Toward a Systematic Evaluation of Warm Conveyor Belts in Numerical Weather Prediction and Climate Models. Part II: Verification of Operational Reforecasts

JAN WANDEL, a JULIAN F. QUINTING, a AND CHRISTIAN M. GRAMS a

a Institute of Meteorology and Climate Research (IMK-TRO), Karlsruhe Institute of Technology, Karlsruhe, Germany

(Manuscript received 21 December 2020, in final form 11 August 2021)

ABSTRACT: Warm conveyor belts (WCBs) associated with extratropical cyclones transport air from the lower troposphere into the tropopause region and contribute to upper-level ridge building and the formation of blocking anticyclones. Recent studies indicate that this constitutes an important source and magnifier of forecast uncertainty and errors in numerical weather prediction (NWP) models. However, a systematic evaluation of the representation of WCBs in NWP models has yet to be determined. Here, we employ the logistic regression models developed in Part I to identify the inflow, ascent, and outflow stages of WCBs in the European Centre for Medium-Range Weather Forecasts (ECMWF) subseasonal reforecasts for Northern Hemisphere winter in the period January 1997 to December 2017. We verify the representation of these WCB stages in terms of systematic occurrence frequency biases, forecast reliability, and forecast skill. Systematic WCB frequency biases emerge already at early lead times of around 3 days with an underestimation for the WCB outflow over the North Atlantic and eastern North Pacific of around 40% relative to climatology. Biases in the predictor variables of the logistic regression models can partially explain these biases in WCB inflow, ascent, or outflow. Despite an overconfidence in predicting high WCB probabilities, skillful WCB forecasts are on average possible up to a lead time of 8–10 days with more skill over the North Pacific compared to the North Atlantic region. Our results corroborate that the current limited forecast skill for the large-scale extratropical circulation on subseasonal time scales beyond 10 days might be tied to the representation of WCBs and associated upscale error growth.

KEYWORDS: Extratropical cyclones; Trajectories; Forecast verification/skill

1. Introduction

Warm conveyor belts (WCBs; e.g., Browning et al. 1973; Harrold 1973; Carlson 1980) are cloudy regions of strong, diabatically enhanced ascent ahead of the cold front of extratropical cyclones. WCBs contribute a major fraction of precipitation in the midlatitudes (Pfahl et al. 2014), and can cause heavy precipitation (e.g., Pfahl et al. 2014; Catto et al. 2015; Flaounas et al. 2016; Agel et al. 2019) and flooding (Grams et al. 2014). From a dynamical perspective, WCBs are an important component of the midlatitude large-scale flow due to their impact on the life cycle of extratropical cyclones (e.g., Wernli and Davies 2014; Binder et al. 2016) and the upper-tropospheric Rossby waveguide (e.g., Pomroy and Thorpe 2000; Grams and Archambault 2016). Thus, an accurate representation of WCBs is desirable in numerical weather prediction (NWP) models.

The WCB airstream is typically identified as coherent bundle of Lagrangian air parcel trajectories which ascend 600 hPa in 48 h (Madonna et al. 2014). The evolution of a WCB can be described in three distinct stages: WCB inflow, WCB ascent, and WCB outflow. During the inflow stage, WCB air is located in the planetary boundary layer of an extratropical cyclone’s warm sector. Though at this stage the air mass can be considered as passive tracer, an inaccurate representation of its moisture content may affect important characteristics of the WCB such as its diabatic heating rate and thus ascent and outflow height (Schäfler and Harnisch 2015). During WCB ascent air rises ahead of the cold front and across the warm front in a region of quasigeostrophic forcing for ascent (Binder et al. 2016) which results in precipitation below the WCB airstream. The strong latent heat release due to condensation further contributes to upward motion through cross-isentropic ascent. On average, this latent heat release amounts to 20 K within 48 h (Madonna et al. 2014).

In the following, we use the potential vorticity (PV)–potential temperature (θ) framework (Hoskins et al. 1985) in order to describe the two major effects of the latent heat release on the further evolution of the WCB and its impact on midlatitude dynamics. First, latent heat release contributes to the formation of a cyclonic PV anomaly below the level of maximum heating. This diabatically generated PV anomaly contributes to the formation of the extratropical cyclone’s PV tower (Rossa et al. 2000) and thus to its overall intensity (Davis et al. 1993; Binder et al. 2016). Second, due to the latent heat release, there is a net cross-isentropic transport of lower-tropospheric low-PV air into the upper troposphere.
Divergent outflow together with the balanced upper-level flow above and downstream of the heating maximum further advects this low-PV air poleward (Teubler and Riemer 2016). When impinging on the marked PV gradient of the midlatitude Rossby waveguide, the diabatic WCB outflow enhances or even triggers upper-tropospheric ridge building resulting in a marked anticyclonic PV anomaly in a region of climatologically high PV (Madonna et al. 2014). Eventually the anticyclonic PV anomaly becomes stationary and persistent, and might contribute to a blocking anticyclone (Pfahl et al. 2015; Grams and Archambault 2016; Steinfeld and Pfahl 2019).

Due to the major dynamical impacts of WCBs described above, errors that occur in NWP models in any of the three WCB stages may affect the representation of the large-scale midlatitude flow. Indeed, recent studies show growing evidence that errors in the representation of WCBs may lead to errors in the downstream Rossby wave pattern (e.g., Lamberson et al. 2016; Martínez-Alvarado et al. 2016; Baumgart et al. 2018; Grams et al. 2018; Rodwell et al. 2018; Berman and Torn 2019; Maddison et al. 2019). This suggests that WCBs act as an active communicator and amplifier of forecast errors and uncertainties from small to large scales (Grams et al. 2018).

On the medium range, the representation of WCBs in NWP models was first evaluated by Madonna et al. (2015) for three winter periods [December–February (DJF)] in the operational high resolution deterministic forecast of the ECMWF Integrated Forecasting System (IFS) model. They used a novel feature-based verification technique that was originally developed to verify precipitation forecasts (Wernli et al. 2008) and found evidence that periods with low forecast skill in terms of geopotential height can be connected to a misrepresentation of the WCB outflow in the forecast model. Furthermore, Madonna et al. (2015) found an improvement in WCB forecast skill from DJF 2002/03 to DJF 2010/11 which they attributed to the overall improvement in model forecast skill over this decade (Magnusson and Källén 2013). Still, a systematic investigation of the representation and forecast skill of WCBs has not been addressed yet. This is due to the high computational costs in calculating 3D trajectories and availability of model data at high horizontal $O(-1^\circ)$, vertical $O(-10\, hPa)$, and temporal resolution $O(-3–6\, h)$, which are usually not archived. Novel approaches are therefore needed in order to systematically investigate the representation of WCBs in NWP models.

This study, for the first time, provides a systematic investigation of the representation of WCBs in the state-of-the-art ECMWF IFS ensemble forecasting system. To this end, we apply the logistic regression models introduced in Quinting and Grams (2021a, Part 1 hereafter) to the IFS extended-range reforecasts available through the subseasonal-to-seasonal (S2S) Prediction Project database (Vitart et al. 2017) in order to identify footprints of WCB inflow, ascent, and outflow. Our study investigates the Northern Hemisphere winter season (DJF).

The paper is organized as follows. The data used, the statistical tools to identify the WCB stages, and the verification methods are introduced in section 2. Section 3 further motivates this work by showing the link between WCB outflow activity and forecast skill in terms of geopotential height. We then evaluate the WCB frequency bias and forecast reliability at different lead times as well as the overall forecast skill horizon for the WCB ensemble forecasts. Section 4 provides a dynamical interpretation of the WCB biases and forecast skill by linking them to biases in the predictor variables and investigating potential sensitivities of the statistical models to bias correction. Furthermore, we investigate the link between WCB forecast and the large-scale midlatitude flow on subseasonal time scales beyond 15 days. In section 5 we discuss our main results and end with concluding remarks.

2. Data and method

a. Forecast model and observations

A systematic investigation of the representation of WCBs in state-of-the-art NWP models requires a consistent dataset that covers several years. Therefore, we employ ECMWF’s subseasonal reforecasts from the S2S Prediction Project database (Vitart et al. 2017) for the period from January 1997 to December 2017. The data are retrieved on a regular 1.5° × 1.5° latitude–longitude grid and remapped to 1° × 1° grid spacing since the regression models were developed for the latter grid spacing. The reforecasts are initialized twice a week from the ECMWF interim reanalysis (ERA-Interim) with 11 ensemble members (1 unperturbed and 10 perturbed forecasts) for a lead time of 46 days. As done in previous studies (e.g., Schiraldi and Roundy 2017; DeFlorio et al. 2018; Büberl et al. 2020), we use different model versions (CY43R1, CY43R3, CY43R5) to increase the overall sample size. It is important to note that these model versions come with the same resolution for ocean and atmosphere and have the same number of model levels. We only consider forecasts initialized in DJF which yields a total of 920 different forecast initialization times. We use temperature $T$, geopotential height $\phi$, specific humidity $q$, horizontal wind components $u$, $v$ (instantaneous at 0000 UTC) on pressure levels (1000, 925, 850, 700, 500, 300, and 200 hPa). These variables are employed to derive quantities identified in Part 1 as logistic regression models’ predictors to identify footprints of WCB inflow, WCB ascent, and WCB outflow without the necessity of running trajectories (section 2b). It is important to note that the operational forecasting system of the ECMWF likely has higher WCB skill than the S2S reforecast dataset. This is due to the higher horizontal and vertical grid spacing, more ensemble members, and a newer data assimilation system (Vitart 2014). However, due to the lack of a comparably long data record with a consistent model version we did not test this hypothesis.

Consistent with the initial conditions of the reforecasts, we use the ERA-Interim data (Dee et al. 2011) between 1997 and 2017 as the observational truth to estimate the overall skill of WCB forecasts. For consistency with the reforecast data, we retrieve ERA-Interim on a 1.5° × 1.5° grid, remap it to 1° × 1° and only use data at 0000 UTC.

b. Logistic regression models

In this section, we briefly introduce the multiple logistic regression models which were set up in Part 1 to identify WCBs in
numerical weather prediction and climate models. For a detailed description of the model development and the selection of the meteorological predictor variables, we refer to Part I. The models are employed for the three stages of the WCB (inflow, ascent, outflow) each using four meteorological predictor variables $x_1$, $x_2$, $x_3$, and $x_4$, respectively [see Eq. (2) in Part I]. The predictor variables for the inflow model are thickness advection at 700 hPa, meridional moisture flux at 850 hPa, moisture flux convergence at 1000 hPa, and moist potential vorticity at 500 hPa (see Table 1 in Part I). The ascent model is based on relative vorticity at 850 hPa and relative humidity at 300 hPa as the two most important predictors. Thickness advection at 300 hPa and the meridional moisture flux at 500 hPa complete the ascent model. Finally, the outflow model uses relative humidity at 300 hPa, the irrotational wind speed at 300 hPa, static stability at 500 hPa, and relative vorticity at 300 hPa.

The logistic regression models yield the conditional probability $P$ of WCB inflow, ascent, and outflow at each grid point where the corresponding climatological frequency reaches at least 1% [Eq. (1) in Part I]. The conditional probability $P$ can be transformed into a deterministic dichotomous prediction by applying a local, gridpoint-based decision threshold which we store as a binary WCB inflow, ascent, and outflow mask. If the conditional probability exceeds the threshold, the respective grid point is considered to feature a WCB inflow, ascent, or outflow and the mask is set to 1, below the threshold the mask is set to 0. For the ensemble forecasts, the dichotomous WCB inflow, ascent, and outflow masks are calculated for each ensemble member. Averaging over the 0 or 1 mask for all members yields an ensemble probability $P$ of WCB inflow, ascent, and outflow occurrence. We use the local threshold defined in Part I that optimizes the replication of the climatological WCB occurrence frequency based on the Lagrangian training data in ERA-Interim reanalyses. It is important to note that the logistic regression models have different skill in reproducing the Lagrangian trajectory definition depending on the region and WCB phase (Figs. 9b,d,f in Part I). In our study, we focus on regions with sufficient skill of at least 0.2 of the Matthews correlation coefficient (MCC; Matthews 1975) giving us confidence that similar results would also be obtained using the Lagrangian definition.

In the following analysis, we apply the logistic regression models to reanalysis and forecast data and discuss the WCB occurrence as detected by the models. Thus, when talking about WCB in the remainder of the manuscript we mean WCB signatures identified by the logistic regression models and not by Lagrangian trajectories. We interpret differences between reanalysis and reforecasts as limitations of the forecast system, but want to stress that differences in the native resolution between the two datasets might have an impact on the respective conditional probabilities, too.

To test the potential effect of different native resolutions, we applied a bias correction to both, the predictor variables (see section 4b) and to the conditional probabilities (not shown). Despite some reduction of the systematic biases, we find no change of the qualitative interpretation of regions with positive or negative biases, and most importantly no effect on skill. This gives us confidence that differences between WCBs in reanalysis and reforecasts mainly indicate deficiencies of the forecast system.

c. Mean error (bias)

At each grid point and for all lead times, we compute the mean error $ME$ (bias) of the ensemble forecast probabilities $P$ compared to the observational truth from ERA-Interim for the period 1997–2017:

$$ ME = \frac{1}{N} \sum_{n=1}^{N} (P_n - o_n),$$  \hspace{1cm} (1)

where $0 \leq P_n \leq 1$ is the ensemble probability of WCB inflow, ascent or outflow at a specific grid point and forecast lead time and $o_n$ the dichotomous observational value from ERA-Interim. The denominator $N$ stands for the number of forecasts that are used to calculate the mean error.

d. FBSS

A common score to verify ensemble probabilities is the Brier score (BS) which is the mean squared error for a probabilistic forecast (Wilks 2011). Due to the relatively low number of ensemble members in the S2S dataset, we use the fair Brier score (FBS; Ferro 2014). The FBS provides an unbiased estimate which one would get for an infinite ensemble of exchangeable members (Ferro et al. 2008):

$$FBS = \frac{1}{N} \sum_{n=1}^{N} \left( \frac{i_n}{m} - o_n \right)^2 - \frac{i_n^2}{m^2(m-1)},$$ \hspace{1cm} (2)

where $i$ is the number of ensemble members that identify a WCB inflow, ascent, or outflow for a given lead time and $m$ is the total number of ensemble members (with $P_n = i_n/m$). The first term compares the ensemble probability $P_n$ to the observations $o_n$ and the second term generally reduces the Brier score for small ensembles. Additionally, we calculate a reference Brier score using the DJF seasonal climatology $\overline{o}$ from the ERA-Interim period 1979–2017 as a predictor variable

$$ BS_{CLIM} = \frac{1}{N} \sum_{n=1}^{N} (\overline{o} - o_n)^2.$$  \hspace{1cm} (3)

The fair Brier skill score (FBSS) then compares the FBS of the ensemble forecasts to the $BS_{CLIM}$ of the climatological reference forecast:

$$ FBSS = 1 - \frac{FBS}{BS_{CLIM}}.$$ \hspace{1cm} (3)

The forecast model performs better than the reference forecast for a fair Brier skill score larger than 0 with higher values indicating a better model forecast.

e. Forecast skill on subseasonal time scales

In contrast to short- and medium-range weather prediction, for which the forecast of a specific event on a given day $M$ is of interest, on subseasonal time scales, a skillful forecast of the mean value of a variable over a certain period of multiple days is of interest (White et al. 2017). We evaluate weekly mean WCB frequencies by computing the fair Brier score ($FBS_M$) for a 7-day WCB frequency forecast with
where the 7-day mean of ensemble probability $P_n (=i_n/m)$ is compared to the 7-day mean of the observations $o_n$. As in Eq. (2), the second term accounts for the small ensemble size. It is simply an average of the Ferro (2014) correction term over all 7 days. The DJF seasonal climatology $\bar{o}_n$ is used again as a predictor for the reference forecast. We then calculate the mean values of biases and skill scores for each of the 1000 subsets at each grid point. This yields a distribution of the mean biases and skill scores. We follow a recent approach from Papritz (2020) and define the biases/scores to be robust on the first level, if their absolute values are greater than the interquartile range of the sample distribution. We define robustness on the second level, if the absolute value of the bias/score is greater than the difference between the 90th and 10th percentiles. For the area-averaged skill scores, we use the same sampling approach but now generate distributions of area-averaged fair Brier skill scores for each lead time. We then define the difference between the 90th and 10th percentile as the variability of the area-averaged fair Brier skill score. Forecasts in two regions have a significantly different skill if the variability bars for the two regions do not overlap (i.e., the 10th percentile for the first region exhibits a higher skill value than the 90th percentiles for the other region).

3. Representation of WCBs in ECMWF’s reforecasts

In this section we assess the representation of WCBs in the reforecasts at medium-range lead times. Before investigating in detail WCB frequency biases, reliability, and forecast skill, we motivate our evaluation by investigating the link between forecast skill of geopotential height and periods with high/lower WCB outflow activity over the North Atlantic.

a. Link between WCB outflow activity and North Atlantic geopotential height forecast skill

The anomaly correlation coefficient (ACC) of 0.6 is used by many forecast centers as a lower limit above which field forecasts are still considered to be useful in an operational context (Hollingsworth et al. 1980). Therefore, we first calculate the area-averaged ACC for all 920 forecasts of 200-hPa geopotential height over the North Atlantic ($20^\circ$–$90^\circ$N, $80^\circ$W–$20^\circ$E) and define the skill horizon as the lead time at which the area-averaged ACC drops below the threshold of 0.6.

Furthermore, we contrast the ACC of the 200-hPa geopotential height forecasts at lead times during which the WCB outflow activity was anomalously high or anomalously low. Here, WCB outflow activity is considered to be high (low) when the area-weighted number of grid points with WCB outflow over the North Atlantic region at the respective lead time lies in the upper (lower) 25th percentile of the climatological distribution of all winters in the period January 1997 to December 2017, respectively.

For all 920 forecasts, the skill horizon is reached around forecast day 11 (black line in Fig. 1). For the forecasts during high WCB outflow activity over the North Atlantic, the 200-hPa geopotential height forecast skill is lower, giving valuable skill only up to forecast day 10.5 (red line in Fig. 1). On the contrary, in periods with low WCB outflow activity, useful skill for geopotential height exists beyond day 11 (blue line in Fig. 1). These results are qualitatively insensitive to the exact definition of the North Atlantic domain, choosing the 300- or 500-hPa level instead of 200 hPa, or using different percentiles in order to differentiate between periods of high and low WCB outflow activity (not shown). The difference in useful forecast skill of slightly more than 1 day indicates that there might be a systematic link between high WCB activity and low forecast skill, and that a better representation of WCBs in the forecast model could lead to an increase in forecast skill for the large-scale extratropical circulation.

b. WCB forecast bias

We now focus on the overall representation of WCBs in the ECMWF reforecasts by investigating WCB frequency biases over the Northern Hemisphere at lead times of 3, 7, and 15 days. For all WCB stages, frequency biases emerge already
at forecast day 3 and remain at a similar level or increase toward later lead times (Fig. 2). The absolute frequency biases for all three WCB stages reach values of up to 5% corresponding to a bias of 30%–100% relative to the climatological frequency (5%–15% in regions of strongest biases). In almost all regions that experience a bias of more than 1%, the findings are robust on the first level. Therefore, we only show robustness on the second level (indicated by the point stippling). In the following, we now focus on the storm track regions over the North Atlantic and North Pacific where the climatological WCB frequency is highest.

In the North Atlantic region, we find an overestimation of WCB inflow on day 3 in most regions (Fig. 2a). In particular over the southeastern United States, the Gulf of Mexico (30°–40°N, 90°–80°W), and farther east along 30°–40°N, the forecast model overestimates the inflow frequency by 1%–3%. Negative biases of around 1% occur north and northeast of the climatological frequency maximum (35°–45°N, 70°–60°W). For WCB ascent, a structurally similar dipole of negative and positive biases is found over the North Atlantic (Fig. 2b). The negative bias extends over a larger area along the North American east coast toward the southern tip of Greenland. WCB outflow is largely underestimated by 1%–3% over the North Atlantic between eastern Canada, Greenland, and Iceland already on day 3 (Fig. 2c). Positive biases of 1%–3% occur only to the south of the climatological maximum over the North Atlantic (30°–60°N, 60°W–0°).

At lead times of 7 and 15 days, the WCB inflow and ascent frequency biases over the North Atlantic remain structurally similar with negative (positive) biases to the north (south) of the regions with climatologically highest WCB activity (Figs. 2d,e,g,h). The magnitude of the biases increases only slightly in most regions except for the southern United States and the Gulf of Mexico where positive biases exceed 1.5% at 15 days lead time (Fig. 2g). For the WCB outflow, the negative bias over the North Atlantic increases to around 3% at 7 and 15 days lead time (Figs. 2f,i).

For the North Pacific region, an area with positive WCB inflow frequency biases of 2%–5% spans from the South China Sea to Japan and farther east toward the climatological frequency maximum (20°–30°N, 120°–170°E) on day 3 of the forecasts (Fig. 2a). Pronounced negative biases of up to 3% are prevalent in and north of the region with the climatological frequency maximum (30°–40°N, 140°–170°E). Recalling that the climatological WCB inflow frequency ranges from 5% to 15%, these absolute biases correspond to a relative frequency bias ranging from 20% to 50%. For WCB ascent, a negative bias of 1%–3% occurs west of the climatological frequency maximum between 120°E and 180° (Fig. 2b). As for the WCB inflow, the absolute WCB ascent biases correspond to a relative frequency bias of 20%–50%. Of all WCB stages, the WCB outflow frequency biases over the Pacific are most pronounced (Fig. 2c). Negative biases of 1%–5% over a large region from the date line along the Bering Sea and over large portions of western North America (30°–60°N, 180°–120°W) correspond to relative biases of 20%–70%.

At later lead times, the magnitude of the WCB inflow biases over the North Pacific increase. With an overestimation of 3%–5%, the

![Fig. 2. Frequency bias of WCB (a), (d), (g) inflow, (b), (e), (h) ascent, (c), (f), (i) outflow for DJF 1997–2017 (shading) at different forecast lead times: (a)–(c) day 3, (d)–(f) day 7, and (g)–(i) day 15. Robustness on the second level is indicated by the stippling. The black contours indicate a climatological frequency of 1%, 5%, 10%, and 15% in the respective layer. For clarity of the figure, all data have been smoothed with a Gaussian filter.](image-url)
inflow bias is most pronounced over the South China Sea (Figs. 2d,g). Recalling that the climatological WCB inflow frequencies range from 5% to 15%, this corresponds to a relative overestimation of up to 100%. For WCB ascent, the negative biases over the western North Pacific weaken with forecast lead time and significant positive biases occur on forecast day 15 (Figs. 2e,h). Furthermore, a negative WCB ascent bias of around 1% develops over the eastern North Pacific for day 15 (30°–60°N, 160°–120°W). In line with the changes of the WCB ascent bias, there is a shift to a positive outflow bias over the western North Pacific on day 15 and an overall increase of the negative bias over the eastern North Pacific (absolute values around 3%–5%; Figs. 2f,i). This negative bias corresponds to a relative underestimation of 50%–100%.

In conclusion, the most pronounced biases of WCB activity in ECMWF’s reforecasts at medium-range lead times are an overestimation of WCB inflow south and southeast of the climatological frequency maxima in the North Pacific and North Atlantic, an underestimation of WCB ascent in the North Atlantic and at later lead times over the eastern North Pacific, and a strong underestimation of WCB outflow in the North Atlantic and eastern North Pacific region.

c. WCB forecast reliability

The agreement between ensemble probabilities \( P \) and the observed frequencies of WCB occurrence is assessed via reliability diagrams at each grid point in the Northern Hemisphere. Since there are 11 ensemble members in the reforecast dataset, we create 12 regular bins \((x\text{ axis})\) for the probabilities predicted by the ensemble. Bin 0 contains those forecasts when zero members predicted the occurrence of a WCB at the respective grid point and at a specific lead time, bin 1 contains those forecasts when one member predicted the occurrence of a WCB, and so on. Accordingly, each bin corresponds to a predicted probability \((\text{e.g.,} \ 0\% \text{ for bin} \ 0 \text{and} \ 64\% \text{ for bin} \ 7)\). For each grid point and each forecast lead time, the forecasts in each bin are then compared against the actual observed frequency \((y\text{ axis})\). In a final step, the reliability curves and the number of forecasts falling into each bin are averaged (area weighted) over all Northern Hemisphere grid points and evaluated for forecast days 3, 7, and 15 (Fig. 3). Due to the relatively small number of forecasts in bins 8–11 for these lead times \((1\%–5\% \text{ of all forecasts})\), we only show results for bins 0–7.

Around 80% of the total number of forecasts fall into bin 0; i.e., none of the ensemble members predicts WCB inflow, ascent, or outflow (Fig. 3). Still we find that in 1%–3% of these forecasts, a WCB is observed in the reanalysis. This underconfidence of the model in a large number of forecasts and at many grid points leads to the overall negative biases found in section 3b. In contrast, if WCB inflow, ascent, or outflow is predicted by at least one of the ensemble members \((\text{bins} \ 1–7)\), the forecast model is mostly overconfident and in reality fewer WCBs occur. The reliability decreases from forecast day 3 to forecast days 7 and 15. The relatively flat curve for forecast day 15 indicates the difficulty of the forecast model in predicting the WCB reliably at this lead time.

d. WCB forecast skill

We now investigate the overall forecast skill as the area average of gridpoint-based FBSS over the Northern Hemisphere, the North Atlantic, and the North Pacific. We subjectively define the forecast skill horizon as the forecast lead time at which the FBSS falls below 0.08 since it only decreases very slowly afterward.

For all three regions, the skill horizon for WCB inflow is reached after 9–10-day forecast lead time with generally higher skill over the North Pacific than over the North Atlantic (Fig. 4b). For lead times up to day 5, the differences in FBSS between the three regions are not statistically significant. For lead times beyond day 6, however, the forecast skill over the North Pacific is significantly higher at the 80% confidence interval than over the North Atlantic as indicated by nonoverlapping error bars. The forecast skill for WCB ascent exhibits a similar evolution with forecast lead time (Fig. 4c). The forecast skill horizon is reached after 8 days for the North Atlantic and 9 days for the North Pacific. For lead times between 4 and...
10 days, the forecast skill over the North Pacific is significantly higher than over the North Atlantic. In particular at early lead times, the forecast skill for WCB outflow is lowest on average (Fig. 4d). However, the overall forecast skill horizon is similar to WCB ascent with 8 days for the North Atlantic and 9 days for the North Pacific. Similar to WCB inflow and ascent, the skill is significantly higher over the North Pacific than over the North Atlantic between forecast days 4 and 10.

Maps of the FBSS on forecast days 3 and 7 reveal that the skill varies not only between the North Atlantic and North Pacific region but also between different subareas within these large regions (Fig. 5). As for the biases, we focus in the following on the forecast skill over the North Atlantic and North Pacific.

On forecast day 3, the FBSS for WCB inflow locally exceeds the North Atlantic mean FBSS of 0.48 (cf. Fig. 4b) in the region with the highest climatological frequency, toward the southeast of this region (25°–35°N, 70°–50°W), and in the Gulf of Mexico (around 0.6, respectively) (Fig. 5a). In contrast, the skill level is lower over the central and northeastern North Atlantic toward the southwest of Iceland. Here, a relatively low skill between 0.1 and 0.5 can be found. For WCB ascent, the areas with highest skill are located between 30° and 60°N and between 80° and 40°W (Fig. 5c). In line with the skill for WCB inflow, we find relatively low skill over the eastern North Atlantic and toward western Europe. Also for WCB outflow highest skill levels are found over the western North Atlantic with values between 0.4 and 0.5 on forecast day 3 (Fig. 5d). Skill tends to be lower over the eastern North Atlantic, the Norwegian Sea, Scandinavia, and the eastern part of Europe.

For forecast day 7, the regions over the North Atlantic with high skill for WCB inflow and ascent are similar to forecast day 3 with relatively high skill over the southwestern part of the climatological frequency maximum (Figs. 5b,d). As for forecast day 3, we also find low skill to the northeast of the climatological maximum. For the WCB outflow, we find relatively high skill in a new region to the southeast of the climatological maximum (35°–40°N, 30°–10°W) (Fig. 5f).

Over the North Pacific, the region east of Japan to the central North Pacific (30°–40°N, 140°E–180°) has the highest skill for the WCB inflow on day 3 (around 0.6, Fig. 5a). Generally, skill is lower over the eastern North Pacific, especially in the region to the north and east of the climatological frequency maximum (40°–50°N, 170°–130°W). For WCB ascent, forecasts exhibit the highest skill on day 3 between 40° and 60°N and between 140°E and 180° (Fig. 4c). Similar to the skill for WCB inflow, lower skill (compared to the area average) is found over the southern part of the western North Pacific and in most regions of the eastern North Pacific. The highest skill for WCB outflow is found east of Japan and over the central North Pacific (40°–55°N, 140°–170°E, Fig. 5e). The skill generally decreases toward the areas with lower climatological WCB outflow frequency.

On forecast day 7, the skill of WCB inflow reaches similar levels over the western and the eastern North Pacific (Fig. 5b); i.e., we find a new area of relatively high skill over the southeastern part of the eastern North Pacific (20°–30°N, 170°–140°W). The large skill difference for WCB ascent between the southern and northern part of the WCB ascent region for forecast day 3, is no longer evident for day 7 (Fig. 5d). The areas with relatively high (low) skill for WCB outflow are similar to forecast day 3 (Fig. 5f).

In conclusion, the WCB forecast skill horizon for DJF in the period January 1997 to December 2017 is around 8–10 days.
with more skill over the North Pacific compared to the North Atlantic region. We find relatively high skill over the western North Atlantic and over the western North Pacific and low skill over the northeastern North Atlantic and eastern North Pacific. In the following section, we investigate potential drivers for differences in WCB biases and forecast skill within the North Atlantic and North Pacific region.

4. Dynamical interpretation of WCB bias and forecast skill

In this section, we discuss potential dynamical reasons for the WCB frequency biases and skill. First we focus on the medium range (0–15 days) and investigate connections between the skill for WCB inflow, ascent, and outflow. Further, we discuss the role of biases in the predictor variables for WCB frequency biases found in section 3b. Second we focus on the subseasonal forecast range and investigate biases in weekly mean WCB frequency and skill, and its potential modulation by the Madden–Julian oscillation (MJO; Madden and Julian 1972).

a. Link between forecast skill of inflow, ascent, and outflow over North Atlantic

The logistic regression models for the different WCB stages are technically independent from each other. However, in reality and in the original Lagrangian definition they represent the time-lagged ascending motion of air parcels within a WCB airstream from the planetary boundary layer to the tropopause within two days. This raises the question if there is a time-lagged relationship in skill between the inflow, ascent, and outflow stage of the WCB. We investigate this based on the skill of WCB inflow on forecast day 4, of WCB ascent on forecast day 5, and of WCB outflow on forecast day 6 relative to the mean skill in the North Atlantic region. Similar results are also found when using other lead times (not shown). It is important to note that we do not aim to quantify the connection between the forecast skill of the three stages in this study, but to show a general overview of possible regions where high/low forecast skill can be consistently found for all three WCB stages.

Climatologically, the inflow of WCBs associated with mid-latitude cyclones over the western North Atlantic is located over the Gulf of Mexico and the subtropical western North Atlantic. From there, the WCBs follow a northeastward pathway, ascend over the Gulfstream, and reach the upper troposphere over the central North Atlantic. When reaching the mid- to upper troposphere, the WCBs may turn cyclonically toward Greenland or anticyclonically toward the subtropics.

FIG. 5. FBSS of WCB forecast for (a),(b) inflow, (c),(d) ascent, (e),(f) outflow for DJF 1997–2017 (shading) at different forecast lead times: (a),(c),(e) day 3 and (b),(d),(f) day 7. The black contours indicate a climatological WCB frequency of 1%, 5%, 10%, and 15% in the respective layer. For clarity of the figure, all data have been smoothed with a Gaussian filter.
Atlantic (25°–45°N, 75°–50°W in Fig. 6a). In contrast, negative FBSS anomalies occur over the eastern half of the North Atlantic. Though the FBSS anomalies are spatially less coherent for WCB ascent (Fig. 6b), a similar dipole structure is found with enhanced regional skill over the western North Atlantic and reduced skill over the eastern North Atlantic. The enhanced skill for the WCB inflow and ascent over the western North Atlantic may suggest that WCBs associated with cyclones in the main storm track region are better represented than WCBs associated with cyclones over the eastern North Atlantic. For WCB outflow, enhanced forecast skill is particularly found over the central North Atlantic between 30° and 60°N (Fig. 6c). At higher latitudes, as well as over central Europe, the skill is anomalously low. The enhanced skill south of 60°N may indicate that in particular the anticyclonic and southward turning branches of the WCBs have higher skill than those that turn cyclonically toward higher latitudes. A detailed investigation of this hypothetical explanation is part of our ongoing research.

b. Bias of predictor variables and impact of bias correction on WCB forecast

Next we investigate to what degree systematic biases in the predictor variables of the logistic regression models are related to systematic biases in WCB inflow, ascent, or outflow frequency (Fig. 2). To identify variables that are potentially related to the frequency biases, we first evaluate the correlation between WCB inflow, ascent, and outflow biases and the biases of the predictor variables at different lead times using the Pearson correlation coefficient (Table 1).

WCB inflow biases correlate relatively strongly with biases in meridional moisture flux at 850 hPa for both the North Atlantic and North Pacific region and all lead times (between 0.47 and 0.72), as well as moisture flux convergence at 1000 hPa (between 0.38 and 0.55) suggesting that these variables are important drivers of WCB inflow biases. The WCB ascent bias can be mostly related to biases in meridional moisture flux at 500 hPa for both the North Atlantic and North Pacific and all lead times (correlation between 0.37 and 0.61). Biases in relative vorticity at 850 hPa and relative humidity at 700 hPa show a smaller correlation to the WCB ascent biases (0.13–0.37 and 0.21–0.36, respectively). For the WCB outflow biases, the correlation to the biases in the predictor variables varies between variables, forecast lead times and ocean basins. For the North Atlantic, the correlation is largest to biases in divergent wind speed at 300 hPa (0.34–0.50) and static stability at 500 hPa (0.31–0.47). For the North Pacific, we find the highest correlation to biases in meridional moisture flux at 300 hPa (0.33–0.52) over the North Pacific and all lead times (correlation between 0.37 and 0.61). Biases in relative vorticity at 850 hPa and relative humidity at 700 hPa show a smaller correlation to the WCB ascent biases (0.13–0.37 and 0.21–0.36, respectively). For the WCB outflow biases, the correlation to the biases in the predictor variables varies between variables, forecast lead times and ocean basins. For the North Atlantic, the correlation is largest to biases in divergent wind speed at 300 hPa (0.34–0.50) and static stability at 500 hPa (0.31–0.47). For the North Pacific, we find the highest correlation to biases in meridional moisture flux at 300 hPa (0.33–0.52) and static stability at 500 hPa (0.37–0.41, respectively). Over the North Pacific, the divergent wind speed bias has a high correlation for forecast day 15 (0.55) but very low correlation for other lead times. It is important to mention that the previously identified relevant correlations between biases in the predictors and WCB biases are independent of the magnitude of the coefficients in the logistic regression models (not shown).

In the following, we investigate in more detail the spatial patterns of the predictor variables with the highest bias correlations on forecast day 15 and evaluate the impact of a bias correction of all four predictor variables on WCB frequency biases and forecast skill. The WCB bias change is discussed in terms of the reduction (negative values) or increase (positive values) of the absolute value of the frequency biases (i.e., negative biases become less negative, and positive biases less positive).

In the North Atlantic region, the bias in meridional moisture flux at 850 hPa has a pronounced tilted dipole pattern with negative biases in the northern central Atlantic, and positive biases to the south. Over the North Pacific, the divergent wind speed bias has a high correlation for forecast day 15 (0.55) but very low correlation for other lead times. It is important to mention that the previously identified relevant correlations between biases in the predictors and WCB biases are independent of the magnitude of the coefficients in the logistic regression models (not shown).

In the following, we investigate in more detail the spatial patterns of the predictor variables with the highest bias correlations on forecast day 15 and evaluate the impact of a bias correction of all four predictor variables on WCB frequency biases and forecast skill. The WCB bias change is discussed in terms of the reduction (negative values) or increase (positive values) of the absolute value of the frequency biases (i.e., negative biases become less negative, and positive biases less positive).

In the North Atlantic region, the bias in meridional moisture flux at 850 hPa has a pronounced tilted dipole pattern with negative biases in the northern central Atlantic, and positive biases to the south. In addition a strong positive bias extends from the Gulf of Mexico into the southeastern United States (Fig. 7a). This pattern correlates well with the bias in WCB
inflow at day 15 (Fig. 2g) and correcting for biases in all predictor fields results in an overall bias reduction in the region (Fig. 7b). Likewise in the western North Pacific a strong positive bias in meridional moisture flux is evident south of Japan (Fig. 7a) and bias correction leads to a strong bias reduction for WCB inflow in that region (Fig. 7b). However, in the central and eastern North Pacific a strong dipole of positive and negative moisture flux biases extends much farther poleward than the corresponding biases in WCB inflow and bias correction only results in moderate bias reduction here.

The biases in meridional moisture flux at 500 hPa show a similar dipole in the North Atlantic albeit shifted somewhat poleward (Fig. 7c). This corresponds to the dipole pattern of WCB ascent biases at day 15 (Fig. 2h) and bias correction in all predictor fields overall reduce these WCB ascent biases (Fig. 7d). The positive moisture flux bias over North America does not affect WCB ascent frequencies. However, the negative moisture flux biases south of Japan and in the eastern North Pacific (Fig. 7c) correlate with WCB ascent biases in these regions (Fig. 2h) and a bias correction in all predictor fields reduces these WCB ascent biases (Fig. 7d).

The bias in divergent wind speed at 300 hPa has the highest correlation with WCB outflow biases (Table 1). However, in the North Atlantic region we see an overall positive bias in divergent wind speed flanked by negative biases at high and low latitudes (Fig. 7e) which does not correlate well with the overall negative WCB outflow bias in that region (Fig. 2i). Consequently, a bias correction in all predictor variables rather results in an increase of negative WCB outflow biases in the North Atlantic region (Fig. 7f). A somewhat better correlation of divergent wind speed biases and WCB outflow biases is evident over northwestern North America (Figs. 7e and 2i). Here a correction of the negative divergent wind speed biases and biases in other predictor fields strongly reduces the negative bias of WCB outflow (Fig. 7f). However, strong positive divergent wind speed biases extend from the central North Pacific to the North American west coast (Fig. 7e). Bias correction in this region only reduces the positive bias over the central Pacific but enhances the negative bias farther east (Fig. 7f).

All in all, our findings show that biases in meridional moisture flux at 850 and 500 hPa have a high correlation to WCB inflow and ascent biases and that a bias correction of all four predictor variables yields a general reduction of the WCB inflow and ascent biases. This is not the case for the WCB outflow, where the correlation to biases in the predictor variables varies between regions and lead times and where the bias correction even leads to an increase in WCB outflow bias in some regions.

c. WCB forecast biases at subseasonal lead times and their link to the MJO

So far we have investigated the instantaneous and local representation of WCBs at a given instantaneous day and specific grid point. This is admittedly a challenging forecasting task and not suitable for subseasonal lead times beyond 15 days. In the following, we explore the bias for weekly averaged WCB outflow frequencies and its relation to the MJO.

On average in week 1 (defined as days 0–6) negative WCB outflow biases extend in the western and northern part of the North Atlantic region, with some positive biases in the southeastern part of the region (Fig. 8a), consistent with the patterns on days 3 and 7 (Figs. 2c,f). Most of the North Pacific exhibits negative WCB outflow biases. In week 2 (defined as days 7–13) the patterns change, in particular in the North Pacific, and saturate toward week 3 and 4 (defined as days 14–20, 21–27, respectively): in the western and central North Pacific a positive WCB outflow bias emerges, while negative biases remain in the eastern North Pacific and western North America (Figs. 8b–d).

Next we focus on week 3, the time when WCB outflow frequency biases saturate (Fig. 9). Similar results are found for

<table>
<thead>
<tr>
<th>Table 1. Correlation (Pearson correlation coefficient) between WCB inflow, ascent, and outflow frequency bias and biases of logistic regression models’ input variables over the North Atlantic and North Pacific region for forecast days 3, 7, and 15 in DJF (1997–2017).</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inflow</strong></td>
</tr>
<tr>
<td>Thickness advection, 700 hPa</td>
</tr>
<tr>
<td>Meridional moisture flux, 850 hPa</td>
</tr>
<tr>
<td>Moisture flux convergence, 1000 hPa</td>
</tr>
<tr>
<td>Moist potential vorticity, 500 hPa</td>
</tr>
<tr>
<td><strong>Ascent</strong></td>
</tr>
<tr>
<td>Relative vorticity, 850 hPa</td>
</tr>
<tr>
<td>Relative humidity, 700 hPa</td>
</tr>
<tr>
<td>Thickness advection, 300 hPa</td>
</tr>
<tr>
<td>Meridional moisture flux, 500 hPa</td>
</tr>
<tr>
<td><strong>Outflow</strong></td>
</tr>
<tr>
<td>Relative humidity, 300 hPa</td>
</tr>
<tr>
<td>Divergent wind speed, 300 hPa</td>
</tr>
<tr>
<td>Static stability, 500 hPa</td>
</tr>
<tr>
<td>Relative vorticity, 300 hPa</td>
</tr>
</tbody>
</table>
weeks 2 and 4 (not shown). First we investigate if a link of WCB outflow biases and biases in the upper-tropospheric large-scale flow, as depicted by geopotential height at 300 hPa, exist. Generally the forecasted large-scale extratropical flow in winter is too zonal and the planetary wave pattern is slightly shifted in week 3 (Fig. 9b): over the eastern North Pacific and western North America a pronounced dipole of more than 50-gpm bias is evident with the negative pole slightly east of the apex of the stationary ridge. This is followed by a weaker positive anomaly centered on the stationary trough over eastern North America and a negative anomaly in the stationary ridge over Europe. Consistently, the negative WCB outflow frequency biases occur just east of the too weakly amplified stationary troughs in both basins, and extend into the region of the stationary ridges downstream, which is more evident over the North Pacific than the North Atlantic (Fig. 9a). It is noteworthy that the geopotential height biases at 300 hPa also reflect a Rossby wave train spanning from the tropical central North Pacific, via North America into the Middle East (Fig. 9b). This raises the question if the found biases can be related to tropical convection like the MJO. The MJO teleconnection is a main driver of subseasonal variability in winter and known to be too weakly represented in models (Vitart 2017).

Interestingly, both WCB outflow biases and biases in 300 hPa geopotential height are enhanced if the MJO is inactive\(^1\) at initialization time (Figs. 9c,d). This suggests that an active MJO with an amplitude $\geq 1$ overall yields a bias reduction. However, investigating biases in week 3 after an active MJO phase 2/3 and 6/7 at initialization time reveals cancelling effects in bias patterns as discussed in the following (Figs. 9e–h).

We focus on MJO phases 2/3 and 6/7 as these have been shown to most effectively establish teleconnections from the tropics to the North Atlantic through Rossby wave propagation. An active MJO phase 2/3 typically leads to an increased probability of the positive phase of the North Atlantic Oscillation (NAO) about 2 weeks later (Cassou 2008; Lin et al. 2009). The model does not fully capture this teleconnection by underestimating the southward extent of the trough and the northward extent of the ridge over the North Atlantic and Europe (Fig. 9f) resulting in a generally too zonal planetary wave pattern. Biases in WCB outflow activity might explain these patterns. Over

\[1 \text{ Here, inactive MJO means that the amplitude of the daily real-time multivariate MJO (RMM; Wheeler and Hendon 2004) index } \sqrt{\text{RMM1}^2 + \text{RMM2}^2} < 1. \]
northeastern North America and the southern tip of Greenland WCB outflow frequencies are strongly overestimated in the week 3 forecast and collocated with regions of overestimated geopotential height (Figs. 9e,f). Except for these regions, negative biases in WCB outflow remain similar to those for inactive MJO at initialization time. In addition, negative WCB outflow biases emerge over western and northern Europe collocated with the apex of the eastward shifted but underestimated ridge during cases with the active MJO phase 2/3 at initialization time (Fig. 9f).

An active MJO phase 6/7 typically leads to a negative phase of the NAO 1–2 weeks later with ridging over Greenland. The model underestimates the northward extension of the ridge (Fig. 9h). This goes along with a strong underestimation of WCB outflow activity in this region (Fig. 9g). Interestingly, biases in 300-hPa geopotential height are weaker over the North Pacific and North America compared to the other stratifications (cf. Figs. 9h,b,d,f). This suggests the model’s ability to establish the teleconnection into North America, but a lack of its continuation into Europe. At the same time WCB outflow biases are less in the eastern North Atlantic but an underestimation of WCB activity emerges in the western North Pacific (Fig. 9g).

In summary, we find a remarkable link of negative/positive biases in WCB outflow and negative/positive geopotential height biases which change depending on the MJO state at forecast initialization time. Although we cannot infer causality, our results indicate that a better representation of WCB activity and its modulation by the MJO could reduce systematic errors in the planetary Rossby wave pattern and large-scale midlatitude flow. The investigation of the cause and effect are not part of this study and are subject of a subsequent investigation.

d. WCB forecast skill at subseasonal lead times and its link to the MJO

As a final step we investigate the forecast skill for weekly mean WCB frequencies and its relation to the MJO. Therefore, we compute the FBSS for the North Pacific and North Atlantic regions during active (inactive) MJO phases at initialization time. As described in section 2e, the skill at day 0 represents the skill for days 0–6, on day 1 for days 1–7, and so on.

For the North Pacific region, there is slightly more skill in predicting WCB inflow, ascent, and outflow for weeks 1 and 2 when there is an active MJO at initialization time compared to the times with an inactive MJO at initial time (Fig. 10), consistent with the tendency to reduced WCB frequency biases in this region discussed before. Contrary, for the North Atlantic region, we generally find higher WCB forecast skill for weeks 1 and 2 when there is an inactive MJO at initial time. This is particularly pronounced for WCB outflow up to week 2 (days 7–13). These findings corroborate the hypothesis stated above that the model is able to establish the MJO teleconnection somewhat into North America but fails in extending it into the North Atlantic and Europe. At the same time it becomes evident that useful skill for weekly mean WCB frequency is already lost in week 2 and thus there is only a little enhanced forecast skill horizon for weekly averages compared to instantaneous WCB forecasts.

Finally, we have a detailed look at the changes in local skill during different states of the MJO at forecast initialization time in week 2 (days 7–13) compared to the skill for all initialization times (Fig. 11).

For the North Pacific, WCB inflow and ascent forecast skill is generally enhanced after an active MJO at initialization time and reduced after inactive MJO (Figs. 11a,b,d,e), consistent with the basinwide averages discussed before (cf. Fig. 10). WCB outflow skill in the North Pacific is not generally enhanced during an active MJO: the eastern North Pacific exhibits a dipole of reduced skill near and east of the date line and enhanced skill farther east (Figs. 11c,f). The reversed
A pattern can be found for forecasts initialized during an inactive MJO. Likewise the picture is less clear over the North Atlantic, but the generally higher WCB outflow skill during inactive MJO at initialization time is confirmed (Figs. 11a–f, cf. Fig. 10). Changes in skill tend to be anticorrelated between active MJO phases 2/3 and 6/7 at initialization time (Figs. 11g–l). This suggests that the little change in skill after active MJO phases stems from a cancelling effect. Most strikingly, skill for WCB outflow in the western North Pacific is strongly enhanced after MJO phase 2/3 but reduced after phase 6/7 (Figs. 11i,l).

In summary, we find that skillful forecast for weekly average WCB frequencies are possible up to week 2. Interestingly, the regional level of skill is strongly modulated by active MJO at initialization time and strongly depends on the phase of the MJO. This points to windows of enhanced predictability for WCBs.

5. Concluding discussion

This study presents a systematic investigation of the representation of warm conveyor belts (WCBs) in terms of frequency bias, forecast reliability, and forecast skill in the ECMWF IFS NWP model based on reforecasts of the S2S database. We use logistic regression models introduced in Part I and which are able to identify the inflow, ascent, and outflow stage of the WCB based on Eulerian input parameters. These regression models are applied to 20 years (1997–2017) of forecast data yielding a total of 920 initial times from the Northern Hemisphere winter season (DJF).

**Fig. 9.** Frequency bias of WCB (a), (c), (e), (g) outflow and (b), (d), (f), (h) bias of geopotential height field at 300 hPa for day 14–20 (week 3) (DJF 1997–2017) (shading) using different forecast initial times: (a), (b) all 920 initial times; (c), (d) all 298 initial times with an inactive MJO at initialization; (e), (f) all 134 initial times with an active MJO in phase 2/3 at initialization, and (g), (h) all 219 initial times with an active MJO in phase 6/7 at initialization. The black contours indicate a climatological frequency of 1%, 5%, 10%, and 15% and a geopotential height of 8400, 8600, 8800, 9000, 9200, and 9400 gpm in the respective layer based on ERA-Interim data. For clarity of the figure, WCB data have been smoothed with a Gaussian filter.
Significant WCB frequency biases occur already at early lead times (forecast day 3) which increase to forecast day 7, and saturate afterward. There is a tendency toward a general overestimation of WCB frequency over the southern part and an underestimation over the northern part of the North Atlantic for all WCB stages. Over the North Pacific, we find a relatively large overestimation of 30%–100% of the WCB inflow in some regions of the western part of the ocean basin and a large underestimation of similar magnitude for the WCB outflow over the eastern part of the ocean basin. An analysis of the forecast reliability reveals that most of the negative biases for the three WCB stages originate from forecasts with a very low ensemble probability of 0% or 9% while in reality a WCB occurred.

Previous studies have shown that an inaccurate representation of WCBs in NWP models can lead to errors in the downstream Rossby wave pattern (e.g., Lamberson et al. 2016; Martínez-Alvarado et al. 2016; Baumgart et al. 2018; Grams et al. 2018; Rodwell et al. 2018; Berman and Torn 2019; Maddison et al. 2019). The systematic underestimation of WCBs in the climatologically most active regions as well as north of these regions may have an impact on the downstream flow evolution and thus contribute to error in the downstream Rossby wave pattern. Since diabatic processes associated with WCBs can be important in the onset and maintenance of blocking anticyclones (Pfahl et al. 2015; Steinfeld and Pfahl 2019), the underestimation of WCB activity over the North Atlantic may at least partly explain negative blocking frequency biases over the North Atlantic–European region (Quinting and Vitart 2019). Our ongoing research aims to corroborate this hypothesis.

Finally, the study at hand, for the first time, quantifies the overall forecast skill horizon for WCBs in NWP models, facilitated by the availability of the reforecast dataset and novel diagnostics. We find a forecast skill horizon of 8–10 days for instantaneous prediction of all three WCB stages. There is significantly higher skill over the North Pacific region compared to the North Atlantic region for most lead times and WCB stages. For the North Atlantic, we find the highest WCB forecast skill in the areas with the highest climatological WCB frequency. Furthermore, there is more skill over the southern and western part of the North Atlantic compared to the northeastern part of the North Atlantic and toward Europe. Similarly, we find generally higher skill over the western North Pacific compared to the eastern part. A possible explanation of the lower skill over the eastern part of the ocean basins could be the relatively higher frequency of secondary cyclones (Schemm et al. 2018; Priestley et al. 2020) which are smaller in scale and potentially harder to predict. A detailed analysis of this hypothesis goes beyond the scope of this paper.

The WCB inflow, ascent, and outflow are calculated with three different logistic regression models which are set up individually. Therefore, there is a nontrivial link between the three stages of the WCB. Still, we find enhanced (reduced) skill in the WCB inflow over the western part (eastern part) of the North Atlantic which is followed by enhanced (reduced) skill for WCB ascent one day later. Even though the picture is less clear for the WCB outflow, this corroborates the hypothesis that skill levels are linked in a Lagrangian sense. We aim to incorporate a dynamical link between the WCB stages in a future version of the statistical WCB models (Quinting and Grams 2021b).

The results for the WCB frequency biases yield the question if WCB biases can be linked to biases in the predictor variables of the logistic regression models and if a correction of the biases would lead to an overall increase in forecast skill. We find strong correlations between WCB inflow (ascent) biases and biases in meridional moisture flux at 850 hPa (500 hPa). These biases in moisture flux, in particular over the North Pacific, correspond to biases in the atmospheric river frequency in ECMWF’s S2S forecasts reported by DeFlorio et al. (2019). For the WCB outflow, the correlation varies between the different predictor variables and the North Atlantic and North Pacific region. The correction of the biases in the predictor variables leads to a reduction in biases for the WCB inflow and ascent, but only in some regions for the WCB outflow. Our findings indicate a potential pathway to a reduced WCB inflow and ascent bias by reducing biases in the meridional moisture flux in the lower to mid troposphere. The impacts of the bias

**Fig. 10.** Area-averaged FBSS for weekly mean WCB frequencies (in observation and forecast) for DJF 1997–2017 dependent on forecast lead time for (a) inflow, (b) ascent, and (c) outflow. For a specific lead time, the value consists of the mean over the day and the following 6 days. We use the same regions as in Fig. 4a and calculate the FBSS for initial times with an active (inactive) MJO at initialization for the North Atlantic in blue (light blue) and the North Pacific in red (yellow).
correction on the overall WCB forecast skill are fairly low (between 1% and 5%) and do not lead to an increase forecast skill horizon beyond 8–10 days.

Since errors in the representation of the WCB can influence the midlatitude large-scale flow, we investigate biases for WCB outflow and geopotential height at 300 hPa on subseasonal time scales. The WCB outflow is underestimated over the eastern North Pacific and western North Atlantic which can be linked to an underestimation of the northward extent of the midlatitude flow at 300 hPa. Furthermore, there is a strong underestimation of the WCB outflow activity and northward extension of the ridge over the eastern North Pacific (western North Atlantic) 2 weeks after MJO phases 2 and 3 (6 and 7), which has also been shown by Vitart (2017). The investigation of the link between the MJO and WCB activity is part of our current research.

Weekly mean WCB frequencies can be skillfully predicted beyond the medium range up to week 3 with again more skill for the North Pacific compared to the North Atlantic (not shown). This is line with Zheng et al. (2019) who found generally higher skill for extratropical cyclones in S2S forecasts over the North Pacific than over the North Atlantic (e.g., their Figs. 4b,c). A hypothetical explanation for the generally higher WCB skill over the North Pacific might be related to a link between WCB activity and the MJO, which exhibits significant forecast skill at even longer lead times (Vitart 2017). In our study, we find higher (lower) skill for the North Pacific region for the WCB inflow, ascent, and outflow for weeks 1 and 2 when there is an active (inactive) MJO at initial time. The differences are especially apparent over the eastern North Pacific where the WCB skill is widely enhanced (reduced) after an active (inactive) MJO. Contrary, for the North Atlantic, there is higher skill in predicting WCB ascent and outflow after inactive MJOs indicating that the MJO signal is lost before reaching the North Atlantic region. Interestingly, we find a strong dependency of the change in skill on the actual phase of the MJO at initialization time, which deserves further inspection.

Overall, our study highlights that process-oriented diagnostic may give hint to synoptic-scale processes that potentially dilute forecast skill on subseasonal time scales. We investigate the representation of WCB forecasts since there is growing evidence that these systems are an important source and
magnifier of forecast uncertainty and errors in NWP models (e.g., Grams et al. 2018; Berman and Torn 2019). We find that WCB forecast skill already vanishes in the medium range with some skill on subseasonal time scales when considering weekly averages of WCB activity. Therefore, the known impact of WCB outflow on the large-scale Rossby wave pattern together with the little skill found in this study might point toward an impact of this synoptic-scale weather system on subseasonal predictability. Still, further work is needed to fully understand whether WCBs are the symptom of model issues or the initial cause of errors that project on the larger-scale Rossby wave pattern. One possible approach for this research problem would be relaxation experiments during which the NWP model is continuously nudged toward observations in the vicinity of the WCB (Magnusson 2017). By making use of the comprehensive S2S database, the link between the representation of WCBs in ECMWF’s subseasonal reforecasts and the prediction of Atlantic–European weather regimes is currently being investigated.

Acknowledgments. This research was funded by the Helmholtz Association as part of the Young Investigator Group “Sub-seasonal Predictability: Understanding the Role of Diabatic Outflow” (SPREADOUT; Grant VH-NG-1243) and was partially embedded in the subprojects A8 and B8 of the Transregional Collaborative Research Center SFB/TRR 165 “Waves to Weather” (https://www.wavestowather.de) funded by the German Research Foundation (DFG). The work is based on S2S data. S2S is a joint initiative of the World Weather Research Programme (WWRP) and the World Climate Research Programme (WCRP). The original S2S database is hosted at ECMWF as an extension of the TIGGE database. We are very grateful to Dominik Büeler for managing the S2S database at KIT and for the various discussions and ideas concerning this study. We also thank Peter Knippertz, Seraphine Hauser, Moritz Pickl, colleagues from ECMWF (Laura Ferranti and Linus Magnusson), Suzanne Gray, and two anonymous reviewers for valuable comments on the project. Furthermore, we acknowledge the ECMWF and Deutscher Wetterdienst for granting access to the ERA-Interim dataset. ECMWF’s subseasonal reforecasts from 1997 to 2017 are freely available at https://apps.ecmwf.int/datasets/data/s2s/ and ERA-Interim data are freely available at https://apps.ecmwf.int/datasets/data/interim-full-daily/. Daily values of the real-time multivariate MJO index were downloaded from http://www.bom.gov.au/climate/mjo/graphics/mmm.74toRealtime.txt.

REFERENCES


—, H. Binder, S. Pfahl, N. Piaget, and H. Wernli, 2014: Atmospheric processes triggering the central European floods

