Northern Hemisphere extratropical cyclone biases in ECMWF subseasonal forecasts

Journal Article

Author(s):
Büeler, Dominik; Sprenger, Michael; Wernli, Heini

Publication date:
2024-01

Permanent link:
https://doi.org/10.3929/ethz-b-000658256

Originally published in:
Quarterly Journal of the Royal Meteorological Society 150(759), https://doi.org/10.1002/qj.4638

Received: 21 February 2023 | Revised: 23 November 2023 | Accepted: 28 November 2023

DOI: 10.1002/qj.4638

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

Northern Hemisphere extratropical cyclone biases in ECMWF subseasonal forecasts

Dominik Büeler | **Michael Sprenger** | **Heini Wernli**

Institute for Atmospheric and Climate Science, ETH Zürich, Zürich, Switzerland

Correspondence
Dominik Büeler, Institute for Atmospheric and Climate Science, ETH Zürich, Universitätsstrasse 16, 8092 Zurich, Switzerland.
Email: dominik.bueeler@env.ethz.ch

Funding information
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: 205419

Extratropical cyclones influence midlatitude surface weather directly via precipitation and wind and indirectly via upscale feedbacks on the large-scale flow. Biases in cyclone frequency and characteristics in medium-range to subseasonal numerical weather prediction might therefore hinder exploitation of potential predictability on these timescales. We thus, for the first time, identify and track extratropical cyclones in 20 years (2000–2020) of subseasonal ensemble reforecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF) in the Northern Hemisphere in all seasons. The reforecasts reproduce the climatology of cyclone frequency and life-cycle characteristics qualitatively well up to six weeks ahead. However, there are significant regional biases in cyclone frequency, which can result from a complex combination of biases in cyclone genesis, size, location, lifetime, and propagation speed. Their magnitude is largest in summer, with the strongest regional deficit of cyclones of more than 30% in the North Atlantic, relatively large in spring, and smallest in winter and autumn. Moreover, the reforecast cyclones reach too-high intensities during most seasons, although intensification rates are captured well. An overestimation of cyclone lifetime might partly but not exclusively explain this intensity bias. While the cyclone bias patterns often appear in lead-time weeks 1 and 2, their magnitudes typically grow further at subseasonal lead times, in some cases up to weeks 5 and 6. Most of the dynamical sources of these biases thus likely appear in the early medium range, but sources on longer timescales probably contribute to the further increase of biases with lead time. Our study provides a useful basis to identify, better understand, and ultimately reduce biases in the large-scale flow and in surface weather in subseasonal weather forecasts. Given the considerable biases during summer, when subseasonal predictions of precipitation and surface temperature will become increasingly important, this season deserves particular attention for future research.

KEYWORDS

cyclone intensity, cyclone frequency, European Centre for Medium-Range Weather Forecasts, extratropical cyclones, Northern Hemisphere, storm track, subseasonal forecast bias, subseasonal predictability

1 | INTRODUCTION

Extratropical cyclones are key weather systems on the synoptic scale. They are partly responsible for the day-to-day variability of and potentially damaging extremes in midlatitude precipitation (e.g., Hawcroft et al., 2012; Pfahl & Wernli, 2012), wind speed (e.g., Wernli et al., 2002; Fink et al., 2009; Martius et al., 2016), and near-surface temperature (e.g., Pfahl, 2014). Moreover, although extratropical cyclone dynamics is determined by the large-scale flow, processes within cyclones can feed back upscale and even actively influence the onset, maintenance, and decay of large-scale flow regimes by modulating downstream Rossby-wave evolution and breaking (e.g., Rivière & Orlanski, 2007; Grams et al., 2011; Hoskins, 2013). Beyond that, cyclone activity integrated over an extended time period can modulate lower-frequency planetary-scale processes such as the stratospheric polar vortex (e.g., Coy et al., 2009; Coy & Pawson, 2015; Attard & Lang, 2019) via poleward heat fluxes that are directly proportional to upward-propagating Rossby-wave activity (e.g., Polvani & Waugh, 2004), which can be an important driver of subseasonal-to-seasonal large-scale flow variability in the midlatitudes in winter (e.g., Baldwin & Dunkerton, 1999; Ambaum & Hoskins, 2002; Domeisen et al., 2020). Considering this vast importance of extratropical cyclones across spatio-temporal scales, it is crucial to represent these weather systems and their life-cycle characteristics correctly in both medium-range (<15 days ahead) and subseasonal numerical weather forecasts (>15 days but less than a season ahead, depending on the forecast center) to provide skillful seamless predictions. This study thus presents—to our knowledge for the first time—a detailed and year-round feature-based verification of extratropical cyclone activity and characteristics in a subseasonal numerical weather forecast system, namely the Integrated Forecasting System (IFS) of the European Centre for Medium-Range Weather Forecasts (ECMWF). We motivate the need for such a verification in more detail hereafter.

A systematic verification of extratropical cyclones and their life-cycle characteristics in numerical weather forecasts has only been reported in a few studies so far. For instance, Hewson (2002) investigated how different extratropical cyclone characteristics are represented in medium-range forecasts of the Unified Model of the Met Office (United Kingdom). He found that the number of weak cyclones and, though slightly less pronounced, also intense cyclones became under-represented with growing lead time. A few years later, Froude et al. (2007) and Froude (2010) performed detailed cyclone verification in the IFS and in a set of further models, with a focus on how far ahead individual cyclones can be predicted at medium-range lead times. They also analyzed biases in cyclone life-cycle characteristics and found that cyclones were predicted as slightly too intense and too slow. On the other hand, Jung et al. (2006) assessed how Northern Hemisphere extratropical cyclones in winter are represented in seasonal IFS forecasts with different spatial resolutions. They found a negative bias in cyclone frequency, which became smaller with increasing horizontal resolution in most but not all regions. In the Gulf Stream region, for instance, the underestimated cyclone frequency was resolution-independent. Apart from these systematic verification studies, various case studies have demonstrated further that the forecast performance for individual cyclones can already be poor at relatively short lead times, typically because of the special dynamical mechanisms driving these cyclones (e.g., Boettcher & Wernli, 2011). Finally, rather than verifying how well cyclones are predicted, other studies highlighted the importance of capturing individual cyclones and the associated uncertainty correctly for subsequent forecast performance (e.g., Grams et al., 2011; Martínez-Alvarado et al., 2016; Baumgart et al., 2018; Grams et al., 2018; Rodwell & Wernli, 2023).

Verification studies of subseasonal forecasts typically do not focus on the evolution of individual synoptic-scale weather systems, mainly because of their intrinsic predictability limit of about two weeks (e.g., Lorenz, 1963; Zhang et al., 2019). Instead, they aim to identify and understand better the biases in the large-scale flow and surface weather, which typically grow with lead time (so-called model drift) and thus hinder exploitation of the potential predictability for subseasonal lead times. For instance, the ECMWF has recently launched an internal project called Understanding systematic error GROWth (UGROW), which aims to UGROW across timescales in the IFS (Balmaseda et al., 2022). So far, the focus of this project has been on the warm midtropospheric bias over large parts of the Northern Hemisphere landmasses in summer (Magnusson et al., 2022), the westward bias in the location of the subtropical North Pacific jet stream in winter (Vitart et al., 2022), and the surface biases in the equatorial eastern Indian Ocean in boreal summer (Mayer et al., 2022). Although various sensitivity experiments in this project have improved our understanding of some of the biases, the sources and consequences of other biases are still unclear.

A few studies have used feature-based approaches to improve understanding of subseasonal forecast biases. For instance, Quinting and Vitart (2019) investigated how well subseasonal models capture the life cycles of Rossby-wave packets (RWPs) as important and relatively predictable drivers of midlatitude weather variability. The most substantial biases they found were an underestimated RWP initiation frequency over the central North

Pacific and an underestimated RWP decay frequency over the North Atlantic–European sector. The latter bias was shown to be linked to an underestimated blocking frequency over Europe. Focusing on specific storm-track regions, Winters (2021) identified significant subseasonal forecast biases in the occurrence frequency of the main North Pacific jet stream regimes in winter, which have a strong impact in North America and further downstream. Some of these biases are consistent with the aforementioned study by Vitart et al. (2022). Likewise, Büeler et al. (2021) highlighted substantial year-round biases in Atlantic–European weather regime frequencies in subseasonal forecasts. The biases were largest in summer and smallest in winter and correlated well with biases in midtropospheric geopotential height.

However, a systematic, feature-based, and year-round analysis of extratropical cyclone biases is, to our knowledge, still missing for subseasonal forecasts. Considering the influence of extratropical cyclones described initially, not only on surface weather but also upscale, this is an important knowledge gap to fill. One study shedding light on the importance of cyclones for subseasonal forecasts by Son et al. (2020) found subseasonal prediction errors in midtropospheric geopotential height to be dominated by eddy rather than zonal mean errors, with similar importance of synoptic-scale and planetary-scale eddies. Zheng et al. (2019) explicitly investigated Northern Hemisphere cyclone activity in subseasonal forecasts during winter, but focused on analyzing the cyclone skill horizon and its associated planetary-scale sources rather than cyclone biases. Moreover, Wandel et al. (2021) quantified subseasonal forecast biases in warm conveyor belts (WCBs)[1] in winter. They found negative WCB frequency biases over the North Atlantic and eastern North Pacific already emerging at early medium-range lead times. Finally, Afargan-Gerstman et al. (2022) were, to our knowledge, the first to identify and track extratropical cyclones in subseasonal forecasts. They focused, however, not on biases but on how well the North Atlantic storm track and individual cyclone characteristics can be predicted following extreme states of the stratospheric polar vortex during winter.

In this study, we continue along this line of research and for the first time analyze the year-round biases of objectively identified and tracked Northern Hemisphere cyclones in subseasonal reforecasts based on recent IFS model versions. The focus will be on the following two research questions. (i) What are the biases in extratropical cyclone frequency, and how do they depend on lead time? (ii) What are the biases in cyclone life-cycle characteristics such as intensity, intensification rate, lifetime, and propagation, and how do they contribute to cyclone frequency biases?

Section 2 describes the data used, the cyclone identification and tracking algorithms, and the verification approach. Results are presented in Section 3, starting with a conceptual description of the potential sources of cyclone frequency biases and then describing the actual biases in cyclone frequency and life-cycle characteristics. Some of these biases are discussed in more detail in Section 4. In Section 5, we summarize and conclude the article. Note that this article is complemented by many supplemental figures, which we make available for the sake of completeness but are not essential for the main aspects of the article.

2 | DATA AND METHODS

2.1 | Reforecast and reanalysis data

We use six-hourly and $1° \times 1°$ sea-level pressure (SLP) fields from 20 years (2000–2020) of subseasonal reforecast (RFC) sets of ECMWF. The 2060 reforecast sets are based on the relatively recent operational forecasts issued between December 2020 and November 2021, which are based on the model versions CY47R1 (December 1, 2020–May 10, 2021), CY47R2 (May 11, 2021–October 11, 2021), and CY47R3 (October 12, 2021–November 30, 2021). They had been initialized from the two reanalysis datasets ERA5 (Hersbach et al., 2020) and ORA5 (for the ocean initial conditions: Zuo et al., 2017) twice a week, with 11 ensemble members and for a lead time of 46 days (of which we analyze 42 days or 6 weeks). Although including the control reforecast would increase the statistical robustness of our analysis slightly, it would also mean that we mix unperturbed members with perturbed members (in which, for instance, the stochastically perturbed parametrisation tendency (SPPT) scheme substantially influences synoptic dynamics). For this reason, we limit our analysis to the 10 perturbed ensemble members only, which yields a total of 20,600 individual reforecast runs. The native horizontal resolution of the reforecasts is TCo639 (about 18 km) up to a lead time of 15 days and TCo319 (about 36 km) beyond, while the native vertical resolution has increased from 91 levels in version CY47R1 to 137 levels in the later versions. Moreover, the third cycle, CY47R3, came with substantial changes in the moist physics, which affected the model performance in the extratropics positively.[2]

For verification, we use SLP fields from ERA5 with the same temporal resolution and extracted on the same horizontal grid. The native horizontal resolution of ERA5 is TL639 (about 31 km) and thus lower than that for the reforecasts in their medium range but similar in their

extended range. The native vertical resolution of ERA5 is 137 levels (Hersbach et al., 2020).

2.2 | Cyclone identification, tracking, and life-cycle characteristics

We identify and track cyclones objectively in the reforecasts and in ERA5 with the algorithm by Wernli and Schwierz (2006) (with further modifications described in the supplementary material of Sprenger et al., 2017). This algorithm first identifies cyclones and their area, respectively, by the outermost closed SLP contour surrounding the local SLP minimum (cyclone centre). In a second step, it obtains cyclone tracks by concatenating the cyclone centres in time based on a search area depending on the previous cyclone trajectory. SLP minima identified over topography higher than 1,500 m above sea level are neglected, as they might be numerical artifacts caused by the pressure reduction to sea level. Since the tracking algorithm allows for gaps of maximum 12 hr (i.e., time steps at which no cyclone has been identified), cyclone tracks can nevertheless cross topography higher than 1,500 m. Note that we use the same topography with the same resolution to identify the cyclone tracks in the reforecasts and in ERA5, which rules out any influence of topography-related artifacts on the cyclone biases. This cyclone identification and tracking method has participated in the intercomparison project by Neu et al. (2013), see also Hewson and Neu (2015), which indicated that results from our method do not appear as outliers when compared with many other methods.

To focus on the Northern Hemisphere extratropics and to neglect short-lived and weak cyclones, we only consider cyclones that reach their maximum intensity (i.e., their lowest SLP minimum along the track) of less than 1,000 hPa between 25° and 80°N (cf. Figure S1a) and have a lifetime of at least 24 hr. Running the same analysis with a higher intensity threshold of 1,020 hPa has not changed any of the key results shown in this study in a qualitative sense, which is why this arbitrary threshold is justified (some further reasons why we consider the threshold of 1,000 hPa more appropriate than 1,020 hPa are discussed in Section 4.8). Furthermore, the 24-hr threshold is widely used in literature (e.g., Neu et al., 2013) and justified in the sense that the characteristic lifetime of baroclinic life cycles is of the order of days. Finally, we note that the minimum latitude criterion filters out some but not all extratropical cyclones that underwent extratropical transition, because the lowest SLP minimum typically reached during the tropical phase of transitioning cyclones (i.e., before the transition) can still occur north of 25°N (e.g., Hart & Evans, 2001).

Apart from obtaining spatial cyclone frequencies by averaging binary cyclone masks that are set to 1 within cyclone areas and 0 outside, we compute further statistics over the cyclone life cycles: cyclone genesis is defined by the first three time steps of a cyclone track, and, consequently, cyclone lysis by the last three time steps. This definition allows for smoother and more robust statistics for genesis and lysis compared with using only the first and last time steps, at which the cyclones are typically small. For each cyclone track, we further determine the following measures: maximum cyclone intensity as the lowest SLP minimum along the track, maximum cyclone intensification rate as the maximum six-hourly SLP deepening, cyclone lifetime as the time between the first and last track time step, mean propagation speed, mean propagation direction, mean latitude, cyclone size as the area (in km^2) of the cyclone mask at the time of maximum intensity, and another measure for cyclone depth as the difference between the outermost closed SLP contour and the SLP at the cyclone center.

2.3 | Cyclone verification approach

We verify cyclone reforecasts for boreal winter (DJF), spring (MAM), summer (JJA), and autumn (SON). The reforecasts are assigned to a specific season if at least parts of their lead times (i.e., valid times) fall into this season. For instance, the winter subset starts with reforecasts initialized at the end of October, from which, however, we only retain the week 6 lead times falling into December. Consequently, the winter subset ends with reforecasts initialized in the last week of February, from which we only retain week 1 lead times up to the end of February. This ensures an equal representation of all lead times in a specific season. Once the reforecasts and corresponding lead times have been assigned to a season, we extract all cyclone tracks that occur during these lead times and truncate the tracks that start earlier or end later than this season. Note that we only retain a truncated track if it still fulfills the lifetime and intensity criteria (cf. Section 2.2) after truncation. Furthermore, the start or end points of the truncated tracks are not used for any genesis or lysis statistics, respectively. We finally verify the extracted cyclone track dataset against ERA5 by treating ERA5 like an additional, "perfect" ensemble member and performing the same truncation of lead times and tracks. This means that the same ERA5 track can be assigned to different consecutive reforecasts and thus appear multiple times, which mimics the reforecast dimension that aims to predict this track multiple times as well. This verification approach eliminates any potential sampling

error in the reforecast bias, as it reduces the ERA5 set to the exact same subset of days per season that are available in the reforecasts due to their irregular initialization frequency.

We perform the seasonal cyclone verification both independently of lead time (i.e, for all 6 weeks) and separately for the three two-weekly chunks with lead times in weeks 1 and 2 (lead-time days 0–13), weeks 3 and 4 (lead-time days 14–27), and weeks 5 and 6 (lead-time days 28–41). The lead-time-dependent analysis allows us to demonstrate whether cyclone biases already occur at medium-range lead times (weeks 1 and 2) and how they evolve further at subseasonal lead times up to weeks 5 and 6. Thereby, the cut between weeks 1 and 2 and beyond is particularly important to detect the potential influence of the drop in native horizontal resolution in reforecasts after 15 days (cf. Section 2.1; note, however, that the first day of weeks 3 and 4, lead-time day 14, thus still comes with the higher resolution). We analyze further whether the cyclone biases depend on the state of the main modes of extratropical large-scale variability in the Northern Hemisphere at reforecast initialization, namely the North Atlantic Oscillation (NAO) and the Pacific–North American (PNA) pattern. For this analysis, the seasonal reforecast subsets are divided into three subsets starting with either a positive (upper tercile), neutral (middle tercile), or negative (lower tercile) phase of the NAO and PNA, respectively. As a basis for this stratification, the year-round definition of daily NAO and PNA indices (based on an empirical orthogonal function (EOF) analysis for every month in the National Centers for Environmental Prediction (NCEP) reanalysis) by the National Oceanic and Atmosphere Administration is used.[3]

3 | RESULTS

3.1 | Conceptual view on potential sources of cyclone frequency biases

To understand potential sources of the cyclone frequency biases described in the next subsection better, we first explain in a conceptual way, using a feature-based perspective, how cyclone frequency biases can occur (Figure 1). The most obvious source is a bias in the number of cyclones (Figure 1a), which is directly related to a bias in cyclone genesis frequency upstream of the observed cyclone tracks. A second source is a bias in cyclone size (Figure 1b), because our cyclones are identified as two-dimensional objects (see Section 2.2). Biases in cyclone location, caused by a biased cyclone propagation direction or "shifted tracks" (e.g., correct cyclone propagation but genesis and lysis are shifted along the track), lead to dipoles of cyclone frequency biases (Figure 1c) and are also associated with biases in the location of cyclone genesis and/or lysis near the start and end of the observed tracks. Compared with these purely spatial bias sources, biases in cyclone lifetime and cyclone speed additionally introduce a temporal bias source. Biases in cyclone lifetime under the assumption of unbiased cyclone speed (Figure 1d) are associated with biases in the location of cyclone genesis and lysis and thus result in cyclone frequency biases around the start and end of the observed tracks. In contrast, biases in cyclone speed under the assumption of unbiased cyclone lifetime (Figure 1e) typically cause dipoles in cyclone frequency biases: too fast cyclones first cause negative cyclone frequency biases due to the reduced local residence time and, later on, positive frequency biases beyond the observed tracks. Vice versa, too slow cyclones first cause positive cyclone frequency biases due to the enhanced local residence time and, later on, negative cyclone frequency biases towards the end of the observed tracks. Finally, biases in cyclone lifetime and speed combined can cause the following cyclone frequency bias patterns (Figure 1f): too short-lived and too fast cyclones can cause negative cyclone frequency biases along the observed tracks, while too long-lived and too slow cyclones can cause positive cyclone frequency biases along the tracks. Furthermore, too short-lived and too slow cyclones result in dipoles of cyclone frequency biases with positive values at the beginning and negative values at the end, while too long-lived and too fast cyclones cause the opposite bias pattern.

Considering Figure 1 as a whole thus indicates that the exact same cyclone frequency bias patterns can occur for completely different reasons: for instance, negative cyclone frequency biases along the storm-track maxima (i.e., along the whole observed tracks in the schematic) can be caused by too few cyclones (second situation in Figure 1a), too small cyclones (second situation in Figure 1b), too fast cyclones (first situation in Figure 1e), too short-lived and too fast cyclones (first situation in Figure 1f), and too long-lived and too fast cyclones (fourth situation in Figure 1e). Quantifying and understanding the reasons for cyclone frequency biases is thus complex, because it first requires identifying these feature-based bias patterns, before trying to understand their dynamical sources, potentially also remote (such as biases in the upper-level forcing, in baroclinicity, etc.). As a first step in this direction, we will first quantify cyclone frequency biases in the following, and then investigate additional cyclone life-cycle characteristics in order to obtain some understanding of the underlying processes.

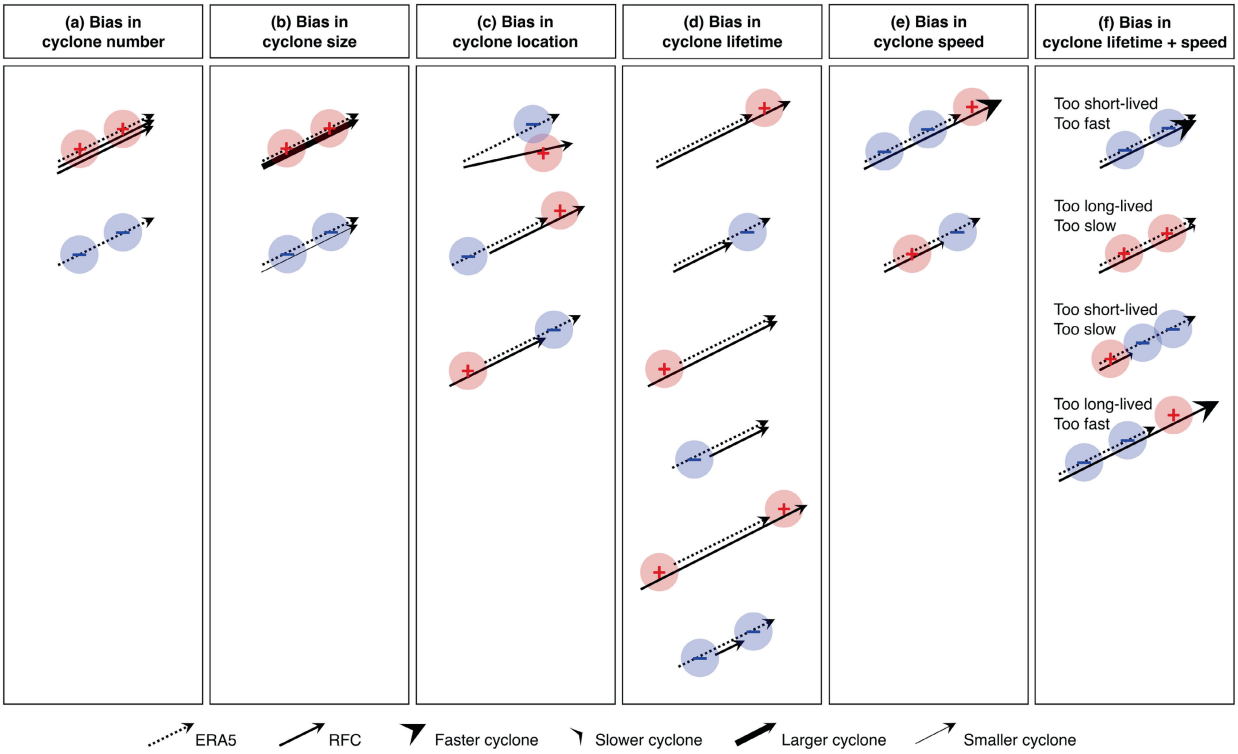

FIGURE 1 Schematic illustrating the spatio-temporal sources of local cyclone frequency biases (red and blue shading indicating regions where positive and negative biases would be expected, respectively; note that a higher number of colored circles does not indicate stronger bias magnitudes) along cyclone tracks (represented by the arrows), considering the fact that cyclone frequency biases are a function of biases in (a) cyclone number, (b) cyclone size, (c) cyclone location, (d) cyclone lifetime (without bias in cyclone speed), (e) cyclone speed (without bias in cyclone lifetime), and (f) cyclone lifetime and speed combined. Note that the combination of the different biases can cause even more bias source patterns or offset existing patterns.

3.2 | Cyclone frequency biases

The seasonal frequencies of the selected Northern Hemisphere cyclone tracks (cf. Section 2.2) in ERA5 reveal the well-known storm-track maxima over the North Atlantic and the North Pacific (Figure 2a,c,e,g; cf., e.g., Whittaker & Horn, 1984; Wernli & Schwierz, 2006; Hoskins & Hodges, 2019). Maximum frequencies are highest in winter (up to 40%), followed by autumn, spring, and summer. Further regional hotspots can be found over the Mediterranean in winter, in the lee of the Rocky Mountains in spring, along the foot of the Himalayas and the Pakistani and Afghan mountain ranges to their west in spring and summer, around the Persian Gulf in summer, and over Hudson Bay in summer and autumn. The North Atlantic and North Pacific hotspots have similar frequency magnitudes in all seasons except for summer, when the maximum is about twice as large in the North Atlantic. Overall, the reforecasts are able to reproduce the structure of these hotspots throughout the year (Figure 2b,d,f,h). Note that the seasonal frequencies of all cyclones, irrespective of any tracking criteria, yield the same qualitative patterns in the main storm-track regions but contain additional small hotspots with frequent short-lived and/or weak cyclones that are not the focus of this analysis (Figure S2).

Despite the overall good agreement, however, the reforecasts have substantial regional biases in cyclone frequency. The magnitude of these biases tends to be smallest in winter (Figure 3a) and autumn (Figure 3j), larger in spring (Figure 3d), and largest in summer (Figure 3g). In all seasons, the bias magnitudes tend to be smallest in weeks 1 and 2 (Figure 3b,e,h,k) and increase substantially in weeks 3 and 4 (Figure 3c,f,i,l). In some cases—for instance in summer—they increase even further in weeks 5 and 6 with regional exceptions (cf. Figure S3). The spatial patterns of the seasonal biases, however, remain similar with lead time. A noteworthy exception is that, in winter, the major positive frequency biases appear at later lead times than the negative frequency biases (Figure 3b,c).

3.2.1 | Winter

Focusing on the individual seasons in more detail reveals some bias hotspots: in winter, the largest positive biases occur west and negative biases east of the North Pacific

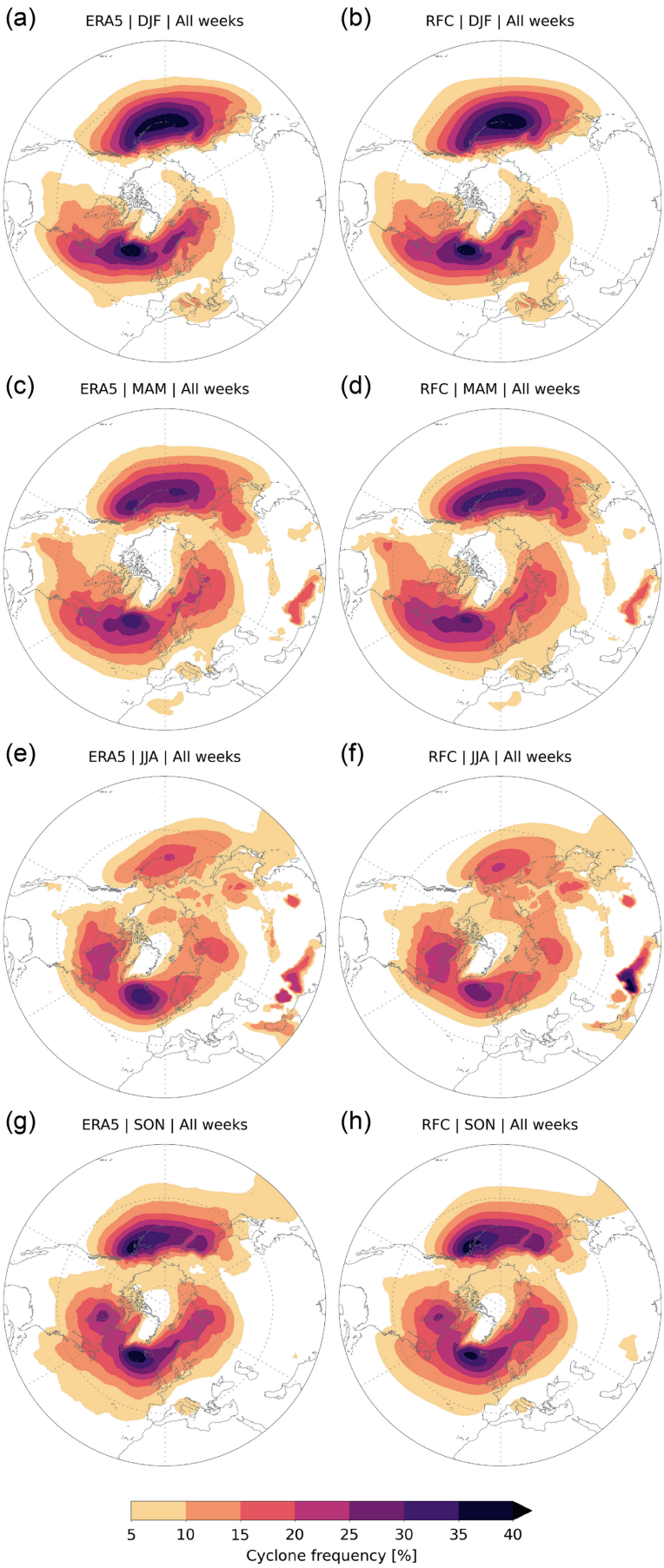

FIGURE 2 Northern Hemisphere seasonal cyclone frequencies in (a,c,e,g) ERA5 and (b,d,f,h) reforecasts of the selected cyclones (see text for details) for (a,b) winter (DJF), (c,d) spring (MAM), (e,f) summer (JJA), and (g,h) autumn (SON). Note that only frequencies of at least 5% are shown.

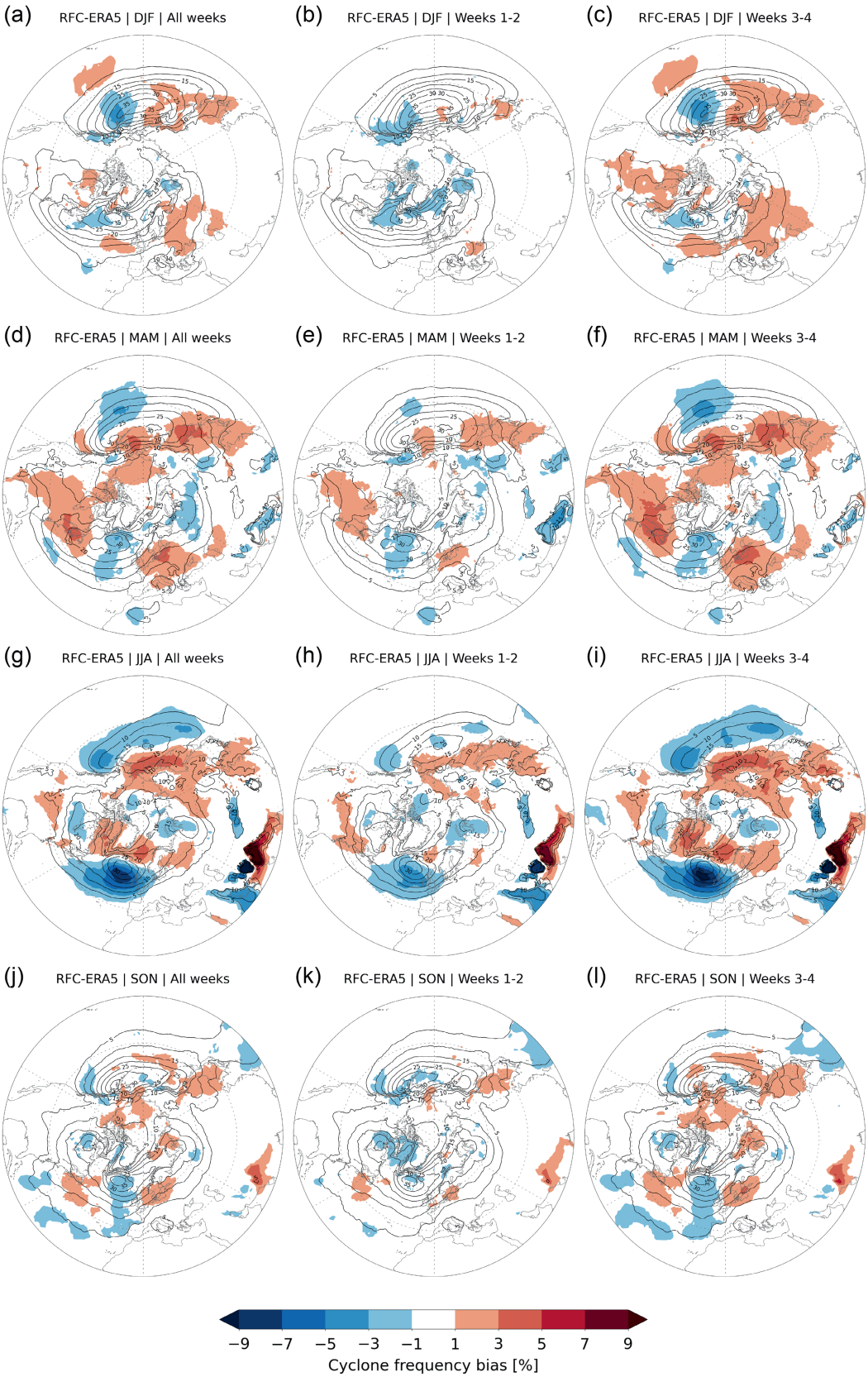

FIGURE 3 Cyclone frequency biases (reforecasts minus ERA5) of the selected cyclones (cf. Figure 2) in color shading for (a–c) winter (DJF), (d–f) spring (MAM), (g–i) summer (JJA), and (j–l) autumn (SON) at (a,d,g,j) all six lead-time weeks, (b,e,h,k) weeks 1 and 2, and (c,f,i,l) weeks 3 and 4. For reference, the corresponding cyclone frequencies in ERA5 (same as in Figure 2) are shown as contours. The biases for lead-time weeks 5 and 6 are shown in Figure S3.

storm-track maximum. Over the North Atlantic, the biases are slightly negative southwest and positive southeast of the storm-track maximum, and positive in Eastern Europe and Western Russia (Figure 3a–c). The largest biases exceed −3% in a region where the climatological frequency exceeds 30% (i.e., the relative frequency biases reach about 10%). To get an impression of what these biases might potentially mean for the surface impact and the large-scale circulation in general, we specifically analyze the tracks and maximum intensity of cyclones in ERA5 that move through some of these biased regions. As a first example, the cyclones passing through the eastern edge of the North Pacific storm-track maximum, where the strongest negative bias occurs, typically form upstream east of Asia, move across the whole North Pacific, and decay either over the Arctic Ocean or along the North American west coast (Figure 4a, left). They reach relatively high maximum intensities (Figure 4a, right), typically somewhere in the eastern North Pacific or even along the North American west coast (see dots in Figure 4a, left). The underestimation of these cyclones could thus affect the performance of subseasonal forecasts in predicting surface weather over Canada and the northern United States, as well as the propagation of Rossby waves downstream into the North Atlantic (e.g., Rivière & Orlanski, 2007), although this hypothesis would require further investigation. As another example, the cyclones that move through the region with a negative bias southwest of the North Atlantic storm-track maximum form either over North America or along its east coast (Figure 4b, left). After reaching relatively high intensities (Figure 4b, right) around the southern tip of Greenland, some of them move into the Norwegian Sea and fewer of them into Northern Europe. Hence, their underestimation might not influence surface weather prediction over Europe directly, but rather indirectly by influencing the evolution of Rossby waves propagating into Europe. Finally, the ERA5 cyclones passing through the positively biased Eastern Europe–Western Russia region form either upstream over the North Atlantic or in the Mediterranean (Figure 4c, left). Although they are relatively weak (Figure 4c, right), they likely have a considerable imprint on Northern and Eastern European surface weather.

3.2.2 | Spring

In spring, the cyclone frequency is too high over the eastern half of North America, over parts of Europe, along the Northeast Asian coast, and over the Bering Sea, and too low in the North Atlantic storm-track maximum and south of the North Pacific maximum (Figure 3d–f). The alternating positive and negative biases over the North American–North Atlantic–Western Eurasian sector resemble a Rossby-wave pattern, which could potentially be maintained by the too many (upstream) cyclones over North America. Indeed, in ERA5, the cyclones contributing to the large positive bias typically form east of the Rocky Mountains and reach their maximum intensity somewhere over the North Atlantic (Figure 4d, left), while the cyclones over Eastern Europe (which are also overestimated in the reforecasts) form further downstream over the eastern North Atlantic or Europe (Figure 4e, left). Whether the overestimated upstream cyclones influence the positive frequency bias of cyclones downstream via Rossby-wave modulation would require further investigation.

3.2.3 | Summer

Summer is characterized by a substantial poleward bias of the storm track, with an underestimation of cyclones at the southern edge of both the North Atlantic and North Pacific storm-track maxima and an overestimation at their northern edge (Figure 3g–i). The underestimation over the North Atlantic is striking, as it already has relatively high magnitudes in weeks 1 and 2 and increases regionally up to more than 9% at late subseasonal lead times (Figure S3f). This is by far the largest bias throughout the year. Considering the absolute cyclone frequencies of 25%–30% in the region of the maximum biases in ERA5 (contours in Figure 3g), this implies a relative cyclone deficit of more than 30% near the North Atlantic storm-track maximum. This is a substantial magnitude and likely indicates this bias to be relevant, although it goes beyond the scope of this study to determine this. According to ERA5, these cyclones typically form over North America or along its east coast, reach their relatively high maximum intensity in the storm-track maximum, and decay over either the Norwegian Sea or Northern Europe (Figure 4f). The strong underestimation of these cyclones might thus be particularly important for the performance of subseasonal surface weather predictions in Northern Europe. Likewise, the overestimated cyclone frequency over the Bering Sea but underestimated cyclone frequency over the eastern North Pacific (Figure 3g–i) might indicate that, according to the corresponding ERA5 cyclones, too many cyclones forming over the western North Pacific move into the Arctic (Figure 4g) and too few cyclones reach the North American west coast (Figure 4h). This likely also affects subseasonal surface weather predictions there. Apart from the main storm-track regions, the summer reforecasts have a large positive cyclone frequency bias along the foot of the Himalayas and a large negative bias to its west over Afghanistan, Pakistan, and

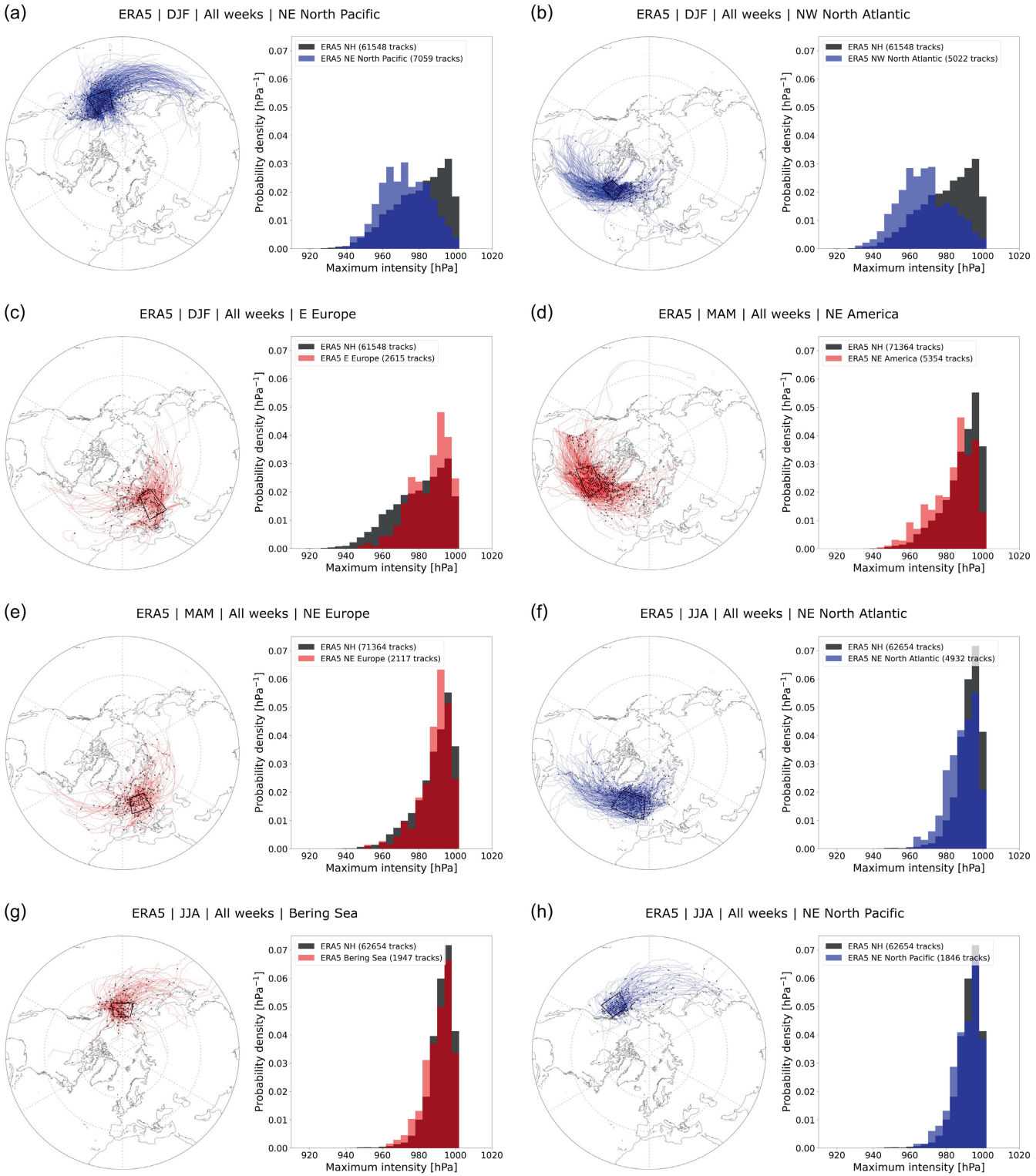

FIGURE 4 ERA5 tracks (maps) and maximum intensity (i.e., lowest SLP minimum; colored histograms) of all cyclones passing through selected subdomains (black rectangles in maps), in which the reforecasts have considerable cyclone frequency biases (cf. Figure 3), during (a–c) winter (DJF), (d,e) spring (MAM), and (f–h) summer (JJA). The black dots in the maps indicate the location of maximum intensity (i.e., lowest SLP minimum) and, as a climatological reference, the black histograms show the distributions for all Northern Hemisphere cyclone tracks. The colors indicate whether the cyclone frequencies in the reforecasts are under- (blue) or overestimated (red) in the respective subdomains. Numbers in brackets in the histogram legends indicate the numbers of cyclones passing through the respective subdomains.

Iran (Figure 3g–i). The fact that these biases are strongly concentrated along the mountain ranges and are relatively high already in weeks 1 and 2 (and even in week 1; cf. Figure S4) indicates potential artificial sources (associated with the pressure reduction to sea level) caused by the higher native spatial resolution of the reforecasts in the medium range (cf. Section 2.1). Whether these biases, including the negative bias around the Persian Gulf, are additionally related to biases in Indian summer monsoon low-pressure systems propagating northwestward along the Himalayas (e.g., Boos et al., 2015; Attada et al., 2019; Deoras et al., 2021) would require further investigation.

3.2.4 | Autumn

In autumn, the largest biases are an underestimation of cyclones in the eastern half of the North Atlantic storm track, and an overestimation in its western half, in parts of Northern Europe, along the East Asian coast, and over parts of the Arctic Sea (Figure 3j–l). The alternating bias pattern over the North Atlantic–European sector again resembles a Rossby-wave pattern similar to the one in spring but slightly shifted and less pronounced. The ERA5 cyclones passing through the North Atlantic storm-track maximum, where the reforecasts have the largest negative bias, show similar characteristics to the ones already discussed for winter (not shown).

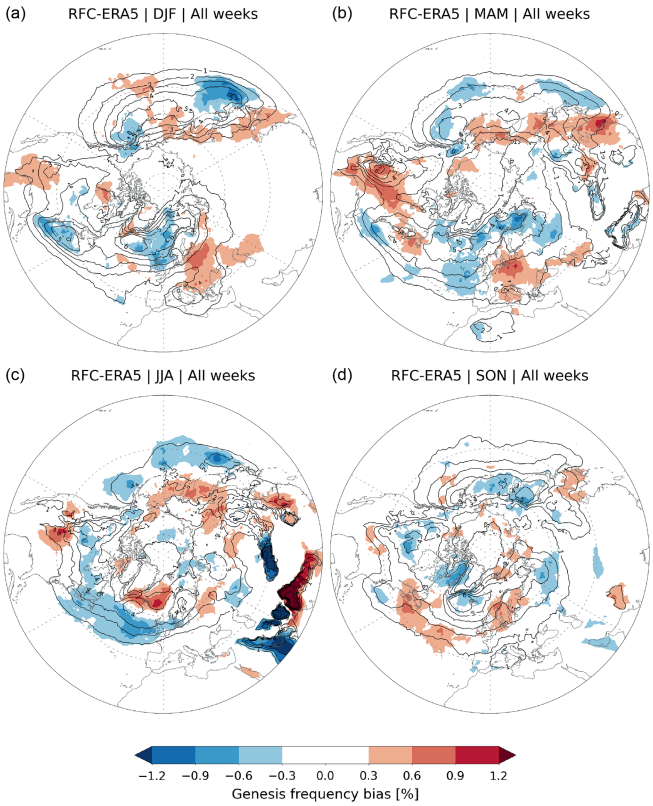

FIGURE 5 Seasonal genesis frequency biases (reforecasts minus ERA5) of the selected cyclones for all six lead-time weeks for (a) winter (DJF), (b) spring (MAM), (c) summer (JJA), and (d) autumn (SON). For reference, the corresponding genesis frequencies in ERA5 are shown as contours.

3.2.5 | Genesis frequency biases as potential source of cyclone frequency biases

As illustrated in Figure 1, the biases described in spatial cyclone frequency can result from a complex combination of spatio-temporal biases associated with cyclone life cycles. In the following, we investigate the spatial pattern of cyclone genesis frequency biases in more detail, because it directly influences the number of cyclones as the most obvious source of cyclone frequency biases (cf. Figure 1a). Spatial patterns of cyclone lysis frequency biases have also been computed, but are not discussed in detail because they do not reveal any obvious connections to the cyclone frequency biases (cf. Figure S5). All the other potential bias sources indicated in Figure 1 will be considered in Section 3.3, where cyclones in individual ocean basins are analyzed in more detail.

The seasonal patterns of the (lead-time-independent, i.e., for all six lead-time weeks) cyclone genesis frequency biases (Figure 5) largely correspond to the patterns of the cyclone frequency biases (Figure 3a,d,g,j), which indicates that a part of the cyclone frequency biases is explained by corresponding in situ genesis biases. However, there are some noteworthy exceptions. The underestimation of cyclones in the eastern North Pacific during winter (Figure 3a) might be related to the underestimation of genesis upstream in the western North Pacific rather than in situ (Figure 5a). Likewise, the underestimated cyclone frequency close to the North Atlantic storm-track maximum during winter (Figure 3a) is probably a consequence of the lack of genesis along the southeastern United States coast (Figure 5a). This is supported by the fact that about 50% of the cyclones in these two biased regions (eastern North Pacific and south of North Atlantic storm-track maximum) typically form in the regions mentioned where we find genesis biases (Wernli & Schwierz, 2006), which also becomes obvious from the cyclone tracks in ERA5 passing through the two biased regions (cf. Figure 4a,b). Interestingly, the genesis deficit east of the North Atlantic storm-track maximum (Figure 5a) does not seem to influence cyclone frequency biases in situ or further downstream (Figure 3a). It could thus be an example of a combination of a bias in genesis location with a bias in cyclone location or lifetime (cf. Figure 1), which does not necessarily lead to cyclone frequency biases downstream.

In spring, there are similar but smaller deficits of genesis upstream of the two storm-track maxima (Figure 5b). Furthermore, the overestimation of cyclones over northeastern North America in spring (Figure 3d) seems to be related to too much genesis in the lee of the Rocky Mountains rather than further upstream over the North Pacific (the overestimation of genesis in the lee of the Rocky Mountains seems to be a nearly year-round problem of the model; cf. Figure 5). This is supported by the fact that almost none of the cyclones in ERA5 passing through northeastern North America form upstream of the Rocky Mountains (cf. Figure 4d)—at least in the cyclone tracking implementation used, in which the elevated topography of the Rocky Mountains seems to act as a blocker, correctly or otherwise (note the very sharp edge in Figure 4d).

The striking negative cyclone frequency bias in the North Atlantic in summer (Figure 3g) seems to be partly the result of an underestimated genesis frequency across the whole North Atlantic storm track (Figure 5c). Again, this hypothesis is supported by the typical track taken by these underestimated cyclones in ERA5 (cf. Figure 4f).

3.3 | Cyclone life-cycle biases

Apart from simulating the number of cyclones correctly, the model also needs to capture cyclone life-cycle characteristics reasonably to properly forecast cyclone interaction with the large-scale Rossby-wave guide and surface weather. Moreover, knowing about biases in cyclone life-cycle characteristics helps us to understand the cyclone frequency biases discussed in the previous section better. We thus verify specific life-cycle measures along all cyclone tracks that reach their maximum intensity in one of the two main storm-track regions, the North Atlantic (80°W–0°E, 25°N–80°N; cf. Figure S1b) and the North Pacific (130°E–120°W, 25°N–80°N; cf. Figure S1c). These measures are maximum cyclone intensity, maximum cyclone intensification, cyclone lifetime, mean propagation speed, mean propagation direction, and mean latitude (cf. Section 2.2). Table 1 shows the total numbers of cyclone tracks these statistics are based on: there are more cyclones in the North Pacific than in the North Atlantic throughout the year, most probably due to the much larger extent of the North Pacific, and the reforecasts both over- and underestimate the total numbers of cyclones by up to 6%. Comparing the number biases in Table 1 with the spatial cyclone frequency biases in Figure 3 indicates that large cyclone frequency biases, such as the underestimation over the North Atlantic during summer (Figure 3g), are not necessarily associated with a consistent bias in the total number of cyclones (+2% in the North Atlantic during summer). This again points to the variety of sources for spatial cyclone frequency biases apart from biases in cyclone number, as discussed in Section 3.1.

TABLE 1 Seasonal numbers of reforecast and ERA5 cyclones (for all six lead-time weeks), the tracks of which reach their maximum intensity inside the North Atlantic and the North Pacific, respectively (see text for details).

Season	North Atlantic	North Pacific
Winter	192193, 19733, −3%	255241, 25188, +1%
Spring	169326, 17973, −6%	207101, 20695, +0%
Summer	116049, 11414, +2%	148309, 15831, −6%
Autumn	172123, 17776, −3%	215078, 21129, +2%

Note: The number of reforecast cyclones is indicated on the left (in total for all 10 ensemble members), the number of ERA5 cyclones in the middle, and the ensemble mean deficit or surplus of reforecast cyclones with respect to the ERA5 cyclones on the right (to obtain this bias, the number of reforecast cyclones is first divided by 10 to get the mean number per ensemble member).

Qualitatively, the winter and summer reforecasts capture the distribution of the different life-cycle measures in the two ocean basins remarkably well, as shown by the thumbnail histograms of Figures 6 and 7. However, a quantitative comparison of the two distributions reveals a set of significant biases (Figures 6 and 7; note that we define a bias to be significant if the black difference curve lies outside the gray confidence interval, which is obtained as explained in Section S.1). They are discussed in some detail in the following paragraphs.

3.3.1 | Winter

During winter, in particular, the intense half of the reforecast cyclones reach significantly higher maximum intensities than in ERA5 in both the North Atlantic and North Pacific (Figure 6a,g; note that the 0%–50% percentiles correspond to the intense half of the cyclone population), although maximum cyclone intensification rates are not considerably biased (Figure 6b,h). This bias in maximum intensity is weakest (but still partly significant) at lead-time weeks 1–4 and grows further at weeks 5 and 6 (cf. Figure S6a,g).

One reason for the maximum intensity bias can be a positive bias in cyclone lifetime: if the cyclones live for too long, they can reach too high maximum intensities even if their intensification rates are unbiased (under the assumption that they keep intensifying during this extended period). In the North Atlantic, the lifetime of the most long-lived cyclones is indeed significantly overestimated (Figure 6c), particularly at subseasonal lead

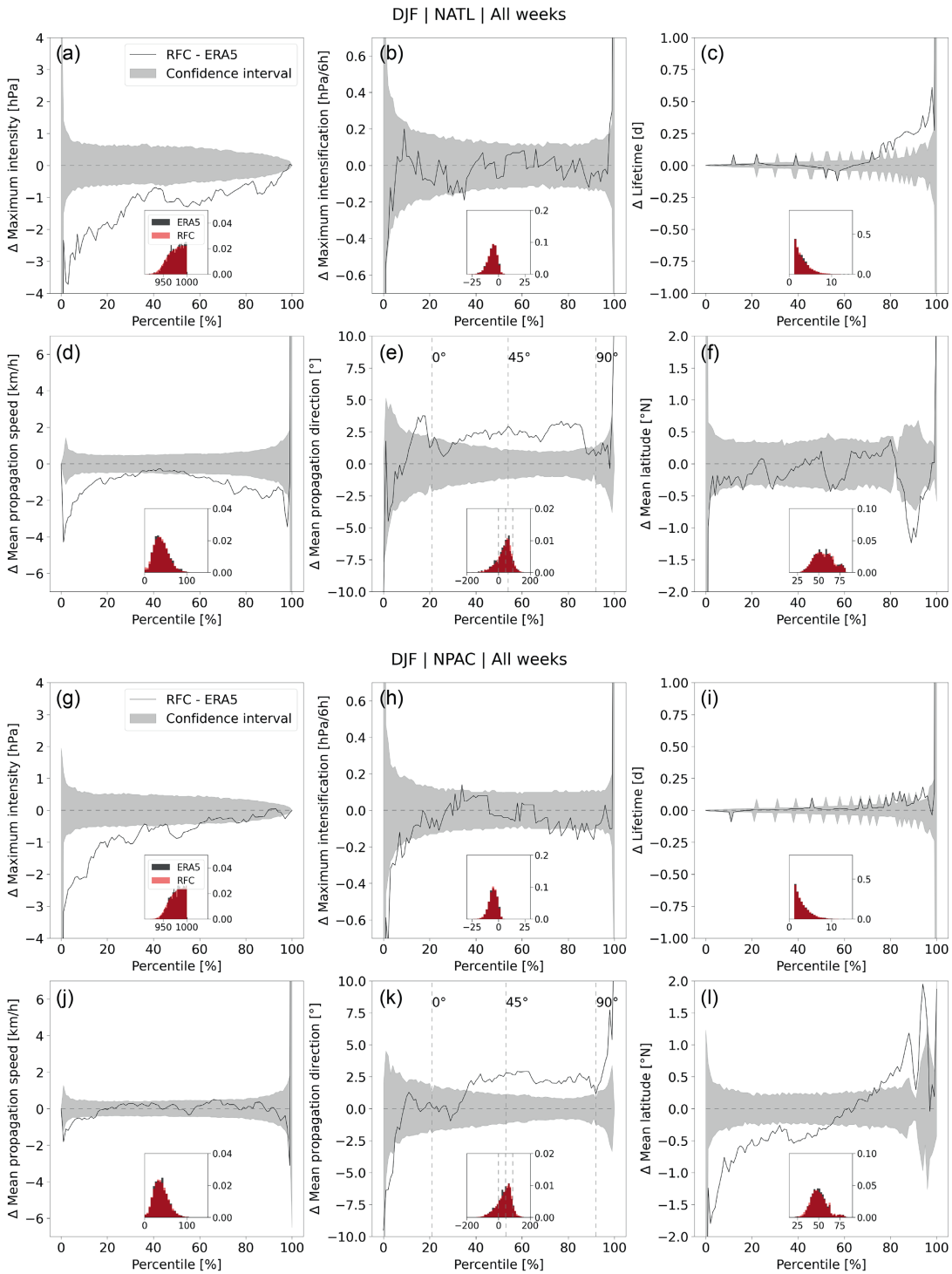

FIGURE 6 Winter (DJF) cyclone life-cycle characteristics (small panels) and corresponding biases (large panels) over all six lead-time weeks associated with those cyclone tracks that reach their maximum intensity in (a–f) the North Atlantic and (g–l) the North Pacific (see Figure S1 for exact domains). The small black and red histograms show the probability density (on the *y*-axis) in ERA5 and in the reforecasts, respectively, of (a,g) the maximum intensity (hPa; on *x*-axis), (b,h) the maximum six-hourly intensification rate (hPa/6hr), (c,i) the lifetime (d), (d,j) the mean propagation speed (km/hr), (e,k) the mean propagation direction (°; in azimuths clockwise from north, that is, 0° is northward, 90° is eastward, etc.), and (f,l) the mean latitude of each track (°N; cf. Section 2.2 for details). The black lines in the large panels show the difference between reforecasts and ERA5 of every *n*th percentile value of the two distributions (with $n = 0, 1, 2, \ldots, 99, 100$) and the gray shading indicates the corresponding 99.9% confidence interval. The confidence intervals are obtained with a bootstrapping, which is explained in some more detail in Section S.1. For ease of interpretation, the vertical lines in the panels of mean propagation direction indicate values corresponding to 0°, 45°, and 90°. The statistics are based on the numbers of cyclone tracks shown in Table 1.

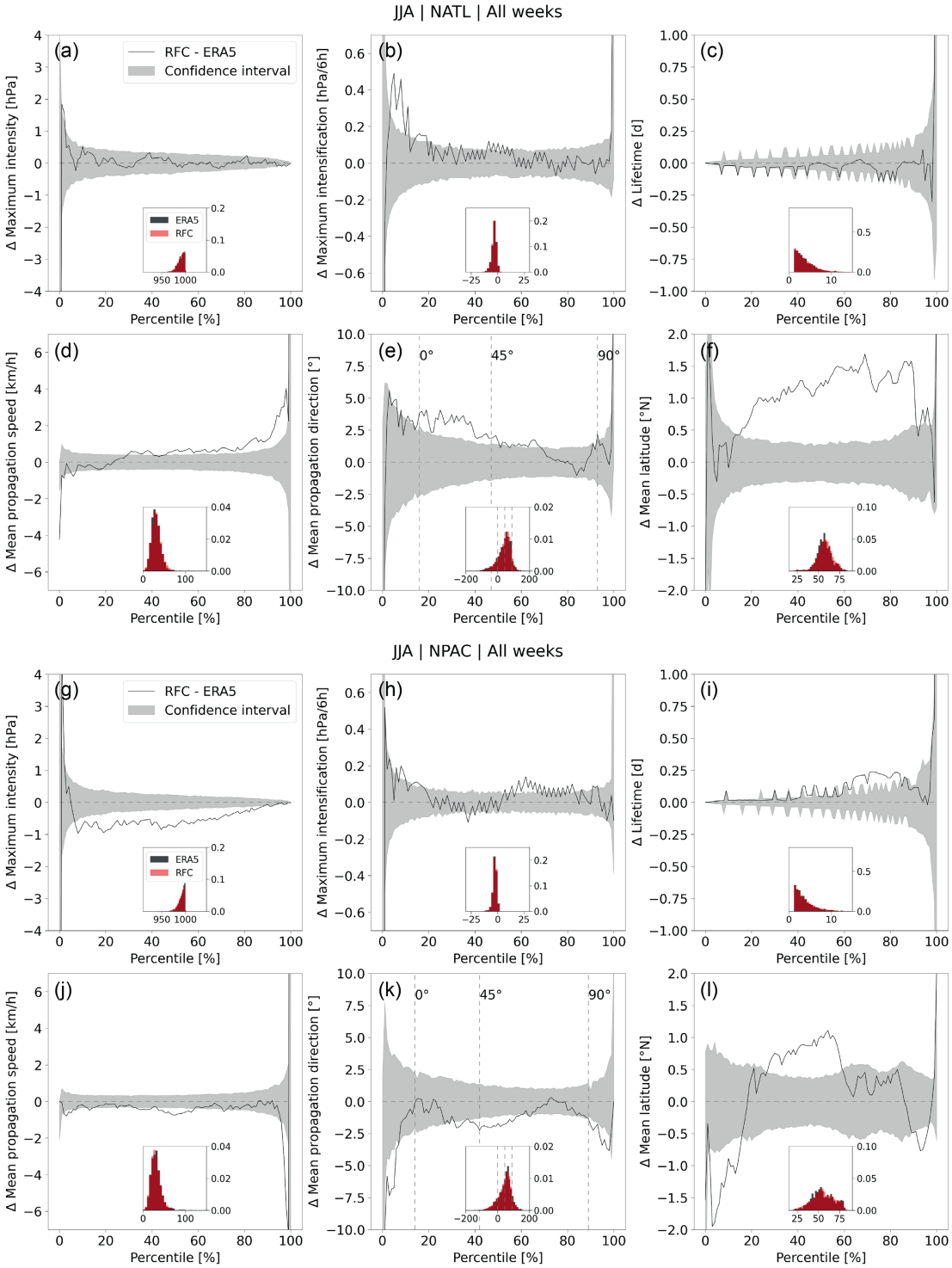

FIGURE 7 Same as Figure 6 but for summer (JJA). The statistics are based on the numbers of cyclone tracks shown in Table 1.

times (cf. Figure S6c), which could contribute to the overestimated maximum intensity of the most intense cyclones. This is supported by the fact that the overestimated cyclone lifetime is dominated by the intensification rather than the decay phase (defining the intensification phase as the time until the maximum intensity is reached and the decay phase as the time beyond; not shown). A further support of this hypothesis is the fact that the 20% most intense reforecast cyclones have a relatively long lifetime (cf. Figure S8a,c), and, vice versa, the 20% most long-lived reforecast cyclones reach relatively high maximum intensities (not shown). In contrast, cyclone

lifetime in the North Pacific is not biased and thus cannot contribute to the maximum intensity bias in this region (Figure 6i).

Another reason for the maximum intensity bias could be a poleward bias in the mean latitude of cyclones (because climatological background SLP decreases towards the poles; cf., e.g., Sinclair, 1995). However, the mean cyclone latitude is unbiased in the North Atlantic (Figure 6f). In the North Pacific, the northernmost cyclones are indeed significantly too far north (Figure 6l). However, statistics over the 20% most intense North Pacific cyclones (i.e., those with a too strong maximum intensity) indicate that they largely occur around the peak latitudes between 40°N and 60°N, where the mean latitude is rather unbiased (cf. Figure S8m,r,x). Biases in cyclone mean latitude thus can hardly explain the bias in maximum cyclone intensity.

A third reason for the maximum intensity bias could be that too intense cyclones start in zones where climatological background SLP is already too low. Indeed, there is a negative climatological SLP bias evolving upstream of both storm-track maxima (Northern Canada and parts of East Asia) and maximizing in weeks 5 and 6 (Figure S26a–d), that is, at the lead times when the bias in maximum intensity also maximizes. However, it is unlikely that all the too intense cyclones start in this region, as it is not a hotspot for genesis during winter (cf. Figure 5a). Furthermore, if the negative climatological SLP bias translated into too low SLP in the cyclone center, it would also have to reduce the SLP equally within the whole cyclone. To test whether this is the case, we additionally compute the bias in the difference between SLP at the cyclone edge (i.e., the outermost closed SLP contour) and the cyclone center at the time of cyclone maximum intensity, as well as the bias in cyclone size at the same time. This analysis shows that the largest SLP differences are significantly overestimated in the reforecasts in both ocean basins (Figure 8b,d), while the cyclone size is not biased in the North Atlantic (Figure 8a) and is slightly underestimated in the North Pacific (Figure 8c). Since an unbiased cyclone size means that the positive bias in the SLP difference goes along with a positive bias in the SLP gradient, the cyclones are truly too deep and we can rule out a major influence of the climatological SLP bias. This is the case for both the North Atlantic cyclones, and, even more so, for the North Pacific cyclones, in which the SLP gradient is even more biased due to the additionally underestimated cyclone size. Figure 8e further indicates that cyclones are generally too deep in most parts of the North Atlantic and North Pacific throughout their life cycle (i.e., not just at the time of maximum intensity; the too shallow cyclones at the southeastern edge of the North Pacific storm track are an exception). The positive bias in maximum intensity thus does not seem to be caused by a specific subset of cyclones (cf. also Figure S8) but rather reflects a systematic problem of overpredicting cyclone intensity in these regions in winter.

Figure 8 allows us further to speculate on the potential relevance of the biases in maximum intensity: for instance, in the North Atlantic during winter, the strongest SLP differences, which are of the order of 50 hPa, are overestimated by about 3 hPa (Figure 8b), which corresponds to a bias of about 6%. Assuming that these are the same cyclones as those with an overestimated maximum intensity of about 3 hPa (Figure 6a), this strengthening of the SLP difference (and thus of the SLP gradient given the unbiased cyclone size) by 6% would approximately manifest in an equally large intensification of near-surface wind speed considering geostrophic–frictional wind balance near the surface.

In addition to the bias in maximum cyclone intensity, cyclone propagation speed is significantly underestimated for both slow and fast cyclones in the North Atlantic (Figure 6d), while it is unbiased in the North Pacific (Figure 6j). The strongest underestimation occurs for the smallest percentiles, which indicates an overestimation of (quasi)stationary cyclones in the reforecasts. Again, this bias does not appear in lead-time weeks 1 and 2 yet but only at subseasonal lead times (cf. Figure S6d). Whether the underestimation of fast cyclone propagation speed is linked to the underestimation of the speed of the upper-tropospheric jet stream previously shown (e.g., Gray et al., 2014; Saffin et al., 2017; Schäfler et al., 2020) would require further investigation. Furthermore, the cyclones propagate slightly too zonally in both ocean basins, as indicated by the overestimated propagation directions larger than 45° (Figure 6e,k; note that this bias does not grow with lead time as shown in Figure S6e,k). Statistics over these too zonal cyclones in the North Atlantic show that they peak between 40°N and 50°N (not shown), which coincides with the positive cyclone frequency bias west of Northern Europe (cf. Figure 3a) and might thus indicate an example of a bias related to cyclone track displacements (cf. Figure 1c). Likewise, the too zonal reforecast cyclones in the North Pacific occur more equatorward (not shown) and might thus be associated with the overestimated cyclone occurrence equatorward of the North Pacific storm-track maximum (Figure 3a), which reflects a similar bias source to that in the North Atlantic.

3.3.2 | Summer

During summer, cyclone life-cycle biases in both ocean basins are generally not larger than during winter (Figure 7), despite the overall larger cyclone frequency

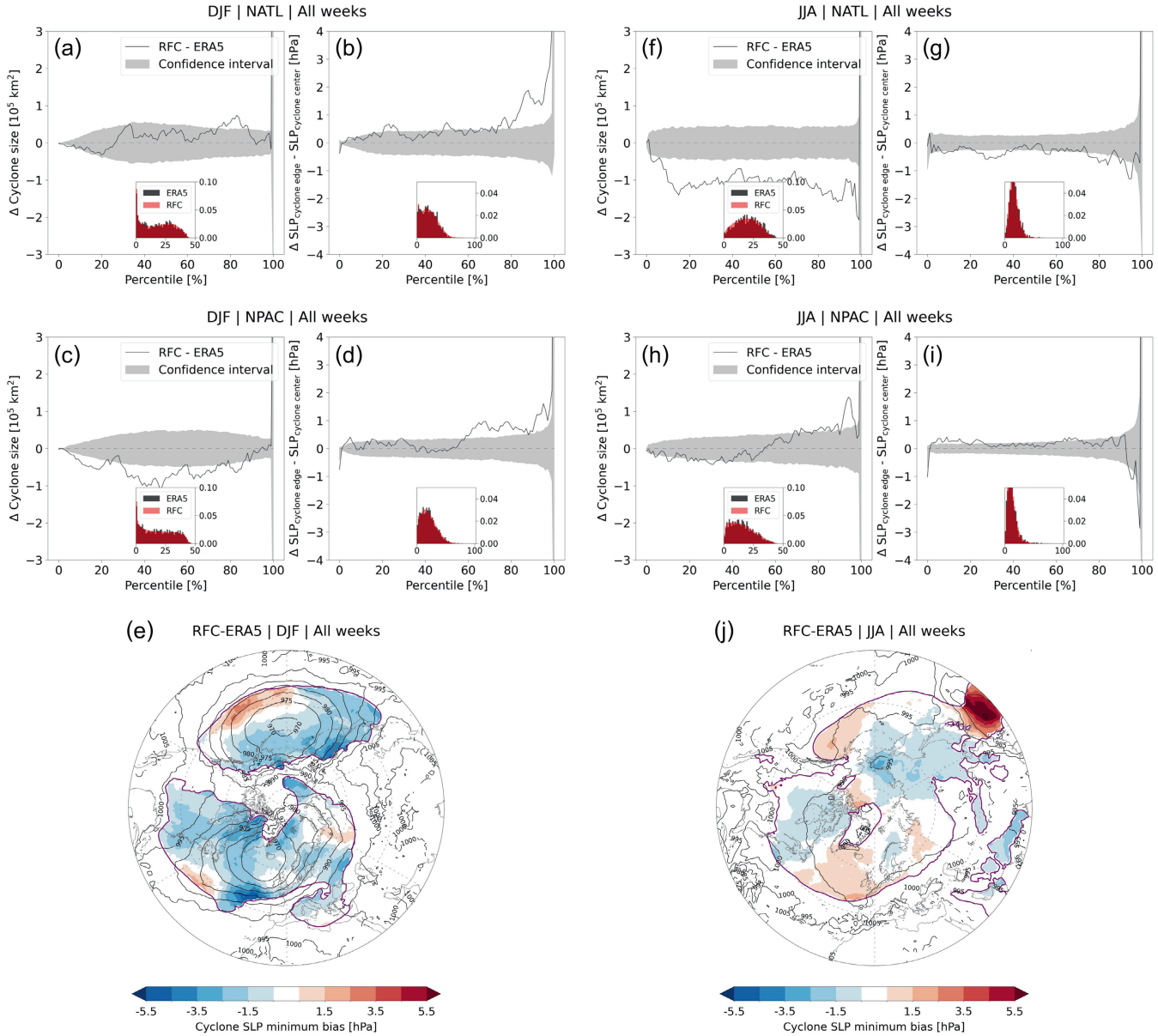

FIGURE 8 (a–d,f–i) Histograms and corresponding differences (at all lead times) as in Figure 6 but for (a,c,f,h) cyclone size (in 10^5 km^2, which is approximately the size of Iceland) and (b,d,g,i) SLP difference inside the cyclone, defined as the difference between the outermost closed SLP contour and the SLP at the cyclone center (a–d: North Atlantic; f–i: North Pacific), and (e,j) gridded cyclone SLP minimum reforecast biases (reforecasts minus ERA5, all lead times) as shading and corresponding absolute field in ERA5 as contours during (a–e) winter (DJF) and (f–j) summer (JJA). The gridded cyclone SLP minimum biases are only shown in regions where ERA5 has a cyclone frequency of at least 5% (indicated by the purple lines). The gridded cyclone SLP minimum fields (in ERA5 and the reforecasts) are computed as follows: for each track time step of each cyclone (i.e., not just at the time of maximum intensity), the current cyclone area is "filled" with the value of the cyclone's current SLP minimum. These "filled" cyclone area fields are summed up over the whole dataset, and the resulting sum is divided by the total number of time steps a grid point is covered by a cyclone. The resulting field can thus be interpreted as the mean depth that cyclones influencing this grid point have (whereby influencing means encapsulating this grid point with their outermost closed SLP contour).

biases in summer (cf. Figure 3g). An exception is the strong and significant northward bias in mean latitude in the North Atlantic (Figure 7f), which becomes significant at subseasonal lead times only (cf. Figure S7f) and is consistent with the considerable underestimation of cyclone frequency at the southern edge of the North Atlantic storm-track maximum (cf. Figure 3g). Furthermore and in contrast to winter, the maximum cyclone intensity in the North Atlantic during summer is nearly perfectly captured (Figure 7a), although the strongest

maximum intensification rates are slightly underestimated (Figure 7b; this bias increases with lead time as shown in Figure S7b). The fact that maximum cyclone intensity itself is not biased in the North Atlantic is consistent with the spatial biases in cyclone depth in this basin, which are generally smaller than in winter as well as both negative and positive and thus compensating over the whole basin, as shown in Figure 8j. It is also consistent with the unbiased cyclone lifetime (Figure 7c) and the unbiased SLP difference (Figure 8g).

In the North Pacific, however, maximum cyclone intensity is slightly but significantly overestimated for a large part of the cyclone population (Figure 7g), although maximum intensification rates are unbiased (Figure 7h). Despite its significance, further investigation would be required to understand the relevance of this bias, considering its small magnitude of 0.5–1 hPa. As during winter, this bias is maximized at subseasonal lead times (cf. Figure S7g). Although there is a positive lifetime bias for a fraction of the longer-lived cyclones as well (Figure 7i), in contrast to winter, this can hardly explain the maximum intensity bias, because it occurs mainly during the decay phase (not shown). Furthermore, the summer reforecasts are hardly affected by negative climatological SLP biases in the East Asian–North Pacific region (Figure S26i–l), which again rules out a substantial influence of climatological SLP biases on the bias in maximum cyclone intensity. On the other hand, there are also no substantial biases in the SLP difference and in cyclone size (Figure 8h,i), which implies that the reforecast cyclones tend to have equally low SLP at their edges as at their centers. The spatial distribution of biases in cyclone depth throughout the whole life cycle indicates that the overestimated maximum intensity might be due primarily to too deep cyclones in large parts of East Asia and the western North Pacific (Figure 8j).

Note that the missing peak of very small cyclones in both the North Atlantic and North Pacific during summer compared with winter (cf. histograms in Figure 8f,h) is due to the intensity threshold of 1,000 hPa used to select the cyclones. A higher threshold of 1,020 hPa changes the cyclone size histograms for summer, in the sense that the probability density peak also shifts to the smallest cyclones, as for winter. Nevertheless, the associated biases remain similar, even with this higher threshold (see also the discussion on the choice of the intensity threshold in Section 4.8).

Furthermore, the fastest North Atlantic reforecast cyclones are too fast (Figure 7d; this bias is only significant at subseasonal lead times as shown in Figure S7d). Statistics over these reforecast cyclones, however, reveal no distinct characteristics (not shown) and the reason for this bias would thus require further investigation. In contrast, the reforecast lacks (or at least underestimates) a set of very fast-moving North Pacific cyclones (negative bias for the largest percentiles in Figure 7j). Statistics over these fastest-propagating (but underestimated) reforecast cyclones reveal that they occur primarily around 40°N and tend to have a more zonally oriented propagation (not shown). Whether this indicates a possible connection to so-called diabatic Rossby waves (DRWs)[4] would require further investigation. Finally, the North Atlantic cyclones are significantly too small during summer (Figure 8f), while the size of the North Pacific cyclones is relatively well captured (Figure 8h).

As illustrated in our conceptual view of cyclone frequency bias sources (cf. Figure 1), the biases in cyclone life-cycle characteristics discussed thus show that the strong underestimation of cyclone frequency along the North Atlantic storm-track maximum during summer might be at least partly a result of the northward-displaced mean cyclone latitude, overestimated cyclone speed, and underestimated cyclone size, as well as the underestimated cyclone genesis across the whole North Atlantic discussed previously.

3.3.3 | Spring and autumn

The cyclone life-cycle biases in spring and autumn are not discussed in detail, but can be seen in Figures S9 and S10, and S11 and S12, respectively. The most striking biases in these seasons are an overestimation of maximum cyclone intensities (particularly of the more intense cyclones) in the North Pacific in both seasons (cf. Figures S9g and S11g) and in the North Atlantic in spring (cf. Figure S9a). This is consistent with the generally too low SLP minimum in the two basins and seasons (cf. Figure S13e,j) and demonstrates that the problem in capturing maximum cyclone intensity correctly is year-round, at least in the North Pacific. Furthermore, in the North Atlantic during autumn, the fastest propagating cyclones are too slow (cf. Figure S11b) and there is a tendency towards too zonal cyclone propagation (cf. Figure S11e). Finally, in the North Pacific during spring, the long-lived cyclones have a substantial positive lifetime bias (cf. Figure S9i).

4 | DISCUSSION

Our analysis of biases in cyclone frequency and properties opens up a variety of follow-up research questions regarding their relevance (because significance does not automatically imply relevance), their dynamical sources, their links with biases in the large- to planetary-scale flow, and their role for surface weather biases in, for

example, precipitation, wind speed, and near-surface temperature. While quantitatively addressing these questions goes beyond the scope of this study, we nevertheless discuss some aspects in the following. Note, however, that some of these aspects are clearly speculative and thus simply aim to provide avenues for further research.

4.1 | Negative cyclone frequency bias in central North Atlantic during summer

The fact that the cyclone frequency biases across the Northern Hemisphere tend to be largest during summer (Figure 3g) is consistent with Büeler et al. (2021), who found a similar seasonality of 500-hPa geopotential height biases in the same model. In particular, the substantial underestimation of cyclones in the central North Atlantic, which is the largest bias throughout the year, is consistent with the co-located positive bias in 500-hPa geopotential height that results in positive frequency biases of the weather regimes associated with ridging over the North Atlantic and Western Europe (i.e., the Atlantic Ridge, Scandinavian Trough, and European Blocking: Büeler et al., 2021). We find some evidence that this negative cyclone frequency might result from a combination of the following factors: the cyclones are located too far to the north, move too fast, are too small, and their genesis frequency is underestimated across the whole North Atlantic.

The dynamical origins of these biases, however, still need to be identified. One potential origin is the positive lower-tropospheric temperature bias found by Magnusson et al. (2022), particularly upstream over North America but also over the North Atlantic, in the same reforecast system. This could result in biases in baroclinicity with impacts on cyclone evolution. In contrast, the cyclone biases themselves might feed back or even maintain the lower-tropospheric temperature biases, particularly over the North Atlantic. Another potential origin might be biases in sea-surface temperature (SST) in the Gulf Stream area, introduced by the relatively coarse spatial resolution of both the ocean and atmospheric model components (e.g., Hewitt et al., 2017). Again, such biases could affect the location and structure of the lower-tropospheric baroclinic zone. Indeed, Roberts et al. (2021) found significantly higher ECMWF subseasonal forecast skill for the North Atlantic and Europe during winter when improving location and structure of the Gulf Stream by bias-correcting SSTs. Since, in summer, SST-related biases might play an even more important role for cyclone dynamics in a relative sense due to the generally weaker forcing from upper-tropospheric Rossby waves (compared with winter), it might be insightful to repeat a similar analysis to that in Roberts et al. (2021) for summer. Identifying the sources of these substantial North Atlantic large-scale biases is ongoing work at ECMWF (Linus Magnusson, personal communication, 2022) and is crucial to improve the notoriously challenging subseasonal summer predictions over Europe further.

4.2 | Negative cyclone frequency bias in part of eastern North Pacific during winter

The underestimated cyclone frequency in the eastern North Pacific (Figure 3a), which is the largest bias in winter, is likely linked to the known bias in the variability of the North Pacific jet stream: Winters (2021) showed that subseasonal forecasts of the ECMWF tend to overpredict the zonally retracted phase of the jet but underpredict its zonally extended phase. This is in line with Vitart et al. (2022), who found the eastern edge of the North Pacific jet to be too far to the west in the same model. The favorable conditions for cyclone development at the left jet exit in the eastern North Pacific might thus be reduced by this westward bias. Although the jet bias might partly be responsible for this underestimation of cyclone frequency, the lack of cyclones could in turn feed back and help to maintain the jet bias due to the reduced sharpening of the tropopause (e.g., Wirth & Szabo, 2007; Chagnon & Gray, 2015). There is also an underestimated initiation frequency of Rossby-wave packets in the central North Pacific in the same model (Quinting & Vitart, 2019), which might as well be linked to the underestimated cyclone frequency. Furthermore, Vitart et al. (2022) showed that the North Pacific jet bias seems to originate from errors in the Rossby-wave source (advection of vorticity by the divergent flow) upstream over the Sea of Japan at early lead times. The underestimated cyclone genesis frequency east of Japan found in our analysis (Figure 5a) likely reflects this error and confirms that the underestimated cyclone frequency in the eastern North Pacific might be the result of both upstream and in situ biases.

On the other hand, the underestimated cyclone frequency is likely responsible for the underestimated WCB outflow further downstream along the North American west coast, as found by Wandel et al. (2021) in the same model. This indicates the potential importance of the underestimated cyclone frequency for predicting the downstream Rossby-wave evolution as well as precipitation along the west coast. Focusing even more downstream, Vitart et al. (2022) argued that the bias in the North Pacific jet, and thus in cyclone frequency, still hinders exploitation of the potential subseasonal predictability of the NAO. This is given by the fact that the NAO can be modulated by the teleconnection from the

Madden–Julian Oscillation via the North Pacific (e.g., Cassou, 2008; Lin et al., 2009).

4.3 | Year-round overestimation of maximum cyclone intensity

The overestimation of maximum cyclone intensity (Figures 6,7, S9, and S11), which can be found year-round and in both ocean basins (expect for summer and autumn in the North Atlantic) and mostly for the intense part of the cyclone population (except for summer in the North Pacific), is in line with the findings of Froude et al. (2007) and Froude (2010) for IFS medium-range forecasts. This is surprising, given the fact that these studies analyzed much older IFS versions than we do. The new aspect of our analysis is that this bias grows with lead time and, in most cases, reaches its maximum value at lead-time weeks 5 and 6 (at least among the evaluated 6 weeks, because the bias might change further if the reforecasts were run for longer). The overestimation does not seem to be related to a specific subset of cyclones but rather is caused by a general overestimation of cyclone depth in large parts of the two ocean basins. Furthermore, we show that in most regions and seasons the cyclones are truly too deep (i.e., their SLP gradient is too strong), which largely rules out a significant influence from any negative biases in background SLP.

One potential source of intensity bias is the overestimated lifetime we find for some long-lived cyclones, which can result in deeper cyclones even with an unbiased intensification rate. However, as this lifetime bias only appears in some regions and seasons, it cannot explain the full extent of the intensity bias. Another source of this bias could be the higher native horizontal resolution in the reforecasts during the first 15 days of lead time compared with ERA5 (cf. Section 2.1), allowing for deeper cyclone centers (even after interpolating the two datasets to the same horizontal resolution for verification, as done in this study). However, as the intensity bias grows further at subseasonal lead times, at which the two native resolutions are basically equal, the mismatch in native resolution can also contribute to the intensity bias only partly. A more detailed investigation is thus required, and we can only speculate about potential further sources. Since the bias grows further at subseasonal lead times, it could be linked to systematic model drifts in large-scale drivers of cyclone intensification such as the jet stream and tropospheric baroclinicity. Moreover, it could be related to the representation of microphysical processes. Pickl et al. (2022) showed that ECMWF's stochastically perturbed parameterisation tendency (SPPT) scheme systematically increases the number of rapidly ascending airstreams, although the SPPT perturbations are symmetric around zero (a similar behavior has recently been found for the stochastically perturbed parametrisations (SPP) scheme, which is currently under development at ECMWF: Deinhard & Grams, 2023). As rapidly ascending airstreams play an important role for cyclone intensification (e.g., Binder et al., 2016), the SPPT could thus contribute to a higher maximum cyclone intensity through the same mechanism. A first step to test this hypothesis could be to investigate cyclone intensity biases in the control hindcasts, in which the SPPT scheme is switched off. Furthermore, Froude et al. (2007) and Froude (2010) found an underestimated cyclone propagation speed, which they hypothesized to result from the overestimated cyclone intensity. The physical argument behind this link is the stronger phase-locking of a stronger surface cyclone with its upper-level trough, which tends to reduce the propagation speed of the cyclone. Indeed, we do find an underestimated cyclone propagation speed in some regions and seasons in which maximum cyclone intensity is underestimated, but not in all.

4.4 | Lead-time dependence of biases

We find that the patterns of most biases in cyclone frequency and life-cycle characteristics already appear at medium-range lead times, but their magnitudes grow further at subseasonal lead times and often reach their maxima in weeks 5 and 6 only (at least among the evaluated six weeks; Figures 3, S3, S6, S7, S10, and S12). It is thus likely that most of these biases originate from systematic model biases at early lead times. Better understanding and reduction of cyclone biases in subseasonal forecasts should thus be considered in concert with identifying and understanding biases in medium-range forecasts. The reason for the continuous increase of the bias magnitudes at subseasonal lead times is probably multifaceted (e.g., Mulholland et al., 2015; Hermanson et al., 2018; Shonk et al., 2018; Voldoire et al., 2019; Merryfield et al., 2020): first of all, it is related to the general and highly complex model drift that results from the accumulation of the aforementioned model biases with lead time as well as from balancing physical imbalances in the ensemble's initial conditions (although the latter should happen primarily during the first few days of the forecast). Furthermore, biases in lower-frequency climate phenomena influence extratropical dynamics via teleconnections on multiweekly timescales and might thus become particularly effective at subseasonal lead times only. Finally, the change from a higher to a coarser horizontal resolution after 15 days reforecast lead time (cf. Section 2.1) might also contribute to the increase of bias magnitudes beyond the medium-range.

4.5 | Flow dependence of biases

It is well-known that subseasonal forecast performance is strongly flow-dependent (e.g., Vitart, 2014; Ferranti et al., 2015, 2018; Büeler et al., 2021), similar to the flow dependence of ensemble prediction already observed at short lead times (e.g., Rodwell et al., 2018). Although a systematic analysis of flow dependence of cyclone biases would go beyond the scope of this study, we shed some light on this question. More specifically, we have investigated how cyclone frequency and its bias depend on the two main modes of Northern Hemisphere extratropical variability, the NAO and PNA, at reforecast initialization (cf. Section 2.3).

The first noteworthy result of this analysis is that, in ERA5, strong NAO or PNA phases coincide with distinct hemispheric cyclone frequency patterns not just 1–2 weeks but up to 5–6 weeks ahead (cf. Figures S14, S15, S16, and S17). For instance, strong NAO− states during winter tend to be followed by more cyclones around the North American east coast, around the Azores, and in the Mediterranean compared with strong NAO+ states up to six weeks ahead (cf. Figure S14c,f,i compared with Figure S14a,d,g). Likewise, strong NAO− states during summer tend to be followed by more cyclones in the North Atlantic storm-track maximum compared with strong NAO+ states up to six weeks ahead (cf. Figure S16c,f,i compared with Figure S16a,d,g). How robust this finding is and to what degree it is driven by very persistent NAO phases with a lifetime of multiple weeks would require further investigation. Nevertheless, it indicates that strong NAO or PNA states might be windows of opportunity for subseasonal predictability of Northern Hemisphere cyclone activity.

The second noteworthy result is the fact that both pattern and magnitude of the cyclone frequency biases, particularly in the North Atlantic and to some degree also in the North Pacific, can be different depending on the state of the NAO or PNA at reforecast initialization (cf. Figures S18, S19, S20, S21, S22, S23, S24, and S25). For instance, the North Atlantic biases tend to be larger following NAO− than following NAO+ throughout the year (cf. Figures S18b,c, S20b,c, S22b,c, and S24b,c). More specifically, the reforecasts produce too few cyclones around the North Atlantic storm-track maximum following NAO−. The model thus likely struggles to capture the general southwestward shift of cyclone genesis associated with NAO− states. One reason could be a bias in the strength or position of the southward-shifted jet stream, which would be associated with a bias in the lower-tropospheric baroclinic zone that is, particularly along the southeastern US coast, driven by the Gulf Stream. Investigating North Atlantic cyclone frequency biases in such a flow-dependent way in SST-bias-corrected reforecasts (cf. Roberts et al., 2021), as proposed earlier in this section, could help us to test this hypothesis and understand better the strong underestimation of North Atlantic cyclones during summer, in particular.

Apart from the dynamical insight, our analysis further indicates the potential risk of correcting for mean biases of subseasonal forecasts. We thus suggest exploring flow-dependent forecast calibration techniques that account for the flow dependence of the model errors. Finally, it is important to mention that this analysis only gives a glimpse of flow dependence: it would be rewarding to also take into account teleconnections from remote and slower modes of variability (such as the Madden–Julian Oscillation, El Niño–Southern Oscillation, or stratospheric polar vortex) when investigating subseasonal forecast performance. Such an analysis, however, goes beyond the scope of this study.

4.6 | Role of biases in cyclone frequency and life-cycle characteristics for surface weather biases

As cyclones are a key driver of surface weather variability, their biases likely influence biases in surface weather fields such as precipitation, wind speed, or temperature. However, it is challenging to link the cyclone frequency biases identified here to surface weather biases found in other studies for the following reasons. The cyclone frequencies over land, where surface weather biases are most relevant, are much smaller than over the main ocean basins. Cyclone frequency biases in these regions might thus be less robust and have to be interpreted with caution. Furthermore, biases in surface weather result from a complex interplay of large-scale and local factors.

A pathway for future research could thus be to decompose local surface weather biases into the contribution from cyclones and the contribution from phenomena other than cyclones (in a methodologically similar way to, e.g., Hawcroft et al., 2012; Owen et al., 2021; Rüdisühli et al., 2020). This would pinpoint those regions where the cyclone biases might be most critical for predicting surface weather at subseasonal lead times. However, we should keep in mind that the effect of cyclone biases on surface weather can also be nonlocal, in the sense that an error in cyclone propagation can lead to erroneous advection of warm and moist air downstream. Last but not least, there is still a lack of studies that identify and describe year-round surface weather biases systematically. An exception is the study by Lavers et al. (2021), which found an overall wet bias in ECMWF's medium-range forecasts, most strikingly over Southeast Asia during the boreal summer half year. It is likely that the overestimated cyclone frequency in

summer in Southeast Asia (Figure 3g) contributes to this precipitation bias, either directly or indirectly by modifying the moisture flux into the biased region (cf. also Lavers et al., 2021, for further details). Note that the same region also exhibits a strong positive lower-tropospheric temperature bias (Magnusson et al., 2022), which might be linked to the positive cyclone frequency bias in different ways (for instance, modified low-level baroclinicity and thus cyclogenesis due to biased temperature, modified warm air advection due to biased cyclone frequency, or modified surface radiation due to biased cyclone frequency). Another exception is Monhart et al. (2018), who identified weather-station-based near-surface temperature and precipitation biases over Europe in subseasonal ECMWF reforecasts. Although they did not find a clear spatial pattern for the precipitation biases, they showed that there are many more locations with negative precipitation biases in Western Europe in summer and autumn than in winter and spring. It is thus likely that the strong underestimation of eastern North Atlantic cyclones in summer contributes to this bias, either directly via a lack of frontal precipitation or indirectly via modifying moisture transport and eventually thermodynamic conditions for convective precipitation (cf., e.g., Rüdisühli et al., 2020).

4.7 | Cyclone frequency biases versus SLP biases

Our analysis is based on cyclones as meteorological objects and thus aims to provide additional dynamical insight to what can be learned from a simple climatological SLP bias analysis. This is obvious when comparing the cyclone frequency biases (Figures 3 and S3) with the SLP biases (Figure S26): as expected, many of the substantial cyclone frequency biases are anticorrelated with the SLP biases (positive cyclone frequency bias causes negative SLP bias and vice versa), such as the negative cyclone frequency bias over the eastern North Pacific in winter or the negative biases over the central North Atlantic and North Pacific in summer. However, there are various cyclone frequency biases that do not manifest as (substantial) SLP biases, such as the positive cyclone frequency biases in the western North Pacific in winter and (particularly) spring or the positive biases over the Bering Sea and the Sea of Okhotsk in summer. Conversely, there are SLP biases that are not or not only caused by the type of cyclones investigated in this analysis, such as the positive SLP bias all over Europe, Northern Africa, and large parts of Asia in autumn. Hence, the SLP bias is a complex sum of biases in baroclinic, thermal, transient, and stationary cyclones and anticyclones, and thus marginally useful for pinpointing the dynamical origins of large-scale circulation and ultimately surface weather biases. In this study, we have thus focused on one contribution to the SLP bias, namely the one from cyclones. Future research could thus perform a similar verification specifically for other contributors such as thermally driven stationary cyclones or transient as well as stationary anticyclones (although this would require adapted or alternative identification and tracking algorithms).

4.8 | Selecting the most appropriate cyclone population

As described in Section 2.2, our bias analysis is based on all cyclones with a maximum intensity stronger than 1,000 hPa. This threshold obviously excludes the non-negligible weak tail of the cyclone population, as the sharp cut in the distributions of maximum intensity indicates (cf. Figures 6 and 7). We have thus repeated our whole analysis with a weaker maximum intensity threshold of 1,020 hPa (as well as for all cyclones independent of any tracking criteria, which, however, only allows computation of biases in spatial cyclone frequency but not in any cyclone life-cycle characteristics). The bias patterns of both cyclone frequency and cyclone life-cycle characteristics hardly change with this new threshold and the key results of this study do not change at all (not shown). Furthermore, the weaker threshold incorporates many more well-known heat lows in summer, for instance over the Gulf of California (e.g., Badan-Dangon et al., 1991) or the Sahara (e.g., Lavaysse et al., 2009) (both of them can also be seen in the track-independent cyclone frequency in Figure S2e), and reveals a strong underestimation of their frequency in the reforecasts. Although these systems would be interesting to study, it would not be meaningful to put them into the same pot as the more "classic" and dynamically very distinct cyclones of the extratropical storm-track maxima. Finally, a weaker maximum intensity threshold also includes more spurious cyclones, because the cyclone algorithm used in this study (but also many other existing methods) is not tailored for the life cycle of very weak SLP minima. For these reasons, we conclude the maximum intensity threshold of 1,000 hPa to be appropriate for the purpose of this study.

5 | CONCLUSIONS

For the first time, we have analyzed the year-round biases of systematically identified and tracked Northern Hemisphere cyclones in subseasonal (re)forecasts of the ECMWF. While, qualitatively, the reforecasts can reproduce the overall storm-track characteristics remarkably

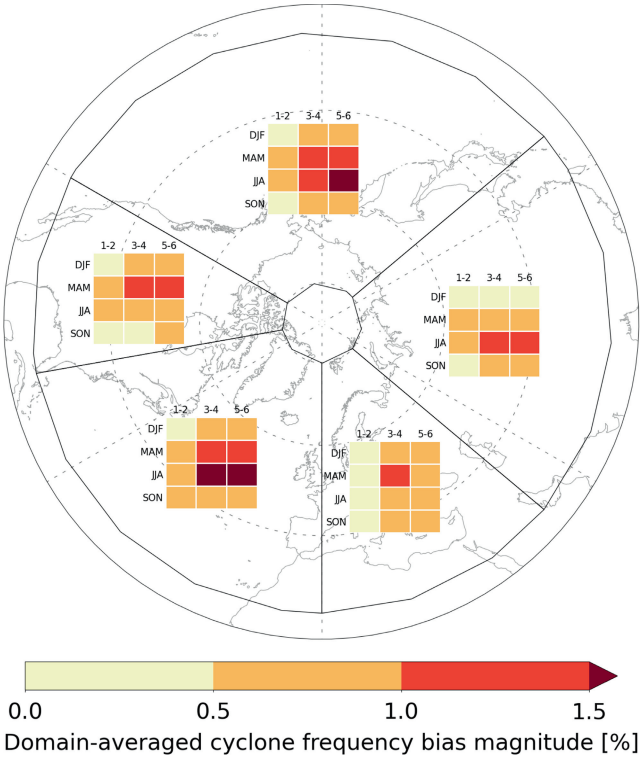

FIGURE 9 Domain-averaged cyclone frequency bias magnitudes depending on lead time and season. Note that the color levels and domains are arbitrary and just aim to summarize the biases in a qualitative way.

well up to six weeks ahead, there are potentially important regional biases. The spatial cyclone frequency biases tend to be largest in summer, relatively large in spring, and smallest in winter and autumn (Figure 9). They are larger over the ocean than over land (Figure 9), due primarily to the higher absolute cyclone frequencies over the ocean. Nevertheless, there are substantial relative cyclone frequency biases over land, the robustness of which, however, is smaller due to much lower absolute cyclone frequencies in these regions. While the patterns of the cyclone frequency biases often already appear at lead-time weeks 1 and 2, their magnitudes grow substantially further at subseasonal lead times, in some cases up to weeks 5 and 6 (Figure 9). The summer season is characterized by a poleward bias of the storm track in both the North Atlantic and the North Pacific. The largest and most remarkable bias in summer is an underestimation of cyclones at the southern edge of the North Atlantic storm-track maximum by more than 30% in relative terms. The biases in spring and autumn partly resemble a Rossby-wave-like pattern over the domain extending from North America to Europe. In winter, the most pronounced bias is an underestimation of cyclones in the eastern North Pacific. Furthermore, there are also biases in cyclone genesis frequency, which often coincide with the cyclone frequency biases. This indicates that some of the cyclone frequency biases are explained by corresponding in situ genesis biases. However, there are also upstream biases such as the underestimated genesis frequency along the North American and Asian east coasts during winter and spring or the overestimated genesis frequency in the lee of the Rocky Mountains almost year-round (but most pronounced during spring), which likely contribute to the cyclone frequency biases downstream.

We have further analyzed how well the reforecasts reproduce a set of important cyclone life-cycle characteristics in the two ocean basins. Although maximum cyclone intensification rates (averaged over 6 hr) are captured relatively well up to 6 weeks ahead, the reforecasts tend to yield too strong maximum cyclone intensities throughout the year (except for summer and autumn in the North Atlantic). This bias in maximum cyclone intensity is most pronounced for the intense half of the cyclones (except for summer in the North Pacific). It already appears at medium-range lead times, but grows further at subseasonal lead times and maximizes at weeks 5 and 6. We have provided evidence that negative biases in background SLP, which emerge in certain storm-track entrance regions, are hardly responsible for the overestimated maximum cyclone intensity. Instead, the overestimated cyclone lifetime as well as the higher native horizontal resolution in the reforecasts at medium-range lead times compared with ERA5 might be responsible for the overestimated maximum cyclone intensity in some regions and seasons. However, there are likely further, so far unidentified, sources for the systematic biases in maximum cyclone intensity, which might be linked to the representation of physical processes in the model. We have provided some arguments for why the SPPT scheme of the model as well as biases in SSTs could contribute to such biases. Furthermore, propagation speed in the North Atlantic is underestimated for the slowest and fastest cyclones during winter, overestimated for the fastest cyclones during summer, and underestimated for the fastest cyclones during autumn. Propagation speed is captured better in the North Pacific, with the exception of an underestimated propagation speed of very fast-moving cyclones during summer. Finally, the cyclones tend to propagate slightly too zonally in both ocean basins during winter and in the North Atlantic during autumn. Although some of the biases in propagation speed and direction grow with lead time, these biases reveal a less systematic lead-time dependence than do the biases in maximum cyclone intensity. Such differences in lead-time dependence might indicate fundamentally different dynamical sources of these biases and might be useful for understanding these biases better in future research.

Apart from lead-time dependence, we have also shed some light on the flow dependence of cyclone frequencies and their biases. First, we have shown that strong NAO or PNA phases can coincide with distinct hemispheric cyclone frequency patterns up to 5 and 6 weeks later—most pronounced in the North Atlantic during summer and slightly less during winter. Although the robustness of this result will have to be investigated further, it indicates that strong NAO or PNA states might be windows of opportunity for subseasonal predictability of Northern Hemisphere cyclone activity. Second, we have shown that both the pattern and magnitude of the cyclone frequency biases, particularly in the North Atlantic and to some degree also in the North Pacific, can be different depending on the state of the NAO or PNA at reforecast initialization. One of the most pronounced differences is the stronger underestimation of cyclone frequency around the North Atlantic storm-track maximum in reforecasts initialized with strong NAO− compared with reforecasts initialized with strong NAO+ in all seasons. This might reflect problems of the model in capturing the southward-shifted jet stream and its link to cyclone formation following NAO− states, and, moreover, could help us to understand better the strong underestimation of North Atlantic cyclones during summer.

The aim of our study has primarily been to identify, quantify, and describe the year-round biases in cyclone frequency and life-cycle characteristics. A logical and important next step would be to identify the dynamical sources of these biases, which, however, is a considerable challenge for the following reasons. First of all, we have illustrated that—regardless of any external dynamical sources—local cyclone frequency biases can result from a complex mix of biases in cyclone number, size, location, lifetime, and propagation speed. For instance, we have demonstrated that the largest cyclone frequency bias throughout the year, the underestimation of cyclones in the North Atlantic during summer, likely results from an underestimated genesis frequency (locally and upstream) and thus number of cyclones, an underestimated cyclone size, a northward-displaced mean cyclone latitude, and an overestimated cyclone propagation speed. Although we have indicated how such sources likely dominate some of the cyclone frequency biases, the sources of other frequency biases remain less understood. Furthermore, tracing these biases in cyclone frequency and life-cycle characteristics back to their dynamical sources will require tailored sensitivity experiments, which goes beyond the scope of this study. However, our analysis of lead-time dependence has provided some evidence that most of the dynamical sources already act in the early medium range, and might thus be linked to biases in the model physics and/or initial analyses. Nevertheless, the further increase of biases at subseasonal lead times might additionally be influenced by biases acting on longer timescales as well as by the increase to a coarser native horizontal resolution in reforecasts after the medium range.

Our year-round analysis is based on a mix of reforecasts from three different IFS model versions, which is clearly a caveat, considering some of the important changes that have been made from one version to the next (such as in the number of vertical levels or moist physics; cf. Section 2.1). This means that some of the biases found in our study could be different if quantified within a single model version only. Likewise, differences among the seasons could partly result from the fact that some seasons in our analysis are dominated by one specific model version. However, there is no way around this caveat, because the ECMWF reforecast strategy does not allow for another way of verifying a whole (recent) year of forecast data in a statistically robust way (unless there is, by chance or by design, a whole year of reforecasts without any major model updates). Furthermore, our findings are based on one specific method to identify and track extratropical cyclones (cf. Section 2.2). Although this method is well-established and has been shown to neither outperform nor perform worse than other well-established methods on average (Neu et al., 2013), some of the biases, particularly in cyclone life cycles, might be different when using different cyclone identification and tracking methods. In summary, however, it is unlikely that some of the substantial and thus key biases of our analysis would completely disappear for single model versions or different cyclone identification and tracking methods. Nevertheless, we would like to encourage the community to look at systematic biases in large forecast datasets by applying different cyclone identification and tracking methods in the future.

In conclusion, our study thus suggests several pathways for future research: the substantially larger cyclone frequency biases during summer than during winter indicate that the subseasonal summer forecasts seem to have fundamental problems in capturing the extratropical large-scale circulation correctly. Considering the increasing importance of skillful subseasonal predictions of precipitation and heatwaves in future summers, our research community should thus put more effort into better understanding and improving predictability during summer. A specific suggestion in this context would be to run summer reforecasts with bias-corrected SSTs (similar to Roberts et al., 2021) to understand the importance of SSTs (and thus of a higher spatial resolution of the ocean model component) for subseasonal predictability during this season. Another suggestion would be to identify extratropical cyclone biases in the two separate sets of medium-range and subseasonal reforecasts, which have

been newly introduced with IFS model version CY48R1 (operational since June 2023)[5]. This would allow us to quantify the influence of the atmospheric horizontal resolution on the behaviour and growth of extratropical cyclone biases in the first two lead-time weeks, because the medium-range (re)forecasts are run with 9 km resolution throughout the 15 days, while the subseasonal (re)forecasts are run with 36 km resolution throughout the 46 days (i.e., with a coarser resolution in the first 15 days compared with the previous model cycles (cf. Section 2.1), which is intended to be compensated by the doubling of the number of operational ensemble members from 51 to 101).

Furthermore, we propose to perform more feature-based verification of forecasts, as done in this study, to improve our dynamical understanding of subseasonal model drift. Given the appearance of many synoptic biases at early lead times, however, such verification studies should be done for medium-range and subseasonal forecasts in concert. More specifically, our cyclone verification could be extended by additionally identifying and tracking moist- and dry-dynamical drivers of cyclone intensification in the vicinity of the cyclones (following the principles of, e.g., Graf et al., 2017; Besson et al., 2021). This would help us to trace the biases in cyclone frequency and life-cycle characteristics back to their dynamical origins. The same type of verification could be done for further synoptic systems such as anticyclones (without focusing on blocking only), which have not been the focus of this study but still contribute to the overall SLP bias. On the other hand, surface weather biases could be decomposed into contributions from cyclones (and other phenomena, respectively), which would, for instance, demonstrate the role of the strong underestimation of North Atlantic summer cyclones for European surface weather forecasts. Moreover, a quantification of biases in heat and momentum fluxes caused by biases in cyclone frequency might help us to understand better the upscale influence of cyclone biases on the evolution of lower-frequency planetary-scale phenomena (such as the stratospheric polar vortex), which are typically important drivers of subseasonal predictability.

Last but not least, all these types of verification analyses should be performed for forecasts from various modeling centers, because this can help us to understand biases better without running expensive sensitivity simulations.

AUTHOR CONTRIBUTIONS

Dominik Büeler: conceptualization; data curation; formal analysis; investigation; methodology; validation; visualization; writing – original draft. **Michael Sprenger:** conceptualization; data curation; methodology; software; supervision; writing – review and editing. **Heini Wernli:** conceptualization; data curation; funding acquisition; methodology; software; supervision; writing – review and editing.

ACKNOWLEDGEMENTS

We thank Tim Hewson (ECMWF) and an anonymous further reviewer for their detailed comments, which helped to improve this article. ECMWF and MeteoSwiss are acknowledged for granting access to the subseasonal IFS reforecast data with six-hourly output. We further thank Linus Magnusson and Frédéric Vitart (ECMWF) for helpful comments at some stages of the project.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

The raw ECMWF hindcast data with six-hourly output are not publicly available and thus have been downloaded from the ECMWF archive with a special access available to the authors (provided by MeteoSwiss). The publicly available raw ERA5 reanalysis data (https://doi.org/10.24381/cds.143582cf) have been downloaded from https://apps.ecmwf.int/data-catalogues/era5. The postprocessed cyclone track data are available from the authors upon reasonable request.

ENDNOTES

[1]Meridionally ascending and strongly diabatically influenced airstreams in extratropical cyclones (Browning et al., 1973; Madonna et al., 2014). They influence cyclone intensification (e.g., Binder et al., 2016) and downstream Rossby-wave evolution (e.g., Grams et al., 2011).

[2]Information retrieved from https://confluence.ecmwf.int/display/S2S/ECMWF+Model, https://confluence.ecmwf.int/display/S2S/ECMWF+model+description, and https://confluence.ecmwf.int/display/FCST/Changes+to+the+forecasting+system on September 26, 2023.

[3]Daily NAO index retrieved from https://ftp.cpc.ncep.noaa.gov/cwlinks/norm.daily.nao.cdas.z500.19500101_current.csv and daily PNA index retrieved from https://ftp.cpc.ncep.noaa.gov/cwlinks/norm.daily.pna.cdas.z500.19500101_current.csv on June 22, 2022.

[4]DRWs are a special type of small-scale, shallow, and strongly diabatically driven summer cyclones, which can intensify strongly and reach higher propagation speed than classic cyclones, because they are, at least initially, not phase-locked with the upper-tropospheric wave guide (Boettcher & Wernli, 2011; Moore & Montgomery, 2004; Wernli et al., 2002). DRWs in the North Pacific typically evolve from

strong mesoscale convective systems (MCSs) forming along the mei-yu front east of China across Taiwan (Boettcher & Wernli, 2013; Chen et al., 2008), where our analysis indicates a strong underestimation of SLP minima by the reforecasts (cf. Figure 8b).

[5]Information retrieved from https://confluence.ecmwf.int/display/FCST/Implementation+of+IFS+Cycle+48r1on September 26, 2023.

ORCID

Dominik Büeler https://orcid.org/0000-0002-9904-6281
Michael Sprenger https://orcid.org/0000-0002-9317-8822
Heini Wernli https://orcid.org/0000-0001-9674-4837

REFERENCES

Afargan-Gerstman, H., Büeler, D., Wulff, C.O., Sprenger, M. & Domeisen, D.I.V. (2022) Stratospheric influence on the winter North Atlantic storm track in subseasonal reforecasts. *Weather and Climate Dynamics Discussions*. Available from: https://doi.org/10.5194/wcd-2022-58

Ambaum, M.H. & Hoskins, B.J. (2002) The NAO troposphere-stratosphere connection. *Journal of Climate*, 15, 1969–1978. Available from: https://doi.org/10.1175/1520-0442(2002)015<1969:TNTSC>2.0.CO;2

Attada, R., Dasari, H.P., Parekh, A., Chowdary, J.S., Langodan, S., Knio, O. et al. (2019) The role of the Indian summer monsoon variability on Arabian peninsula summer climate. *Climate Dynamics*, 52, 3389–3404. Available from: https://doi.org/10.1007/s00382-018-4333-x

Attard, H.E. & Lang, A.L. (2019) Troposphere-stratosphere coupling following tropospheric blocking and extratropical cyclones. *Monthly Weather Review*, 147, 1781–1804. Available from: https://doi.org/10.1175/MWR-D-18-0335.1

Badan-Dangon, A., Dorman, C.E., Merrifield, M.A. & Winant, C.D. (1991) The lower atmosphere over the Gulf of California. *Journal of Geophysical Research*, 96, 16877–16896. Available from: https://doi.org/10.1029/91jc01433

Baldwin, M.P. & Dunkerton, T.J. (1999) Propagation of the Arctic oscillation from the stratosphere to the troposphere. *Journal of Geophysical Research*, 104, 30937–30946. Available from: https://doi.org/10.1029/1999JD900445

Balmaseda, M.A., Magnusson, L., Vitart, F., Mayer, M., Johnson, S., Emerton, R. et al. (2022) Understanding systematic error growth across timescales. *ECMWF Newsletter*, 172, 10–11. Available from: https://www.ecmwf.int/sites/default/files/elibrary/072022/20443-newsletter-no-172-summer-2022_1.pdf

Baumgart, M., Riemer, M., Wirth, V., Teubler, F., Lang, S.T.K., Baumgart, M. et al. (2018) Potential-vorticity dynamics of forecast errors: a quantitative case study. *Monthly Weather Review*, 146, 1405–1425. Available from: https://doi.org/10.1175/MWR-D-17-0196.1

Besson, P., Fischer, L.J., Schemm, S. & Sprenger, M. (2021) A global analysis of the dry-dynamic forcing during cyclone growth and propagation. *Weather and Climate Dynamics*, 2, 991–1009. Available from: https://doi.org/10.5194/wcd-2-991-2021

Binder, H., Boettcher, M., Joos, H. & Wernli, H. (2016) The role of warm conveyor belts for the intensification of extratropical cyclones in Northern Hemisphere winter. *Journal of the Atmospheric Sciences*, 73, 3997–4020. Available from: https://doi.org/10.1175/JAS-D-15-0302.1

Boettcher, M. & Wernli, H. (2011) Life cycle study of a diabatic Rossby wave as a precursor to rapid cyclogenesis in the North Atlantic-dynamics and forecast performance. *Monthly Weather Review*, 139, 1861–1878. Available from: https://doi.org/10.1175/2011MWR3504.1

Boettcher, M. & Wernli, H. (2013) A 10-yr climatology of diabatic Rossby waves in the Northern Hemisphere. *Monthly Weather Review*, 141, 1139–1154. Available from: https://doi.org/10.1175/MWR-D-12-00012.1

Boos, W.R., Hurley, J.V. & Murthy, V.S. (2015) Adiabatic westward drift of Indian monsoon depressions. *Quarterly Journal of the Royal Meteorological Society*, 141, 1035–1048. Available from: https://doi.org/10.1002/qj.2454

Browning, K.A., Hardman, M.E., Harrold, T.W. & Pardoe, C.W. (1973) The structure of rainbands within a mid-latitude depression. *Quarterly Journal of the Royal Meteorological Society*, 99, 215–231. Available from: https://doi.org/10.1002/qj.49709942002

Büeler, D., Ferranti, L., Magnusson, L., Quinting, J.F. & Grams, C.M. (2021) Year-round sub-seasonal forecast skill for Atlantic-european weather regimes. *Quarterly Journal of the Royal Meteorological Society*, 147, 4283–4309. Available from: https://doi.org/10.1002/qj.4178

Cassou, C. (2008) Intraseasonal interaction between the Madden-Julian Oscillation and the North Atlantic Oscillation. *Nature*, 455, 523–527. Available from: https://doi.org/10.1038/nature07286

Chagnon, J.M. & Gray, S.L. (2015) A diabatically generated potential vorticity structure near the extratropical tropopause in three simulated extratropical cyclones. *Monthly Weather Review*, 143, 2337–2347. Available from: https://doi.org/10.1175/MWR-D-14-00092.1

Chen, G.T.J., Wang, C.C. & Chang, S.W. (2008) A diagnostic case study of mei-yu frontogenesis and development of wavelike frontal disturbances in the subtropical environment. *Monthly Weather Review*, 136, 41–61. Available from: https://doi.org/10.1175/2007MWR1966.1

Coy, L., Eckermann, S. & Hoppel, K. (2009) Planetary wave breaking and tropospheric forcing as seen in the stratospheric sudden warming of 2006. *Journal of the Atmospheric Sciences*, 66, 495–507. Available from: https://doi.org/10.1175/2008JAS2784.1

Coy, L. & Pawson, S. (2015) The major stratospheric sudden warming of January 2013: analyses and forecasts in the GEOS-5 data assimilation system. *Monthly Weather Review*, 143, 491–510. Available from: https://doi.org/10.1175/MWR-D-14-00023.1

Deinhard, M. & Grams, C.M. (2023) Towards a process-oriented understanding of the impact of stochastic perturbations on the model climate. *EGUsphere [Preprint]*. Available from: https://doi.org/10.5194/egusphere-2023-1938

Deoras, A., Hunt, K.M. & Turner, A.G. (2021) Comparison of the prediction of Indian monsoon low pressure systems by subseasonal-to-seasonal prediction models. *Weather and Forecasting*, 36, 859–877. Available from: https://doi.org/10.1175/WAF-D-20-0081.1

Domeisen, D.I.V., Butler, A.H., Charlton-Perez, A.J., Ayarzagüena, B., Baldwin, M.P., Dunn-Sigouin, E. et al. (2020) The

role of the stratosphere in subseasonal to seasonal prediction: 2. Predictability arising from stratosphere-troposphere coupling. *Journal of Geophysical Research: Atmospheres*, 125, e2019JD030923. Available from: https://doi.org/10.1029/2019JD030923

Ferranti, L., Corti, S. & Janousek, M. (2015) Flow-dependent verification of the ECMWF ensemble over the Euro-Atlantic sector. *Quarterly Journal of the Royal Meteorological Society*, 141, 916–924. Available from: https://doi.org/10.1002/qj.2411

Ferranti, L., Magnusson, L., Vitart, F. & Richardson, D.S. (2018) How far in advance can we predict changes in large-scale flow leading to severe cold conditions over Europe? *Quarterly Journal of the Royal Meteorological Society*, 144, 1788–1802. Available from: https://doi.org/10.1002/qj.3341

Fink, A.H., Ermert, V. & Pinto, J.G. (2009) The European storm Kyrill in January 2007: synoptic evolution, meteorological impacts and some considerations with respect to climate change. *Natural Hazards and Earth System Sciences*, 9, 405–423. Available from: https://doi.org/10.5194/nhess-9-405-2009

Froude, L.S. (2010) TIGGE: comparison of the prediction of northern hemisphere extratropical cyclones by different ensemble prediction systems. *Weather and Forecasting*, 25, 819–836. Available from: https://doi.org/10.1175/2010WAF2222326.1

Froude, L.S., Bengtsson, L. & Hodges, K.I. (2007) The prediction of extratropical storm tracks by the ECMWF and NCEP ensemble prediction systems. *Monthly Weather Review*, 135, 2545–2567. Available from: https://doi.org/10.1175/MWR3422.1

Graf, M.A., Wernli, H. & Sprenger, M. (2017) Objective classification of extratropical cyclogenesis. *Quarterly Journal of the Royal Meteorological Society*, 143, 1047–1061. Available from: https://doi.org/10.1002/qj.2989

Grams, C.M., Magnusson, L. & Madonna, E. (2018) An atmospheric dynamics perspective on the amplification and propagation of forecast error in numerical weather prediction models: a case study. *Quarterly Journal of the Royal Meteorological Society*, 144, 2577–2591. Available from: https://doi.org/10.1002/qj.3353

Grams, C.M., Wernli, H., Böttcher, M., Čampa, J., Corsmeier, U., Jones, S.C. et al. (2011) The key role of diabatic processes in modifying the upper-tropospheric wave guide: a North Atlantic case-study. *Quarterly Journal of the Royal Meteorological Society*, 137, 2174–2193. Available from: https://doi.org/10.1002/qj.891

Gray, S.L., Dunning, C.M., Methven, J., Masato, G. & Chagnon, J.M. (2014) Systematic model forecast error in Rossby wave structure. *Geophysical Research Letters*, 41, 2979–2987. Available from: https://doi.org/10.1002/2014GL059282

Hart, R.E. & Evans, J.L. (2001) A climatology of the extratropical transition of Atlantic tropical cyclones. *Journal of Climate*, 14, 546–564. Available from: https://doi.org/10.1175/1520-0442(2001)014<0546:ACOTET>2.0.CO;2

Hawcroft, M.K., Shaffrey, L.C., Hodges, K.I. & Dacre, H.F. (2012) How much Northern Hemisphere precipitation is associated with extratropical cyclones? *Geophysical Research Letters*, 39, L24809. Available from: https://doi.org/10.1029/2012GL053866

Hermanson, L., Ren, H.L., Vellinga, M., Dunstone, N.D., Hyder, P., Ineson, S. et al. (2018) Different types of drifts in two seasonal forecast systems and their dependence on ENSO. *Climate Dynamics*, 51, 1411–1426. Available from: https://doi.org/10.1007/s00382-017-3962-9

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J. et al. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049. Available from: https://doi.org/10.1002/qj.3803

Hewitt, H., Bell, M.J., Chassignet, E.P. & Czaja, A. (2017) Will high-resolution global ocean models benefit coupled predictions on short-range to climate timescales? *Ocean Modelling*, 120, 120–136. Available from: https://doi.org/10.1016/j.ocemod.2017.11.002

Hewson, T. (2002) A comparison of cyclone spectra in forecasts from operational (G1) and new dynamics trial (NT) versions of the unified model-Aug to Dec 2001. *Met Office Forecasting Research Technical Report*, 376, 1–14 https://library.metoffice.gov.uk/Portal/Default/en-GB/DownloadImageFile.ashx?objectId=1258&ownerType=0&ownerId=47720

Hewson, T.D. & Neu, U. (2015) Cyclones, windstorms and the IMILAST project. *Tellus A*, 67, 27128. Available from: https://doi.org/10.3402/tellusa.v67.27128

Hoskins, B. (2013) The potential for skill across the range of the seamless weather-climate prediction problem: a stimulus for our science. *Quarterly Journal of the Royal Meteorological Society*, 139, 573–584. Available from: https://doi.org/10.1002/qj.1991

Hoskins, B.J. & Hodges, K.I. (2019) The annual cycle of Northern Hemisphere storm tracks. Part I: seasons. *Journal of Climate*, 32, 1743–1760. Available from: https://doi.org/10.1175/JCLI-D-17-0870.1

Jung, T., Gulev, S.K., Rudeva, I. & Soloviov, V. (2006) Sensitivity of extratropical cyclone characteristics to horizontal resolution in the ECMWF model. *Quarterly Journal of the Royal Meteorological Society*, 132, 1839–1857. Available from: https://doi.org/10.1256/qj.05.212

Lavaysse, C., Flamant, C., Janicot, S., Parker, D.J., Lafore, J.P., Sultan, B. et al. (2009) Seasonal evolution of the west African heat low: a climatological perspective. *Climate Dynamics*, 33, 313–330. Available from: https://doi.org/10.1007/s00382-009-0553-4

Lavers, D.A., Harrigan, S. & Prudhomme, C. (2021) Precipitation biases in the ECMWF integrated forecasting system. *Journal of Hydrometeorology*, 22, 1187–1198. Available from: https://doi.org/10.1175/JHM-D-20-0308.1

Lin, H., Brunet, G. & Derome, J. (2009) An observed connection between the North Atlantic Oscillation and the Madden-Julian Oscillation. *Journal of Climate*, 22, 364–380. Available from: https://doi.org/10.1175/2008JCLI2515.1

Lorenz, E.N. (1963) Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20, 130–141. Available from: https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2

Madonna, E., Wernli, H., Joos, H. & Martius, O. (2014) Warm conveyor belts in the ERA-interim dataset (1979–2010). Part I: climatology and potential vorticity evolution. *Journal of Climate*, 27, 3–26. Available from: https://doi.org/10.1175/JCLI-D-12-00720.1

Magnusson, L., Alonso-Balmaseda, M., Dahoui, M., Forbes, R., Haiden, T., Lavers, D. et al. (2022) Summary of the UGROW subproject on tropospheric temperature bias during JJA over the Northern Hemisphere. *ECMWF Technical Memo*, 891, 1–15. Available from: https://doi.org/10.21957/vsy2iiph

Martínez-Alvarado, O., Madonna, E., Gray, S.L. & Joos, H. (2016) A route to systematic error in forecasts of Rossby waves. *Quarterly Journal of the Royal Meteorological Society*, 142, 196–210. Available from: https://doi.org/10.1002/qj.2645

Martius, O., Pfahl, S. & Chevalier, C. (2016) A global quantification of compound precipitation and wind extremes. *Geophysical Research Letters*, 43, 7709–7717. Available from: https://doi.org/10.1002/2016GL070017

Mayer, M., Balmaseda, M.A., Johnson, S., Magnusson, L., Roberts, C. & Zuo, H. (2022) Outcomes from UGROW-IO: forecast errors in the eastern Indian Ocean across lead times. *ECMWF Technical Memo*, 898, 1–25. Available from: https://doi.org/10.21957/q4v6n81vl

Merryfield, W.J., Baehr, J., Batté, L., Becker, E.J., Butler, A.H., Coelho, C.A. et al. (2020) Current and emerging developments in subseasonal to decadal prediction. *Bulletin of the American Meteorological Society*, 101, E869–E896. Available from: https://doi.org/10.1175/BAMS-D-19-0037.1

Monhart, S., Spirig, C., Bhend, J., Bogner, K., Schär, C. & Liniger, M.A. (2018) Skill of subseasonal forecasts in Europe: effect of bias correction and downscaling using surface observations. *Journal of Geophysical Research: Atmospheres*, 123, 7999–8016. Available from: https://doi.org/10.1029/2017JD027923

Moore, R.W. & Montgomery, M.T. (2004) Reexamining the dynamics of short-scale, diabatic Rossby waves and their role in midlatitude moist cyclogenesis. *Journal of the Atmospheric Sciences*, 61, 754–768. Available from: https://doi.org/10.1175/1520-0469(2004)061<0754:RTDOSD>2.0.CO;2

Mulholland, D.P., Laloyaux, P., Haines, K. & Balmaseda, M.A. (2015) Origin and impact of initialization shocks in coupled atmosphere-ocean forecasts. *Monthly Weather Review*, 143, 4631–4644. Available from: https://doi.org/10.1175/MWR-D-15-0076.s1

Neu, U., Akperov, M.G., Bellenbaum, N., Benestad, R., Blender, R., Caballero, R. et al. (2013) IMILAST: a community effort to intercompare extratropical cyclone detection and tracking algorithms. *Bulletin of the American Meteorological Society*, 94, 529–547. Available from: https://doi.org/10.1175/BAMS-D-11-00154.1

Owen, L.E., Catto, J.L., Stephenson, D.B. & Dunstone, N.J. (2021) Compound precipitation and wind extremes over Europe and their relationship to extratropical cyclones. *Weather and Climate Extremes*, 33, 100342. Available from: https://doi.org/10.1016/j.wace.2021.100342

Pfahl, S. (2014) Characterising the relationship between weather extremes in Europe and synoptic circulation features. *Natural Hazards and Earth System Science*, 14, 1461–1475. Available from: https://doi.org/10.5194/nhess-14-1461-2014

Pfahl, S. & Wernli, H. (2012) Quantifying the relevance of cyclones for precipitation extremes. *Journal of Climate*, 25, 6770–6780. Available from: https://doi.org/10.1175/JCLI-D-11-00705.1

Pickl, M., Lang, S.T., Leutbecher, M. & Grams, C.M. (2022) The effect of stochastically perturbed parametrisation tendencies (SPPT) on rapidly ascending air streams. *Quarterly Journal of the Royal Meteorological Society*, 148, 1242–1261. Available from: https://doi.org/10.1002/qj.4257

Polvani, L.M. & Waugh, D.W. (2004) Upward wave activity flux as a precursor to extreme stratospheric events and subsequent anomalous surface weather regimes. *Journal of Climate*, 17, 3548–3554. Available from: https://doi.org/10.1175/1520-0442(2004)017<3548:UWAFAA>2.0.CO;2

Quinting, J.F. & Vitart, F. (2019) Representation of synoptic-scale Rossby wave packets and blocking in the S2S prediction project database. *Geophysical Research Letters*, 46, 1070–1078. Available from: https://doi.org/10.1029/2018GL081381

Rivière, G. & Orlanski, I. (2007) Characteristics of the Atlantic storm-track eddy activity and its relation with North Atlantic Oscillation. *Journal of the Atmospheric Sciences*, 64, 241–266. Available from: https://doi.org/10.1175/JAS3850.1

Roberts, C.D., Vitart, F. & Balmaseda, M.A. (2021) Hemispheric impact of North Atlantic SSTs in subseasonal forecasts. *Geophysical Research Letters*, 48, e2020GL091446. Available from: https://doi.org/10.1029/2020GL091446

Rodwell, M.J., Richardson, D.S., Parsons, D.B. & Wernli, H. (2018) Flow-dependent reliability: a path to more skillful ensemble forecasts. *Bulletin of the American Meteorological Society*, 99, 1015–1026. Available from: https://doi.org/10.1175/BAMS-D-17-0027.1

Rodwell, M.J. & Wernli, H. (2023) Uncertainty growth and forecast reliability during extratropical cyclogenesis. *Weather and Climate Dynamics*, 4, 591–615. Available from: https://doi.org/10.5194/wcd-4-591-2023

Rüdisühli, S., Sprenger, M., Leutwyler, D., Schär, C. & Wernli, H. (2020) Attribution of precipitation to cyclones and fronts over Europe in a kilometer-scale regional climate simulation. *Weather and Climate Dynamics*, 1, 675–699. Available from: https://doi.org/10.5194/wcd-1-675-2020

Saffin, L., Methven, J., Gray, S.L. & Williams, K.D. (2017) Processes maintaining tropopause sharpness in numerical weather prediction models. *Journal of Geophysical Research: Atmospheres*, 122, 9611–9627. Available from: https://doi.org/10.1002/2017JD026879

Schäfler, A., Harvey, B., Methven, J., Doyle, J.D., Rahm, S., Reitebuch, O. et al. (2020) Observation of jet stream winds during NAWDEX and characterization of systematic meteorological analysis errors. *Monthly Weather Review*, 148, 2889–2907. Available from: https://doi.org/10.1175/MWR-D-19-0229.1

Shonk, J.K., Guilyardi, E., Toniazzo, T., Woolnough, S.J. & Stockdale, T. (2018) Identifying causes of Western Pacific ITCZ drift in ECMWF system 4 hindcasts. *Climate Dynamics*, 50, 939–954. Available from: https://doi.org/10.1007/s00382-017-3650-9

Sinclair, M.R. (1995) A climatology of cyclogenesis for the Southern Hemisphere. *Monthly Weather Review*, 123, 1601–1619. Available from: https://doi.org/10.1175/1520-0493(1995)123<1601:ACOCFT>2.0.CO;2

Son, S.W., Kim, H., Song, K., Kim, S.W., Martineau, P., Hyun, Y.K. et al. (2020) Extratropical prediction skill of the subseasonal-to-seasonal (S2S) prediction models. *Journal of Geophysical Research: Atmospheres*, 125, e2019JD031273. Available from: https://doi.org/10.1029/2019JD031273

Sprenger, M., Fragkoulidis, G., Binder, H., Croci-Maspoli, M., Graf, P., Grams, C.M. et al. (2017) Global climatologies of Eulerian and Lagrangian flow features based on ERA-Interim. *Bulletin of the American Meteorological Society*, 98, 1739–1748. Available from: https://doi.org/10.1175/BAMS-D-15-00299.1

Vitart, F. (2014) Evolution of ECMWF sub-seasonal forecast skill scores. *Quarterly Journal of the Royal Meteorological Society*, 140, 1889–1899. Available from: https://doi.org/10.1002/qj.2256

Vitart, F., Emerton, R., Rodwell, M., Balmaseda, M., Haiden, T., Johnson, S. et al. (2022) Investigating biases in the representation of the Pacific sub-tropical jet stream and associated teleconnections (a UGROW sub-project). *ECMWF Technical Memo*, 889, 1–20. Available from: https://doi.org/10.21957/jcfqpdr2

Voldoire, A., Exarchou, E., Sanchez-Gomez, E., Demissie, T., Deppenmeier, A.L., Frauen, C. et al. (2019) Role of wind stress in

driving SST biases in the tropical Atlantic. *Climate Dynamics*, 53, 3481–3504. Available from: https://doi.org/10.1007/s00382-019-04717-0

Wandel, J., Quinting, J.F. & Grams, C.M. (2021) Towards a systematic evaluation of warm conveyor belts in numerical weather prediction and climate models. Part II: verification of operational reforecasts. *Journal of the Atmospheric Sciences*, 78, 3965–3982. Available from: https://doi.org/10.1175/JAS-D-20-0385.1

Wernli, H., Dirren, S., Liniger, M.A. & Zillig, M. (2002) Dynamical aspects of the life cycle of the winter storm 'Lothar' (24–26 December 1999). *Quarterly Journal of the Royal Meteorological Society*, 128, 405–429. Available from: https://doi.org/10.1256/003590002321042036

Wernli, H. & Schwierz, C. (2006) Surface cyclones in the ERA-40 dataset (1958–2001). Part I: novel identification method and global climatology. *Journal of the Atmospheric Sciences*, 63, 2486–2507. Available from: https://doi.org/10.1175/JAS3766.1

Whittaker, L.M. & Horn, L.H. (1984) Northern Hemisphere extratropical cyclone activity for four mid-season months. *Journal of Climatology*, 4, 297–310. Available from: https://doi.org/10.1002/joc.3370040307

Winters, A.C. (2021) Subseasonal prediction of the state and evolution of the North Pacific jet stream. *Journal of Geophysical Research: Atmospheres*, 126, e2021JD035094. Available from: https://doi.org/10.1029/2021JD035094

Wirth, V. & Szabo, T. (2007) Sharpness of the extratropical tropopause in baroclinic life cycle experiments. *Geophysical Research Letters*, 34, L02809. Available from: https://doi.org/10.1029/2006GL028369

Zhang, F., Sun, Y.Q., Magnusson, L., Buizza, R., Lin, S.-J., Chen, J.-H. et al. (2019) What is the predictability limit of midlatitude weather? *Journal of the Atmospheric Sciences*, 76, 1077–1091. Available from: https://doi.org/10.1175/jas-d-18-0269.1

Zheng, C., Chang, E.K.-M., Kim, H. & Zhang, M. (2019) Subseasonal to seasonal prediction of wintertime Northern Hemisphere extratropical cyclone activity by S2S and NMME models. *Journal of Geophysical Research Atmospheres*, 124, 12057–12077. Available from: https://doi.org/10.1029/2019JD031252

Zuo, H., Balmaseda, M.A. & Mogensen, K. (2017) The new eddy-permitting ORAP5 ocean reanalysis: description, evaluation and uncertainties in climate signals. *Climate Dynamics*, 49, 791–811. Available from: https://doi.org/10.1007/s00382-015-2675-1

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Büeler, D., Sprenger, M. & Wernli, H. (2024) Northern Hemisphere extratropical cyclone biases in ECMWF subseasonal forecasts. *Quarterly Journal of the Royal Meteorological Society*, 150(759), 1096–1123. Available from: https://doi.org/10.1002/qj.4638