consistency in the typical blocking location, whereas day-to-day lateral shifts capture seasonal changes in where blocking most frequently occurs. The frequency stack complements this view by revealing bands of elevated occupancy over time: thin, faint layers denote scattered, short-lived events, while thicker, brighter layers mark periods when blocking repeatedly forms in similar regions.

These visualizations also have limitations: projecting a 3D object into 2D can obscure structure. In particular, the 3D frequency stack may hide interior features due to occlusion. In practice, interpretation is most effective with interactive exploration—rotating and slicing through selected dates or latitude–longitude planes—and linking the stacks with the corresponding 2D visualizations. A video demonstration for exploring the 3D temporal stacks is available in the GitHub repository.

Monthly blocking event frequency. Beyond day-to-day variability, a monthly view offers a more robust picture of seasonal behavior for blocking events. Although our parameters are tuned using data from the summer months (the JJA window with ground-truth labels), the agreement with DG83 outside JJA (Tab. 3) remains comparable, so we extend the analysis to a monthly basis.

Fig. 8 shows the resulting frequency heatmaps with a monthly ensemble: winter months exhibit substantially higher blocking frequencies than the summer months, with December–January maxima focused over northern Europe and the adjacent North Atlantic, and a more scattered distribution by February–March as circulation transitions toward spring. This is in line with the expected increase in weather variability compared to December and January, for example due to the weakening of the stratospheric polar vortex and related changes in the frequency of sudden stratospheric warmings (SSWs, major events commonly defined by a reversal of the zonal-mean zonal wind at 60° N and 10 hPa). SSWs are known to be associated with changes in tropospheric westerlies and blocking occurrence [2]. More generally, blocking is closely associated with the North Atlantic jet: shifts in jet latitude and changes in jet strength are typically accompanied by changes in where blocking occurs and how often it occurs [20, 52, 57]. In summer, detected blocking events concentrate at higher latitudes (notably north of Europe) with comparatively low frequencies along the Mediterranean coast, reflecting the seasonal shift of blocking events.

8 CONCLUSION AND DISCUSSION

We present an uncertainty visualization framework for detecting and characterizing atmospheric blocking events. We compare our method against two baseline blocking indices, as summarized in Tab. 4. We
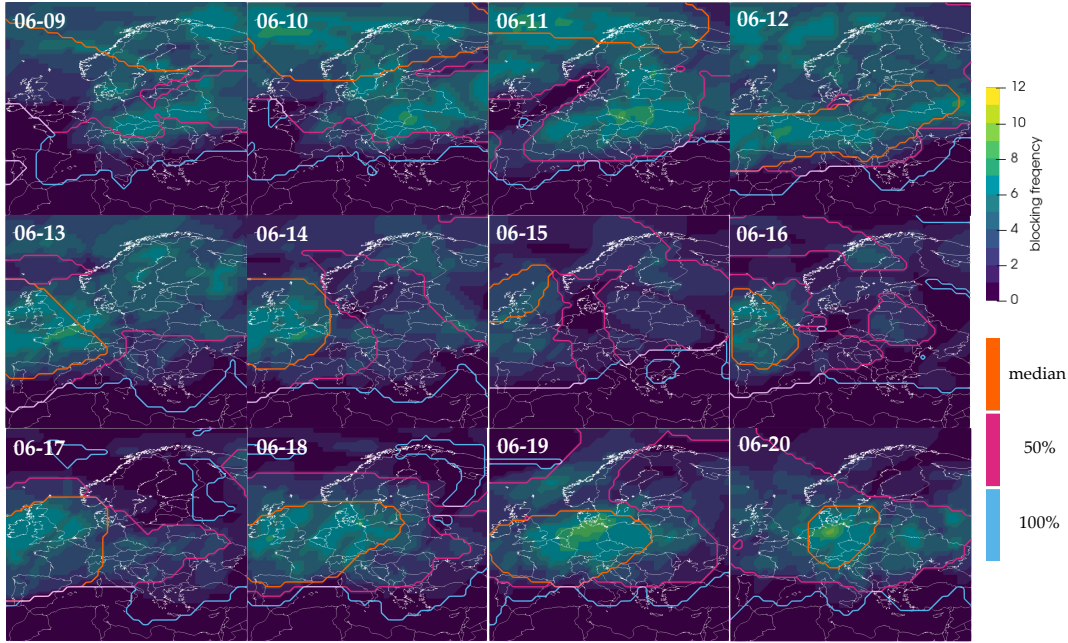
Fig. 7: Daily contour boxplots for June 9-20 (ERA5, 1979-2019). In each panel, the background frequency heatmap shows blocking frequency across the analysis period; overlaid contour boxplots summarize the ensemble of detected blocking boundaries for that date: the median boundary, the 50-percentile envelope, and the 100-percentile envelope. The contributing years vary by day, since not every year has a blocking event on each date.
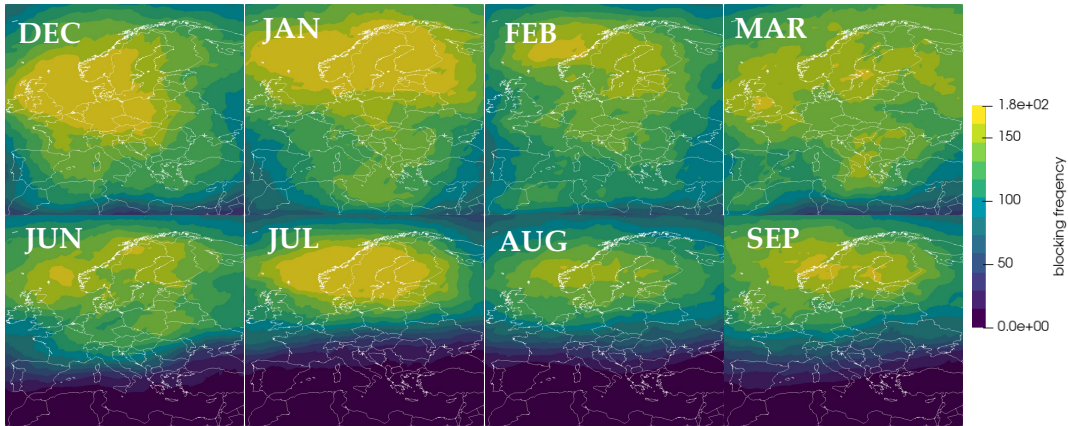

Fig. 8: Frequency heatmaps for a monthly ensemble (ERA5, 1979–2019). A 2×4 grid of heatmaps: top row includes the winter months (December, January, February, and March); bottom row includes the summer months (June, July, August, and September). For each month, the heatmap shows the frequency of blocked days aggregated across the 41-year dataset (using the same threshold and domain as in the case studies). Contour boxplots are omitted here: monthly ensembles contain many more events than a single day, making their computation expensive.

Table 4: Comparison summary of DG83, SOM-BI, and our method for atmospheric blocking event detection and characterization.

Method	Event geometry & trajectory	Uncertainty summary	F1-score (ERA5)	F1-score (UKESM)
DG83	Yes[*]	Frequency heatmap	0.72	0.75
SOM-BI	No	None	0.74	0.71
ours	Yes	Contour boxplot; frequency heatmap; 3D temporal stack	**0.77**	**0.82**

[*] DG83 requires additional post-processing for objects detection and tracking.

introduce a geometry-based detection and tracking method applied to the normalized Z_{500} anomaly field that enforces quasi-stationarity and returns explicit spatial footprints for blocking events. We evaluate the method on both pre-industrial climate simulations (UKESM) and reanalysis data (ERA5). Compared with two strong baselines—DG83 and SOM-BI—our approach achieves higher F1-scores, yielding improved detection performance relative to expert-labeled ground truth. Beyond the tuning season (JJA), monthly comparisons with DG83 show no notable drop in agreement, indicating stable performance across seasons. We further demonstrate results on the well-documented 2003 European heatwave, where the diagnosed blocked days and footprints closely match independent analyses of the event's evolution.

Another main contribution of our work is a suite of uncertainty-aware summaries for visualizing blocking events. Given an ensemble of detections, we employ contour boxplots to represent typical boundaries and their spread, frequency heatmaps to reveal spatial and temporal occurrence patterns, and 3D temporal stacks to highlight seasonal evolution at a glance. Together, these visualizations make the behavior of blocking events interpretable in both space and time without increasing model complexity.

Our framework has several limitations. The 3D temporal stacks are most effective when explored interactively or through slicing, as static volume renderings may obscure internal structures. In addition, our detection method shows a slight positive bias on the ERA5 dataset, reflected in an imbalance between precision and recall; this bias could be mitigated through further parameter tuning or additional quasi-stationarity constraints while maintaining transparency and reproducibility.

We envision this framework as a valuable tool for climate scientists and meteorologists: by quantifying how blocking frequency, duration, and intensity vary across regions and climate scenarios, it supports more accurate assessment of present and future risks associated with prolonged extreme weather events.

REFERENCES

[1] J. Ahrens, B. Geveci, and C. Law. Paraview: An end-user tool for large-data visualization. *The Visualization Handbook*, p. 717, 2005. doi: 10.1016/B978-012387582-2/50038-1 5

[2] M. P. Baldwin, B. Ayarzagüena, T. Birner, N. Butchart, A. H. Butler, A. J. Charlton-Perez, D. I. V. Domeisen, C. I. Garfinkel, H. Garny, E. P. Gerber, M. I. Hegglin, U. Langematz, and N. M. Pedatella. Sudden stratospheric warmings. *Reviews of Geophysics*, 59(1):e2020RG000708, 2021. doi: 10.1029/2020RG000708 8

[3] D. Barriopedro, R. García-Herrera, J. F. González-Rouco, and R. M. Trigo. Application of blocking diagnosis methods to General Circulation Models. Part II: Model simulations. *Climate Dynamics*, 35(7):1393–1409, 2010. doi: 10.1007/s00382-010-0766-6 2

[4] A. Biswas, G. Lin, X. Liu, and H.-W. Shen. Visualization of time-varying weather ensembles across multiple resolutions. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):841–850, 2017. doi: 10.1109/TVCG.2016.2598869 2

[5] E. Black, M. Blackburn, G. Harrison, B. Hoskins, and J. Methven. Factors contributing to the summer 2003 European heatwave. *Weather*, 59(8):217–223, 2004. doi: 10.1256/wea.74.04 7

[6] C. Cassou. Intraseasonal interaction between the Madden–Julian oscillation and the North Atlantic oscillation. *Nature*, 455:523–527, 2008. doi: 10.1038/nature07286 2

[7] M. Croci-Maspoli, C. Schwierz, and H. C. Davies. Atmospheric blocking: Space-time links to the NAO and PNA. *Climate Dynamics*, 29(7-8):713–725, 2007. doi: 10.1007/s00382-007-0259-4 2

[8] R. M. Dole and N. D. Gordon. Persistent anomalies of the extratropical Northern Hemisphere wintertime circulation: Geographical distribution and regional persistence characteristics. *Monthly Weather Review*, 111(8):1567 – 1586, 1983. doi: 10.1175/1520-0493(1983)111<1567:PAOTEN>2.0.CO;2 1, 2, 5, 6, 12

[9] H. Doraiswamy, V. Natarajan, and R. S. Nanjundiah. An exploration framework to identify and track movement of cloud systems. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2896–2905, 2013. doi: 10.1109/TVCG.2013.131 2

[10] F. Fabiano, V. L. Meccia, P. Davini, P. Ghinassi, and S. Corti. A regime view of future atmospheric circulation changes in northern mid-latitudes. *Weather and Climate Dynamics*, 2(1):163–180, 2021. doi: 10.5194/wcd-2-163-2021 2

[11] G. Farnebäck. Two-frame motion estimation based on polynomial expansion. In J. Bigun and T. Gustavsson, eds., *Image Analysis*, pp. 363–370. Springer, Berlin, Heidelberg, 2003. doi: 10.1007/3-540-45103-X_50 2

[12] Z. Feng, J. Hardin, H. C. Barnes, J. Li, L. R. Leung, A. Varble, and Z. Zhang. PyFLEXTRKR: A flexible feature tracking Python software for convective cloud analysis. *Geoscientific Model Development*, 16(10):2753–2776, 2023. doi: 10.5194/gmd-16-2753-2023 2

[13] E. M. Fischer, S. I. Seneviratne, P. L. Vidale, D. Lüthi, and C. Schär. Soil moisture–atmosphere interactions during the 2003 European summer heat wave. *Journal of Climate*, 20(20):5081 – 5099, 2007. doi: 10.1175/JCLI4288.1 7

[14] R. García-Herrera, J. Díaz, R. M. Trigo, and E. Hernández. Extreme summer temperatures in Iberia: health impacts and associated synoptic conditions. *Annales Geophysicae*, 23(2):239–251, 2005. doi: 10.5194/angeo-23-239-2005 7

[15] R. Grotjahn and R. Zhang. Synoptic analysis of cold air outbreaks over the California central valley. *Journal of Climate*, 30(23):9417 – 9433, 2017. doi: 10.1175/JCLI-D-17-0167.1 3

[16] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, G. De Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy, R. J. Hogan, E. Hólm, M. Janisková, S. Keeley, P. Laloyaux,

P. Lopez, C. Lupu, G. Radnoti, P. de Rosnay, I. Rozum, F. Vamborg, S. Villaume, and J.-N. Thépaut. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020. doi: 10.1002/qj.3803 1, 3

[17] M. Iturbide, J. M. Gutiérrez, L. M. Alves, J. Bedia, R. Cerezo-Mota, E. Cimadevilla, A. S. Cofiño, A. Di Luca, S. H. Faria, I. V. Gorodetskaya, M. Hauser, S. Herrera, K. Hennessy, H. T. Hewitt, R. G. Jones, S. Krakovska, R. Manzanas, D. Martínez-Castro, G. T. Narisma, I. S. Nurhati, I. Pinto, S. I. Seneviratne, B. van den Hurk, and C. S. Vera. An update of IPCC climate reference regions for subcontinental analysis of climate model data: Definition and aggregated datasets. *Earth System Science Data*, 12(4):2959–2970, 2020. doi: 10.5194/essd-12-2959-2020 3

[18] A. Kamal, P. Dhakal, A. Y. Javaid, V. K. Devabhaktuni, D. Kaur, J. Zaientz, and R. Marinier. Recent advances and challenges in uncertainty visualization: a survey. *Journal of Visualization*, 24(5):861–890, 2021. doi: 10.1007/s12650-021-00755-1 2

[19] C. Karamperidou. Extracting paleoweather from paleoclimate through a deep learning reconstruction of Last Millennium atmospheric blocking. *Communications Earth & Environment*, 5(1):535, 2024. doi: 10.1038/s43247-024-01687-y 2

[20] L.-A. Kautz, O. Martius, S. Pfahl, J. G. Pinto, A. M. Ramos, P. M. Sousa, and T. Woollings. Atmospheric blocking and weather extremes over the Euro-Atlantic sector – a review. *Weather and Climate Dynamics*, 3(1):305–336, 2022. doi: 10.5194/wcd-3-305-2022 1, 8

[21] T. Kohonen. *Self-Organizing Maps*, vol. 30 of *Springer Series in Information Sciences*. Springer, 3 ed., 2001. doi: 10.1007/978-3-642-56927-2 2

[22] F. Lan, B. Gamelin, L. Yan, J. Wang, B. Wang, and H. Guo. Topological characterization and uncertainty visualization of atmospheric rivers. *Computer Graphics Forum*, 43(3):e15084, 2024. doi: 10.1111/cgf.15084 2

[23] H. Lejenas and H. Okland. Characteristics of Northern Hemisphere blocking as determined from a long time series of observational data. *Tellus, Series A*, 35 A(5):350–362, 1983. doi: 10.3402/tellusa.v35i5.11446 2

[24] M. Li, D. Chatterjee, F. Glassmeier, F. Senf, and B. Wang. Tracking low-level cloud systems with topology. In *2025 IEEE Workshop on Topological Data Analysis and Visualization (TopoInVis)*, pp. 89–99, 2025. doi: 10.1109/TopoInVis68599.2025.00013 2

[25] X. Liu, B. He, L. Guo, L. Huang, and D. Chen. Similarities and differences in the mechanisms causing the European summer heatwaves in 2003, 2010, and 2018. *Earth's Future*, 8(4):e2019EF001386, 2020. doi: 10.1029/2019EF001386 7

[26] A. R. Lupo. Atmospheric blocking events: A review. *Annals of the New York Academy of Sciences*, 1504(1):5–24, 2021. doi: 10.1111/nyas.14557 1

[27] P.-A. Michelangeli, R. Vautard, and B. Legras. Weather regimes: Recurrence and quasi stationarity. *Journal of Atmospheric Sciences*, 52(8):1237 – 1256, 1995. doi: 10.1175/1520-0469(1995)052<1237:WRRAQS>2.0.CO;2 2

[28] M. Mirzargar, R. T. Whitaker, and R. M. Kirby. Curve boxplot: Generalization of boxplot for ensembles of curves. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2654–2663, 2014. doi: 10.1109/TVCG.2014.2346455 2

[29] G. Muszynski, Prabhat, J. Balewski, K. Kashinath, M. Wehner, and V. Kurlin. Atmospheric blocking pattern recognition in global climate model simulation data. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 677–684, 2021. doi: 10.1109/ICPR48806.2021.9412736 2

[30] A. T. Pang, C. M. Wittenbrink, and S. K. Lodha. Approaches to uncertainty visualization. *The Visual Computer*, 13(8):370–390, 1997. doi: 10.1007/s003710050111 2

[31] J. L. Pelly and B. J. Hoskins. A new perspective on blocking. *Journal of the Atmospheric Sciences*, 60(5):743–755, 2003. doi: 10.1175/1520-0469(2003)060<0743:ANPOB>2.0.CO;2 2

[32] M. C. Pinheiro, P. A. Ullrich, and R. Grotjahn. Atmospheric blocking and intercomparison of objective detection methods: Flow field characteristics. *Climate Dynamics*, 53(7–8):4189–4216, Apr 2019. doi: 10.1007/s00382-019-04782-5 2, 3, 5, 6, 8, 12

[33] K. Potter, P. Rosen, and C. R. Johnson. From quantification to visualization: A taxonomy of uncertainty visualization approaches. In A. M. Dienstfrey and R. F. Boisvert, eds., *Uncertainty Quantification in Scientific Computing*, pp. 226–249. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. 2

[34] K. Potter, A. Wilson, P.-T. Bremer, D. Williams, C. Doutriaux, V. Pascucci, and C. R. Johnson. Ensemble-Vis: A framework for the statistical visualization of ensemble data. In *2009 IEEE International Conference on Data Mining Workshops*, pp. 233–240, 2009. doi: 10.1109/ICDMW.2009.55 2

[35] F. M. Ralph, J. J. Rutz, J. M. Cordeira, M. Dettinger, M. Anderson, D. Reynolds, L. J. Schick, and C. Smallcomb. A scale to characterize the strength and impacts of atmospheric rivers. *Bulletin of the American Meteorological Society*, 100(2):269 – 289, 2019. doi: 10.1175/BAMS-D -18-0023.1 2

[36] D. F. Rex. Blocking action in the middle troposphere and its effect upon regional climate. *Tellus*, 2(3):137–236, 1950. doi: 10.1111/j.2153-3490. 1950.tb00331.x 1

[37] N. Roberto, A. Toniazzo, S. Puca, G. Panegrossi, L. P. D'adderio, M. Petracca, L. Ciabatta, L. Brocca, D. Gencic, I. Smiljanic, and N. S. Mahovic. Extreme floods in Libya and central Greece. EUMETSAT (European Organisation for the Exploitation of Meteorological Satellites), 2023. EUMETSAT, Case studies. 1, 2

[38] J.-M. Robine, S. L. K. Cheung, S. L. Roy, H. V. Oyen, C. Griffiths, J.-P. Michel, and F. R. Herrmann. Death toll exceeded 70,000 in Europe during the summer of 2003. *Comptes Rendus Biologies*, 331(2):171–178, 2008. doi: 10.1016/j.crvi.2007.12.001 7

[39] J. J. Rutz, C. A. Shields, J. M. Lora, A. E. Payne, B. Guan, P. Ullrich, T. O'Brien, L. R. Leung, F. M. Ralph, M. Wehner, S. Brands, A. Collow, N. Goldenson, I. Gorodetskaya, H. Griffith, K. Kashinath, B. Kawzenuk, H. Krishnan, V. Kurlin, D. Lavers, G. Magnusdottir, K. Mahoney, E. McClenny, G. Muszynski, P. D. Nguyen, Y. Qian, A. M. Ramos, C. Sarangi, S. Sellars, T. Shulgina, R. Tome, D. Waliser, D. Walton, G. Wick, A. M. Wilson, and M. Viale. The atmospheric river tracking method intercomparison project (ARTMIP): Quantifying uncertainties in atmospheric river climatology. *Journal of Geophysical Research: Atmospheres*, 124(24):13777–13802, 2019. doi: 10.1029/2019JD030936 2

[40] J. Sanyal, S. Zhang, J. Dyer, A. Mercer, P. Amburn, and R. Moorhead. Noodles: A tool for visualization of numerical weather model ensemble uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1421–1430, 2010. doi: 10.1109/TVCG.2010.181 2

[41] C. Schär, P. L. Vidale, D. Lüthi, C. Frei, C. Häberli, M. A. Liniger, and C. Appenzeller. The role of increasing temperature variability in European summer heatwaves. *Nature*, 427:332–336, 2004. doi: 10.1038/ nature02300 7

[42] R. Schiemann, M.-E. Demory, L. C. Shaffrey, J. Strachan, P. L. Vidale, M. S. Mizielinski, M. J. Roberts, M. Matsueda, M. F. Wehner, and T. Jung. The resolution sensitivity of Northern Hemisphere blocking in four 25-km atmospheric global circulation models. *Journal of Climate*, 30(1):337 – 358, 2017. doi: 10.1175/JCLI-D-16-0100.1 1

[43] A. A. Sellar, C. G. Jones, J. P. Mulcahy, Y. Tang, A. Yool, A. Wiltshire, F. M. O'Connor, M. Stringer, R. Hill, J. Palmieri, S. Woodward, L. de Mora, T. Kuhlbrodt, S. T. Rumbold, D. I. Kelley, R. Ellis, C. E. Johnson, J. Walton, N. L. Abraham, M. B. Andrews, T. Andrews, A. T. Archibald, S. Berthou, E. Burke, E. Blockley, K. Carslaw, M. Dalvi, J. Edwards, G. A. Folberth, N. Gedney, P. T. Griffiths, A. B. Harper, M. A. Hendry, A. J. Hewitt, B. Johnson, A. Jones, C. D. Jones, J. Keeble, S. Liddicoat, O. Morgenstern, R. J. Parker, V. Predoi, E. Robertson, A. Siahaan, R. S. Smith, R. Swaminathan, M. T. Woodhouse, G. Zeng, and M. Zerroukat. UKESM1: Description and evaluation of the U.K. earth system model. *Journal of Advances in Modeling Earth Systems*, 11(12):4513–4558, 2019. doi: 10.1029/2019MS001739 1, 3

[44] G. A. Sokolowsky, S. W. Freeman, W. K. Jones, J. Kukulies, F. Senf, P. J. Marinescu, M. Heikenfeld, K. N. Brunner, E. C. Bruning, S. M. Collis, R. C. Jackson, G. R. Leung, N. Pfeifer, B. A. Raut, S. M. Saleeby, P. Stier, , and S. C. van den Heever. *tobac* v1.5: Introducing fast 3D tracking, splits and mergers, and other enhancements for identifying and analysing meteorological phenomena. *Geoscientific Model Development*, 17(13):5309–5330, 2024. doi: 10.5194/gmd-17-5309-2024 2

[45] K. Strommen, I. Mavilia, S. Corti, M. Matsueda, P. Davini, P.-L. Vidale, and R. Mizuta. The sensitivity of Euro-Atlantic regimes to model horizontal resolution. *Geophysical Research Letters*, 46(13):7810–7818, 2019. doi: 10.1029/2019GL082843 2

[46] C. Thomas, A. Voulgarakis, G. Lim, J. Haigh, and P. Nowack. An unsupervised learning approach to identifying blocking events: The case of European summer. *Weather and Climate Dynamics*, 2(3):581–608, 2021. doi: 10.5194/wcd-2-581-2021 1, 2, 3, 4, 5, 6, 8

[47] S. Tibaldi and F. Molteni. On the operational predictability of blocking.

[48] R. M. Trigo, R. García-Herrera, J. Díaz, I. F. Trigo, and M. A. Valente. How exceptional was the early August 2003 heatwave in France? *Geophysical Research Letters*, 32(10), 2005. doi: 10.1029/2005GL022410 7

[49] A. Ullmann, B. Fontaine, and P. Roucou. Euro-Atlantic weather regimes and Mediterranean rainfall patterns: present-day variability and expected changes under CMIP5 projections. *International Journal of Climatology*, 34(8):2634–2650, 2014. doi: 10.1002/joc.3864 2

[50] R. Vautard. Multiple weather regimes over the North Atlantic: Analysis of precursors and successors. *Monthly Weather Review*, 118(10):2056 – 2081, 1990. doi: 10.1175/1520-0493(1990)118<2056:MWROTN>2.0. CO;2 2

[51] J. Wang, S. Hazarika, C. Li, and H.-W. Shen. Visualization and visual analysis of ensemble data: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 25(9):2853–2872, 2019. doi: 10.1109/TVCG. 2018.2853721 2

[52] H. Wazneh, P. Gachon, R. Laprise, A. de Vernal, and B. Tremblay. Atmospheric blocking events in the North Atlantic: trends and links to climate anomalies and teleconnections. *Climate Dynamics*, 56(7):2199–2221, 2021. doi: 10.1007/s00382-020-05583-x 8

[53] R. T. Whitaker, M. Mirzargar, and R. M. Kirby. Contour boxplots: A method for characterizing uncertainty in feature sets from simulation ensembles. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2713–2722, 2013. doi: 10.1109/TVCG.2013.143 2, 4

[54] W. Widanagamaachchi, A. Jacques, B. Wang, E. Crosman, P.-T. Bremer, V. Pascucci, and J. Horel. Exploring the evolution of pressure-perturbations to understand atmospheric phenomena. *IEEE 10th Pacific Visualization Symposium (PacificVis)*, 2017. doi: 10.1109/PACIFICVIS.2017.8031584 2

[55] S. Wilkinson, P. Nowack, and M. Joshi. Process-based machine learning observationally constrains future regional warming projections. *Journal of Geophysical Research: Machine Learning and Computation*, 2(2):e2025JH000698, 2025. doi: 10.1029/2025JH000698 1

[56] T. Woollings, D. Barriopedro, J. Methven, S.-W. Son, O. Martius, B. Harvey, J. Sillmann, A. R. Lupo, and S. Seneviratne. Blocking and its response to climate change. *Current Climate Change Reports*, 4:287–300, 2018. doi: 10.1007/s40641-018-0108-z 1

[57] T. Woollings, A. Hannachi, and B. Hoskins. Variability of the North Atlantic eddy-driven jet stream. *Quarterly Journal of the Royal Meteorological Society*, 136(649):856–868, 2010. doi: 10.1002/qj.625 8

[58] L. Yan, H. Guo, T. Peterka, B. Wang, and J. Wang. TROPHY: A topologically robust physics-informed tracking framework for tropical cyclones. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):1249–1259, 2024. doi: 10.1109/TVCG.2023.3326905 2

[59] L. Yan, P. A. Ullrich, L. P. Van Roekel, B. Wang, and H. Guo. Multilevel robustness for 2D vector field feature tracking, selection and comparison. *Computer Graphics Forum*, 42(6):e14799, 2023. doi: 10.1111/cgf.14799 2

[60] H. Zhang, J. Finkel, D. S. Abbot, E. P. Gerber, and J. Weare. Using explainable AI and transfer learning to understand and predict the maintenance of atlantic blocking with limited observational data. *Journal of Geophysical Research: Machine Learning and Computation*, 1(4):e2024JH000243, 2024. doi: 10.1029/2024JH000243 2

A RUNTIME PERFORMANCE AND ANALYSIS

In this section, we report the runtime performance of the blocking-detection and uncertainty-visualization components of our framework. The benchmarking hardware configuration is described in Sec. 6.

Blocking detection. We report the runtime of blocking detection on both ERA5 and UKESM datasets in Tab. 5. For comparison, we include the runtime of the DG83 pipeline with the modifications of Pinheiro et al. [8, 32]; see Sec. 7.1 for details.

Tab. 5 shows that both DG83 and our method require less than one minute to detect blocking events for the JJA days across both datasets. Both methods are highly efficient, as they do not involve a training process to learn blocking patterns. Compared with DG83, our implementation incurs an additional runtime of 43.1%–46.7%, primarily due to the extra step of searching for large region overlaps.

Table 5: Runtime (in seconds) for blocking detection during the JJA period using DG83 and our method.

Dataset	Years	Method	Total runtime (sec)	Average runtime per year (sec)
ERA5	41	DG83	31.02	0.757
		Ours	45.51	1.110
UKESM	101	DG83	18.17	0.180
		Ours	26.00	0.257

Uncertainty visualization. Recall that we visualize uncertainty using blocking frequency heatmaps, contour boxplots, and 3D temporal stacks. Computing frequency maps and temporal stacks involves straightforward accumulation, whereas contour boxplots are more computationally demanding: for a given calendar date i with d_i blocked days across all years, constructing the contour boxplot requires $O(d_i^2)$ operations.

For all data within the JJA period, our implementation takes 35.18 seconds on ERA5 (41 years) and 140.31 seconds on UKESM (101 years) for all uncertainty-visualization computations. The higher cost for UKESM is expected, as the typical d_i is larger than in ERA5. In both datasets, the runtime remains sufficiently low for practical use and downstream analysis.

Potential improvement. Although our implementation is already efficient, further performance gains can be achieved through parallelization. In blocking detection, both region-overlap tests and trajectory linkage are independent across time steps and thus well suited to parallel execution. For uncertainty visualization, particularly contour boxplots, the pairwise band computations are also parallelizable. These opportunities make it straightforward to scale our framework to larger datasets.

B EXTENDED EVALUATION

In this section, we extend the blocking-detection evaluation to assess temporal stability across the JJA season. As in Sec. 7.2, we evaluate performance against the expert GTD labels. We omit SOM-BI here due to the unavailability of per-day outputs in our environment.

Seasonal pattern of daily outcomes. Fig. 9 aggregates, for each calendar date, the 41 JJA instances from ERA5; all bars have equal height, and the color composition indicates the proportions of **TN**, **TP**, **FP**, and **FN** (true negatives, true positives, false positives, and false negatives, respectively).

In early summer (May 28–June 24), our method (bottom) shows a larger share of **FP** than **FN**, reflecting a more inclusive tendency during this period. From mid- to late-summer (June 25–September 4), **FP** decrease noticeably, while a modest but persistent number of **FN** appear from mid-July through late-August. Because **FP** and **FN** remain of similar magnitude during this interval, the behavior is broadly balanced rather than biased toward misses or false alarms. In contrast, DG83 (top) records fewer **FP** early in the season but substantially more **FN** overall, especially through mid- and late-summer, consistent with a more conservative detector. Both methods exhibit a seasonal pattern within JJA: **FP** generally decline from early to late summer, while **FN** increase. These represent multi-week patterns observable at the 41-year aggregate level; we do not interpret day-scale fluctuations.

Fig. 10 presents the corresponding per-date aggregation for UKESM (101 JJA seasons). Unlike ERA5, the TN/TP/FP/FN composition remains comparatively uniform throughout the season, showing little systematic intra-seasonal drift—consistent with the pre-industrial simulation setting, which lacks pronounced seasonal variation. Our method (bottom) maintains a balanced mix of **FP** and **FN** across dates, indicating no strong bias toward false alarms or misses. Compared with DG83 (top), our detector produces a similar number of **FP** but noticeably fewer **FN**, consistent with its higher overall detection performance on UKESM.

Using a single parameter configuration, our detector exhibits stable performance across both JJA seasons and datasets, while accommodating mild seasonal variation. In ERA5, early summer shows a higher proportion of **FP**, an inclusive phase consistent with our previously noted bias toward higher recall, followed by mid- to late-summer, where **FP** and **FN** remain broadly balanced. In UKESM, the per-date composition is nearly flat, maintaining a consistent balance between **FP** and **FN** throughout the season. Together, these results indicate that our detector operates stably at the seasonal scale, slightly favoring inclusivity in early JJA and achieving balanced performance thereafter, while maintaining the precision levels reported in the main evaluation.
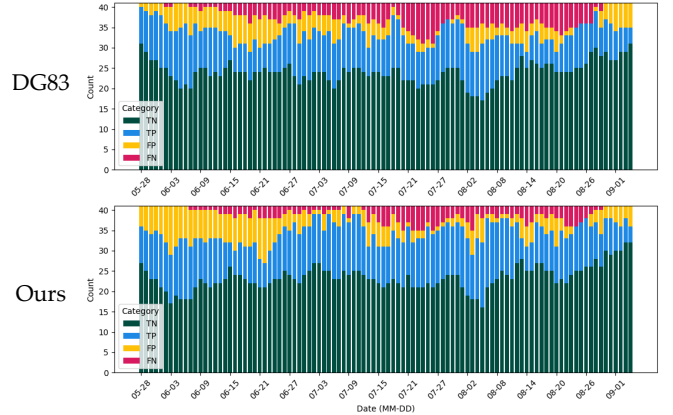

Fig. 9: Temporal breakdown of daily detection outcomes during the JJA period in ERA5 (41 years). Top: DG83; bottom: ours. For each calendar date (x-axis), the stacked bar aggregates the 41 daily instances (one per year), so all bars have equal height. Colors indicate counts of TN (dark green), TP (blue), FP (yellow), and FN (magenta). This visualization compares how the composition of correct and error outcomes evolves through the season for the two methods.
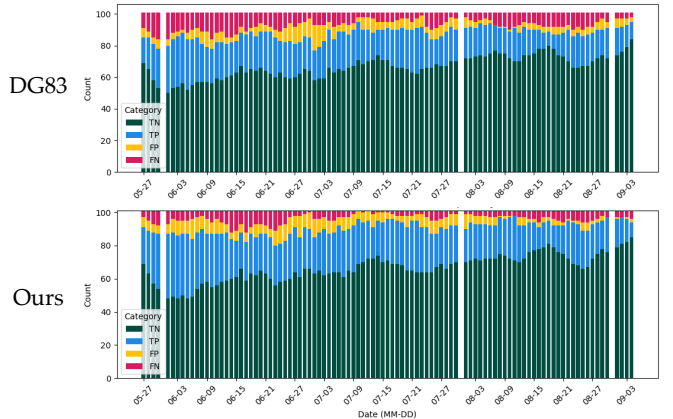

Fig. 10: Temporal breakdown of daily detection outcomes during the JJA period in UKESM (101 years). Top: DG83; bottom: ours. For each calendar date (x-axis), the stacked bar aggregates the 101 daily instances (one per year), so all bars have equal height. Colors indicate counts of TN (dark green), TP (blue), FP (yellow), and FN (magenta). Columns for May 31, July 31, and August 31 are blank because the UKESM simulation calendar does not include these dates.