1
2
3
4
5
6
7
8
9 **Subseasonal prediction of the state and evolution of the North Pacific jet stream**
10
11 *By*
12
13 ANDREW C. WINTERS[1][*]
14
15
16
17 [1]Department of Atmospheric and Oceanic Sciences
18 University of Colorado Boulder
19 Boulder, CO 80309
20
21
26
27
28
30
31 **Key Points:**
32
33 **1)** Skillful predictions of the prevailing North Pacific jet regime extend into the week 3

34   forecast period.

35 **2)** Bias-corrected forecasts verifying during jet retraction or equatorward shift regimes

36   feature the largest errors at subseasonal lead times.

37 **3)** The worst 21-day forecasts from each model are associated with the development,

38   maintenance, and decay of upper-tropospheric ridges.

[*] *Corresponding author address:* Andrew C. Winters, Dept. of Atmospheric and Oceanic Sciences, University of Colorado Boulder, 311 UCB, Boulder, CO 80309. E-mail: andrew.c.winters@colorado.edu

39 **Abstract**
40
41 The state and evolution of the North Pacific jet (NPJ) stream strongly influences the character of

42 the downstream synoptic-scale flow pattern over North America. This study employs data from

43 nine models within the Subseasonal-to-Seasonal Reforecast Database hosted by the European

44 Centre for Medium-Range Weather Forecasts to examine the subseasonal (2 weeks–1 month)

45 predictability of the NPJ through the lens of an NPJ phase diagram. The NPJ phase diagram

46 provides a visual representation of the state and evolution of the NPJ with respect to the two

47 leading modes of NPJ variability. The first mode of NPJ variability corresponds to a zonal

48 extension or retraction of the climatological jet-exit region, whereas the second mode

49 corresponds to a poleward or equatorward shift of the climatological jet-exit region. The analysis

50 reveals that ensemble forecasts of the prevailing NPJ regime, as determined from the NPJ phase

51 diagram, are skillful into week 3 of the forecast period. Forecasts initialized during a jet

52 retraction, or verifying during a jet retraction and equatorward shift, generally feature the largest

53 errors during the forecast period. Examination of the worst-performing 21-day forecasts from

54 each model demonstrates that the worst-performing forecasts are uniformly associated with

55 development, maintenance, and decay of upper-tropospheric ridges over the high-latitude North

56 Pacific. These results demonstrate that bias-corrected NPJ phase diagram forecasts have the

57 potential to identify periods that may exhibit enhanced forecast skill at subseasonal lead times

58 based on the anticipated NPJ evolution.

59

60

61

62

63   **Plain Language Summary**

64   The jet stream is a ribbon of rapidly moving air that circumnavigates the globe approximately 12

65   km above the Earth's surface. The evolution of a segment of the jet stream over the North

66   Pacific, hereafter referred to as the North Pacific jet (NPJ), exerts an important influence on

67   downstream weather conditions over North America. Consequently, this study examines the

68   extent to which forecast models can accurately capture the state and evolution of the NPJ 2–4

69   weeks in advance. The analysis reveals that an elongated or poleward shifted NPJ is generally

70   characterized by enhanced forecast accuracy, whereas a wavier or split NPJ is generally

71   characterized by reduced forecast accuracy. Recognition of these NPJ configurations within a

72   real time forecast environment can provide "windows of opportunity", in which conditions over

73   the North Pacific and North America can be forecasted with a higher degree of precision

74   compared to climatology up to 4 weeks in advance.

75

76

77

78

79

80

81

82

83

84

85

1. Introduction

The improvement of subseasonal (2 weeks to 1 month) forecasts has been a priority for the meteorological community and its partners (NRC, 2010; NAS, 2018) given that this time scale is characterized by a forecast skill "gap" within numerical weather prediction models. In particular, skillful forecasts on shorter (i.e., weather) time scales predominantly arise from atmospheric initial conditions, whereas skillful forecasts on longer (i.e., seasonal) time scales predominantly arise from low frequency climate variations such as sea-surface temperature and soil moisture fluctuations (e.g., NRC, 2010; NAS, 2018; Vitart et al., 2017; Pegion et al., 2019; Meehl et al., 2021). Consequently, the subseasonal time scale lies in a transition period during which forecast skill is not as effectively derived from atmospheric initial conditions or low frequency climate variations. Nevertheless, subseasonal forecasts offer considerable value to stakeholders, including individuals in emergency management, agriculture, water management, and public health (White et al., 2017; Pegion et al., 2019), who can act to mitigate risks from the occurrence of anomalous weather conditions.

The identification and prediction of "weather regimes", which are defined as reoccurring and/or persistent large-scale atmospheric patterns maintained by synoptic-scale weather systems (e.g., Reinhold & Pierrehumbert, 1982; Vautard, 1990; Ferranti et al., 2015, 2018; Straus et al., 2017; Vigaud et al. 2018; Lee et al. 2019; Winters et al. 2019a; Robertson et al., 2020), represent burgeoning areas of research relevant to the subseasonal time scale. Weather regimes can be defined over a spectrum of spatial domains, such as the Northern Hemisphere (e.g., Mo & Ghil, 1988; Kimoto & Ghil, 1993; Corti et al., 1999), the Euro–Atlantic sector (e.g., Vautard, 1990; Michelangeli et al., 1995; Cassou, 2008; Dawson & Palmer, 2014; Ferranti et al., 2015, 2018; Grams et al., 2017; Matsueda & Palmer, 2018), and the Pacific–North American sector (e.g.,

109 Robertson & Ghil, 1999; Straus et al., 2007; Riddle et al., 2013; Matsueda & Kyouda, 2016;

110 Vigaud et al., 2018; Amini & Straus, 2019; Lee et al., 2019; Winters et al., 2019a; Robertson et

111 al., 2020). Knowledge of the prevailing or forecasted weather regime subsequently provides

112 insight into the character of the large-scale flow pattern over a region as well as the relative

113 likelihood for anomalous sensible weather to develop in conjunction with that regime.

114   Examinations into the predictability of weather regimes have been predominantly focused

115 on the Euro-Atlantic sector (e.g., Ferranti et al. 2015, 2018; Matsueda & Palmer, 2018). A

116 common thread among these examinations is that forecast models have difficulty capturing the

117 onset, maintenance, and decay of upper-tropospheric blocking events, which has implications for

118 the occurrence of high-impact weather events over Europe, such as cold-air outbreaks and heat

119 waves (e.g., Jung et al., 2011; Ferranti et al., 2018; Quandt et al., 2019). Evaluation of the

120 predictability of weather regimes over North America has recently received greater attention. In

121 particular, Vigaud et al. (2018) and Robertson et al. (2020) demonstrate that the predictability of

122 North American weather regimes, as defined from a k-means clustering analysis of 500-hPa

123 geopotential height, is generally on the order of two weeks. Robertson et al. (2020) observe,

124 however, that there are "forecasts of opportunity" in which the prevailing weather regime may be

125 predicted with skill up to four weeks in advance. These forecasts of opportunity were found to

126 coincide with periods influenced by low frequency modes of climate variability such as the El

127 Niño–Southern Oscillation and the Madden–Julian Oscillation.

128   The North Pacific jet (NPJ) stream represents a synoptic-scale feature whose state and

129 evolution serves as a conduit between the aforementioned modes of low frequency climate

130 variability and the character of the downstream large-scale flow pattern over North America

131 (e.g., Cordeira & Bosart, 2010; Archambault et al., 2015; Bosart et al., 2017; Griffin & Martin,

132   2017; Vigaud et al. 2018; Winters et al., 2019a,b; Robertson et al., 2020). Therefore, accurate

133   forecasts of the state and evolution of the NPJ may also exhibit the potential to inform

134   predictions of weather conditions over North America. Winters et al. (2019a) developed an NPJ

135   phase diagram on the basis of this observation to objectively track the state and evolution of the

136   NPJ using output from reanalysis products and numerical weather prediction models. The NPJ

137   phase diagram is constructed from the two-leading empirical orthogonal functions (EOFs) of

138   250-hPa zonal wind anomalies over the North Pacific during September–May. The first EOF

139   corresponds to a zonal extension or retraction of the climatological exit region of the NPJ,

140   whereas the second EOF corresponds to a poleward or equatorward shift of the climatological

141   exit region of the NPJ. Figure 1 shows the characteristic large-scale flow patterns associated with

142   the four primary NPJ regimes derived from the NPJ phase diagram and reveals that each NPJ

143   regime is associated with distinct temperature and sea-level pressure anomaly patterns across the

144   Pacific–North American sector. Winters et al. (2019b) and Turasky (2019) further demonstrate

145   that the frequencies of continental U.S. extreme temperature events and landfalling atmospheric

146   river events along the U.S. west coast are significantly modulated by the antecedent state and

147   evolution of the NPJ as determined from the NPJ phase diagram.

148   Predicated on the relationship between each NPJ regime and the large-scale flow pattern

149   over North America, Winters et al. (2019a) conducted an evaluation of the medium-range (6–10-

150   day) forecast skill associated with each NPJ regime by calculating 9-day ensemble forecasts of

151   the state and evolution of the NPJ in the context of the NPJ phase diagram using the GEFS

152   Reforecast Version 2 dataset (Hamill et al., 2013). Their analysis found that ensemble mean

153   forecasts verifying during jet retraction and equatorward shift regimes were associated with

154   larger medium-range forecast errors than forecasts verifying during jet extension and poleward

155 shift regimes. Consideration of the worst-performing 9-day NPJ phase diagram forecasts also

156 found that the worst forecasts occurred in conjunction with rapid NPJ regime transitions towards

157 an equatorward shift regime as well as the development of North Pacific blocking ridges.

158   A limitation to the Winters et al. (2019a) analysis is that it focuses on one ensemble

159 prediction system and does not consider the extent to which the forecast skill of the NPJ extends

160 into subseasonal time scales. Furthermore, prior work on North American weather regimes does

161 not quantify the forecast skill associated with regime transitions, which are periods that can often

162 lead to substantial downstream impacts over North America (e.g., Bosart et al., 2017). The NPJ

163 phase diagram is well suited for such an analysis, much like those diagrams used by Matsueda &

164 Palmer (2018) and Ferranti et al. (2018) to describe Euro-Atlantic weather regimes, since it

165 allows for observable transitions between regimes and serves as an objective tool to evaluate the

166 ability of models to simulate the cumulative upper-tropospheric flow response to tropical and

167 midlatitude forcing. The remainder of this study is organized as follows. Section 2 discusses the

168 data and methodology used to construct the NPJ phase diagram and NPJ phase diagram

169 forecasts. Section 3 examines the biases and multi-model skill of NPJ phase diagram forecasts.

170 Section 4 considers the evolution of the synoptic-scale flow pattern associated with the best- and

171 worst-performing NPJ phase diagram forecasts from each model, and section 5 provides a

172 discussion of the main conclusions from this work.

173

174 **2. Data and methods**

175 *2.1. Data*

176   This study uses data at 6-h intervals during September–May 1979–2019 from the

177 National Centers for Environmental Prediction (NCEP) Climate Forecast System Reanalysis

7

178   (CFSR; Saha et al., 2010, 2014) as well as data during September–May from the Subseasonal-to-

179   Seasonal (S2S) Reforecast Database hosted by the European Centre for Medium-Range Weather

180   Forecasts (ECMWF; Vitart et al., 2017). The CFSR features 0.5° horizontal grid spacing and 64

181   vertical levels that extend from the surface to 0.26 hPa. The S2S Reforecast Database consists of

182   reforecasts from 11 operational centers, each with a different reforecast period, ensemble size,

183   forecast frequency, forecast length, and model version. Reforecast data are stored on 10 pressure

184   levels and a 1.5° × 1.5° latitude-longitude grid, are initialized at 0000 UTC, and are available at

185   forecast lead times as long as 32–61 days at 24-h intervals. Exceptions to this format are

186   reforecasts from the Australian Bureau of Meteorology (BoM), which are stored on a gaussian

187   grid, and reforecasts from the Japan Meteorological Agency (JMA), which are initialized at 1200

188   UTC.

189   To ensure uniformity in the forthcoming analyses, this study does not consider

190   reforecasts from the BoM and JMA, and only uses reforecasts from the nine operational centers

191   identified in Table 1. These centers include Environment and Climate Change Canada (ECCC),

192   Météo-France/Centre National de Recherche Meteorologiques (CNRM), the Institute of

193   Atmospheric Sciences and Climate of the National Research Council (ISAC), the Korea

194   Meteorological Administration (KMA), NCEP, the UK Met Office (UKMO), the

195   Hydrometeorological Center of Russia (HMCR), ECMWF, and the China Meteorological

196   Administration (CMA). The reforecasts from a particular center are constructed using either a

197   "fixed" version of a forecast model or "on the fly" using the current version of a forecast model

198   on the date reforecasts were conducted. For this study, the most recent version of a forecast

199   model prior to 2019 is used to acquire "fixed" reforecast data, and those reforecasts that were

200   conducted during 2019 represent reforecast data that was compiled "on the fly". Some "on the

8

201   fly" reforecasts from the CMA model were also conducted during 2020 to ensure that reforecasts

202   are available throughout September–May during the CMA's reforecast period. Full details on the

203   characteristics of each reforecast dataset are discussed at length in Vitart et al. (2017).

204   *2.2. The NPJ phase diagram*

205   The NPJ phase diagram is constructed in an identical manner as in Winters et al. (2019a)

206   with slight modifications to align with the format of the S2S Reforecast Database. Therefore, the

207   forthcoming discussion in this subsection mirrors that from Winters et al. (2019a). First, CFSR

208   data are regridded to 1.5° horizontal grid spacing to match the grid spacing of the reforecast data.

209   Next, 300-hPa zonal wind anomalies from the CFSR are calculated at 6-h intervals during

210   September–May 1979–2019 for each grid point within a North Pacific domain (10.5–79.5°N;

211   100.5–240°E) that aligns with those used in prior work on NPJ variability (e.g., Jaffe et al., 2011;

212   Griffin & Martin, 2017; Winters et al., 2019a,b). 300-hPa zonal wind anomalies are determined

213   with respect to the CFSR climatology, which is calculated at 6-h intervals for each grid point by

214   retaining the first four harmonics of the mean annual cycle. Note that S2S reforecast data are

215   only available at 300 hPa and 200 hPa. Therefore, the use of 300-hPa zonal wind anomalies in

216   this study represents a departure from the 250-hPa zonal wind anomalies that Winters et al.

217   (2019a) employ in their development of the NPJ phase diagram.

218   A traditional EOF analysis (Wilks, 2011) is performed on the aforementioned 300-hPa

219   zonal wind anomaly data from the CFSR to reveal the two leading modes of NPJ variability

220   (Figs. 2a,b). EOF 1 explains 9.9% of the variance and corresponds to a zonal extension or

221   retraction of the climatological jet-exit region. EOF 2 explains 7.2% of the variance and

222   corresponds to a poleward or equatorward shift of the climatological jet-exit region. The two

223   leading EOFs, and their explained variance, are similar to those found in prior work (e.g.,

9

224   Athanasiadis et al., 2010; Jaffe et al., 2011; Griffin & Martin, 2017; Winters et al., 2019a), and

225   are statistically well separated (North et al., 1982). To instill confidence that the identified NPJ

226   regimes are robust, the same modes of NPJ variability found using 6-h data from the CFSR were

227   also observed when EOF analyses were performed on monthly-averaged zonal wind anomaly

228   data as well as on 6-h data from ERA-Interim (Dee et al., 2011). In particular, the correlation and

229   median absolute difference between the principal component (PC) time series obtained from

230   separate EOF analyses on CFSR and ERA-Interim data were 0.99 and 0.03, respectively.

231   The temporal evolution of the NPJ with respect to the two leading EOFs is characterized

232   using the PC time series that are returned from the traditional EOF analysis. For this study, 6-h

233   PC data are normalized to unit variance and are averaged over a 5-day period centered on each

234   analysis time. This 5-day average of the PCs removes the high frequency variability of the jet on

235   daily time scales but retains the lower frequency variability of the jet on synoptic time scales.

236   The PCs at a particular analysis time can be visualized by plotting them on the NPJ phase

237   diagram shown in Fig. 2c. The distance along the *x*-axis in the NPJ phase diagram identifies how

238   strongly 300-hPa zonal wind anomalies at that time project onto EOF 1, where positive values

239   represent a jet extension and negative values represent a jet retraction. The distance along the *y*-

240   axis in the NPJ phase diagram identifies how strongly 300-hPa zonal wind anomalies at that time

241   project onto EOF 2, where positive values represent a poleward shift and negative values

242   represent an equatorward shift. The projection of PCs onto the two leading EOFs over a selected

243   time period produces a trajectory within the NPJ phase diagram that describes the NPJ evolution

244   in the context of the two leading EOFs.

245   The NPJ phase diagram is subsequently used to classify the state of the NPJ into four NPJ

246   regimes based on whether the magnitude of PC 1 or PC 2 is larger and whether the NPJ resides

247 at a distance of greater than 1 PC unit from the origin. A projection that falls within a radius of 1

248 PC unit of the origin of the NPJ phase diagram represents an NPJ that does not project well onto

249 the two leading EOFs or that resembles climatology. For reference, the NPJ typically resides

250 within each of the four primary NPJ regimes approximately 15% of the time and within the unit

251 circle centered on the origin approximately 40% of the time (Winters et al., 2019a). There are

252 generally no preferred transitions between NPJ regimes (i.e., cross correlations between PC 1

253 and PC 2 are close to zero at all time lags), and the autocorrelation functions for PC 1 and PC 2

254 drop below 0.5 after 1 week (Fig. 3a), which can serve as a benchmark for the forthcoming

255 analysis in section 3 that evaluates the forecast skill added by each S2S model (e.g., Pegion et al.

256 2019; Domeisen & Butler, 2020; Feng et al. 2021).

257 *2.3. NPJ phase diagram reforecasts and verification*

258 300-hPa zonal wind anomalies from the nine reforecast datasets identified in Table 1 are

259 used to construct ensembles of NPJ phase diagram forecasts with forecast lead times as long as

260 32–61 days, depending on the model. To start, 300-hPa zonal wind anomalies are calculated for

261 each ensemble member and at every forecast lead time based on the CFSR climatology. This is

262 done to provide a baseline quantification of forecast skill for each model and to identify any

263 biases in each model's representation of the NPJ. The zonal wind anomalies associated with each

264 ensemble member forecast are then projected onto the two leading modes of NPJ variability

265 shown in Fig. 2 to construct an ensemble of trajectories within the NPJ phase diagram that

266 describe the forecast evolution of the NPJ (e.g., Fig. 3b). As with the CFSR data, the forecast

267 PCs within a 5-day window centered on each forecast lead time are averaged together to remove

268 high frequency variations of the NPJ on daily time scales. The 5-day average forecast PCs at 0-h,

269 24-h, and 48-h lead times are specifically calculated by appending CFSR PCs 48-h, 24-h, and 0-h

11

270   prior to the start of the forecast period onto the beginning of the forecast PC time series

271   associated with each ensemble member. All ensemble member NPJ phase diagram forecasts

272   initialized at the same time from a particular model are then averaged together to produce an

273   ensemble mean NPJ phase diagram forecast.

274   NPJ phase diagram forecasts are evaluated by calculating the Euclidean distance between

275   the ensemble mean forecast position of the NPJ within the NPJ phase diagram at a particular

276   forecast lead time and the verifying position of the NPJ at that same forecast lead time using the

277   CFSR. These Euclidean distance statistics are calculated for individual ensemble member NPJ

278   phase diagram forecasts, as well. Note that a reanalysis product must be used for verification

279   given that 0-h forecasts are not available at a daily frequency for each model within the S2S

280   dataset. Forecasts are then classified based on the NPJ regime at the time of forecast initialization

281   as well as the observed NPJ regime at the time of forecast verification using the position of the

282   NPJ within the NPJ phase diagram according to Fig. 2c. This classification of forecasts permits

283   an examination of the extent to which forecast performance varies across models and the four

284   primary NPJ regimes. Forecasts verifying during the month of June are excluded from any

285   calculated forecast statistics given that the NPJ phase diagram is derived solely from zonal wind

286   anomaly data during September–May.

287   Once these baseline statistics are obtained, the analyses described above are repeated by

288   calculating forecast 300-hPa zonal wind anomalies using each model's lead-dependent

289   climatology rather than the CFSR climatology. These analyses account for biases in each

290   model's representation of the NPJ as a function of forecast lead time and allow for a

291   quantification of whether bias correction improves the predictive skill of the NPJ on S2S time

292   scales. Each model's lead-dependent climatology is constructed by averaging all forecasts at the

293   same forecast lead time that were initialized within 10 days of a selected calendar day during that

294   model's reforecast period. Similar to Robertson et al. (2020), no cross validation is used in the

295   calculation of model climatologies. EOF analyses performed on the bias-corrected forecast

296   anomalies from each model consistently reproduce the same two leading modes of NPJ

297   variability shown in Figs. 2a,b (not shown).

298   The present study also identifies the synoptic-scale flow patterns and evolutions that are

299   associated with the best- and worst-performing NPJ phase diagram forecasts from each model.

300   For this purpose, the bias-corrected NPJ forecasts are used. The best- and worst-performing

301   forecasts are identified in a similar manner as Winters et al. (2019a, see their Fig. 10 for a

302   schematic) using both (1) the cumulative ensemble mean Euclidean distance error in the context

303   of the NPJ phase diagram during days 3–21 of the forecast period and (2) the cumulative

304   ensemble member Euclidean distance error during the same period. This forecast period is

305   selected to remove the influence of CFSR PCs on short-lead forecast errors. The best-performing

306   forecasts are those forecasts that rank in the lowest 10% in terms of both the cumulative

307   ensemble mean error and the cumulative ensemble member error for a particular model, whereas

308   the worst-performing forecasts are those forecasts that rank in the highest 10% in terms of both

309   the cumulative ensemble mean error and the cumulative ensemble member error for a particular

310   model. The use of both criteria identifies the best-performing forecasts as those that are accurate

311   and confident (i.e., small ensemble spread). The worst-performing forecasts based on both

312   criteria are those that are inaccurate and uncertain (i.e., large ensemble spread) *or* inaccurate but

313   confident (i.e., small ensemble spread). Put another way, the worst-performing forecasts are

314   those that are the most inaccurate, regardless of the ensemble spread.

315

3. Multi-model performance of NPJ phase diagram forecasts

316

317   The total number of valid NPJ phase diagram forecasts from each model (i.e., only those

318   forecasts that are initialized *and* verified during September–May) is shown as a function of

319   forecast lead time in Fig. 4a. Each model is associated with at least 500 valid NPJ phase diagram

320   forecasts at every forecast lead time, with those models that feature a greater forecast frequency

321   (i.e., the NCEP, ISAC, and ECMWF) characterized by larger sample sizes. While each model is

322   initialized at different frequencies and over different years (Table 1), Fig. 4a reveals that there is

323   a suitable sample size of reforecasts from each model from which to draw conclusions

324   concerning the predictability of the NPJ on subseasonal timescales.

325   First, the analysis considers the baseline skill of NPJ phase diagram forecasts without the

326   application of bias correction. The average ensemble mean distance error of NPJ phase diagram

327   forecasts increases exponentially during week 1 of the forecast period, increases linearly during

328   week 2, and levels off during week 3 for all models. The average ensemble mean error remains

329   approximately constant thereafter, suggesting that any differentiable skill of NPJ phase diagram

330   forecasts diminishes after 21 days (Fig. 4b). The difference in the average ensemble mean error

331   between models at any forecast lead time is also no larger than 0.5 PC units, with the ECWMF

332   model exhibiting the lowest average ensemble mean error at all forecast lead times for which it

333   features a valid forecast. Note that the larger ensembles (e.g., ECMWF, CNRM, HMCR) aren't

334   uniformly associated with lower average ensemble mean errors, as the HMCR model ranks in the

335   top 50% of all models in terms of its average ensemble mean error at every forecast lead time.

336   Figure 4c shows the percent of ensemble member forecasts from each model that correctly

337   identify the verifying NPJ regime at each forecast lead time and reveals that all models are

338   significantly more skillful at identifying the prevailing NPJ regime compared to random chance

339   at lead times shorter than 21–28 days. The largest difference in performance between forecast

340   models is maximized near the end of week 1 (Fig. 4c), where some models are approximately

341   20% less accurate at identifying the prevailing NPJ regime than the best-performing model (i.e.,

342   ECMWF).

343   Motivated by the observation that NPJ phase diagram forecasts exhibit skill compared to

344   climatology into weeks 3 and 4 of the forecast period (Figs. 4b,c), the forthcoming analysis

345   considers the extent to which NPJ phase diagram forecast errors vary based on the initial NPJ

346   configuration. Figures 5a,c,e,g show the number of forecasts from each model that are initialized

347   within each of the four primary NPJ regimes as a function of forecast lead time. In contrast to

348   earlier analyses, forecast error (Figs. 5b,d,f,h) is now expressed as a percentage relative to the

349   average ensemble mean error of all forecasts from a particular model that are initialized within

350   one of the four primary NPJ regimes. Forecasts that are initialized within the origin of the NPJ

351   phase diagram are not factored into this analysis since the NPJ does not project strongly onto one

352   of the leading modes of NPJ variability. Positive percentages indicate that ensemble mean

353   forecast errors are larger than average when a model is initialized during a certain NPJ regime,

354   whereas negative percentages indicate that ensemble mean forecast errors are smaller than

355   average.

356   Figure 5d reveals that forecasts initialized during a jet retraction feature an ensemble

357   mean forecast error that is 10–20% greater than each model's average at a 7-day lead time,

358   whereas forecasts initialized during a poleward shift feature errors that are 5–15% less than each

359   model's average at the same lead time (Fig. 5f). Forecasts initialized during a jet extension are

360   characterized by errors that are between 10% less and 5% greater than each model's average at

361   0–2-week lead times (Fig. 5b), and forecasts initialized during an equatorward shift are

362   characterized by errors that are between 5% less and 10% greater than each model's average at

363   0–2-week lead times. At lead times beyond 2 weeks, the forecast errors associated with each NPJ

364   regime are comparable to one another. Consequently, there does not appear to be a systematic

365   difference in forecast performance based on the initial NPJ regime at lead times longer than 2

366   weeks as the forecasts are further removed from the influence of the model's initial conditions.

367   Figure 6 considers the ensemble mean forecast error associated with each model based on

368   the NPJ regime at the time of forecast verification. This approach evaluates the extent to which

369   forecast performance varies based on the character of the NPJ evolution following forecast

370   initialization. The number of forecasts associated with each model as a function of the verifying

371   NPJ regime are shown in Figs. 6a,c,e,g. Overall, NPJ phase diagram forecasts that verify during

372   a jet retraction (Fig. 6d) or equatorward shift (Fig. 6h) exhibit systematically larger ensemble

373   mean forecast errors than forecasts that verify during a jet extension (Fig. 6b) or poleward shift

374   (Fig. 6f) at lead times less than 7 days. This result aligns with that found by Winters et al.

375   (2019a) using the GEFS Reforecast Version 2 dataset and implies that forecasts associated with

376   the development of a North Pacific ridge (Figs. 1c,f) during week 1 feature greater ensemble

377   mean forecast errors across all models.

378   At lead times longer than 7 days, the performance of NPJ phase diagram forecasts

379   verifying during each NPJ regime is dependent on the model. In particular, the forecasts with the

380   largest errors at lead times exceeding 2 weeks verify during an equatorward shift regime for the

381   ECCC, CNRM, HMCR, ECMWF, and CMA models, during a poleward shift for the KMA and

382   UKMO models, during a jet retraction for the NCEP model, and during a jet extension for the

383   ISAC model (cf. Figs. 6b,d,f,h). Similar variability across models is also observed when

384   considering the verifying NPJ regimes that exhibit the lowest forecast errors at lead times

16

385 exceeding 2 weeks. Since the preceding analysis does not yet account for forecast model biases,

386 the observed differences in forecast model performance based on the verifying NPJ regime at

387 lead times exceeding 2 weeks may be related to frequency biases in the prediction of each NPJ

388 regime.

389   To this aim, Figures 7a–d depict the percent frequency that each NPJ regime is

390 overforecast or underforecast in each model with respect to verification. Note that the ISAC

391 model is not included in this initial analysis and will be discussed separately. For this analysis,

392 each ensemble member initialized using a particular model is treated as a separate forecast of the

393 NPJ regime. Figure 7a reveals that the NCEP model overforecasts the occurrence of jet

394 extensions by approximately 30–40% compared to verification at lead times exceeding 2 weeks,

395 whereas jet extensions are underforecast by all other models by as much as 20%. Conversely, all

396 models overforecast the occurrence of jet retractions by as much as 30% at lead times exceeding

397 2 weeks, except for the NCEP model, which underforecasts the occurrence of jet retractions by

398 approximately 30% (Fig. 7b).

399   The frequency of poleward shift and equatorward shift forecasts compared to verification

400 is more variable across models compared to jet extension and jet retraction forecasts. In

401 particular, the HMCR, ECCC, and CNRM models overforecast the occurrence of poleward shifts

402 at lead times exceeding 2 weeks, with an overforecast of poleward shifts by as much as 70–90%

403 during week 4 in the HMCR model (Fig. 7c). Poleward shifts are underforecast by the CMA,

404 ECMWF, NCEP, KMA, and UKMO models by as much as 30% compared to verification at lead

405 times exceeding 2 weeks. Last, equatorward shifts are overforecast by 10–50% in the NCEP,

406 CMA, UKMO, and KMA models, while the ECCC and HMCR models underforecast the

407 occurrence of equatorward shifts by 20–60% (Fig. 7d). Notably, the frequency of CNRM and

17

408   ECMWF forecasts of equatorward shifts is comparable to verification throughout the forecast

409   period. The ISAC model is a particularly interesting case (Fig. 7e), in which jet extensions are

410   underforecast by close to 80% during weeks 2–4, and jet retractions are overforecast by 140–

411   200%.

412   Biases in the forecast frequency of each NPJ regime in Fig. 7 are associated with the

413   forecast errors identified in Fig. 6. Namely, the largest ensemble mean forecast errors during

414   weeks 2–4 in the ECCC, HMCR, KMA, UKMO, CFSR, and ISAC models are associated with

415   the same verifying NPJ regime for which those models exhibit a low forecast frequency bias (cf.

416   Figs. 6–7). As suggested by Ferranti et al. (2015), this observation implies that the reduced

417   performance of model forecasts that verify in those respective NPJ regimes may be due to the

418   misrepresentation of physical processes that lead to the development of those NPJ regimes. For

419   the ECMWF, CNRM, and CMA models, which feature their largest forecast errors during

420   periods that verify during an equatorward shift, there is not a clear low forecast frequency bias

421   for equatorward shifts. In fact, the CMA exhibits a high forecast frequency bias for equatorward

422   shifts compared to verification. This result implies that these three models are able to represent

423   the physical processes that lead to the development of equatorward shifts with fidelity, but that

424   equatorward shifts may be characterized by low intrinsic predictability.

425   The same analyses described above are repeated with bias-corrected NPJ phase diagram

426   forecasts that utilize 300-hPa zonal wind anomalies based on each model's lead-dependent

427   climatology rather than the CFSR. The use of bias-corrected forecasts substantially reduces the

428   regime frequency biases shown in Fig. 7. While not shown explicitly, all bias-corrected forecast

429   statistics are similar to those shown in Fig. 3 and feature slightly reduced ensemble mean errors

430   at lead times exceeding 2 weeks. Additionally, the classification of bias-corrected forecast errors

431   based on the NPJ regime at the time of forecast initialization is similar to the results shown in

432   Fig. 5 (not shown). Substantial differences are noted in comparison to the baseline forecasts,

433   however, when classifying bias-corrected forecast errors based on the verifying NPJ regime (cf.,

434   Figs. 6 and 8). Namely, bias-corrected forecast errors (Fig. 8) are not as substantial as those

435   shown in Fig. 6, and there is a general agreement between model errors associated with each

436   verifying NPJ regime at all forecast lead times. In particular, forecasts verifying during jet

437   retractions and equatorward shifts typically exhibit larger than normal forecast errors, whereas

438   forecasts verifying during jet extensions and poleward shifts typically exhibit reduced forecast

439   errors compared to each model's climatology. The general agreement among bias-corrected

440   model errors for each verifying NPJ regime suggests that NPJ evolutions towards a jet retraction

441   or equatorward shift may be characterized by a lower degree of intrinsic predictability than NPJ

442   evolutions towards a jet extension or poleward shift.

443   Reliability diagrams that evaluate the probabilistic detection of the verifying NPJ regime

444   for the three largest ensembles (i.e., CNRM, ECMWF, HMCR) further demonstrate that bias-

445   corrected NPJ phase diagram forecasts are underdispersive at forecast lead times exceeding 7

446   days (Fig. 9). Consequently, ensemble forecasts from these three models tend to be

447   overconfident in the development of a particular NPJ regime at medium-range and subseasonal

448   lead times. In particular, both CNRM (Fig. 9a) and ECMWF (Fig. 9b) forecast probabilities

449   exceeding 50% are overconfident by 5–20% at forecast lead times exceeding 14 days, whereas

450   HMCR forecast probabilities exceeding 50% are overconfident by 20–40% (Fig. 9c). The

451   reduced performance of HMCR forecasts compared to CNRM and ECMWF forecasts is also

452   apparent in Fig. 9d, which reveals that the Brier Skill Score (Wilks 2011) for HMCR forecasts is

453   less than that for the CNRM and ECMWF models at all forecast lead times.

454

4. Synoptic-scale flow patterns associated with the best- and worst-performing forecasts

Results from the previous section suggest that the best- and worst-performing subseasonal NPJ phase diagram forecasts are associated with different NPJ regimes (e.g., Figs. 5 and 8). Consequently, the forthcoming analysis considers the synoptic-scale characteristics of the 21-day period following the initiation of a best- and worst-performing forecast from each model. As mentioned in section 2.3, the best-performing forecasts are those bias-corrected forecasts in which there is both a low cumulative ensemble mean distance error in the context of the NPJ phase diagram (i.e., an accurate forecast) and a low cumulative ensemble member distance error (i.e., a confident forecast) during days 3–21 of the forecast period. The worst-performing forecasts are those in which there is both a high cumulative ensemble mean distance error in the context of the NPJ phase diagram and a high cumulative ensemble member distance error (i.e., the most inaccurate forecasts).

The average position of the NPJ within the NPJ phase diagram on the date a best-performing forecast is initialized from each forecast model is shown in Fig. 10a and reveals that the NPJ is generally displaced towards a poleward shift regime. The models are clustered near the origin, however, which suggests that the NPJ may also be close to its climatological state or exhibit considerable variability in its state at the time a best-performing forecast is initialized. The state of the NPJ at the start of a worst-performing forecast period is displaced towards a jet retraction or equatorward shift (Fig. 10b). This result aligns well with Figs. 5d,h, which indicate that forecast errors are often higher than each model's average during the first 2 weeks of the forecast period when a model is initialized during those two NPJ regimes.

476   Figures 10c,d illustrate the composite evolution of the NPJ during the 21-day period

477   following the initialization of a best- and worst-performing forecast from each model. The

478   composite evolution of the NPJ associated with each model is calculated by projecting 300-hPa

479   zonal wind anomalies from the CFSR onto the NPJ phase diagram during the 21-day period

480   following the initialization of each best- or worst-performing forecast, resulting in a series of

481   trajectories within the NPJ phase diagram. These trajectories are then shifted so that they all

482   begin at the origin of the NPJ phase diagram and the PCs corresponding to the same day after

483   forecast initialization are averaged together to construct a composite trajectory. Note that the

484   trajectories shown in Figs. 10c,d do not show forecast trajectories, but instead depict the how the

485   NPJ evolved in reality following a best- or worst-performing forecast.

486   The composite CFSR trajectories during the 21-day period following a best-performing

487   forecast from each model are clustered near the origin and exhibit a slight transition towards a jet

488   extension or poleward shift during the first few days of the forecast period (Fig. 10c). The 21-day

489   period following a worst-performing forecast, on the other hand, exhibits an opposite character

490   (Fig. 10d). Namely, the worst-performing forecast periods generally feature an NPJ that evolves

491   towards an equatorward shift or a jet retraction during the first half of the forecast period before

492   returning towards the origin. Given that the NPJ is already displaced towards a jet retraction or

493   equatorward shift at the time a worst-performing forecast is initialized (Fig. 10b), the NPJ

494   trajectories shown in Fig. 10d indicate that the NPJ amplifies its projection onto these two NPJ

495   regimes during the subsequent 21-day period.

496   Figure 10e shows a composite of the ensemble mean NPJ phase diagram forecast

497   trajectory associated with a worst-performing forecast from each model. Overall, each model's

498   forecast trajectory exhibits errors in the forecasted NPJ regime transition and/or in the amplitude

21

499   of a particular NPJ regime (cf., Figs. 10d,e). Furthermore, the forecast trajectories (Fig. 10e) are

500   more biased towards a jet extension and poleward shift at the end of the 21-day forecast period

501   than observations (Fig. 10d), which suggests the models may be too quick to transition the jet out

502   of a jet retraction or equatorward shift regime. Given that both jet retractions and equatorward

503   shifts feature upper-tropospheric ridging over the North Pacific (Figs. 1c,g), these differences

504   between the forecast and observed trajectories indicate that forecast errors may be related to each

505   model's representation of physical processes that govern the extent and duration of North Pacific

506   flow amplification. These physical processes can include the magnitude of diabatic heating and

507   upper-level irrotational outflow associated with midlatitude cyclogenesis events along the North

508   Pacific storm track (e.g., Torn & Hakim 2015; Teubler & Riemer 2016; Martinez-Alvarado et

509   al., 2016; Bosart et al., 2017).

510   The synoptic-scale flow patterns associated with the worst-performing forecasts from

511   each model are examined further by compositing CFSR mass and wind fields 0 days (Fig. 11),

512   10 days (Fig. 12), and 20 days (Fig. 13) following the initialization of a worst-performing

513   forecast. At the time of forecast initialization, every model features some degree of anomalous

514   upper-tropospheric ridging over the central North Pacific (Fig. 11). For some models, such as the

515   ECCC, CNRM, KMA, UKMO, ECMWF, and CMA (Figs. 11a,b,d,f,h,i), the North Pacific ridge

516   is more anomalous, suggesting that the worst-performing forecasts for those models may be

517   preferentially initialized during or immediately following ridge amplification rather than prior to

518   ridge amplification. Ten days after forecast initialization, the synoptic-scale flow pattern features

519   a well-developed upper-tropospheric ridge across the high-latitude North Pacific within each

520   model (Fig. 12). The presence of a high-latitude ridge is consistent with both a jet retraction and

521   equatorward shift regime (Figs. 1c,g), which are the same NPJ regimes that are generally

22

522   characterized by the greatest forecast errors at the time of forecast verification during the week

523   1–2 forecast period (Figs. 8b,d).

524   Twenty days after the initialization of a worst-performing forecast, the composite upper-

525   tropospheric flow patterns feature considerable differences across models (Fig. 13). In particular,

526   the ECCC, CNRM, NCEP, UKMO, ECMWF, and CMA models (Figs. 13a,b,e,f,h,i) continue to

527   feature an amplified upper-tropospheric ridge over the North Pacific, albeit slightly farther west

528   than observed in Fig. 12 in some cases. Conversely, the composite flow patterns following the

529   worst-performing forecasts from the ISAC, KMA, and HMCR models (Figs. 13c,d,g) indicate

530   that the upper-tropospheric ridge over the central North Pacific decays more rapidly than in the

531   other models 20 days after forecast initialization. All models also exhibit considerable

532   differences with respect to the character of the resultant flow pattern over North America. To

533   synthesize the composite evolutions shown in Figs. 11–13, the largest NPJ phase diagram

534   forecast errors from each model are clearly associated with North Pacific ridge amplification

535   during the week 1–2 period. After that, the variable synoptic-scale flow patterns that prevail 20

536   days after forecast initialization imply that aspects of the life cycle of North Pacific ridges, such

537   as their persistence, retrogression, and decay, may hinder model performance.

538   Last, Winters et al. (2019b, their Fig. 13) demonstrate that periods in which the NPJ

539   evolves towards an equatorward shift (similar to those trajectories shown in Fig. 10d) increase

540   the likelihood of extreme cold events across the continental U.S. Indeed, the composite upper-

541   tropospheric flow pattern 10 days after a worst-performing forecast from each model features an

542   anomalous trough over central Canada (Fig. 12). The longitudinal juxtaposition of a high-latitude

543   ridge over the North Pacific and trough over central Canada subsequently favors the

544   development of an anomalous surface anticyclone across Alaska and western Canada in the

545   aforementioned composites (Fig. 14). To the east of this anticyclone, perturbation northerly

546   geostrophic flow is conducive to the equatorward transport of anomalously cold air towards

547   southern Canada and the northern U.S. Therefore, the composite lower-tropospheric temperature

548   patterns following a worst-performing forecast suggest that the worst-performing forecasts may

549   coincide with the occurrence of North American cold-air outbreaks during the week 2 period,

550   potentially limiting the prediction of those events.

551

552   **5. Conclusions**

553   This study examines the subseasonal predictability of the state and evolution of the NPJ

554   across nine models within the S2S Reforecast Database hosted by ECMWF (Vitart et al., 2017).

555   The state and evolution of the NPJ is specifically examined in the context of an NPJ phase

556   diagram (Winters et al., 2019a), which identifies periods during which the NPJ is characterized

557   by an extended or retracted state, and during which the NPJ is poleward or equatorward shifted

558   relative to its climatological position. 300-hPa zonal wind anomaly data from the S2S Reforecast

559   Database are then projected onto the NPJ phase diagram to construct ensemble forecasts

560   describing the state and evolution of the NPJ at subseasonal time scales. NPJ phase diagram

561   forecasts are evaluated by considering the Euclidean distance between the forecast position of

562   the NPJ within the NPJ phase diagram at a particular lead time and the verification position of

563   the NPJ in the CFSR. Forecasts are also partitioned based on whether a forecast is initialized or

564   verified during a particular NPJ regime to determine the extent to which verification statistics

565   vary depending on those metrics. Last, the best- and worst-performing forecasts associated with

566   each model are identified to examine the synoptic-scale flow evolution that characterizes the 21-

567   day period following a best- or worst-performing forecast from each model.

568   An evaluation of NPJ phase diagram forecasts reveals that skillful predictions of the state

569   and evolution of the NPJ can extend into the week 3 forecast period, with the ECMWF model

570   featuring the lowest forecast errors among all models at every forecast lead time. The fact that

571   the skill of NPJ phase diagram forecasts extends into the week 3 period is consistent with prior

572   work on North American weather regimes, which suggest that skillful predictions are generally

573   possible at lead times of 15 days (e.g., Vigaud et al., 2018; Robertson et al., 2020). NPJ phase

574   diagram forecasts of the verifying NPJ regime from the three largest ensembles considered as

575   part of this study (i.e., ECMWF, CNRM, HMCR) are also generally reliable at forecast lead

576   times extending into weeks 2–3, but are uniformly underdispersive, and thus overconfident in the

577   development of a particular NPJ regime.

578   Forecast errors in the context of the NPJ phase diagram vary depending on the NPJ

579   regime at the time of forecast initialization during the first two weeks of the forecast period.

580   Thereafter, forecast errors do not show much dependence on the initial NPJ regime as the model

581   forecast is further removed from knowledge of atmospheric initial conditions. Overall, forecasts

582   initialized during a jet retraction feature 7-day forecast errors that are 10–20% larger than all

583   forecasts that are initialized during one of the four primary NPJ regimes, whereas forecasts

584   initialized during a poleward shift feature forecast errors that are 5–15% smaller. Forecasts

585   verifying during jet retractions and equatorward shifts also exhibit larger errors during the first

586   two weeks of the forecast period compared to forecasts verifying during jet extensions and

587   poleward shifts. Notably, both jet retractions and equatorward shifts are associated with the

588   development of an upper-tropospheric North Pacific ridge, which can be strongly influenced by

589   diabatic processes that occur within midlatitude cyclones along the Pacific storm track or in

590   conjunction with tropical convection (e.g., Torn & Hakim, 2015; Teubler & Riemer, 2016;

25

591   Martinez-Alvarado et al., 2016; Bosart et al., 2017; Breeden et al., 2020). The inability for

592   models to represent the extent, magnitude, and cumulative influence of these diabatic processes

593   on the upper-tropospheric flow pattern is hypothesized to contribute to the larger-than-average

594   forecast errors associated with jet retractions and equatorward shifts during the first two weeks of

595   the forecast period.

596   At lead times longer than two weeks, forecast errors associated with each NPJ regime

597   appear to be strongly influenced by biases in each model's representation of the jet at

598   subseasonal lead times. Namely, NPJ regimes that were characterized by a low forecast

599   frequency bias at subseasonal lead times within a particular model were often the same NPJ

600   regimes that were associated with the largest forecast errors at the time of verification for that

601   model. The use of bias-corrected forecasts resolved these forecast frequency biases and resulted

602   in stronger agreement between forecast model errors at subseasonal lead times. Namely,

603   forecasts verifying during jet retractions or equatorward shifts were generally associated with the

604   largest forecast errors at subseasonal lead times. These results indicate that bias-corrected NPJ

605   phase diagram forecasts have the potential to identify periods that may exhibit enhanced skill

606   compared to each model's climatology at subseasonal lead times based on the anticipated NPJ

607   evolution.

608   The best-performing forecasts associated with each model occurred during periods in

609   which the NPJ featured a slight poleward shift, whereas the worst-performing forecasts featured

610   an NPJ that evolved towards a jet retraction or equatorward shift. Composites of the 21-day

611   period following the initiation of a worst-performing forecast from each model indicated that the

612   largest NPJ forecast errors coincided with the development of an upper-tropospheric North

613   Pacific ridge during the first 10 days after forecast initialization and the subsequent maintenance,

614 retrogression, or decay of that ridge over the next 10 days. This result generalizes the analysis

615 from Winters et al. (2019a), who found a similar flow pattern was associated with the worst-

616 performing forecasts on medium-range time scales in the GEFS Reforecast Version 2 dataset,

617 and reaffirms that the life cycle of upper-tropospheric blocks remains a considerable

618 predictability challenge at subseasonal lead times (e.g., D'Andrea et al., 1998; Pelly & Hoskins,

619 2003; Ferranti et al., 2015; Matsueda & Palmer, 2018).

620   The results from this study motivate new avenues for future work. First, differences in the

621 forecast frequency of NPJ regimes at lead times exceeding two weeks within the baseline

622 forecasts from each model motivate further investigation into each model's representation of

623 physical processes that lead to the development of each NPJ regime (i.e., diabatic heating from

624 midlatitude and tropical sources and its subsequent influence on the character of the upper-

625 tropospheric flow pattern). Second, the present results do not consider the extent to which

626 forecast errors associated with each NPJ regime translate to forecast errors over the North

627 American continent. Therefore, a study that considers the relationship between the prevailing

628 NPJ regime and downstream forecast errors would be a worthwhile endeavor. Finally, North

629 American weather is also influenced by the state and evolution of the synoptic-scale flow pattern

630 over the North Atlantic. A similar approach as used in this study can be applied to the North

631 Atlantic jet to examine the ability of models to accurately capture the state and evolution of that

632 jet.

633

634 **Acknowledgments**

637   work. The author is also appreciative to the Editor and two anonymous reviewers, whose

638   constructive comments greatly improved the manuscript. Support for the author to perform this

639   work was provided by start-up funding furnished by the University of Colorado Boulder.

640

641   **Data Availability Statement**

642   CFSR data utilized for this study is publicly available from the NCAR Research Data Archive

643   (https://doi.org/10.5065/D69K487J). This work is based on S2S Reforecast data available from

644   ECMWF. S2S is a joint initiative of the World Weather Research Programme (WWRP) and the

645   World Climate Research Programme (WCRP). The original S2S database is hosted at ECMWF

646   as an extension of the TIGGE database (Vitart et al., 2017). A database of NPJ phase diagram

647   forecasts derived from the S2S Reforecast Database is archived at the University of Colorado

648   Boulder (Winters, 2021; https://scholar.colorado.edu/concern/datasets/0v838153k) Any

649   computer programs necessary to reproduce the results shown in this study are available from the

650   author upon request.

651

652

653

654

655

656

657

658

659

660   **References**

661   Amini, S., & Straus, D. M. (2019). Control of storminess over the Pacific and North America by

662   circulation regimes. *Climate Dynamics, 52,* 4749–4770. https://doi.org/10.1007/s00382-

663   018-4409-7.

664   Archambault, H. M., Keyser, D., Bosart, L. F., Davis, C. A., & Cordeira, J. M. (2015). A

665   composite perspective of the extratropical flow response to recurving western North

666   Pacific tropical cyclones. *Monthly Weather Review, 143,* 1122–1141.

667   https://doi.org/10.1175/ MWR-D-14-00270.1.

668   Athanasiadis, P. J., Wallace, J. M., & Wettstein, J. J. (2010). Patterns of wintertime jet stream

669   variability and their relation to the storm tracks. *Journal of the Atmospheric Sciences, 67,*

670   1361–1381. https://doi.org/10.1175/2009JAS3270.1.

671   Bosart, L. F., Moore, B. J., Cordeira, J. M., & Archambault, H. M. (2017). Interactions of North

672   Pacific tropical, midlatitude, and polar disturbances resulting in linked extreme weather

673   events over North America in October 2007. *Monthly Weather Review, 145,* 1245–1273.

674   https://doi.org/10.1175/MWR-D-16-0230.1.

675   Breeden, M.L., Hoover, B. T., Newman, M., & Vimont, D. J. (2020). Optimal North Pacific

676   blocking precursors and their deterministic subseasonal evolution during boreal winter.

677   *Monthly Weather Review, 148,* 739–761. https://doi.org/10.1175/MWR-D-19-0273.1.

678   Cassou, C. (2008). Intraseasonal interaction between the Madden–Julian Oscillation and the

679   North Atlantic Oscillation. *Nature, 455,* 523–527. https://doi.org/10.1038/nature07286.

680   Cordeira, J. M., & Bosart, L. F. (2010). The antecedent large-scale conditions of the ''Perfect

681   Storms'' of late October and early November 1991. *Monthly Weather Review, 138,*

682   2546–2569. https://doi.org/10.1175/2010MWR3280.1.

683   Corti, S., Molteni, F. & Palmer, T. N. (1999). Signature of recent climate change in frequencies

684   of natural atmospheric circulations. *Nature*, *398,* 799–802. https://doi.org/10.1038/19745.

685   D'Andrea, F., Tibaldi, S., Blackburn, M., Boer, G., Déqué, M., Dix, M. R. (1998). Northern

686   Hemisphere atmospheric blocking as simulated by 15 atmospheric general circulation

687   models in the period 1979–1988. *Climate Dynamics, 14,* 385–407.

688   https://doi.org/10.1007/s003820050230.

689   Dawson, A., & Palmer, T. N. (2014). Simulating weather regimes: Impact of model resolution

690   and stochastic parameterization. *Climate Dynamics, 44,* 2177–2193.

691   https://doi.org/10.1007/s00382-014-2238-x.

692   Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al., (2011).

693   The ERA-Interim reanalysis: Configuration and performance of the data assimilation

694   system. *Quarterly Journal of the Royal Meteorological Society*, *137,* 553–597.

695   http://doi.org/10.1002/qj.828.

696   Domeisen, D. I. V. & Butler, A. H., (2020). Stratospheric drivers of extreme events at the Earth's

697   surface. *Communications Earth & Environment.*, *1*, 59, http://doi.org/10.1038/s43247-

698   020-00060-z.

699   Feng, P.-N., Lin, H., Derome, J., & Merlis, T. M. (2021). Forecast skill of the NAO in the

700   subseasonal-to-seasonal prediction models. *Journal of Climate, 34,* 4757–4769.

701   https://doi.org/ 10.1175/JCLI-D-20-0430.1.

702   Ferranti, L., Corti, S., & Janousek, M. (2015). Flow-dependent verification of the ECMWF

703   ensemble over the Euro-Atlantic sector. *Quarterly Journal of the Royal Meteorological*

704   *Society, 141,* 916–924. https://doi.org/10.1002/qj.2411.

705   Ferranti, L., Magnusson, L., Vitart, F., & Richardson, D. S. (2018). How far in advance can we

706   predict changes in large-scale flow leading to severe cold conditions over Europe?

707   *Quarterly Journal of the Royal Meteorological Society, 144,* 1788–1802.

708   https://doi.org/10.1002/qj.3341.

709   Grams, C. M., Beerli, R., Pfenninger, S., Staffell, I., & Wernli, H. (2017). Balancing Europe's

710   wind-power output through spatial deployment informed by weather regimes. *Nature*

711   *Climate Change, 7,* 557–562. https://doi.org/10.1038/nclimate3338.

712   Griffin, K. S., & Martin, J. E. (2017). Synoptic features associated with temporally coherent

713   modes of variability of the North Pacific jet stream. *Journal of Climate, 30,* 39–54.

714   https://doi.org/10.1175/JCLI-D-15-0833.1.

715   Hamill, T. M., Bates, G. T., Whitaker, J. S., Murray, D. R., Fiorino, M., Galarneau Jr., T. J., Zhu,

716   Y., & Lapenta, W. (2013). NOAA's second-generation global medium-range ensemble

717   reforecast dataset. *Bulletin of the American Meteorological Society, 94,* 1553–1565.

718   https://doi.org/10.1175/BAMS-D-12-00014.1.

719   Jaffe, S. C., Martin, J. E., Vimont, D. J., & Lorenz, D. J. (2011). A synoptic climatology of

720   episodic, subseasonal retractions of the Pacific jet. *Journal of Climate, 24,* 2846–2860.

721   https://doi.org/10.1175/2010JCLI3995.1.

722   Jung, T., Vitart F., Ferranti, L., & Morcrette, J.-J. (2011). Origin and predictability of the

723   extreme negative NAO winter of 2009/10. *Geophysical Research Letters, 38,* L07701.

724   https://doi.org/10.1029/2011GL046786.

725   Kimoto, M., & Ghil, M. (1993). Multiple flow regimes in the Northern Hemisphere winter. Part

726   I: Methodology and hemispheric regimes. *Journal of the Atmospheric Sciences, 50,*

727   2625–2644. https://doi.org/10.1175/1520-0469(1993)050<2625:MFRITN>2.0.CO;2.

728   Lee., S. H., Furtado, J. C., & Charlton-Perez, A. J. (2019). Wintertime North American weather

729   regimes and the Arctic stratospheric polar vortex. *Geophysical Research Letters*, *46*,

730   14892–14900. https://doi.org/10.1029/2019GL085592.

731   Martínez-Alvarado, O., Gray, S. L., & Methven, J. (2016). Diabatic processes and the evolution

732   of two contrasting summer extratropical cyclones. *Monthly Weather Review, 144,* 3251–

733   3276. https://doi.org/10.1175/MWR-D-15-0395.1.

734   Matsueda, M., & Kyouda, M. (2016). Wintertime east Asian flow patterns and their

735   predictability on medium-range timescales. *Scientific Online Letters on the Atmosphere,*

736   *12,* 121–126. https://doi.org/10.2151/sola.2016-027.

737   Matsueda, M., & Palmer, T. N. (2018). Estimates of flow-dependent predictability of wintertime

738   Euro-Atlantic weather regimes in medium-range forecasts. *Quarterly Journal of the*

739   *Royal Meteorological Society, 144,* 1012–1027. https://doi.org/10.1002/qj.3265.

740   Meehl, G. A., Richter, J. H., Teng, H., Capotondi, A., Cobb, K., Doblas-Reyes, F., et al.

741   (2021). Initialized earth system prediction from subseasonal to decadal timescales.

742   *Nature Reviews Earth & Environment, 2,* 340–357. https://doi.org/10.1038/s43017-021-

743   00155-x.

744   Michelangeli, P.-A., Vautard, R., & Legras, B. (1995). Weather regimes: Recurrence and quasi

745   stationarity. *Journal of the Atmospheric Sciences, 52,* 1237–1256.

746   https://doi.org/10.1175/1520-0469(1995)052<1237:WRRAQS>2.0.CO;2.

747   Mo, K., & Ghil, M. (1988). Cluster analysis of multiple planetary flow regimes. *Journal of*

748   *Geophysical Research: Atmospheres, 93,* 10927–10952.

749   https://doi.org/10.1029/JD093iD09p10927.

750   National Academies of Sciences, Engineering, and Medicine [NAS]. (2018). *Thriving on Our*

751   *Changing Planet: A Decadal Strategy for Earth Observation from Space.* Washington,

752   DC: National Academies Press. https://doi.org/10.17226/24938.

753   National Research Council [NRC]. (2010). *Assessment of Intraseasonal to Interannual Climate*

754   *Prediction and Predictability.* Washington, DC: National Academies Press.

755   https://doi.org/10.17226/12878.

756   North, G. R., Bell, T. L., Cahalan, R. F., & Moeng, F. J. (1982). Sampling errors in the

757   estimation of empirical orthogonal functions. *Monthly Weather Review, 110,* 699–706.

758   https://doi.org/10.1175/1520-0493(1982)110<0699:SEITEO>2.0.CO;2.

759   Pegion, K., Kirtman, B. P., Becker, E., Collins, D. C., LaJoie, E., Burgman, R., et al. (2019). The

760   subseasonal experiment (SubX): A multimodel subseasonal prediction experiment.

761   *Bulletin of the American Meteorological Society*, *100,* 2043–2060.

762   https://doi.org/10.1175/BAMS-D-18-0270.1.

763   Pelly, J. L., & Hoskins, B. J. (2003). How well does the ECMWF Ensemble Prediction System

764   predict blocking? *Quarterly Journal of the Royal Meteorological Society, 129,* 1683–

765   1702. https://doi.org/10.1256/qj.01.173.

766   Quandt, L.-A., Keller, J. H., Martius, O., Pinto, J. G., & Jones, S. C. (2019). Ensemble sensitivity

767   analysis of the blocking system over Russia in summer 2010. *Monthly Weather Review,*

768   *147,* 657–675. https://doi.org/10.1175/MWR-D-18-0252.1.

769   Reinhold, B., & Pierrehumbert, R. (1982). Dynamics of weather regimes: Quasi-stationary

770   waves and blocking. *Monthly Weather Review, 110,* 1105–1145.

771   https://doi.org/10.1175/1520-0493(1982) 110,1105:DOWRQS.2.0.CO;2.

772   Riddle, E. E., Stoner, M., Johnson, N., L'Heureux, M. (2013). The impact of the MJO on clusters

773   of wintertime circulation anomalies over the North American region. *Climate Dynamics,*

774   *40,* 1749–1766. https://doi.org/10.1007/s00382-012-1493-y.

775   Robertson, A. W., & Ghil, M. (1999). Large-scale weather regimes and local climate over the

776   western United States. *Journal of Climate, 12,* 1796–1813. https://doi.org/10.1175/1520-

777   0442(1999)012<1796:LSWRAL>2.0.CO;2.

778   Robertson, A. W., Vigaud, N., Yuan, J., & Tippett, M. K. (2020). Toward identifying

779   subseasonal forecasts of opportunity using North American weather regimes. *Monthly*

780   *Weather Review, 148,* 1861–1875. https://doi.org/10.1175/MWR-D-19-0285.1.

781   Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., et al. (2010). The NCEP Climate

782   Forecast System Reanalysis. *Bulletin of the American Meteorological Society, 91,* 1015–

783   1057. https://doi.org/10.1175/2010BAMS3001.1.

784   Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., et al. (2014). The NCEP Climate

785   Forecast System version 2. *Journal of Climate*, *27*, 2185–2208.

786   https://doi.org/10.1175/JCLI-D-12-00823.1.

787   Straus, D. M., Corti, S., & Molteni, F. (2007). Circulation regimes: Chaotic variability versus

788   SST-forced predictability. *Journal of Climate, 20,* 2251–2272.

789   https://doi.org/10.1175/JCLI4070.1.

790   Straus, D. M., Molteni, F., & Corti, S. (2017). Atmospheric regimes: The link between weather

791   and the large-scale circulation. In C. L. E. Franzke and T. J. O'Kane (Eds.), *Nonlinear*

792   *and Stochastic Climate Dynamics* (pp. 105–135). Cambridge, UK: Cambridge University

793   Press.

794 Teubler, F., & Riemer, M. (2016). Dynamics of Rossby wave packets in a quantitative potential

795   vorticity–potential temperature framework. *Journal of the Atmospheric Sciences, 73,*

796   1063–1081. https://doi.org/10.1175/JAS-D-15-0162.1.

797 Torn, R. D., & Hakim, G. J. (2015). Comparison of wave packets associated with extratropical

798   transition and winter cyclones. *Monthly Weather Review, 143,* 1782–1803.

799   https://doi.org/10.1175/MWR-D-14-00006.1.

800 Turasky, E. (2019). *An examination of North Pacific jet regimes conducive to landfalling*

801   *atmospheric rivers along the west coast of North America* (M.S. Thesis). Retrieved from

802   ProQuest Dissertations and Theses

803   (https://search.proquest.com/docview/2283452686/BFB9D80CD67A438EPQ/2?accounti

804   d=14503). Albany, NY: University at Albany, SUNY.

805 Vautard, R. (1990). Multiple weather regimes over the North Atlantic: Analysis of precursors

806   and successors. *Monthly Weather Review, 118,* 2056–2081. https://doi.org/10.1175/1520-

807   0493(1990)118<2056:MWROTN>2.0.CO;2.

808 Vigaud, N., Robertson, A. W., & Tippett, M. K. (2018). Predictability of recurrent weather

809   regimes over North America during winter from submonthly reforecasts. *Monthly*

810   *Weather Review, 146,* 2559–2577. https://doi.org/10.1175/MWR-D18-0058.1.

811 Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., et al. (2017). The

812   subseasonal to seasonal (S2S) prediction project database. *Bulletin of the American*

813   *Meteorological Society, 98,* 163–173. https://doi.org/10.1175/BAMS-D-16-0017.1.

814 White, C. J., Carlsen, H., Robertson, A. W., Klein, R. J. T., Lazo, J. K., Kumar, A., et al. (2017).

815   Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteorological*

816   *Applications, 24,* 315–325. https://doi.org/10.1002/met.1654.

817 Wilks, D. S. (2011). *Statistical Methods in the Atmospheric Sciences* (3rd ed). Amsterdam,

818   Netherlands: Elsevier.

819 Winters, A. C., Keyser, D., & Bosart, L. F. (2019a) The development of the North Pacific Jet

820   Phase Diagram as an objective tool to monitor the state and forecast skill of the upper-

821   tropospheric flow pattern. *Weather and Forecasting, 34,* 199–219.

822   https://doi.org/10.1175/WAF-D-18-0106.1.

823 Winters, A. C., Bosart, L. F., & Keyser, D. (2019b) Antecedent North Pacific jet regimes

824   conducive to the development of continental U.S. extreme temperature events during the

825   cool season. *Weather and Forecasting, 34,* 393–414. https://doi.org/10.1175/WAF-D-18-

826   0168.1.

827 Winters, A. C. (2021) Subseasonal forecasts of the state and evolution of the North Pacific jet

828   stream [Data set]. University of Colorado Boulder.

829   https://scholar.colorado.edu/concern/datasets/0v838153k

830

831 **Tables**

Model	Horizontal Resolution	Reforecast Type	Model Version	Reforecast Period	Reforecast Frequency	Forecast Length	Ensemble Members
ECCC	1.5°×1.5°	On the fly	2019	1998–2017	Every 7 days	32 days	4
CNRM	1.5°×1.5°	Fixed	12/01/14	1993–2014	4 / month	61 days	15
ISAC	1.5°×1.5°	Fixed	06/08/17	1981–2010	Every 5 days	32 days	5
KMA	1.5°×1.5°	On the fly	2019	1991–2010	4 / month	60 days	3
NCEP	1.5°×1.5°	Fixed	03/01/11	1999–2010	Daily	44 days	4
UKMO	1.5°×1.5°	On the fly	2019	1993–2016	4 / month	60 days	7
HMCR	1.5°×1.5°	On the fly	2019	1985–2010	Every 7 days	61 days	10
ECMWF	1.5°×1.5°	On the fly	2019	1999–2018	2 / week	46 days	11
CMA	1.5°×1.5°	On the fly	2019–2020	2005–2018	2 / week	60 days	4

832

833 **TABLE 1.** Characteristics of the nine forecast models within the S2S Reforecast Database that
834 are utilized as part of this study.
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856

Figures
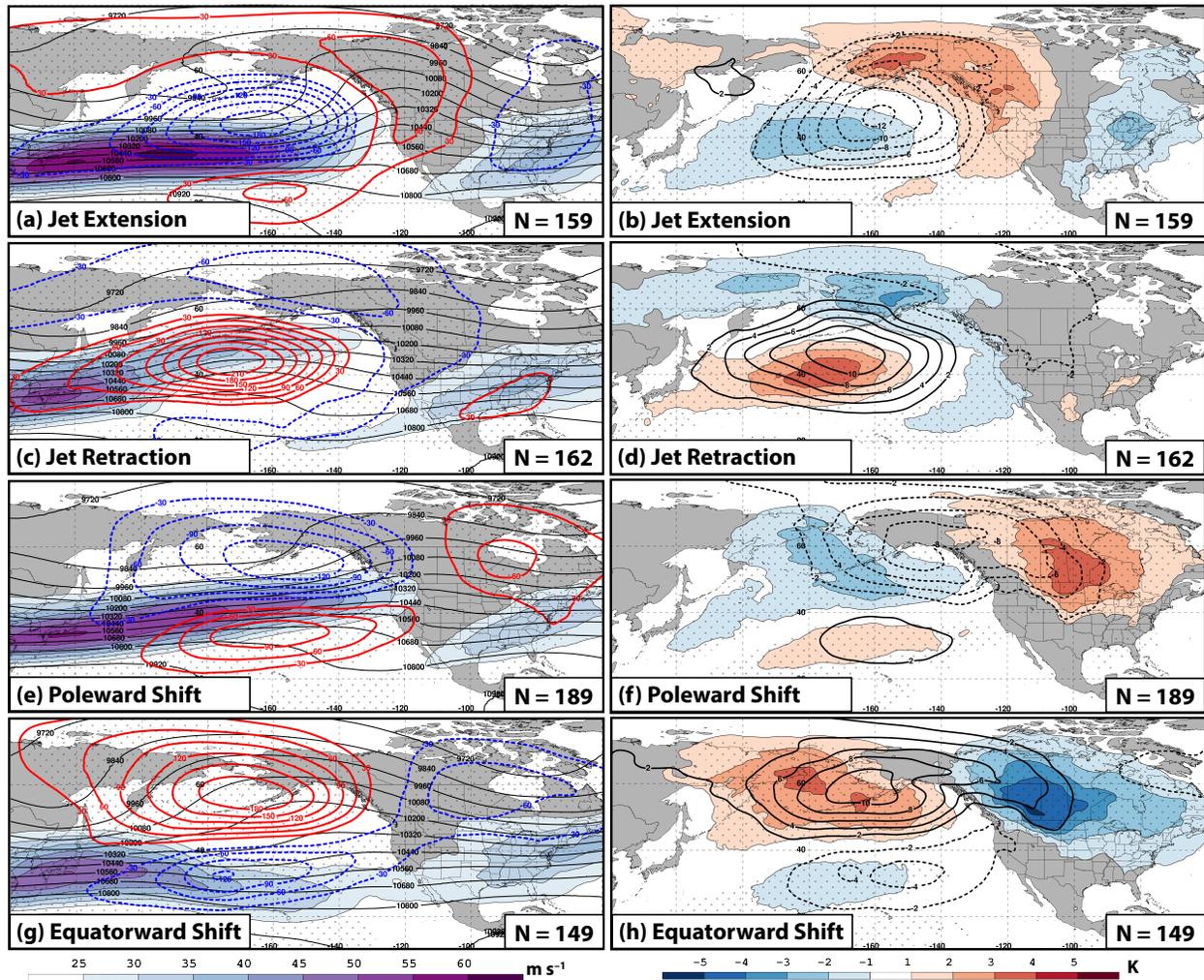
FIG. 1. Composite mean 250-hPa wind speed (shaded according to the fill pattern; m s^{-1}), 250-hPa geopotential height (contoured in black every 120 m), and 250-hPa geopotential height anomalies (contoured every 30 m in red where positive and in dashed blue where negative) 4 days following the initiation of (a) a jet extension, (c) a jet retraction, (e) a poleward shift, and (g) an equatorward shift NPJ regime. Composite anomalies of mean sea-level pressure (contoured every 2 hPa in solid black where positive and in dashed black where negative) and 850-hPa temperature (shaded according to the legend every 1 K) 4 days following the initiation of (b) a jet extension, (d) a jet retraction, (f) a poleward shift, and (h) an equatorward shift NPJ regime. The numbers in the bottom right of each panel indicate the number of cases included in each composite. Stippled areas represent locations where the 250-hPa geopotential height anomalies or 850-hPa temperature anomalies are statistically distinct from climatology at the 99% confidence level based on a two-sided Student's *t* test. Figure and caption adapted from Winters et al. (2019a; their Fig. 5). © American Meteorological Society. Used with permission.
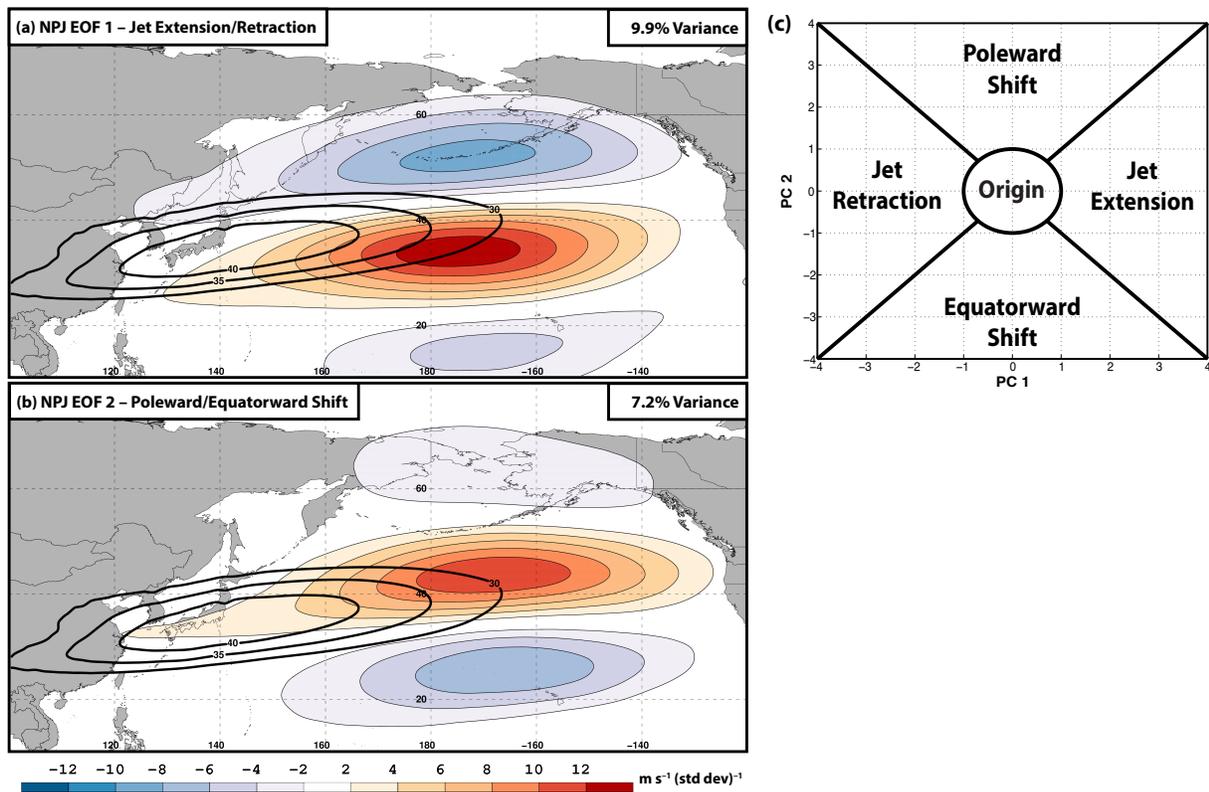
877 **FIG. 2.** September–May 300-hPa mean zonal wind is contoured in black every 5 m s^{-1} above 30
878 m s^{-1}, and the regression of 300-hPa zonal wind anomaly data onto standardized PC 1 (i.e., EOF
879 1) is shaded. The variance of 300-hPa zonal wind anomalies during September–May that is
880 explained by EOF 1 is listed in the top right of the panel. (b) As in (a), but for the regression of
881 300-hPa zonal wind anomaly data onto standardized PC 2 (i.e., EOF 2). (c) Schematic depicting
882 the NPJ phase diagram and the method used to classify the NPJ into an NPJ regime.

883
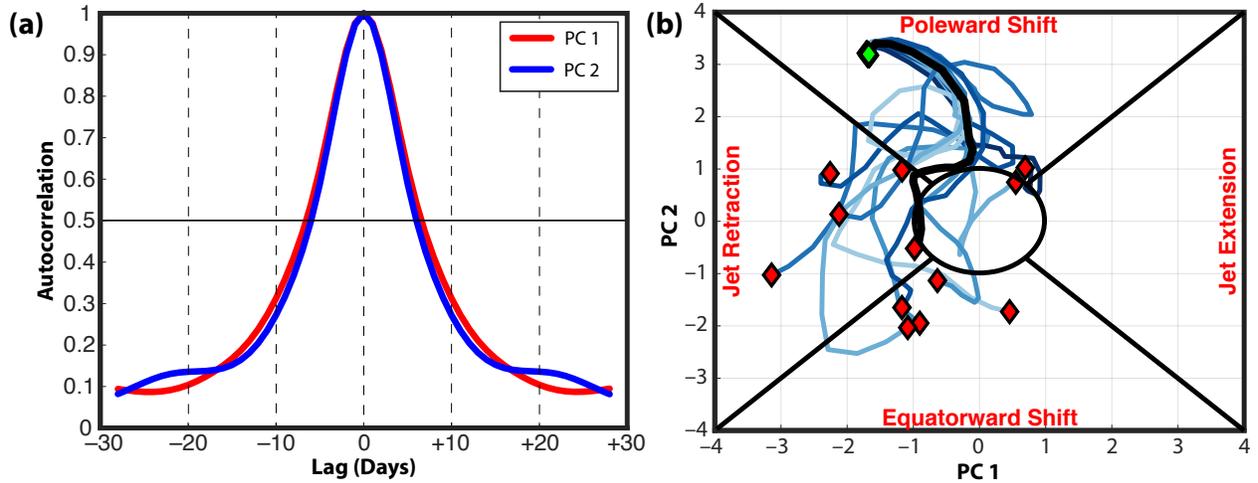
FIG. 3. (a) Autocorrelation functions for PC 1 and PC 2 that are derived from an EOF analysis of 300-hPa zonal wind anomalies over the North Pacific during September–May within the CFSR. The thin horizontal black line corresponds to an autocorrelation of 0.5. (b) A sample 21-day NPJ phase diagram ensemble forecast initialized at 0000 UTC 4 February 1999. Blue lines correspond to individual ensemble member forecasts and the thick black line corresponds to the ensemble mean forecast. The green diamond identifies the state of the NPJ at the time of forecast initialization and red diamonds identify the state of the NPJ at the end of the 21-day forecast period for each ensemble member and the ensemble mean.

895
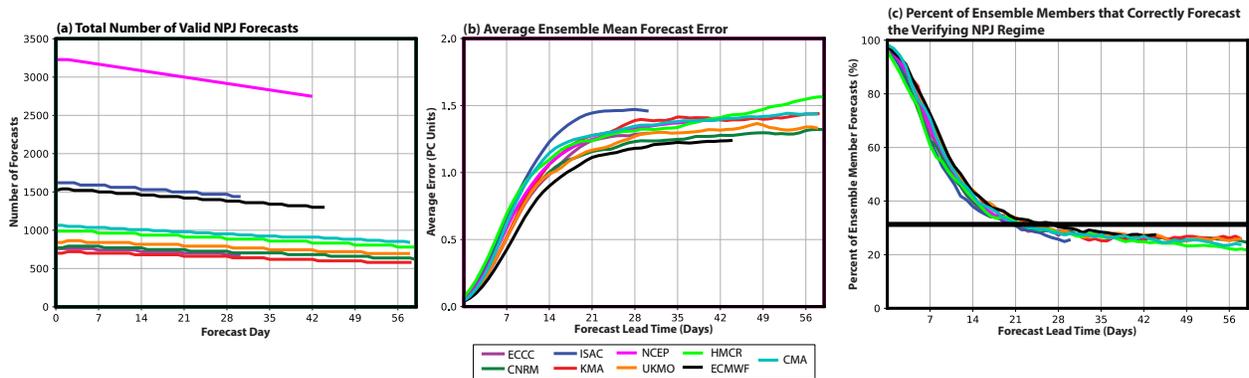
896
897
898 **FIG. 4.** (a) The total number of valid NPJ phase diagram forecasts initialized by each model at
899 each forecast lead time. (b) The average Euclidean distance error (in principal component (PC)
900 units) of ensemble mean NPJ phase diagram forecasts from each model as a function of forecast
901 lead time. (c) The percent of ensemble member forecasts initialized from each model that
902 correctly forecasted the verifying NPJ regime as a function of forecast lead time. The horizontal
903 black bar identifies percentages that are statistically significant at the 99% confidence interval
904 compared to random chance based on a bootstrap resampling test with replacement.
905
906
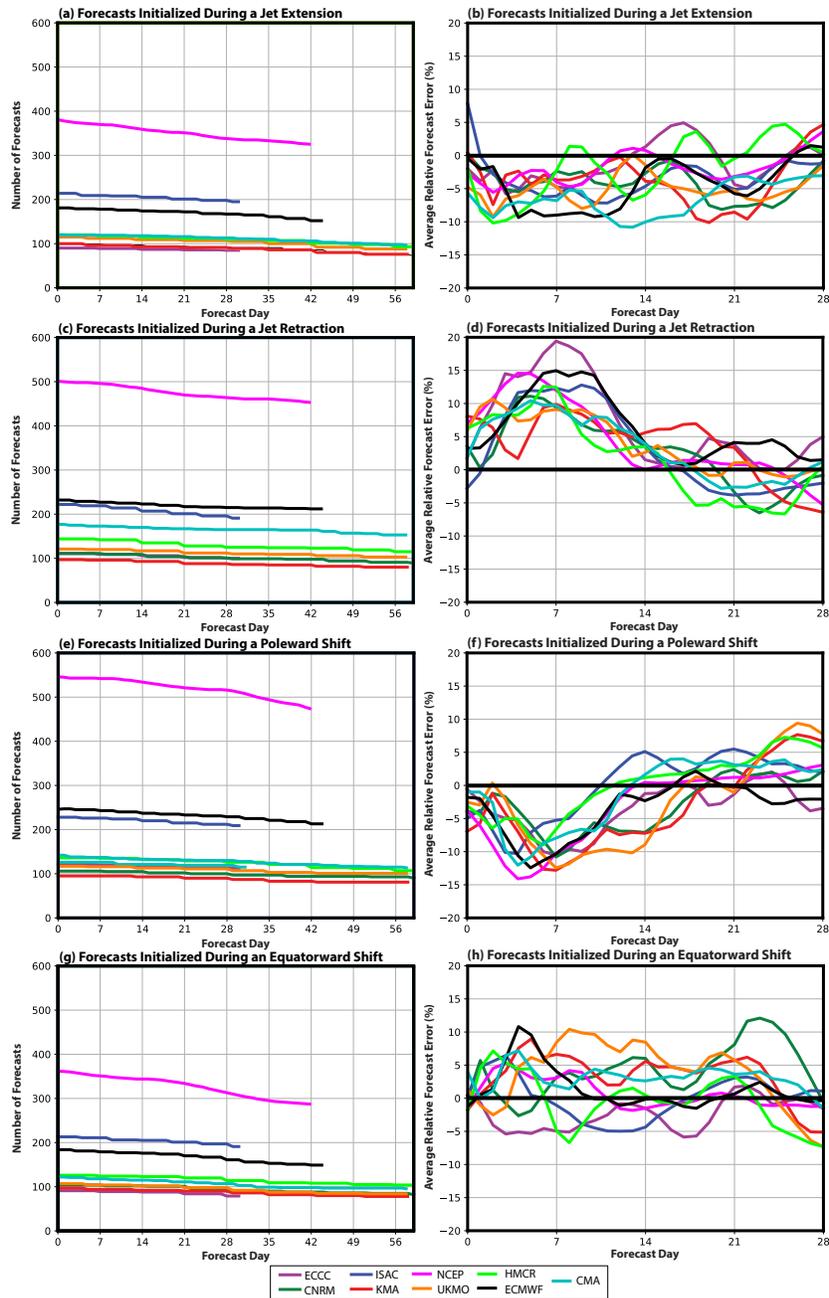
FIG. 5. The number of NPJ phase diagram forecasts from each model that were initialized during (a) a jet extension, (c) a jet retraction, (e) a poleward shift, and (g) an equatorward shift as a function of forecast lead time. The average Euclidean distance error of ensemble mean NPJ phase diagram forecasts from each model that were initialized during (b) a jet extension, (d) a jet retraction, (f) a poleward shift, and (h) an equatorward shift. All forecast model errors in (b,d,f,h) are expressed as a percentage greater or less than the average error of all NPJ phase diagram forecasts from that model that were initialized within one of the four primary NPJ regimes.
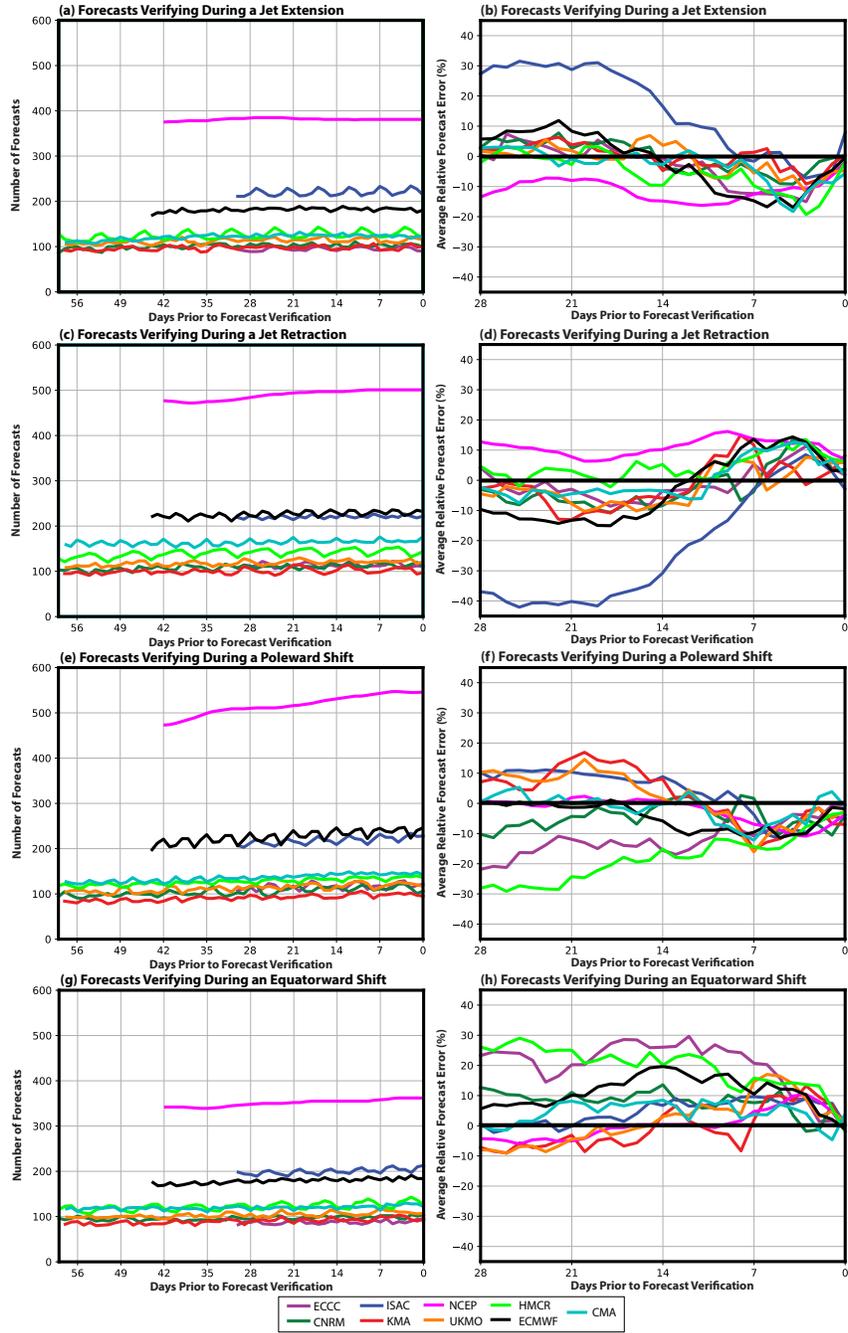
42

FIG. 6. The number of NPJ phase diagram forecasts from each model that verified during (a) a jet extension, (c) a jet retraction, (e) a poleward shift, and (g) an equatorward shift as a function of forecast lead time prior to verification. The average Euclidean distance error of ensemble mean NPJ phase diagram forecasts from each model that verified during (b) a jet extension, (d) a jet retraction, (f) a poleward shift, and (h) an equatorward shift. All forecast model errors in (b,d,f,h) are expressed as a percentage greater or less than the average error of all NPJ phase diagram forecasts from that model that verified within one of the four primary NPJ regimes.
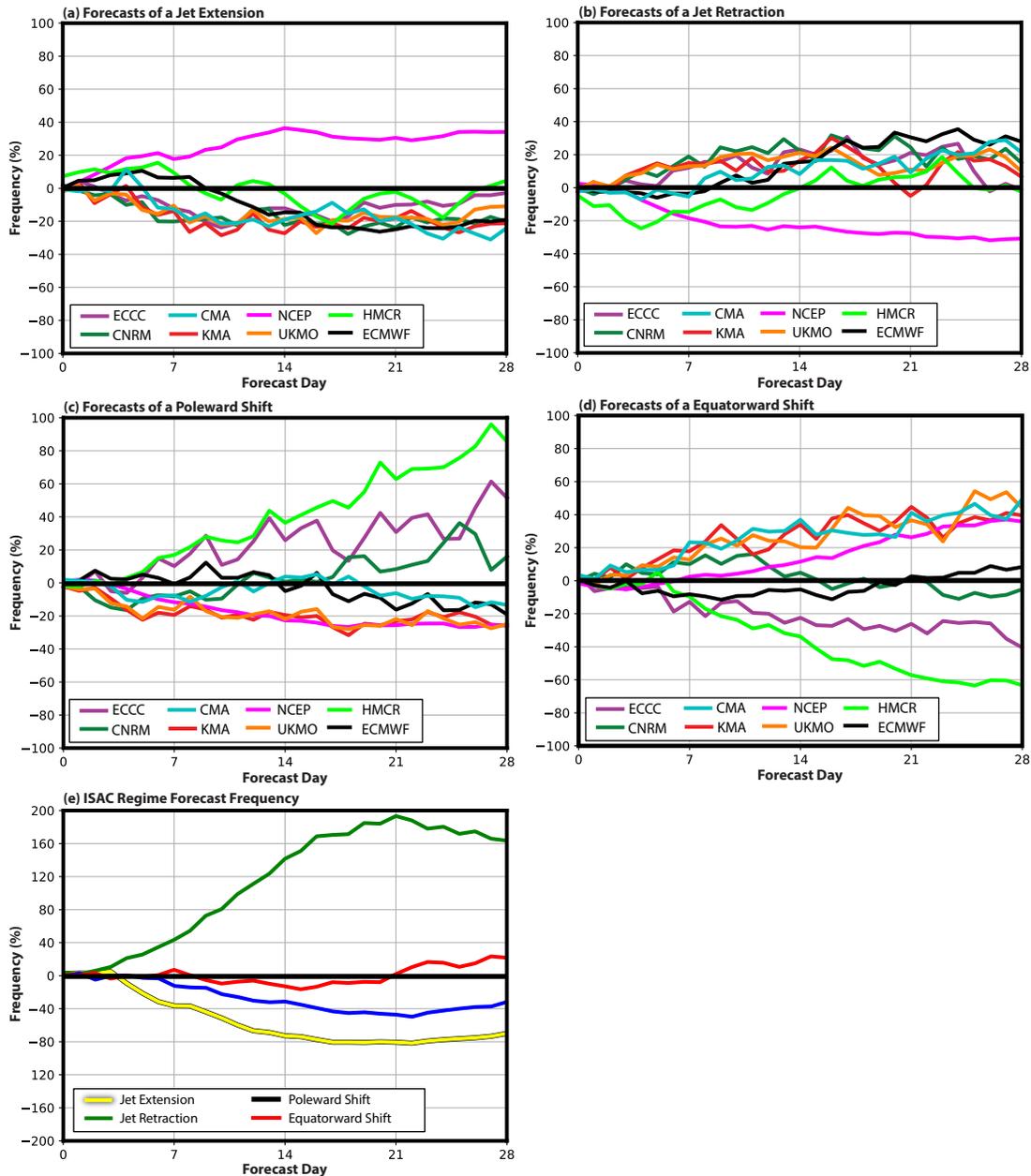
FIG. 7. The percent frequency that (a) a jet extension, (b) a jet retraction, (c) a poleward shift, and (d) an equatorward shift is overforecast (positive percentages) or underforecast (negative percentages) by ensemble member NPJ phase diagram forecasts from each model relative to verification at every forecast lead time. (e) The percent frequency that each NPJ regime is overforecast or underforecast relative to verification at each forecast lead time for the ISAC model.

(a) Forecasts Verifying During a Jet Extension

(b) Forecasts Verifying During a Jet Retraction

(c) Forecasts Verifying During a Poleward Shift

(d) Forecasts Verifying During an Equatorward Shift

Legend: ECCC, CNRM, ISAC, KMA, NCEP, UKMO, HMCR, ECMWF, CMA
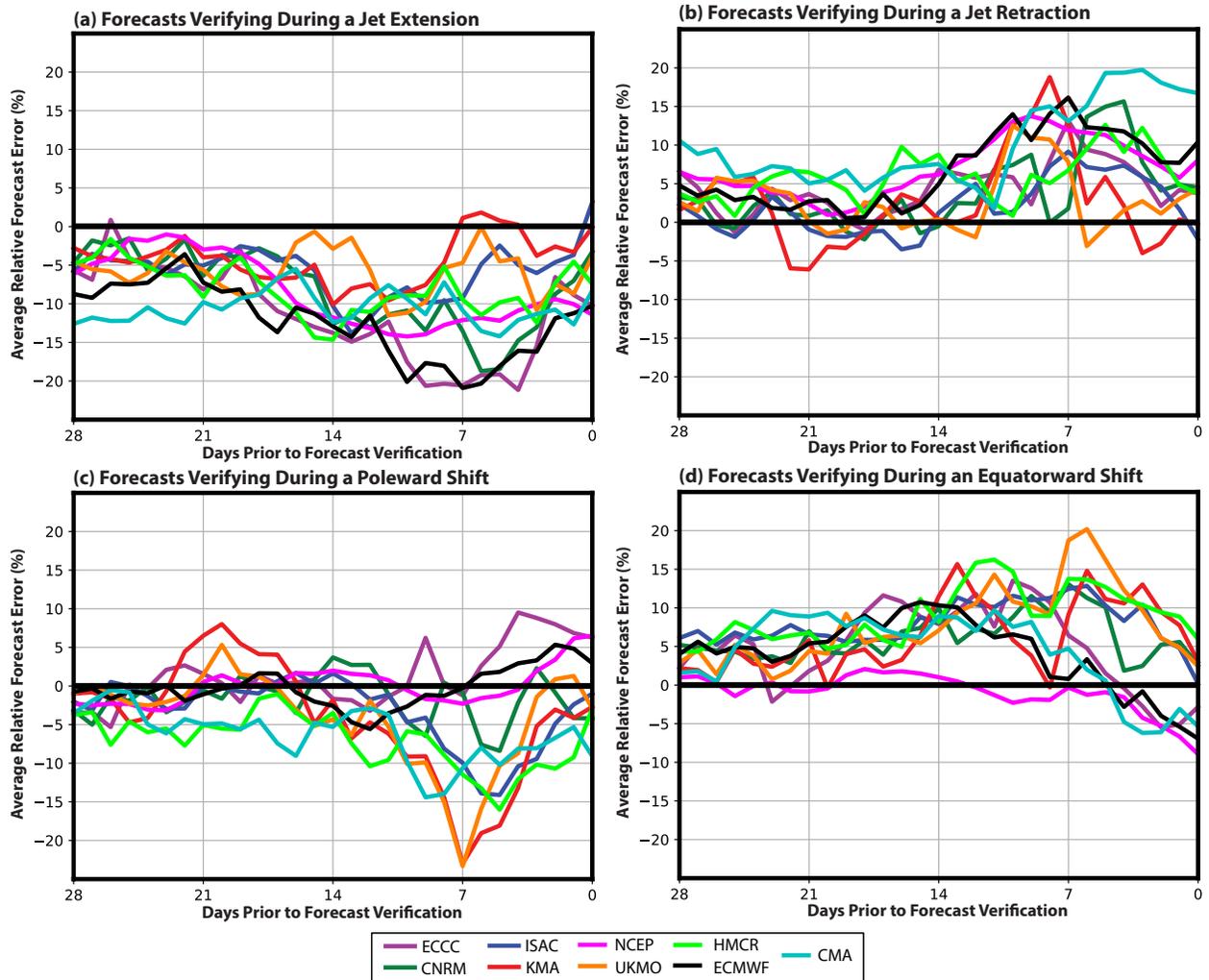
972
973
974 **FIG. 8.** As in Figs. 6b,d,f,h, but showing the errors associated with bias-corrected NPJ phase
975 diagram forecasts from each model as a function of forecast lead time for forecasts that verify
976 during (a) a jet extension, (b) a jet retraction, (c) a poleward shift, and (d) an equatorward shift.
977
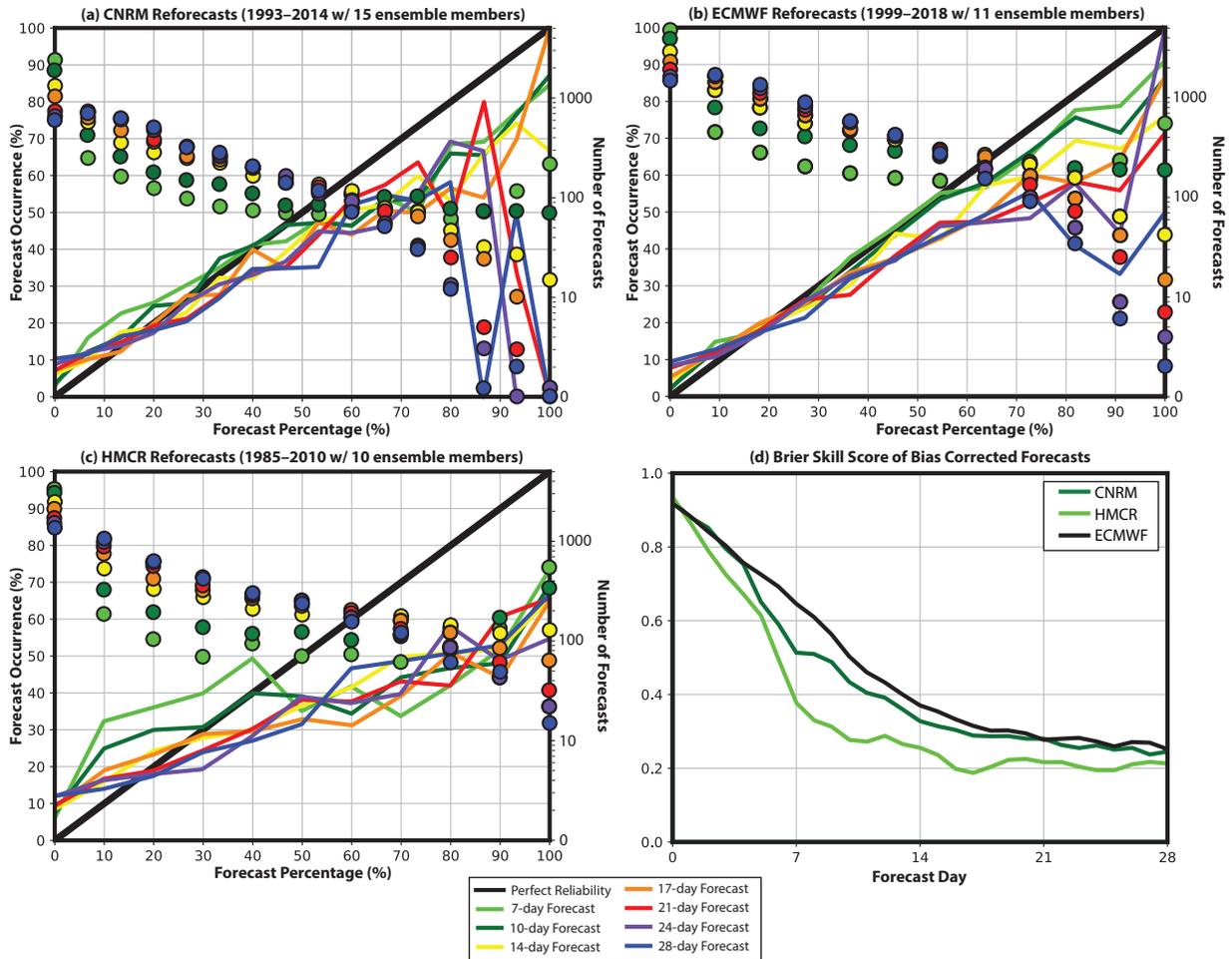
FIG. 9. Reliability diagrams at a variety of forecast lead times for the (a) CNRM, (b) ECMWF, and (c) HMCR bias-corrected ensembles. Shown in these diagrams are the probability that a particular NPJ regime is forecast to occur at a given lead time versus the percent of time that the forecasted NPJ regime verified. The thick black line represents a perfectly reliable forecast, and the colored dots show the number of forecasts within each probabilistic bin on a log scale as a function of forecast lead time. (d) The Brier Skill Scores associated with CNRM, HMCR, and ECMWF probabilistic forecasts as a function of forecast lead time.
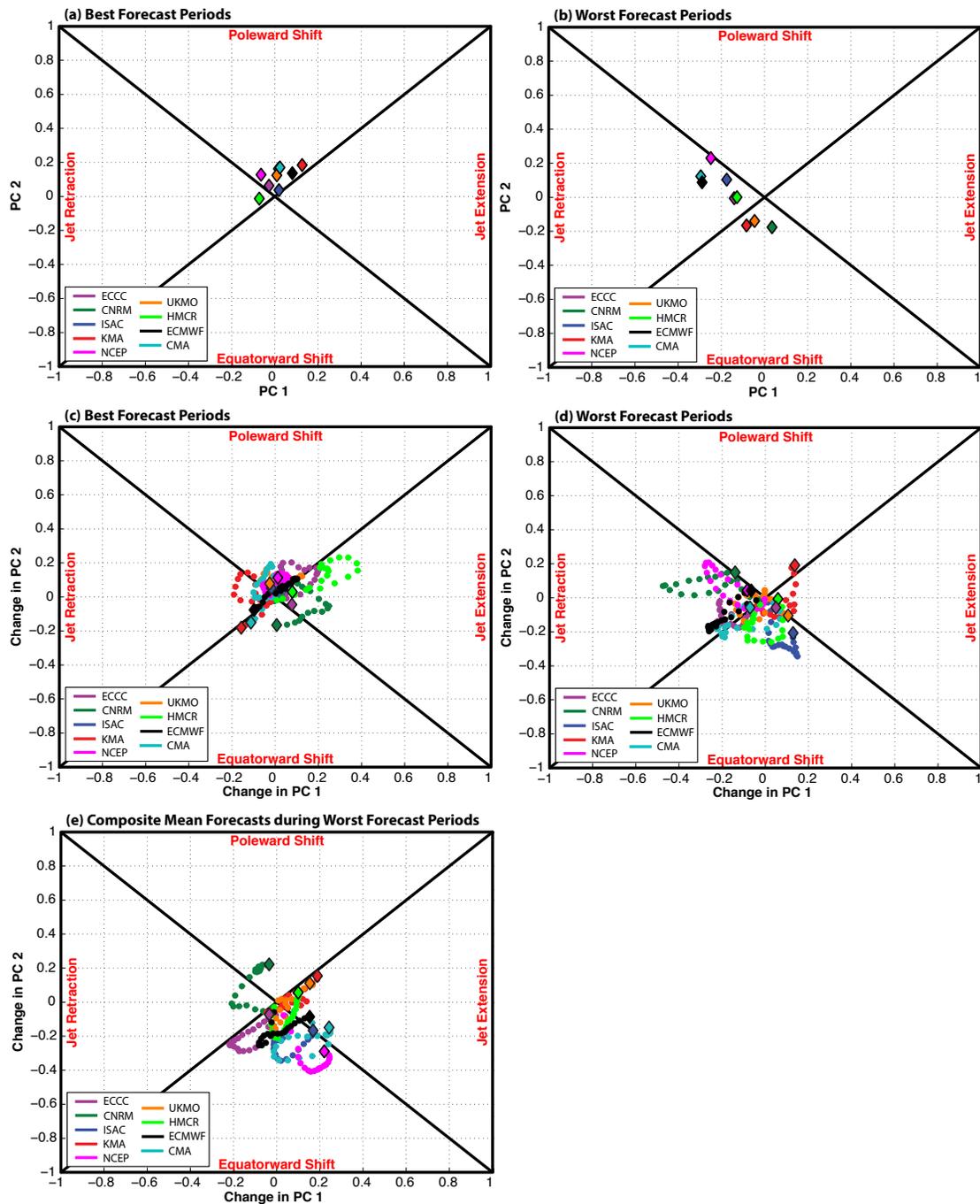
FIG. 10. The average position of the NPJ within the NPJ phase diagram at the time (a) a best-performing forecast and (b) a worst-performing forecast is initialized from each model. Trajectories showing the composite evolution of the NPJ within the NPJ phase diagram during the 21-day period after the initiation of (c) a best-performing forecast and (d) a worst-performing forecast from each model. (e) The composite ensemble mean 21-day forecast trajectories constructed from the worst-performing NPJ phase diagram forecasts from each model.
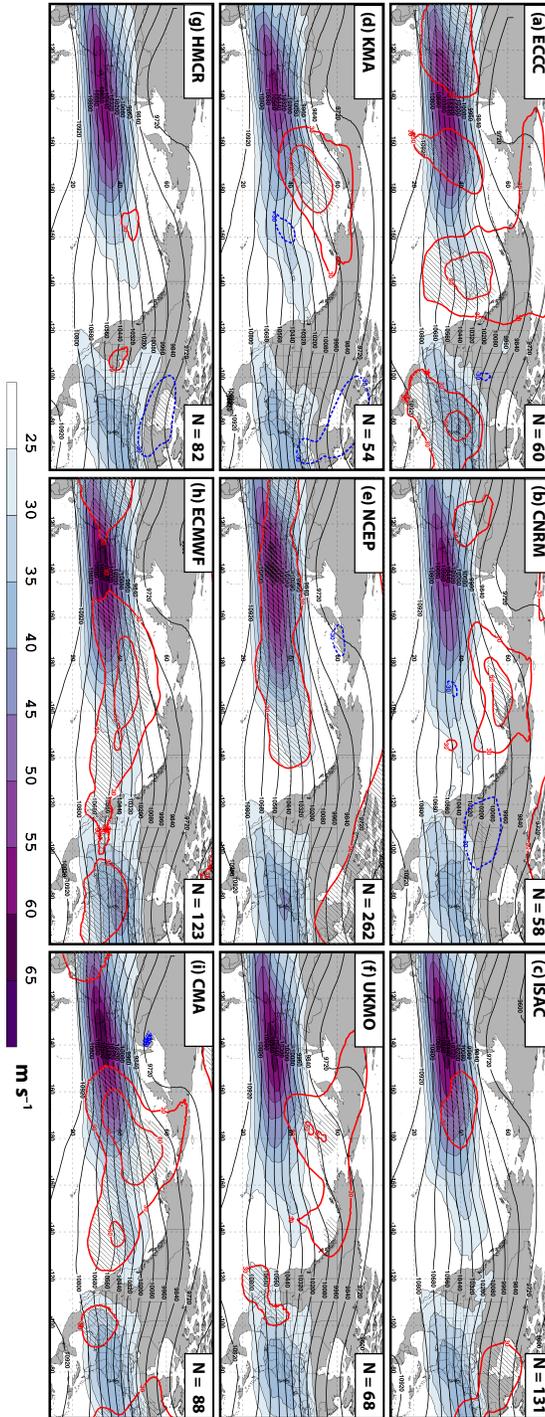
996
997
998 **FIG. 11.** Composite mean 250-hPa wind speed (shaded according to the fill pattern; m s^{-1}), 250-
999 hPa geopotential height (contoured in black every 120 m), and 250-hPa geopotential height
1000 anomalies (contoured every 30 m in red where positive and in dashed blue where negative) from
1001 the CFSR at the time a worst-performing forecast is initialized from the (a) ECCC, (b) CNRM,
1002 (c) ISAC, (d) KMA, (e) NCEP, (f) UKMO, (g) HMCR, (h) ECMWF, and (i) CMA model.
1003 Hatched regions indicate geopotential height anomalies that are statistically distinct from
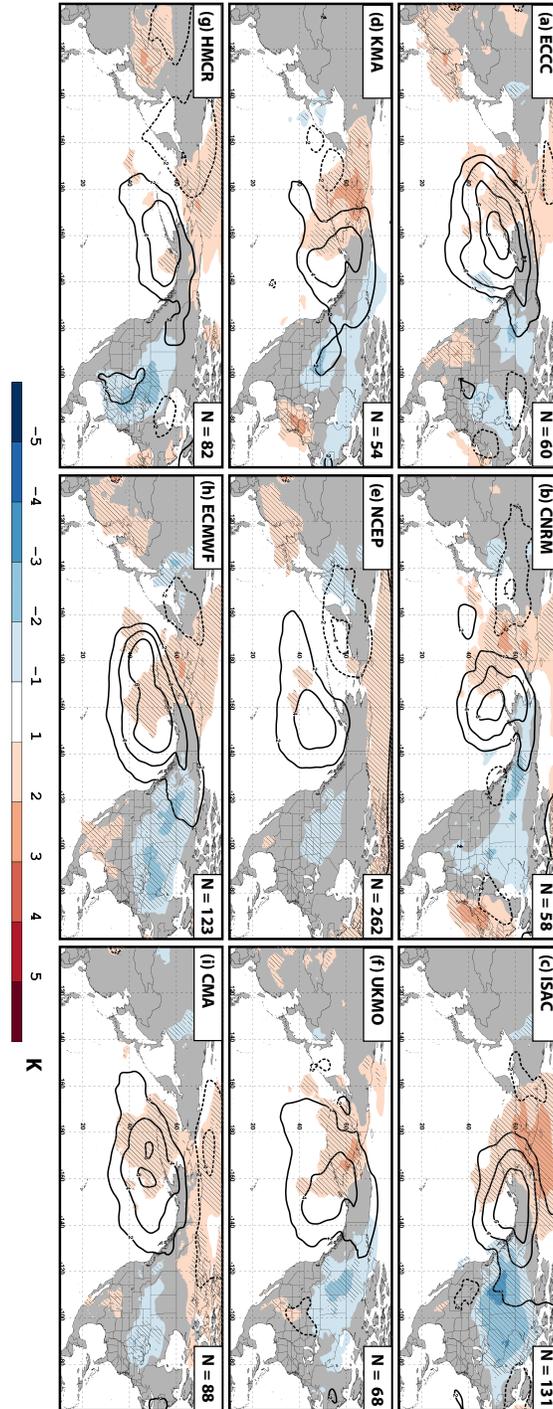1004 climatology at the 95% confidence interval using a two-sided Student's *t* test.

48
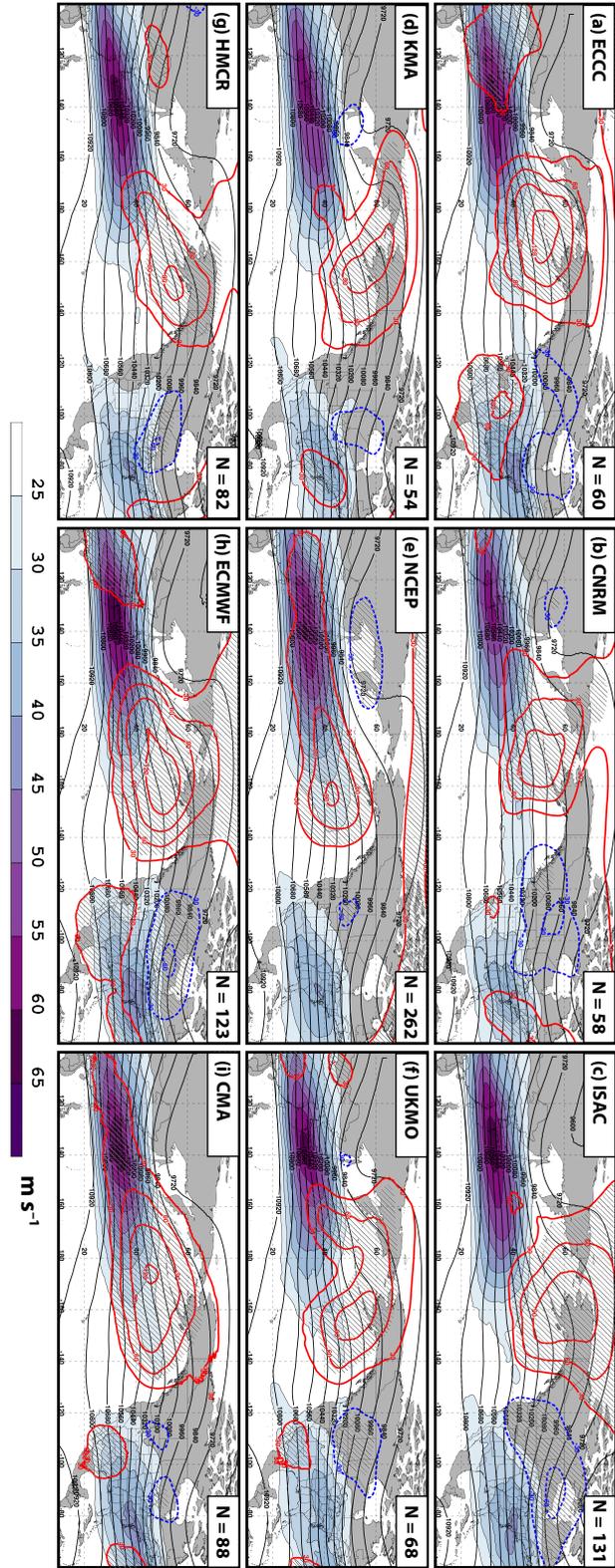
1005
1006
1007 **FIG. 12.** As in Fig. 11, but showing composites from the CFSR 10 days after the initialization of
1008 a worst-performing forecast from each model.
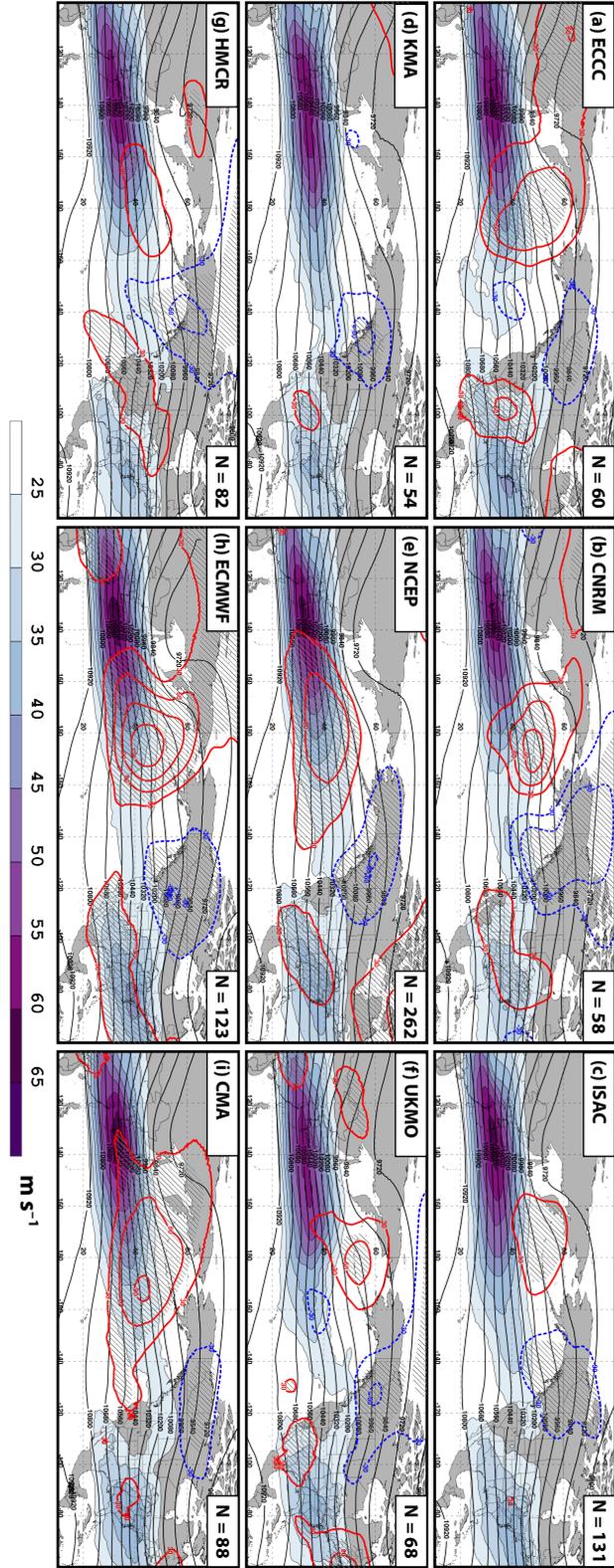
1009
1010
FIG. 13. As in Fig. 11, but showing composites from the CFSR 20 days after the initialization of
a worst-performing forecast from each model.

50

1013
1014
1015 **FIG. 14.** Composite mean 850-hPa temperature anomalies (shaded according to the legend every
1016 1 K), and mean sea-level pressure anomalies (contoured every 2 hPa in solid black where
1017 positive and in dashed black where negative) from the CFSR 10 days after a worst-performing
1018 forecast is initialized from the (a) ECCC, (b) CNRM, (c) ISAC, (d) KMA, (e) NCEP, (f) UKMO,
1019 (g) HMCR, (h) ECMWF, and (i) CMA model. Hatched regions indicate 850-hPa temperature
1020 anomalies that are statistically distinct from climatology at the 95% confidence interval using a
1021 two-sided Student's t test.