Dear Prof. Talagrand:

We are really grateful to the editor for the careful reviews. Our point-by-point response to the review comment follows.

=== Response to the editor

Dear Dr. Kondo,

I just have one final correction to suggest

L. 373, ... provide the minimum variance linear estimator ...

Response: Revised as suggested (l. 373).

Non-Gaussian statistics in global atmospheric dynamics: a study with a 10240-member ensemble Kalman filter using an intermediate AGCM

Keiichi KONDO[1, 2] and Takemasa MIYOSHI[1, 3, 4, 5, 6]

[1]RIKEN Center for Computational Science, Kobe, Japan

[2]Meteorological Research Institute, Japan Meteorological Agency, Tsukuba, Japan

[3]RIKEN Cluster for Pioneering Research, Kobe, Japan

[4]RIKEN interdisciplinary Theoretical and Mathematical Sciences Program, Kobe, Japan

[5]Department of Atmospheric and Oceanic Science, University of Maryland, College Park,

  Maryland, USA

[6]Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan

Correspondence to: Keiichi Kondo (Email: keiichi.kondo@riken.jp)

Abstract.

We previously performed local ensemble transform Kalman filter (LETKF) experiments with up to 10240 ensemble members using an intermediate atmospheric general circulation model (AGCM). While the previous study focused on the impact of localization on the analysis accuracy, the present study focuses on the probability density functions (PDFs) represented by the 10240-member ensemble. The 10240-member ensemble can resolve the detailed structures of the PDFs and indicates that non-Gaussianity is caused in those PDFs by multimodality and outliers. The results show that the spatial patterns of the analysis errors are similar to those of non-Gaussianity. While the outliers appear randomly, large multimodality corresponds well with large analysis error, mainly in the tropical regions and storm track regions where highly nonlinear processes appear frequently. Therefore, we further investigate the lifecycle of multimodal PDFs, and show that the multimodal PDFs are mainly generated by the on-off switch of convective parameterization in the tropical regions and by the instability associated with advection in the storm track regions. Sensitivity to the ensemble size suggests that approximately 1000 ensemble members be necessary in the intermediate AGCM-LETKF system to represent the detailed structures of the non-Gaussian PDF such as skewness and kurtosis; the higher-order non-Gaussian statistics are more vulnerable to the sampling errors due to a smaller ensemble size.

1 Introduction

Data assimilation is a statistical approach to estimate a posterior probability density function (PDF) using information of a prior PDF and observations. Based on the posterior PDF estimate, the optimal initial state is given for numerical weather prediction (NWP). The ensemble Kalman filter (EnKF; Evensen 1994) is an ensemble data assimilation method based on the Kalman filter (Kalman 1960) and approximates the background error covariance matrix by an ensemble of forecasts. The EnKF can explicitly represent the PDF of the model state, where the ensemble size is essential because the sampling error contaminates the PDF represented by the ensemble. Although the sampling error is reduced by increasing the ensemble size, the EnKF is usually performed with a limited ensemble size up to $O(100)$ due to the high computational cost of ensemble model runs. Recently, EnKF experiments with a large ensemble have been performed using powerful supercomputers. Miyoshi et al. (2014; hereafter MKI14) implemented a 10240-member EnKF with an intermediate atmospheric general circulation model (AGCM) known as the Simplified Parameterizations, Primitive Equation Dynamics model (SPEEDY; Molteni 2003), and found meaningful long-range error correlations. In addition, they reported that sampling errors in the error correlation were reduced by increasing the ensemble size. Further, Miyoshi et al. (2015) assimilated real atmospheric observations with a realistic model known as the Nonhydrostatic Icosahedral Atmospheric Model (NICAM; Tomita and Satoh 2004; Satoh et al. 2008; 2014) using an EnKF with 10240 members. Kondo and Miyoshi (2016; hereafter KM16) investigated the impact of covariance localization on the accuracy of analysis using a modified version of the MKI14 system.

56   MKI14 also focused on the PDF and reported strong non-Gaussianity, such as a bimodal PDF.

57   Previous studies investigated the impact of non-Gaussianity on the EnKF. Anderson (2010) reported

58   that an *N*-member ensemble could contain an outlier and a cluster of *N*-1 ensemble members under

59   nonlinear scenarios using the ensemble adjustment Kalman filter (EAKF; Anderson 2001). Anderson

60   (2010) called this phenomenon ensemble clustering (EC), which leads to degradation of analysis

61   accuracy. Amezcua et al. (2012) investigated EC with the ensemble transform Kalman filter (ETKF;

62   Bishop et al. 2001) and local ensemble transform Kalman filter (LETKF; Hunt et al. 2007), and found

63   that random rotations of the ensemble perturbations could avoid EC. Posselt and Bishop (2012)

64   explored the non-Gaussian PDF of microphysical parameters using an idealized one-dimensional

65   (1D) model of deep convection and showed that the non-Gaussianity of the parameter was generated

66   by nonlinearity between the parameters and model output.

67   Using the precious dataset of KM16 with 10240 ensemble members, we can make various

68   investigations such as non-Gaussian statistics and sampling errors in the background error covariance.

69   Here we focus on the non-Gaussian statistics in this study. Since the Gaussian assumption makes the

70   minimum variance estimator of the EnKF coincide with the maximum likelihood estimator, the non-

71   Gaussian PDF may bring some negative impacts on the LETKF analysis. KM16 showed that the

72   improvement in the tropics was relatively small by increasing the ensemble size up to 10240, and

73   suggested that the small improvement be related to the convectively dominated tropical dynamics.

74   This study aims to investigate the non-Gaussian statistics of the atmospheric dynamics in more detail

75   to investigate the relationship between the analysis error and the non-Gaussian PDF, as well as the

behavior and lifecycle of the non-Gaussian PDF. To the best of the authors' knowledge, this is the

first study investigating the non-Gaussian PDF using a 10240-member ensemble of an intermediate

AGCM. This study also discusses how many ensemble members are necessary to represent non-

Gaussian PDF without contamination by the sampling error, since in general higher-order non-

Gaussian statistics are more vulnerable to the sampling error due to a limited ensemble size. This

paper is organized as follows. Section 2 describes measures for the non-Gaussian PDF. Section 3

describes experimental settings, and Section 4 presents the results. Finally, summary and discussions

are provided in Section 5.

2   Non-Gaussian measures

Sample skewness $\beta_1^{1/2}$ and sample excess kurtosis β_2 are well-known parametric properties of a

non-Gaussian PDF, and are defined as follows:

$$\beta_1^{1/2} = \frac{N}{(N-1)(N-2)} \frac{\sum_{i=1}^{N}(x_i - \bar{x})^3}{\sigma^3} \tag{1}$$

$$\beta_2 = \frac{N(N+1)}{(N-1)(N-2)(N-3)} \frac{\sum_{i=1}^{N}(x_i - \bar{x})^4}{\sigma^4} - \frac{3(N-1)^2}{(N-2)(N-3)} \tag{2}$$

where x_i and $\bar{x}$ denote the ith ensemble member and N-member ensemble mean, respectively; σ

denotes the sample standard deviation, i.e., $\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$, and skewness $\beta_1^{1/2}$ represents

the asymmetry of the PDF. Positive (negative) skewness $\beta_1^{1/2}$ corresponds to the PDF with the

longer tail on the right (left) side. Positive (negative) kurtosis β_2 corresponds to the PDF with a more

92 pointed (rounded) peak and longer (shorter) tails on both sides. When the PDF is Gaussian, both

93 skewness $\beta_1^{1/2}$ and kurtosis β_2 go to zero in the limit of infinite sample size. In addition, we also

94 use Kullback–Leibler divergence (KL divergence, Kullback and Leibler 1951) from the Gaussian

95 PDF. KL divergence is a direct measure of the difference between two PDFs. Let $p(x)$ and $q(x)$ be two

96 PDFs. The KL divergence D_{KL} between the two PDFs is defined as

$$D_{KL} = \int p(x)\log\frac{p(x)}{q(x)}\,dx \qquad (3)$$

97 Here, we obtain $p(x)$ from the histogram based on the ensemble, and $q(x)$ from the theoretical

98 Gaussian function with the ensemble mean $\bar{x}$ and standard deviation σ, respectively. D_{KL} measures

99 the difference between the ensemble-based histogram and the fitted Gaussian function. Figure 1

100 shows examples of ensemble-based histograms and corresponding skewness $\beta_1^{1/2}$, kurtosis β_2, and

101 KL divergence D_{KL} with 10240 samples. Here, the Scott's choice method (Scott 1979) is applied to

102 decide the bin width for histograms. The histogram with KL divergence $D_{KL} = 0.01$ looks

103 approximately Gaussian while the other three histograms with larger D_{KL} values show significant

104 discrepancies from the Gaussian function. The skewness and kurtosis measure the degrees of

105 symmetry and tailedness, respectively, while the KL divergence D_{KL} is more suitable for measuring

106 the degrees of difference between a given PDF and the fitted Gaussian function. Based on the

107 subjective observation of Fig. 1, hereafter, the PDF is considered to be non-Gaussian when $D_{KL} >$

108 0.01.

109   A non-Gaussian PDF can also be caused by outliers. Although detailed results are shown in

110 Section 4, one or several ensemble members are detached from the main cluster; this also results in

111 the large KL divergence D_{KL}, as well as large skewness and kurtosis, shown in Fig. 2b. We tested two

112 outlier detection methods: the standard deviation-based method (SD method) and the local outlier

113 factor method (LOF method; Breunig et al. 2000). Here, univariate PDFs are considered, so that SD

114 and LOF methods are computed for each variable at each grid point separately.

115 In the SD method, the ensemble members beyond a prescribed threshold in the unit of SD are

116 defined as outliers. If we make 10240 independent random draws from a Gaussian PDF, statistically

117 27.6, 0.65, and 0.0059 samples (0.270, 0.00633, and 0.0000573 %) are expected beyond the $\pm 3\sigma$, $\pm 4\sigma$,

118 and $\pm 5\sigma$ thresholds, respectively. Namely, with the threshold of $\pm 3\sigma$, we would expect to detect 27.6

119 outliers at every grid point. With the $\pm 4\sigma$ threshold, we would expect to detect 1.3 outliers in two grid

120 points (20480 random draws). With the $\pm 5\sigma$ threshold, we would expect to detect 1.18 outliers in 200

121 grid points (2048000 random draws). Since outliers appear frequently with $\pm 3\sigma$ and $\pm 4\sigma$ thresholds,

122 we choose the $\pm 5\sigma$ threshold for the SD method in this study.

123 Unlike the SD method, the LOF method is based on the local density, not on the distance from

124 the sample mean. For a given two-dimensional dataset D, let $d(p, o)$ denote the distance between two

125 objects $p \in D$ and $o \in D$. For any positive integer k, define $k\text{-}distance(p)$ to be the distance

126 between the object p and the kth nearest neighbor. The $k\text{-}distance\ neighborhood$ of p, or simply N_k

127 (p), is defined as the k nearest objects:

$$N_k(p) = \{q \in D \mid q \neq p, d(p, q) \leq k\text{-}distance(p)\} \tag{4}$$

128 The cardinality of $N_k(p)$, or $|N_k(p)|$, is greater than or equal to the number of objects (except for the

129 object p itself) within $k\text{-}distance(p)$. We define the $reachability\ distance$ of p with respect to the object

130   o as

$$reach\text{-}dist_k\,(p,\,o)\ =\ \max\{k\text{-}distance(o),\,d(p,\,o)\} \qquad (5)$$

131   That is, if the object p is sufficiently distant from the object o, *reach-dist$_k$* $(p,\,o)$ is $d(p,\,o)$. If they are

132   sufficiently close to each other, *reach-dist$_k$* $(p,\,o)$ is replaced by *k-distance(o)* instead of $d(p,\,o)$. Figure

133   3 shows a schematic diagram of *reach-dist$_k$* $(p,\,o)$ with $k=\ 3$. $N_k\,(p)$ includes o_1, o_2, o_3, and o_4, and

134    $|N_k\,(p)|$ is 4. In Fig. 3 (a), *reach-dist$_k$* $(p,\,o_1)$ is *k-distance(o_1)* $= d(o_1,\,o_4)$ because *k-distance(o_1)* is

135   greater than $d(p,\,o_1)$. In contrast, in Fig. 3 (b), *reach-dist$_k$* $(p,\,o_1)$ is $d(p,\,o_1)$. We further define the

136   *local reachability density* of p, or simply *lrd$_k$* (p), as the inverse of the average of *reachability distance*

137   of p:

$$lrd_k(p) = \frac{|N_k(p)|}{\sum_{o\,\in\,N_k(p)} reach\text{-}dist_k(p,o)} \qquad (6)$$

139   Finally, the *local outlier factor* of p, denoted as *LOF$_k$* (p), is defined as:

$$LOF_k(p) = \frac{\sum_{o\,\in\,N_k(p)} \dfrac{lrd_k(o)}{lrd_k(p)}}{|N_k(p)|}. \qquad (7)$$

141   Given a lower *local reachability density* of p and a higher *local reachability density* of p's *k*-nearest

142   neighbors, *LOF$_k$* (p) becomes higher. *LOF$_k$* (p) or simply *LOF* is approximately 1 for an object deep

143   within a cluster, and *LOF* becomes larger around the edge of the cluster due to sparse objects on the

144   far side from the cluster. To summarize, the LOF method focuses on the local densities of objects,

145   and outliers are detected by comparing the local densities. For instance, when $k = 3$ in Fig. 3a, the

146   local densities of the objects p and $o_{1,\,2,\,3,\,4,\,5}$ have all similar values because the *k-distance(p)* is similar

147   to the *k-distance*s$(o_{1,\,2,\,3,\,4,\,5})$. Therefore, they are not identified as outliers. In contrast, in Fig. 3b the

148   object p has a smaller local density than the other objects $o_{1, 2, 3, 4, 5}$ because $k\text{-}distance(p) > k\text{-}$

149     $distances(o_{1, 2, 3, 4, 5})$. Therefore, the object p has a larger *LOF* and is identified as an outlier. An object

150   with *LOF* much larger than 1 may be categorized as an outlier, but it is not clear how to determine

151   the threshold for outliers because the threshold also depends on the dataset. The threshold of *LOF* is

152   chosen to be 8.0 in this study, and Section 4 shows the results with different values of the threshold

153   and discusses why we choose this value. k is a control parameter for the LOF method and depends on

154   the dataset, as shown by Breunig et al. (2000), who suggested that choosing k from 10 to 20 work

155   well for most of the datasets. If we choose k too small, some objects deeply inside a cluster have a

156   large *LOF*, and the LOF method does not work. In fact, using the dataset of KM16, $k = 10$ showed

157   this problem, while $k = 20$ did not. Therefore, we chose $k = 20$ in this study. Similar to the SD method,

158   the LOF method is applied to a one-dimensional dataset consisted of 10240 ensemble members.

159   The statistics of the KL divergence, SD and LOF methods with 10240 samples are evaluated

160   numerically with 1 million trials of 10240 random draws from the standard normal distribution by

161   the Box-Muller's method (Box and Muller 1958). The results show that the expected value of KL

162   divergence D_{KL} is 0.0025, and its standard deviation is 0.00048. As for outlier detections, 5767 and

163   16088 trials have at least one outlier for SD and LOF methods, respectively. Namely, the probabilities

164   to detect at least one outlier at a grid point are 0.58 % for the SD method and 1.6 % for the LOF

165   method. Here, the threshold for the SD method is $\pm 5\sigma$. For the LOF method, we choose $k = 20$ and,

166   as discussed below in Section 4, the threshold value *LOF* = 8.0.

167

3 Experimental settings

We use the 10240-member global atmospheric analysis data from an idealized LETKF experiment of KM16. That is, the experiment was performed with the SPEEDY-LETKF system (Miyoshi 2005) consisting of the SPEEDY model (Molteni 2003) and the LETKF (Hunt et al. 2007; Miyoshi and Yamane 2007). The SPEEDY model is an intermediate AGCM based on the primitive equations at T30/L7 resolution, which corresponds horizontally to 96×48 grid points and vertically to seven levels, and has simplified forms of physical parametrization schemes including large-scale condensation, cumulus convection (Tiedtke 1993), clouds, short- and long-wave radiation, surface fluxes, and vertical diffusion. Due to the very low computational cost, the SPEEDY model has been used in many studies on data assimilation (e.g., Miyoshi 2005; Greybush et al. 2011; Miyoshi 2011; Amezcua et al. 2012; Miyoshi and Kondo 2013; Kondo et al. 2013; MKI14; KM16).

The LETKF applies the ETKF (Bishop et al. 2001) algorithm to the local ensemble Kalman filter (LEKF; Ott et al. 2004). The LETKF can assimilate observations at every grid point independently, which is particularly advantageous in high-performance computation. In fact, Miyoshi and Yamane (2007) showed that the parallelization ratio reached 99.99% on the Japanese Earth Simulator supercomputer, and KM16 performed 10240-member SPEEDY-LETKF experiments within 5 minutes for one execution of LETKF, not including the forecast part on 4608 nodes of the Japanese K supercomputer. The LETKF is computed as follows. Let $\mathbf{X}$ ($\delta\mathbf{X}$) denote an $n \times m$ matrix, whose columns are composed of m ensemble members (deviations from the mean of the ensemble) with the system dimension n. The superscripts a and f denote the analysis and forecast, respectively. The

188   analysis ensemble $\mathbf{X}^a$ is written as:

$$\mathbf{X}^a = \bar{\mathbf{x}}^f \mathbf{1} + \delta\mathbf{X}^f \left[\widetilde{\mathbf{P}}^a (\mathbf{H}\delta\mathbf{X}^f)^{\mathrm{T}} \mathbf{R}^{-1}(\boldsymbol{y}^o - \mathbf{H}\bar{\mathbf{x}}^f)\mathbf{1} + \sqrt{m-1}\left(\widetilde{\mathbf{P}}^a\right)^{1/2} \right] \tag{8}$$

189   [cf. Eqs. (6) and (7) of Miyoshi and Yamane 2007]. Here, $\bar{\mathbf{x}}^f$, $\boldsymbol{y}^o$, $\mathbf{H}$, and $\mathbf{R}$ denote the background

190   ensemble mean, observations, linear observation operator, and observation error covariance matrix,

191   respectively. $\mathbf{1}$ is an m-dimensional row vector with all elements being 1. The $m \times m$ analysis error

192   covariance matrix $\widetilde{\mathbf{P}}^a$ in the ensemble space is given as

$$\widetilde{\mathbf{P}}^a = [(m-1)\mathbf{I}/\rho + (\mathbf{H}\delta\mathbf{X}^f)^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{X}^f)]^{-1} = \mathbf{U}\mathbf{D}^{-1}\mathbf{U}^{\mathrm{T}} \tag{9}$$

193   [cf. Eqs. (3) and (9) of Miyoshi and Yamane 2007]. Here, ρ denotes the covariance inflation factor.

194   As $\widetilde{\mathbf{P}}^a$ is real symmetric, $\mathbf{U}$ is composed of the orthonormal eigenvectors, such that $\mathbf{U}\mathbf{U}^{\mathrm{T}} = \mathbf{I}$. The

195   diagonal matrix $\mathbf{D}$ is composed of the non-negative eigenvalues.

196   KM16 performed a perfect-model twin experiment for 60 days from 0000 UTC 1 January in the

197   second year of the nature run, which was initiated at 0000 UTC 1 January from the standard

198   atmosphere at rest (zero wind). The first year of the nature run was discarded as spin-up. To resolve

199   detailed PDF structures, the ensemble size was fixed to 10240. No localization was applied, yielding

200   the best analysis accuracy as shown by KM16 who compared five 10240-member experiments with

201   different choices of localization: step functions with 2000-km, 4000-km and 7303-km localization

202   radii, a Gaussian function with a 7303-km localization radius, and no localization. The observations

203   for horizontal wind components (U, V), temperature (T), specific humidity (Q), and surface pressure

204   (Ps) were simulated by adding observational errors to the nature run every 6 h at radiosonde-like

205   locations (cf. Fig. 8, crosses) for all seven vertical levels, but the observations of specific humidity

206   were simulated from the bottom to the fourth model level (about 500 hPa). The observational errors

207   were generated from independent Gaussian random numbers, and the observational error standard

208   deviations were fixed at 1.0 m s^{-1}, 1.0 K, 0.1 g kg^{-1}, and 1.0 hPa for U/V, T, Q, and Ps, respectively.

209   The non-Gaussian measures, skewness $\beta_1^{1/2}$, kurtosis β_2, and KL divergence D_{KL}, are calculated

210   at each grid point for each variable. Outliers are diagnosed similarly at each grid point for each

211   variable with the SD method and LOF method.

212

213   **4   Results**

214   Figure 4 shows the spatial distributions of the analysis absolute error, ensemble spread, background

215   skewness $\beta_1^{1/2}$, kurtosis β_2, and KL divergence D_{KL} for temperature at the fourth model level (~500

216   hPa) at 0600 UTC 22 February. When the analysis absolute error is large, the background non-

217   Gaussian measures also tend to be large, especially in the tropics. The peaks for skewness $\beta_1^{1/2}$,

218   kurtosis β_2, and KL divergence D_{KL} tend to coincide. Although grid point A (16.7°S, 90.0°E) has a

219   large KL divergence D_{KL} with large analysis absolute error, at grid point B (35.3°N, 146.3°E) with a

220   large KL divergence D_{KL} the analysis absolute error is small (< 0.08 K). This result shows that the

221   large analysis error is not always associated with the strong non-Gaussianity at a specific time. The

222   PDFs at grid points A and B are shown in Fig. 2a, b, respectively. The histogram at the grid point A

223   is clearly a multimodal PDF with KL divergence $D_{KL} > 0.01$, and the right mode captures the truth

224   (yellow star). At grid point B, although the PDF seems to be closer to Gaussian, skewness $\beta_1^{1/2}$ and

225 kurtosis β_2 are much larger than those at grid point A. In fact, the PDF does not fit to the Gaussian

226 function calculated by the ensemble mean and standard deviation. Zooming in on the left side of Fig.

227 2b shows a small cluster composed of 76 members detached from the main cluster; 74 members of

228 the small cluster exceed -5σ and are categorized as outliers in the SD method. This small cluster

229 causes the standard deviation to become large and results in the Gaussian function having a longer

230 tail than the histogram. The small cluster should not be considered as consisting of outliers because

231 it may have some physical significance. Scatter diagrams of *LOF* versus distance from ensemble

232 mean for all ensemble members at grid points A and B are shown in Fig. 5a, b, respectively. At grid

233 point A, *LOF* is not so large even at the edge of the cluster (< 4), and the multimodal PDF does not

234 influence *LOF*. In addition, all members are within $\pm 3\sigma$. Therefore, there are no clear outliers at grid

235 point A. At grid point B, although most of the small cluster exceeds -5σ, the maximum *LOF* in the

236 small cluster is still smaller than 3. This indicates that all members of the small cluster should not be

237 outliers in the LOF method. As an outlier case, we pick up the grid point C (35.3°N, 112.5°W) in Fig.

238 4. The PDF at the grid point C fits the Gaussian function well, and the non-Gaussian measures are

239 quite small (Fig. 2c). A member on the left edge of the scatter diagram in Fig. 5c has the largest *LOF*

240 > 8.0, but the member is within $\pm 3\sigma$. As mentioned in Section 2, the threshold of *LOF* for outliers

241 depends on the dataset. Figure 6 shows the number of outliers for thresholds of 5.0, 8.0, and 11.0 at

242 0600 UTC 22 February. There are too many outliers with threshold $= 5.0$, but in contrast, the number

243 of outliers decreases markedly with threshold $= 8.0$ or 11.0. Based on these results, and as already

244 mentioned in Section 2, we adopt *LOF* $= 8.0$ as a threshold for outliers.

245   Figure 7 shows the spatial distributions of the time-mean analysis RMSE, ensemble spread, the

246   background absolute skewness $\beta_1^{1/2}$, absolute kurtosis β_2, and KL divergence D_{KL}. As mentioned in

247   KM16, the time-mean ensemble spread corresponds well to the RMSE, which is larger in the tropics.

248   The pattern correlation between the RMSE and ensemble spread is 0.97. Moreover, the distributions

249   of non-Gaussian measures are similar to each other and also correspond well to the RMSE and

250   ensemble spread. The RMSE and non-Gaussian measures differ in that the non-Gaussianity is large

251   in storm tracks, such as the North Pacific Ocean and the North Atlantic Ocean. This may be because

252   the LETKF inhibits growing errors well in storm tracks regardless of the strong non-Gaussianity. To

253   investigate the non-Gaussianity in more detail, Figs. 8 and 9 show the frequencies of non-Gaussian

254   PDF with high KL divergence $D_{KL} > 0.01$ and identifying at least one outlier with high $LOF > 8.0$

255   on a 10240-member ensemble, respectively. The frequency of non-Gaussian PDF is defined as the

256   ratio of non-Gaussianity appearance at every grid point during the 36-day period from 0000 UTC 25

257   January to 1800 UTC 1 March. The spatial distribution of frequency of non-Gaussianity for

258   temperature is similar to that of the time mean RMSE and D_{KL} (Figs. 7 a, e, and 8 b), and the pattern

259   correlation between the spatial distribution of mean RMSE and D_{KL} is 0.68. We find high frequency

260   of non-Gaussian PDF in the tropics and storm track regions for temperature, specific humidity, and

261   surface pressure, although non-Gaussian PDF seldom appears in the densely observed regions. In the

262   tropics, the frequency reaches up to 90%, and in South America the frequency reaches the highest

263   value over 95%, i.e., the non-Gaussian PDF appears for 34 days out of 36 days. In contrast, the non-

264   Gaussian PDF for zonal wind hardly appears (Fig. 8 a), and the intensity of the non-Gaussianity, as

265   evaluated by other measures, is also weak (not shown). On the other hand, the outliers appear almost

266   randomly and do not clearly depend on the region for any of the variables (Fig. 9), and most outliers

267   disappear within only one or a few analysis steps. Moreover, there are no correlations between the

268   frequency of outliers and analysis RMSE.

269   To investigate how the non-Gaussian PDF is generated, we plot the forecast and analysis update

270   processes at 1.9°N, 168.7°E for 256 members chosen randomly from 10240 members from the

271   analysis at 0000 UTC 9 February (157th analysis cycle) to the forecast at 0000 UTC 10 February

272   (161st analysis cycle, Fig. 10a). That is, Fig. 10a shows the lifecycle of the non-Gaussian PDF. As the

273   vertical axis, we introduce the convective instability $d\theta_e$, which is defined as a difference between

274   equivalent potential temperature θ_e at the fourth model level (~500 hPa) and θ_e at the second model

275   level (~850 hPa). Negative (Positive) $d\theta_e$ indicates a convectively unstable (stable) atmosphere. The

276   non-Gaussian PDF appears in the background at the 159th cycle (1200 UTC 9 February), and the

277   model forecast increases the KL divergence D_{KL} for $d\theta_e$ up to 0.154 with a bimodal PDF of clusters

278   A and B. We find many lines crossing in the forecast step from the analyses at the 158th cycle to the

279   background at the 159th cycle. Namely, many of the upper side cluster A at the 159th cycle come from

280   the lower side analyses in the previous 158th cycle, generally reducing the instability (increasing

281   values of $d\theta_e$) in the forecast step, and vice versa for the lower side cluster B. In the background

282   temperature at the fourth model level, the KL divergence D_{KL} also increases from 0.003 to 0.299 for

283   6 h (Figs. 10b, c). Finally, the non-Gaussian PDF almost disappears at the 161st cycle (0000 UTC 10

284   February). Figure 11 shows a scatter diagram of 0600 UTC versus 1200 UTC 9 February for

285   background temperature in the fourth model level for each member at 1.9°N, 168.7° E, and also shows

286   histograms corresponding to the scatter diagrams. The PDF at 0600 UTC is almost Gaussian.

287   However, at 1200 UTC, the bimodal structure with KL divergence $D_{KL} = 0.299$ appears. The dot

288   colors show $d\theta_e'$ evaluated from 0600 UTC to 1200 UTC 9 February, namely, $d\theta_e' =$

289    $(d\theta_{e\ 1200\ UTC} - d\theta_{e\ 0600\ UTC}) - (d\bar{\theta}_{e\ 1200\ UTC} - d\bar{\theta}_{e\ 0600\ UTC})$, where $\bar{\theta}_e$ indicates the equivalent

290   potential temperature calculated from the ensemble mean. That is, a red (blue) dot shows more

291   stability (instability) than the ensemble mean. The red and blue dots are clearly divided into the right

292   and left side modes, respectively. Most members with mitigated (enhanced) instability move to the

293   right (left) side mode. The members with larger (smaller) temperature values at 1200 UTC correspond

294   to larger (smaller) values of stability as shown by the warmer (colder) color. In addition, both right

295   and left modes correspond to the opposite side modes in the specific humidity, respectively (not

296   shown). That is, the members with higher (lower) temperature have lower (higher) humidity than the

297   ensemble mean. The instability is driven by precipitation. Figure 12 is similar to Fig. 11, but for

298   precipitation. The 10240 members are clearly divided into three clusters at 1200 UTC by the

299   instability. The three clusters indicate the number of times cumulus parameterization is triggered.

300   Most members in the right (left) cluster are red (blue) and show mitigation (enhancement) of the

301   instability. Figure 13 is also similar to Fig. 11, but for zonal wind at the fourth model level. In

302   agreement with what has been seen on Fig. 8a, the non-Gaussianity of zonal wind is weak, and the

303   bimodal structure appearing in temperature and humidity seldom affects the PDF of zonal wind. We

304   found no relationship between the atmospheric instability and zonal wind. Therefore, the genesis of

305 non-Gaussian PDF in the tropics is deeply related to precipitation process, which is driven by

306 convective instability through cumulus parameterization in the SPEEDY model. As a result, the

307 precipitation process mitigates the instability, with rising temperature and decreasing humidity.

308 Similar results are generally obtained at other grid points with non-Gaussian PDF.

309   In the extratropics, non-Gaussian PDF is generated differently. To investigate the genesis of non-

310 Gaussian PDF in the extratropics, we focus on a case around an extratropical cyclone over the Atlantic

311 Ocean. A non-Gaussian PDF appears at 0600 UTC 15 February at 42.7°N, 48.8°W, and the KL

312 divergence D_{KL} of background temperature increases from 0.003 to 0.460 (Fig. 14, crosses). Figure

313 15 is similar to Fig. 11, but for background specific humidity at the second model level (~850 hPa)

314 versus precipitation at 42.7°N, 48.8°W at 0006 UTC 15 February. Trimodal PDFs appear in both

315 specific humidity and precipitation. The three modes of specific humidity are clearly separated by the

316 color, i.e., instability $d\theta_e'$. Namely, modes with larger humidity has colder colors (smaller $d\theta_e'$

317 corresponding to more instability). However, the three modes of precipitation show no clear

318 dependence on $d\theta_e'$. Therefore, the trimodal PDF of specific humidity would not be driven by the

319 cumulus parameterization. Next, the relationship between background specific humidity and

320 meridional wind at the second model level (~850 hPa) is shown in Fig. 16. The members in the left

321 mode have lower specific humidity with relatively stronger northerly wind. If we look at the fourth

322 model level (~500 hPa) for these members with lower humidity, they have relatively weaker northerly

323 wind and warm temperature (not shown). Namely, instabilities are mitigated by the northerly

324 advection of dry air at the lower troposphere and by warm temperature at the mid troposphere. In this

325   case study, the non-Gaussianity genesis in the extratropics is associated with the advections. This is

326   only an example, and the non-Gaussianity genesis in the extratropics is generally more complicated

327   and would be affected by not only vertical stratification but also larger-scale atmospheric phenomena

328   such as extratropical cyclones and advections. Here, we do not go into details for different cases of

329   non-Gaussianity genesis, but instead, this is further discussed in Section 5.

330   The non-Gaussian measures are sensitive to the ensemble size due to sampling errors. Figure 17

331   shows the spatial distributions of the skewness $\beta_1^{1/2}$, kurtosis β_2, and KL divergence D_{KL} for

332   temperature at the fourth model level (~500 hPa) at 0600 UTC 22 February with 80, 320, and 1280

333   subsamples from 10240 members, respectively. Skewness $\beta_1^{1/2}$, kurtosis β_2, and KL divergence D_{KL}

334   with 80 members contain high levels of contaminating errors originating from sampling errors, and

335   the non-Gaussian measures are difficult to distinguish from the contaminating errors. With increasing

336   the ensemble size up to 1280, the sampling errors become smaller by gradation. With 1280 members,

337   the sampling errors are essentially removed, and the distributions are comparable to those with 10240

338   members (see Fig. 4). Therefore, a sample size of about 1000 members is necessary to represent non-

339   Gaussian PDF. The outliers also depend on the sample size. Figure 18 shows *LOF* with 80, 320, 1280,

340   and 5120 subsamples from 10240 members for temperature at the fourth model level at the grid point

341   B (35.3°N, 146.3°E), as in Fig. 5b. With 80 members, there are no outliers as the *LOF* of each member

342   is much smaller than the outlier threshold of 8.0. When the ensemble size is 320, four members with

343   high *LOF* > 8.0 are identified as outliers. With the ensemble sizes of 1280 and 5120, 13 and 41

344   members construct a small cluster, respectively, but they are not outliers with the threshold of *LOF* =

345 8.0. With increasing the ensemble size up to 10240, the *LOF*s of the small cluster and main cluster

346 show almost the same value (Fig. 5b).

347   We saw a good agreement between the RMSE and ensemble spread (Figs. 7a, b), but it is useful

348 to further evaluate the 10240-member ensemble using ranked probability scores. The rank histogram

349 (Hamill and Collucci 1997, Talagrand et al. 1999, Anderson 1996, Hamill 2001) evaluates the

350 reliability of ensemble statistically. Figure 19 shows almost flat rank histograms at all grid points and

351 the grid points with non-Gaussian PDF. The truth is known in this study and used as a verifying

352 analysis. The flat rank histograms correspond to healthy background ensemble distributions. The

353 continuous ranked probability score (CRPS, Hersbach 2000) is another method to evaluate ensemble

354 distributions, decomposed into reliability, resolution and uncertainty as

$$\text{CRPS} = \text{Reli} - \text{Resol} + U. \tag{10}$$

355 Here, the reliability Reli becomes zero under the perfectly reliable system. The resolution Resol

356 indicates the degree to which the ensemble distinguishes situations with different frequencies of

357 occurrence, and is associated with the accuracy or sharpness. The uncertainty U measures the

358 climatological variability. The reliability, resolution and uncertainty are given on the prescribed area

359 as

$$\text{Reli} = \sum_{i=0}^{N} \bar{g}_i (\bar{o}_i - p_i)^2 \,, \tag{11}$$

$$p_i = \frac{i}{N},$$

$$U - \text{Resol} = \sum_{i=0}^{N} \bar{g}_i \bar{o}_i (1 - \bar{o}_i), \tag{12}$$

$$U = \sum_{k,l<k} w_k \, w_l |y^k - y^l|, \tag{13}$$

360   [cf. Eqs 36, 37 and 19 in Hersbach 2000, respectively]. Here, $\bar{g}_i$ is the area-weighted average width

361   of the bin i between consecutive ensemble members x_i and x_{i+1}, and $\bar{o}_i$ is the area-weighted average

362   frequency that the verifying analysis is less than $(x_{i+1} + x_i)/2$. N denotes an ensemble size. In this

363   study, y^k and y^l indicate the anomalies between the background ensemble mean and monthly

364   climatology computed from a 30-year nature run at the grid points k and l, respectively. The weights

365   w_k, w_l are proportional to the cosine of latitude. Table 1 shows that the reliability is closer to zero and

366   that the resolution is much higher at all grid points than at the grid points with non-Gaussian PDF.

367   Therefore, the non-Gaussian PDF has a negative impact on updating the state variables for the LETKF.

368   The smaller uncertainty at the grid points with non-Gaussian PDF reflects generally smaller variations

369   in the tropics where the non-Gaussian PDFs frequently appear. Similar results are obtained for the

370   other variables.

371

5   Summary and discussions

373   Kalman filters provide the minimum variance linear estimator, which coincides with the maximum

374   likelihood estimator if the PDFs are Gaussian. This study investigated the non-Gaussian PDF and its

375   behavior using the SPEEDY-LETKF system with 10240 members. Non-Gaussian PDFs appear

376 frequently in the areas where the RMSE and ensemble spread are larger. Moreover, an ensemble size

377 of about 1000 is necessary to identify the possible non-Gaussianity of PDFs, which may be difficult

378 to detect in the presence of sampling error.

379   The non-Gaussian PDF appears frequently in the tropics and the storm track regions over the

380 Pacific and Atlantic Oceans, particularly for temperature and specific humidity, but not for winds.

381 With the SPEEDY model, the genesis of non-Gaussian PDF in the tropics is mainly associated with

382 the convective instability. These results suggest that the non-Gaussianity be mainly caused by

383 precipitation processes such those associated with cumulus convection, but much less by dynamic

384 processes. Generally, the atmosphere in the tropics tends to become unstable, and the convective

385 instability is mitigated by vertical convection with precipitation. In the SPEEDY model, a simplified

386 mass-flux scheme developed by Tiedtke (1993) is applied. Convection occurs when either the specific

387 or relative humidity exceeds a prescribed threshold (Molteni 2003). The members that hit the

388 threshold have precipitation, and this process mitigates their own convective instability resulting in a

389 temperature rise and humidity decrease. In contrast, the members with no or little precipitation

390 enhance or cannot mitigate their own convective instability. Therefore, convective instability is a key

391 to non-Gaussianity genesis in the tropics in the SPEEDY model.

392   In the extratropics, non-Gaussianity is generally weak and seldom appears except in the storm

393 track regions, where the genesis of non-Gaussian PDF is also associated with instabilities, but with

394 different processes from the tropics. This study focused on a case near the extratropical cyclone in

395 the North Atlantic, and the results showed that the instability was associated with the horizontal

advections. The members with reduced instabilities had lower humidity at the lower troposphere and

higher temperature at the mid troposphere by meridional advections. In contrast, the members with

higher humidity at the lower troposphere and lower temperature at the mid troposphere enhanced

their instability. Moreover, the precipitation process through the cumulus parameterization did not

explain the non-Gaussian PDF. Precipitation associated with extratropical cyclones is usually caused

by synoptic-scale baroclinic instabilities and does not mitigate the local instability completely.

As mentioned in Section 4, to generalize the process of non-Gaussianity genesis in the extratropics

is not simple. The non-Gaussianity genesis is generally associated with instability from various

processes such as the convection, advection and larger-scale atmospheric phenomena, so that it is

very difficult to find general mechanisms of the non-Gaussianity genesis in the extratropics even for

the simple SPEEDY model. Furthermore, if we use more realistic models with complex physics

schemes, the process of non-Gaussianity genesis would be much more diverse and complicated. This

is partly why we did not go into details to investigate different cases of non-Gaussianity genesis with

the SPEEDY model.

The non-Gaussianity is less frequent in the wind components not only in the time scale of 1 month

but also for the snapshot, although the dynamic process of the atmosphere is a nonlinear system.

Moreover, the non-Gaussian PDFs of temperature and specific humidity seldom affect the PDFs of

the wind components. We hypothesize that the model complexity may be a reason for this. The

SPEEDY model could not resolve some local interactions between wind components and other

variables due to its coarse resolution and simplified processes. With more realistic models, physical

416   processes are much more complex, and the local interactions can also be represented. Indeed, we

417   obtained widely distributed non-Gaussianity with a 10240-member NICAM-LETKF system with

418   112-km horizontal resolution assimilating real observations from the National Centers for

419   Environmental Prediction (NCEP) known as PREPBUFR from 0000 UTC 1 November to 0000 UTC

420   8 November (Miyoshi et al. 2015). Figure 20 shows the spatial distributions of background KL

421   divergence of zonal wind and temperature at the second model level (~850 hPa) for SPEEDY at 0000

422   UTC 1 March and zonal wind and temperature at the eighth model level (~850 hPa) for the NICAM

423   at 0000 UTC 8 November 2011. With NICAM, the non-Gaussianity appears globally not only in the

424   temperature field but also in the zonal wind although we should account for the model errors of

425   NICAM. This result implies that the NICAM has various sources of non-Gaussianity such as smaller

426   scale physical and dynamical processes with various interactions among different model variables,

427   and suggests the limitation of this study using the SPEEDY model. In the realistic situation, we would

428   presumably have more frequent occurrence of non-Gaussianity.

429   The outliers appear almost randomly regardless of locations, levels, and variables, and the lifetime

430   is about a few analysis steps. When the outliers appear, the number of outliers is basically one per

431   grid point, but sometimes the number is more than one. Anderson (2010) also reported similar results

432   using a low-order dry atmospheric model. These results seem not to be consistent with Amezcua et

433   al. (2012) who reported that just one outlier appeared with the ensemble square root filters in low-

434   dimensional models and that the outlier did not rejoin the cluster easily. These properties of their

435   outlier and our outliers in the SPEEDY model are somewhat different. In the low-dimensional models,

a certain ensemble member tends to become an outlier at all grid points and all variables. In contrast, the outliers in the SPEEDY model appear at just some grid points but not all grid points and do not appear in all variables simultaneously. In addition, the negative influence of outliers on the analysis accuracy may be sufficiently small in high-dimensional models due to the randomness and short longevity of outliers. In fact, the results showed no clear correspondence between the outlier frequency and analysis accuracy. These are the results from the simple SPEEDY model. It remains to be a subject of future research how the outliers behave with a more realistic model and real observations.

As measures of non-Gaussianity, skewness, kurtosis, and KL divergence for the non-Gaussianity, and the SD and LOF methods for outliers, are introduced and compared with each other. The KL divergence is a more suitable measure because it measures the direct difference between the ensemble-based histogram and the fitted Gaussian function. The LOF method is better than the SD method because it can detect the outliers depending on the density of objects. Although it is easy to detect the outliers using the SD method, misdetection of outliers is possible because this method categorizes a small cluster far from the main cluster into outliers. The small cluster may be generated through physical processes and have physical significance; this should not be treated as outliers. The measures of non-Gaussianity are evaluated in the univariate field in this study. An extension to multivariate fields with multivariate analysis remains as a subject of future research.

Non-Gaussian measures tend to be more sensitive to the sampling error due to the limited ensemble size (see Figs. 17, 18). When the ensemble size is small, it is difficult to determine whether

456   a split member is a real outlier or a sample from a small cluster. Amezcua et al. (2012) discussed the

457   outliers by skewness using the 20-member SPEEDY-LETKF and reported that the skewness is clearly

458   large in the tropics and the Southern Hemisphere for the temperature and humidity fields. These

459   results were not consistent with those of the present study because the outliers appear randomly.

460   However, this inconsistency may have been due to the small ensemble size. The large skewness of

461   Amezcua et al. (2012) could possibly indicate the non-Gaussianity rather than the outliers with a large

462   ensemble size. Having a sufficient ensemble size, suggested to be about 1000 according to this study,

463   would be essential when discussing about non-Gaussianity and outliers.

464

465   **Data availability**

466   All data and source code are archived in RIKEN Center for Computational Science and are available

467   upon request from the corresponding authors under the license of the original providers. The original

468   source code of the SPEEDY-LETKF is available at https://github.com/takemasa-miyoshi/letkf.

469

470   **Acknowledgments**

471   The authors are grateful to the members of the Data Assimilation Research Team, RIKEN R-CCS and

472   of the Meteorological Research Institute for fruitful discussions. The SPEEDY-LETKF code is

473   publicly available at https://github.com/takemasa-miyoshi/letkf. Part of the results was obtained using

474 the K computer at the RIKEN R-CCS through proposal numbers ra000015 and hp150019. This study

475 was partly supported by JST CREST Grant number JPMJCR1312, and JSPS KAKENHI Grant

476 number JP16K17806.

477

478

479   **References**

480   Anderson, J. L.: A method for producing and evaluating probabilistic forecasts from ensemble

481   model integrations, J. Climate, 9, 1518–1530, 1996.

482   Anderson, J. L.: An ensemble adjustment Kalman filter for data assimilation, Mon. Wea. Rev., 129,

483   2884-2903, 2001.

484   Anderson, J. L.: A non-Gaussian ensemble filter update for data assimilation, Mon. Wea. Rev., 138,

485   4186-4198, 2010.

486   Amezcua, J., Ide, K., Bishop, C. H., and Kalnay, E.: Ensemble clustering in deterministic ensemble

487   Kalman filters, Tellus, 64A, 1-12, 2012.

488   Bishop, C. H., Etherton, B. J. and Majumdar, S. J.: Adaptive sampling with the ensemble transform

489   Kalman filter. Part I: Theoretical aspects. Mon. Wea. Rev., 129, 420-436, 2001.

490   Box, G. E. P. and Muller, Mervin E.: A note on the generation of random normal deviates, Ann.

491   Math. Statist., 29, 610-611, doi:10.1214/aoms/1177706645.

492   Breunig, M. M, Kriegel,H. P. R., Ng, T., and Sander, J.: LOF: Identifying density-based local

493   outliers, Proceedings of the 2000 ACM SIGMOD International Conference on Management of

494   Data, 93-104, doi: 10.1145/335191.335388, 2000.

495   Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte

496   Carlo methods to forecast error statistics, J. Geophys. Res. 99C5, 10143-10162, 1994.

497   Greybush, S. J., Kalnay, E., Miyoshi, T., Ide, K., and Hunt, B. R.: Balance and ensemble Kalman

498   filter localization techniques, Mon. Wea. Rev., 139, 511-522, 2011.

499   Hamill, T. M.: Interpretation of rank histograms for verifying ensemble forecasts, Mon. Wea. Rev.,

500   129, 550-560, 2001.

501   Hamill, T., and Colucci, S. J.: Verification of Eta–RSM short- range ensemble forecasts, Mon. Wea.

502   Rev., 125, 1312–1327, 1997.

503   Hersbach, H.: Decomposition on the continuous ranked prob- ability score for ensemble prediction

504   systems, Wea. Forecasting, 15, 559–570, 2000.

505   Hunt, B. R., Kostelich, E. J., and Syzunogh, I.: Efficient data assimilation for spatiotemporal chaos:

506   A local ensemble transform Kalman filter, Physica D, 230, 112-126, 2007.

507   Kalman, R. E.: A new approach to linear filtering and predicted problems, J. Basic Eng.. 82, 35-45,

508   1960.

509   Kondo, K. and Miyoshi, T.: Impact of removing covariance localization in an ensemble Kalman

510   filter: experiments with 10240 members using an intermediate AGCM, Mon. Wea. Rev., 144,

511   4849-4865, 2016.

512   Kondo, K. and Miyoshi, T., and Tanaka, H. L.: Parameter sensitivities of the dual-localization

513   approach in the local ensemble transform Kalman filter, SOLA, 9, 174-178, 2013.

514   Kullback, S., and Leibler, R. A.: On information and sufficiency, The Annals of Mathematical

515   Statistics, 22, 79-86, 1951.

516   Miyoshi, T.: *Ensemble Kalman Filter Experiments with a Primitive-equation Global Model*. PhD

517   Thesis, University of Maryland, College Park, 226 pp., 2005.

518   Miyoshi, T.: The Gaussian approach to adaptive covariance inflation and its implementation with

the local ensemble transform Kalman filter, Mon. Wea. Rev., 139, 1519–1535, doi:

10.1175/2010MWR3570.1, 2011.

Miyoshi, T. and Yamane, S.: Local ensemble transform Kalman filtering with an AGCM at a

T159/L48 resolution, Mon. Wea. Rev., 135, 2841-3861, 2007.

Miyoshi, T. and Kondo, K.: A multi-scale localization approach to an ensemble Kalman filter,

SOLA, 9, 170-173, 2013.

Miyoshi, T., Kondo, K., and Imamura, T.: 10240-member ensemble Kalman filtering with an

intermediate AGCM. Geophys. Res. Lett., 41, 5264–5271, doi: 10.1002/2014GL060863, 2014.

Miyoshi, T., Kondo, K., and Terasaki, K.: Big Ensemble Data Assimilation in Numerical Weather

Prediction, Computer, 48, 15-21, doi:10.1109/MC.2015.332, 2015.

Molteni, F.: Atmospheric simulations using a GCM with simplified physical parameterizations. I:

model climatology and variability in multi-decadal experiments, Clim. Dyn., 20, 175-191, 2003.

Ott, E., and Coauthors: A local ensemble Kalman filter for atmospheric data assimilation, Tellus, 56A,

415–428, 2004.

Pearson, Egon S.: Note on tests for normality, Biometrika, 22, 423-424, 1931.

Posselt, D., and Bishop, C. H.: Nonlinear parameter estimation: comparison of an ensemble Kalman

smoother with a Markov chain Monte Carlo algorithm, Mon. Wea. Rev., 140, 1957-1974,

2012.

Satoh, M., Matsuno, T., Tomita, H., Miura, H., Nasuno, T., and Iga, S.: Nonhydrostatic icosahedral

atmospheric model (NICAM) for global cloud resolving simulations, Journal of Computational

539   Physics, the special issue on Predicting Weather, Climate and Extreme events, 227, 3486-3514,

540   doi:10.1016/j.jcp.2007.02.006, 2008.

541 Satoh, M., Tomita, H., Yashiro, H., Miura, H., Kodama, C., Seiki, T., Noda, A. T., Yamada, Y., Goto,

542   D., Sawada, M., Miyoshi, T., Niwa, Y., Hara, M., Ohno, T., Iga, S., Arakawa, T., Inoue, T., and

543   Kubokawa, H.: The non-hydrostatic icosahedral atmospheric model: Description and

544   development, Progress in Earth and Planetary Science, 1, 18, doi:10.1186/s40645-014-0018-1,

545   2014.

546 Scott, D. W.: On optimal and data-based histograms, Biometrika, 66, 605-610,

547   doi:10.1093/biomet/66.3.605, 1979.

548 Talagrand, O., Vautard, R., and Strauss, B.: Evaluation of Probabilistic Prediction Systems,

549   Proceedings of Workshop on *Predictability*, European Centre for Medium-range Weather

550   Forecasts, Reading, England, October 1997, 1-25, 1999.

551 Tiedtke, M: A comprehensive mass flux scheme for cumulus parameterization in large-scale

552   models, Mon. Wea. Rev., 117, 1779-1800, 1993.

553 Tomita, H., and Satoh, M.: A new dynamical framework of nonhydrostatic global model using the

554   icosahedral grid, Fluid Dyn. Res., 34, 357-400, 2004.

555

556

557

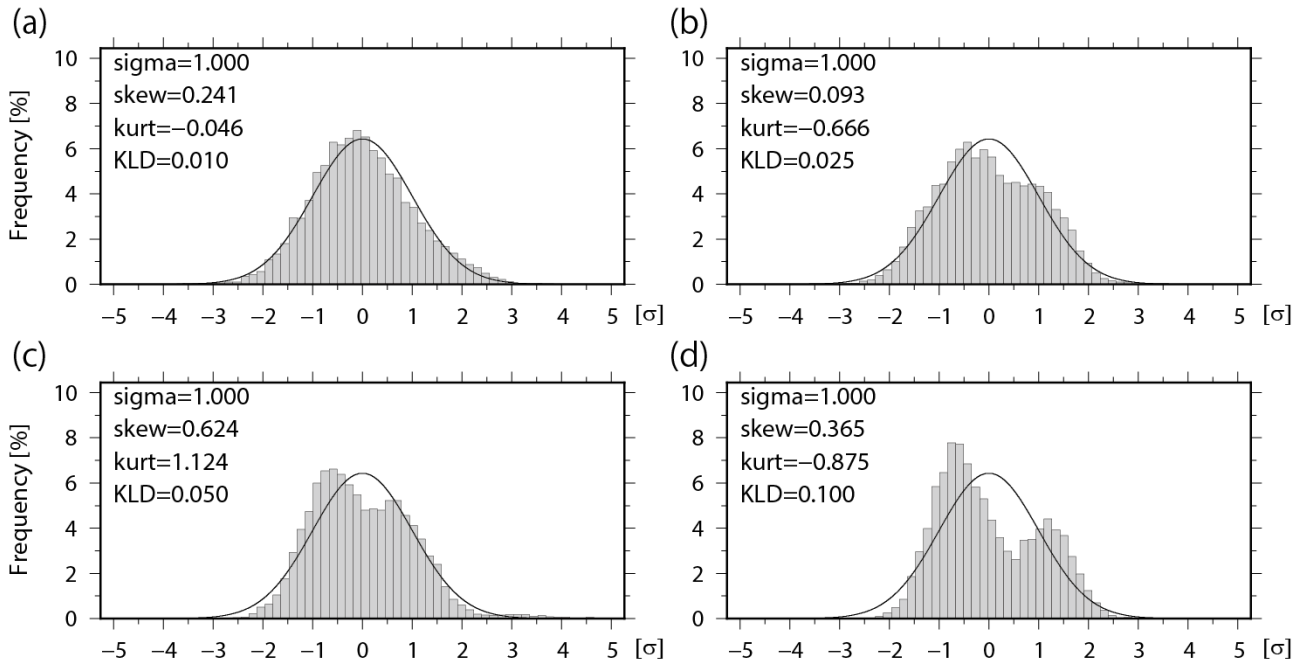



Figure 1: Ensemble-based histograms with 10240 ensemble members when the Kullback–Leibler

(KL) divergence D_{KL} = (a) 0.010, (b) 0.025, (c) 0.050, and (d) 0.100. Solid lines indicate fitted

Gaussian functions. Skewness (skew) and kurtosis (kurt) are also shown in the figure.

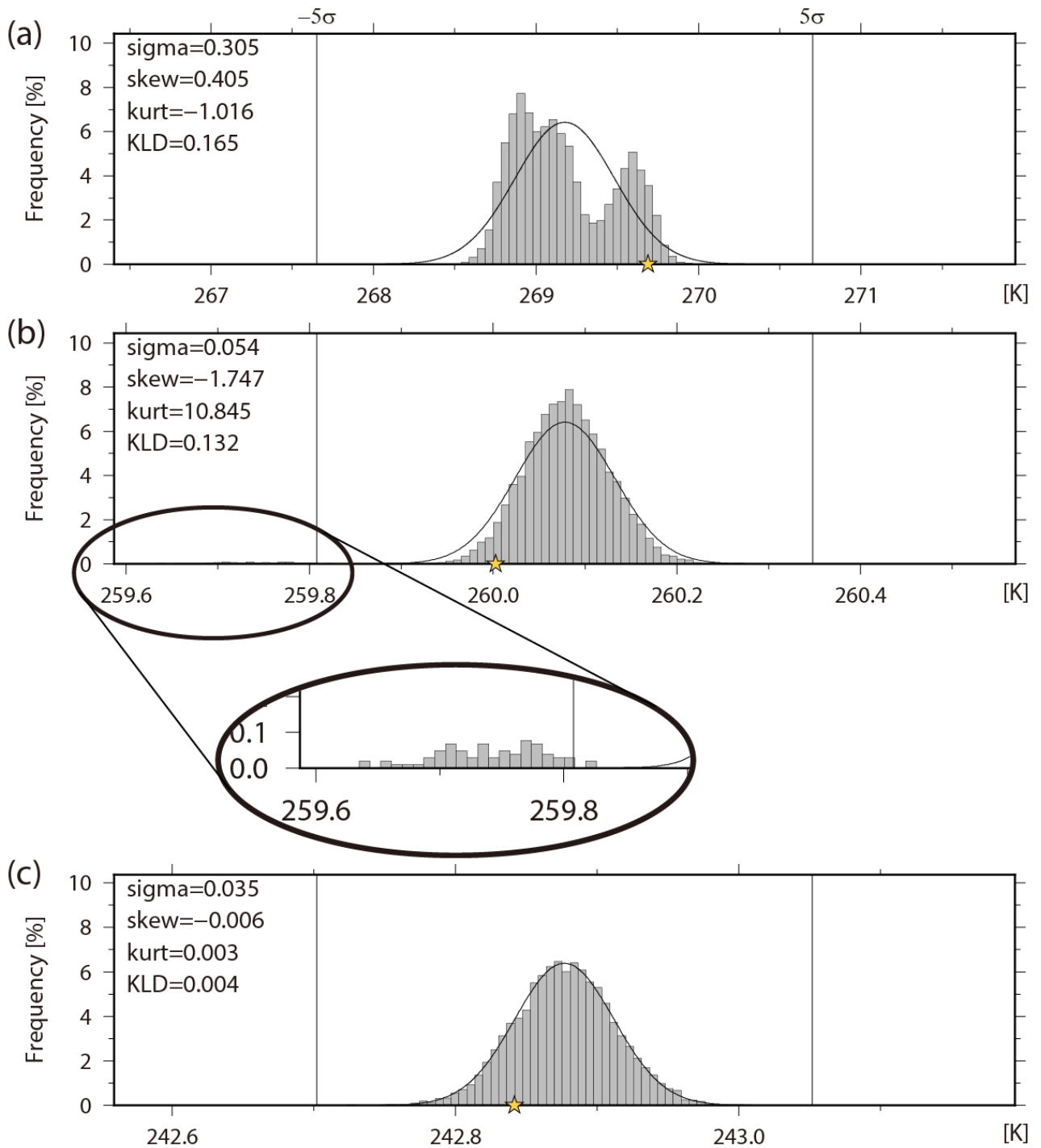
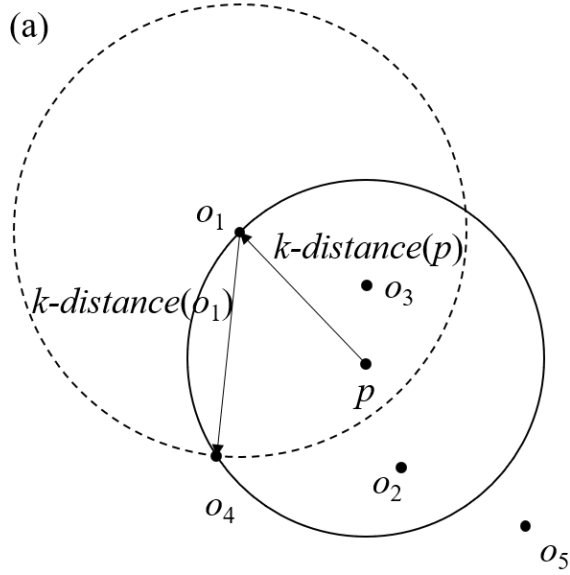
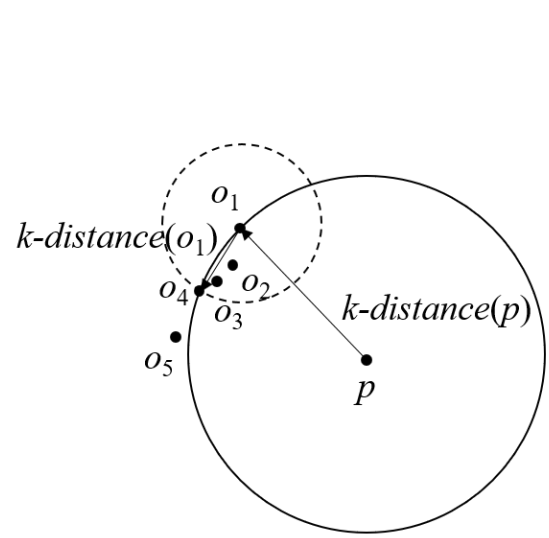



Figure 2: Histograms of background temperature (K) at the fourth model level (~500 hPa) at (a) grid point A (16.7°S, 90.0°E), (b) grid point B (35.3°N, 146.3°E), and (c) grid point C (35.3°N, 112.5°W). The yellow star shows the truth.

$reach\text{-}dist_k(p, o_1)$
    $= \max\{k\text{-}distance(o_1), d(p, o_1)\}$
    $= k\text{-}distance(o_1)$
    $= d(o_1, o_4)$

$reach\text{-}dist_k(p, o_1)$
    $= \max\{k\text{-}distance(o_1), d(p, o_1)\}$
    $= d(p, o_1)$

567

568   Figure 3: Schematic diagrams of $reach\text{-}dist_k$ (p, o) with $k = 3$ for (a) uniformly distributed data and

569   (b) data with an asymmetrical distribution.

570

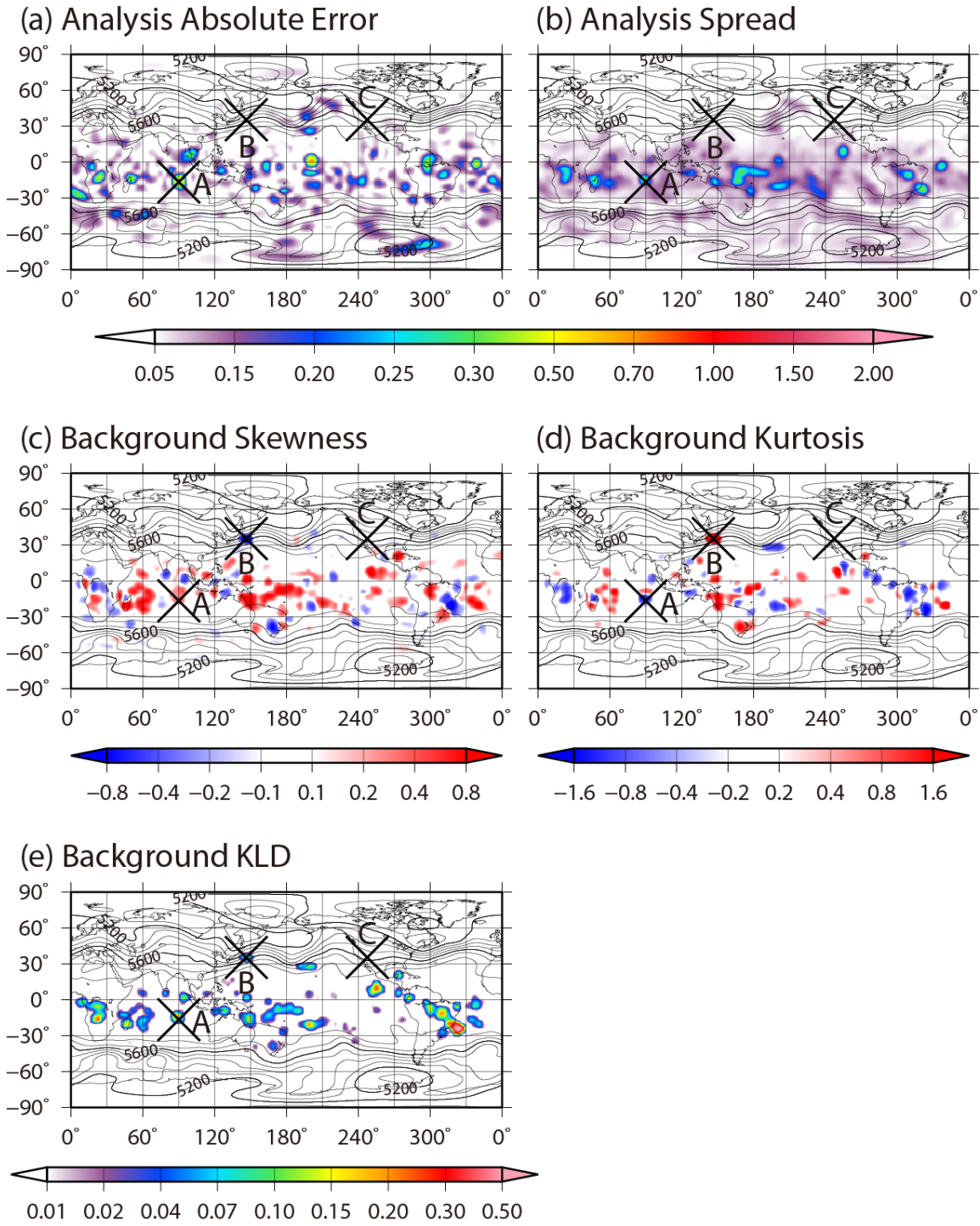



Figure 4: Spatial distributions of (a) analysis absolute error, (b) analysis ensemble spread, (c) background skewness, (d) background kurtosis, and (e) background KL divergence for temperature at the fourth model level (~500 hPa) at 0600 UTC 22 February. Contours indicate geopotential height of the ensemble mean at the 500 hPa level.

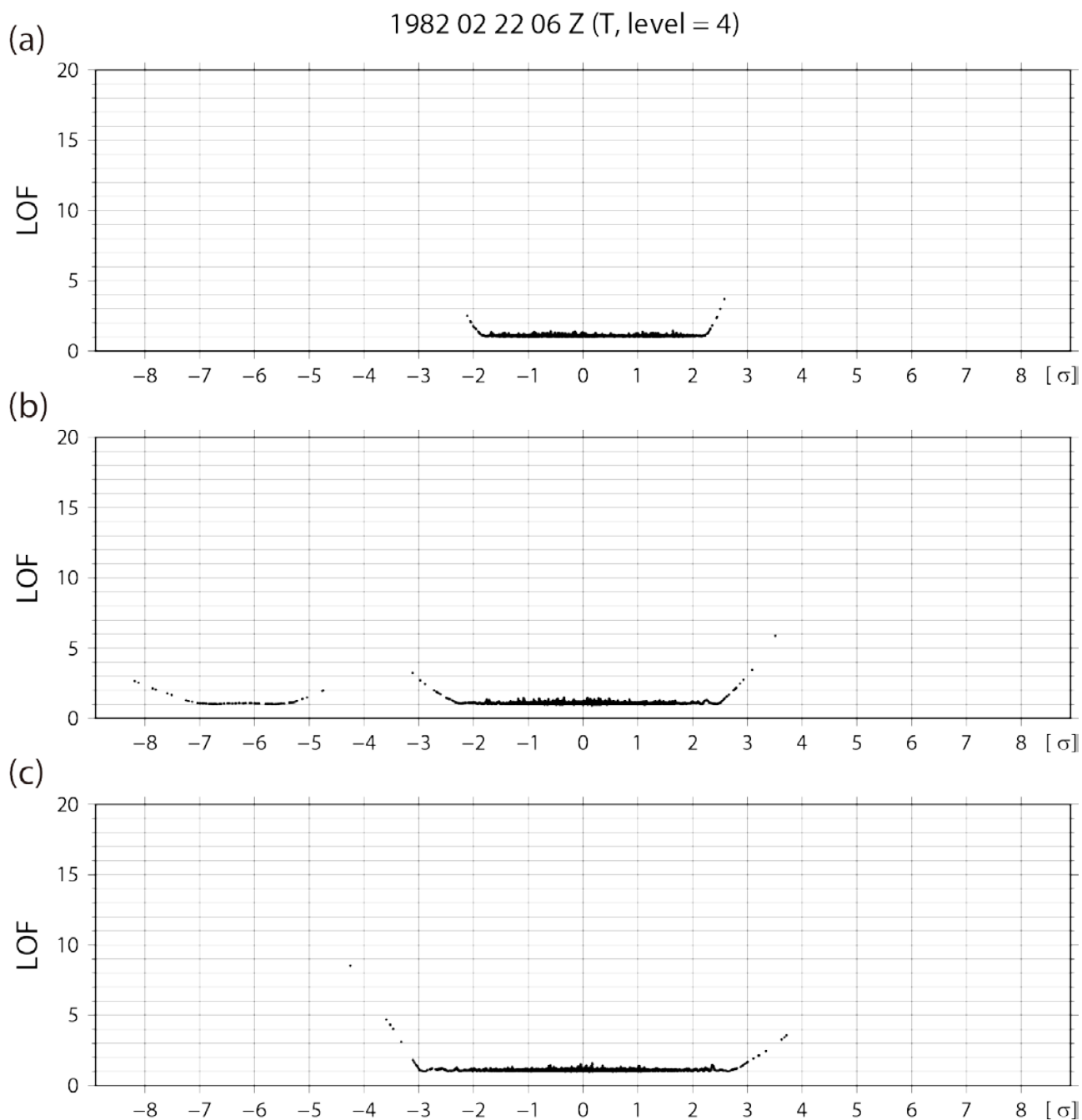



Figure 5: Scatter diagrams of the local outlier factor method (*LOF*) versus distance from the

ensemble mean for all ensemble members for background temperature at the fourth model level

(~500 hPa) at (a) grid point A (16.7°S, 90.0°E), (b) grid point B (35.3°N, 146.3°E), and (c) grid

point C (35.3°N, 112.5°W).

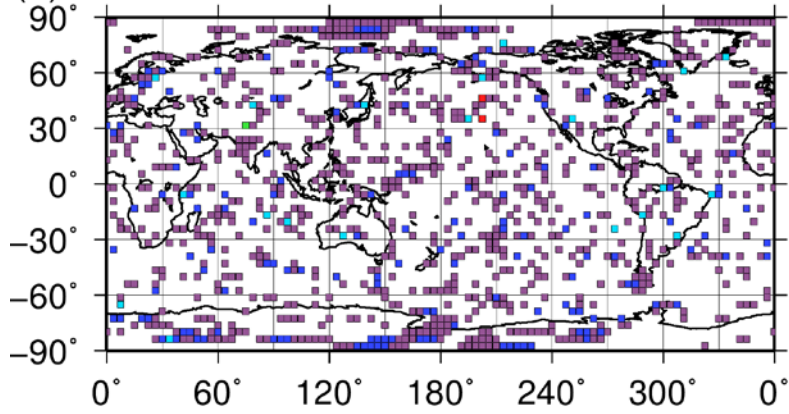
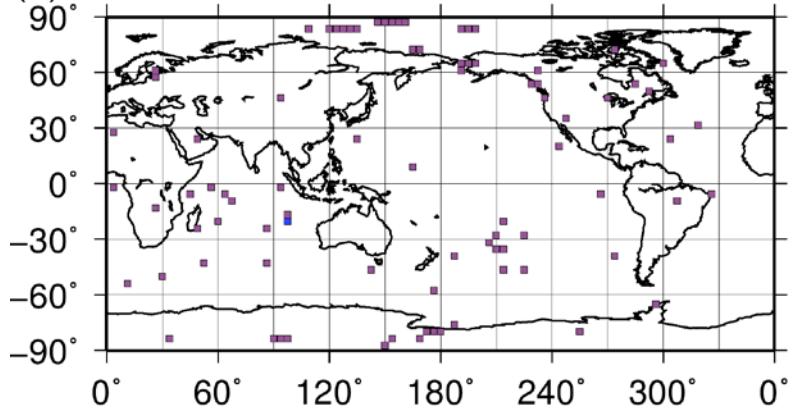
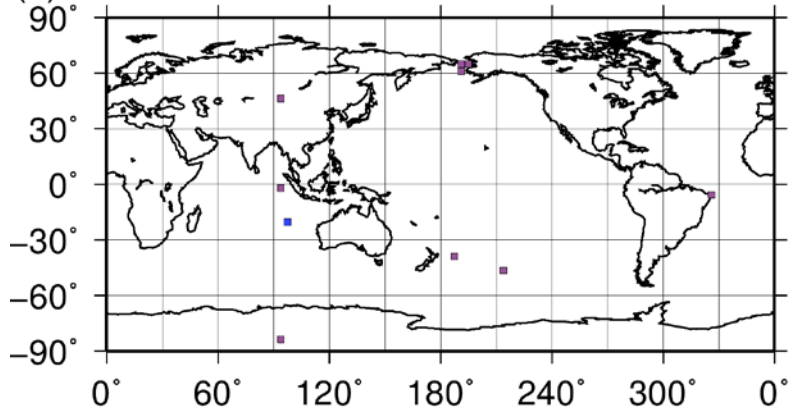
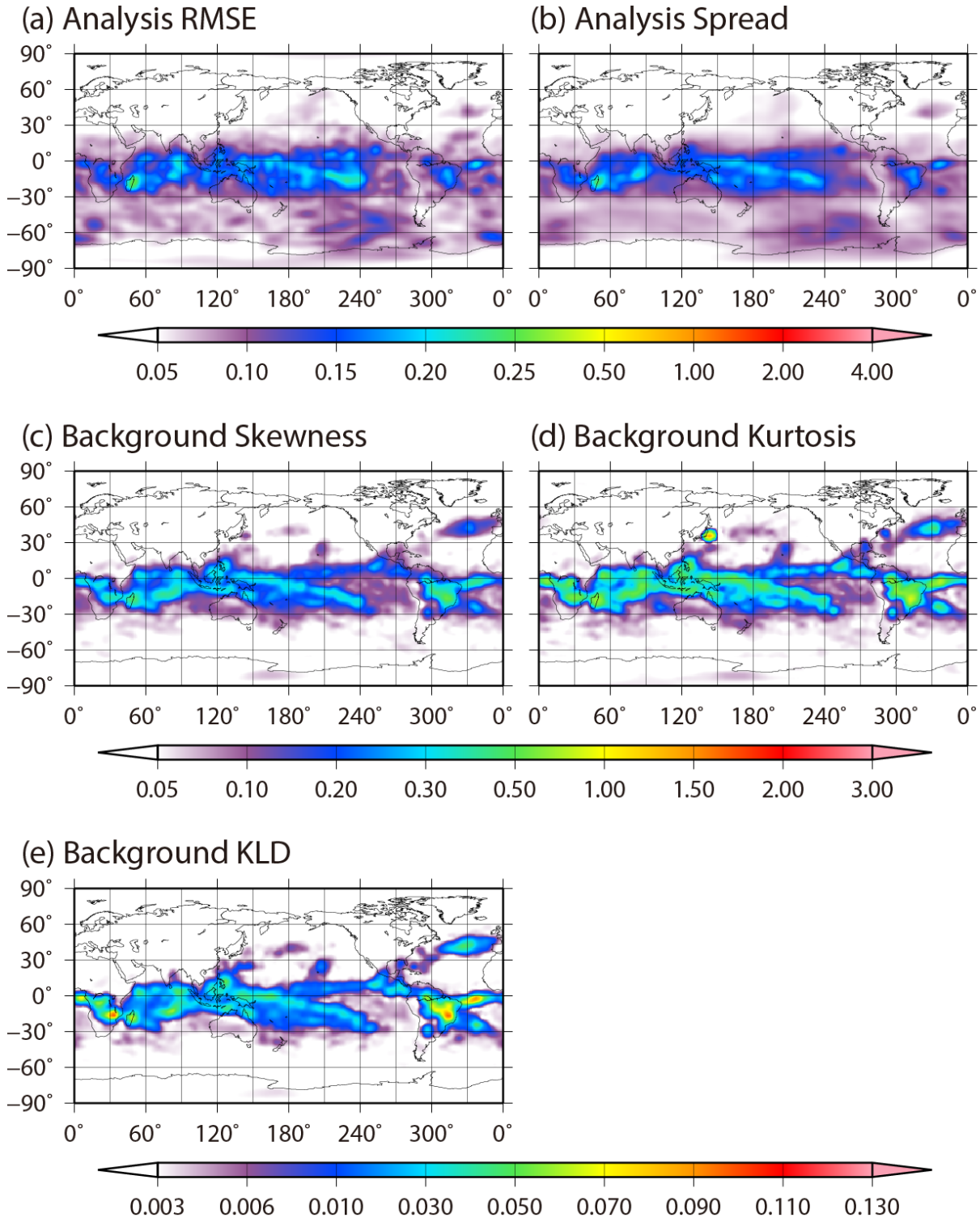



Figure 6: Spatial distributions of the number of outliers for background temperature at the fourth

model level (~500 hPa) at 0600 UTC 22 February for *LOF* thresholds of (a) 5.0, (b) 8.0, and (c)

11.0.

587 Figure 7: Spatial distributions of the time-mean (a) analysis RMSE, (b) analysis ensemble spread,

588 (c) background absolute skewness, (d) background absolute kurtosis, and (e) background KL

589 divergence for temperature at the fourth model level (~500 hPa) from 0000 UTC 25 January to

590 1800 UTC 1 March.

591

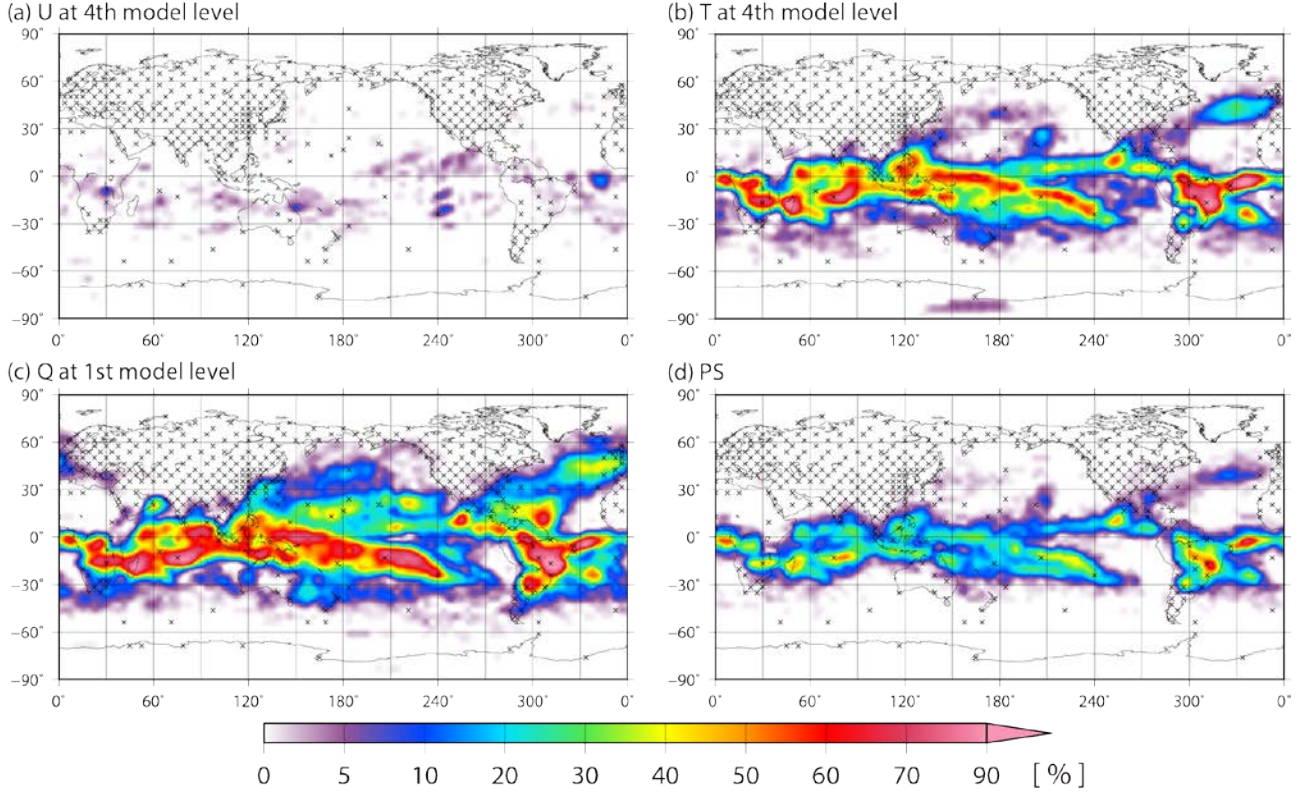



Figure 8: Spatial distributions of frequency of non-Gaussian PDF with high KL divergence $D_{KL} >$

0.01 for (a) zonal wind at the fourth model level, (b) temperature at the fourth model level, (c)

specific humidity at the lowest model level, and (d) surface pressure. The frequency is defined as a

ratio of high KL divergence D_{KL} appearance from 0000 UTC 25 January to 1800 UTC 1 March.

The crosses indicate the radiosonde-like locations.

597

Frequency of Outlier

(a) U at 4th model level

(b) T at 4th model level

(c) Q at 1st model level

(d) PS

598

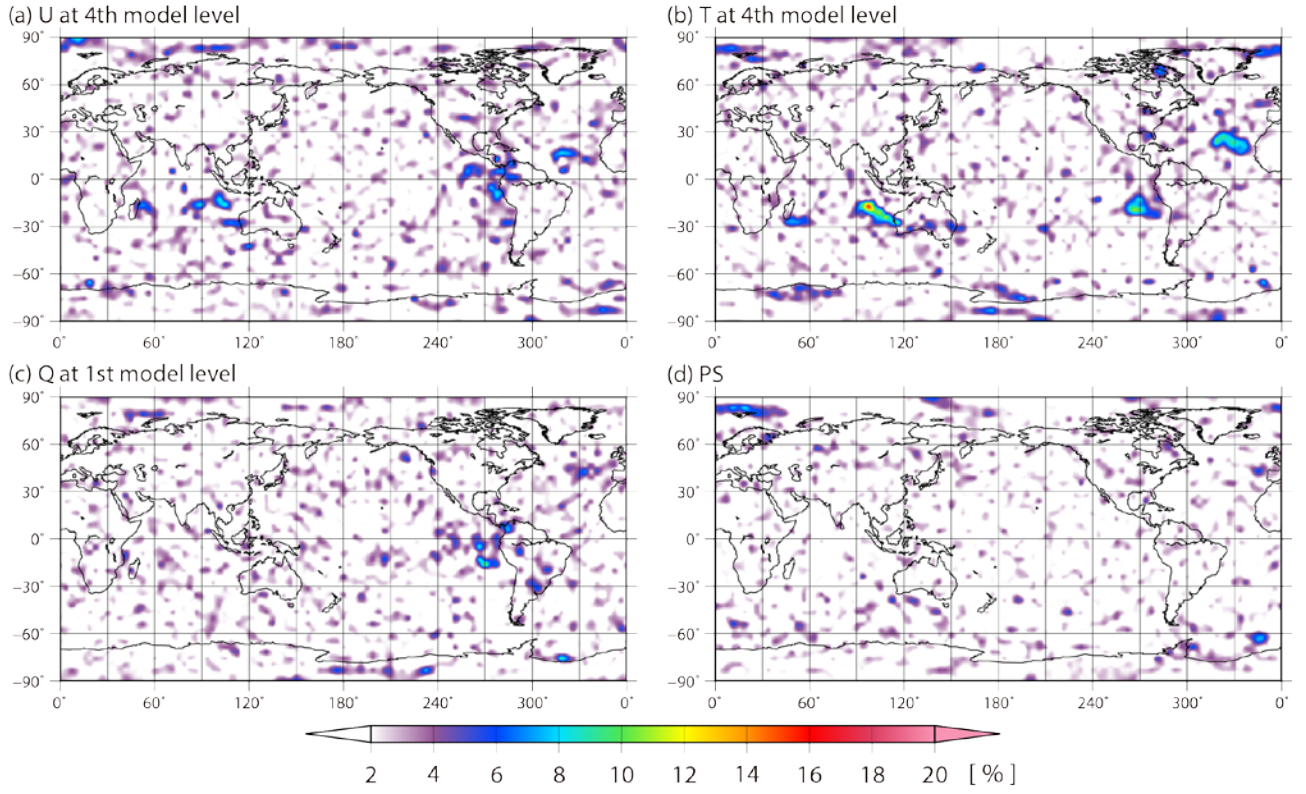



Figure 9: Similar to Fig. 8, but showing the frequency of identifying at least one outlier with high

LOF > 8.0 on a 10240-member ensemble.

601

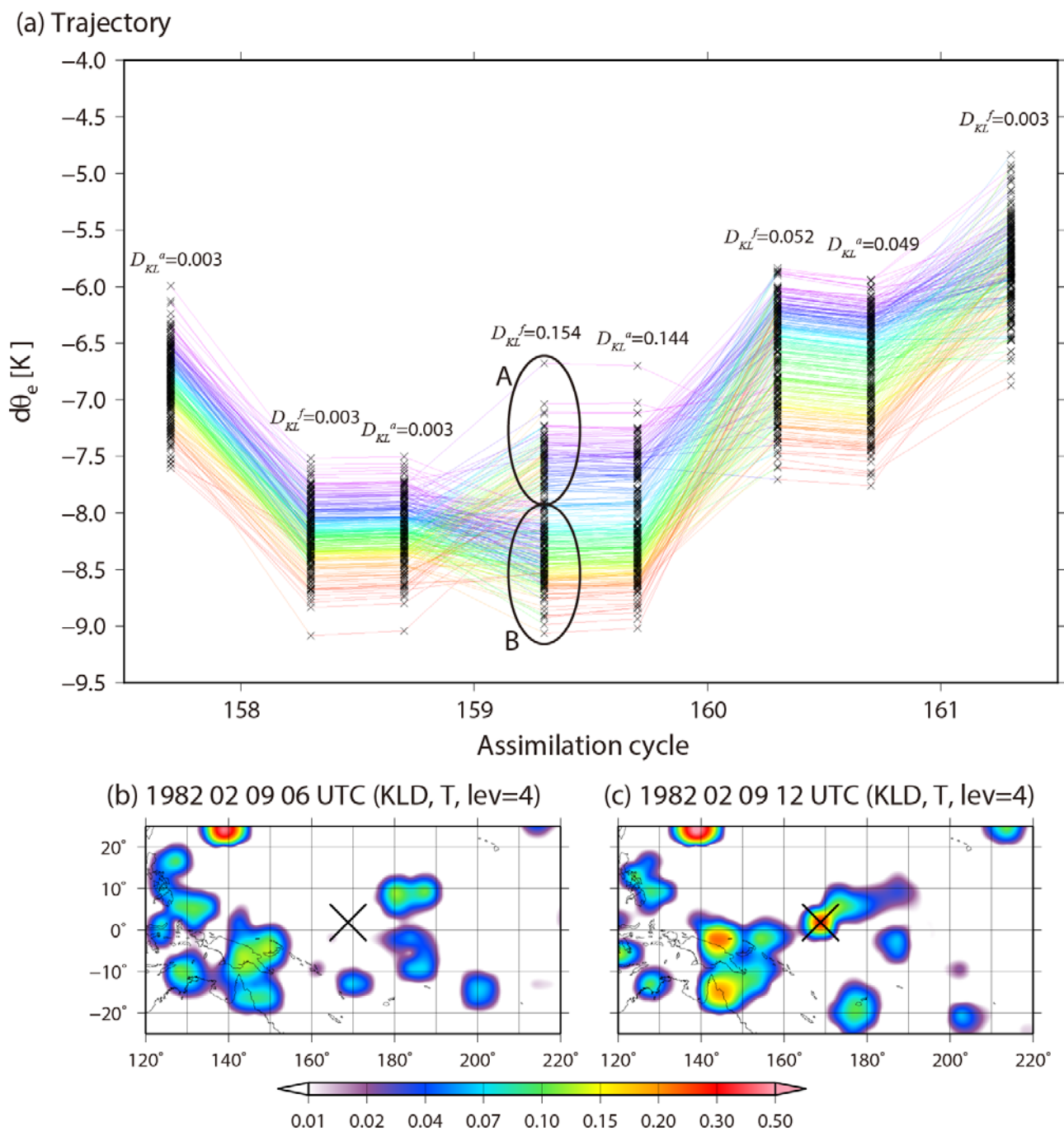



Figure 10: Lifecycle of non-Gaussianity at 1.9°N, 168.7°E. (a) Trajectories of 256 randomly chosen

members from 10240 members for $d\theta_e$ (see text for definition) from analysis at the 157th analysis

cycle (0000 UTC 9 February) to forecast the 161st analysis cycle (0000 UTC 10 February). The

colors show the order of $d\theta_e$ for every analysis. D_{KL} shows KL divergence for $d\theta_e$, and the

superscripts a and f indicate analysis and forecast, respectively. (b, c) Spatial distributions of KL

608   divergence for background temperature at the fourth model level (~500 hPa) at the 158th analysis

609   cycle (0600 UTC 9 February) and the 159th analysis cycle (1200 UTC 9 February), respectively.

610   The cross shows the location of the point considered in panel a.

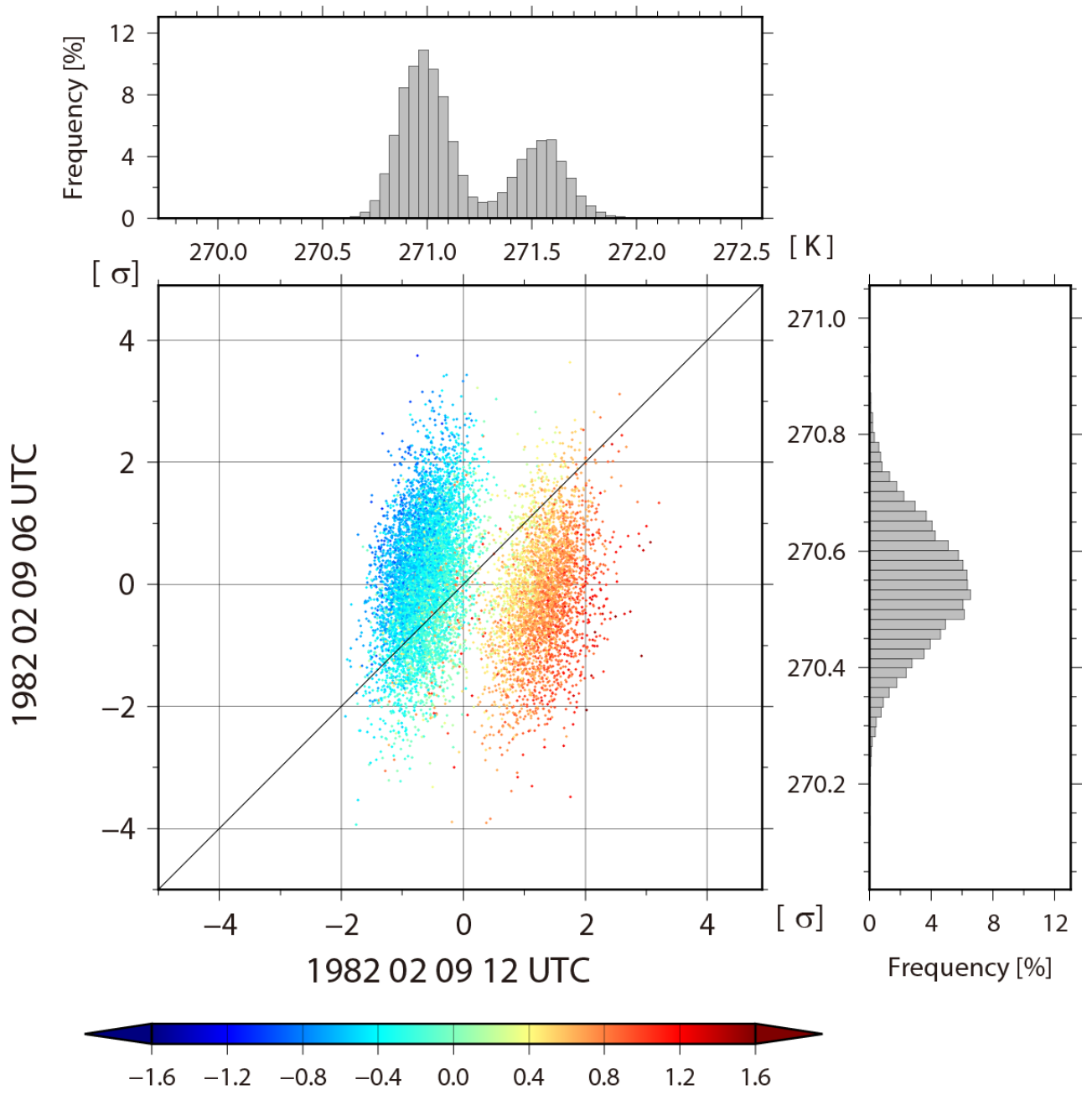



Figure 11: Scatter diagram of 0600 UTC versus 1200 UTC 9 February for the background

temperature at the fourth model level (~500 hPa) at 1.9°N, 168.7°E. The colors show $d\theta'_e =$

$(d\theta_{e\ 1200\ UTC} - d\theta_{e\ 0600\ UTC}) - (d\bar{\theta}_{e\ 1200\ UTC} - d\bar{\theta}_{e\ 0600\ UTC})$. The histograms on the right side

and upper side show the background temperature at the same grid point.

617

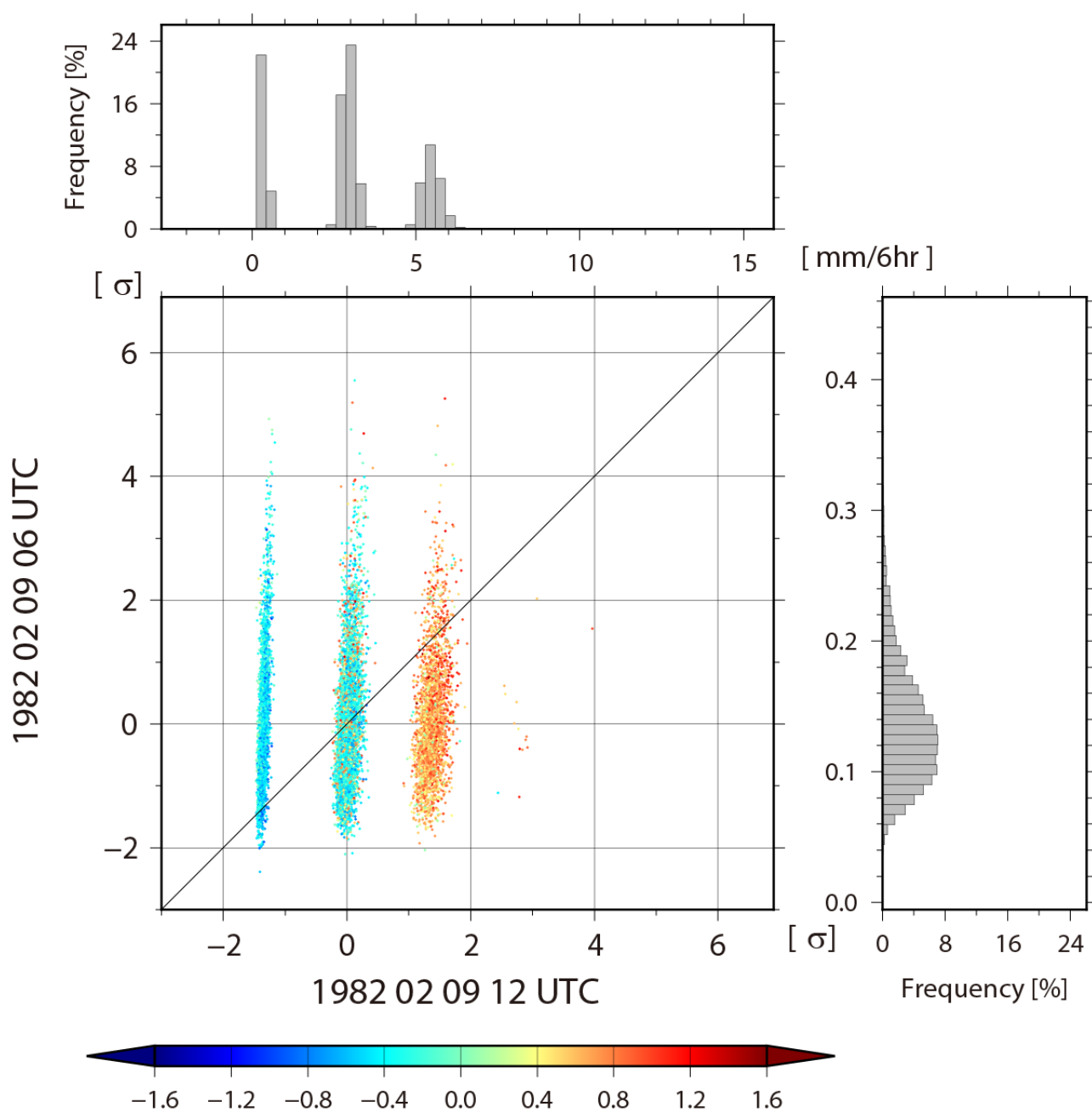



Figure 12: Similar to Fig. 11, but for 0600 UTC versus 1200 UTC 9 February for background

precipitation.

618

619

620

621

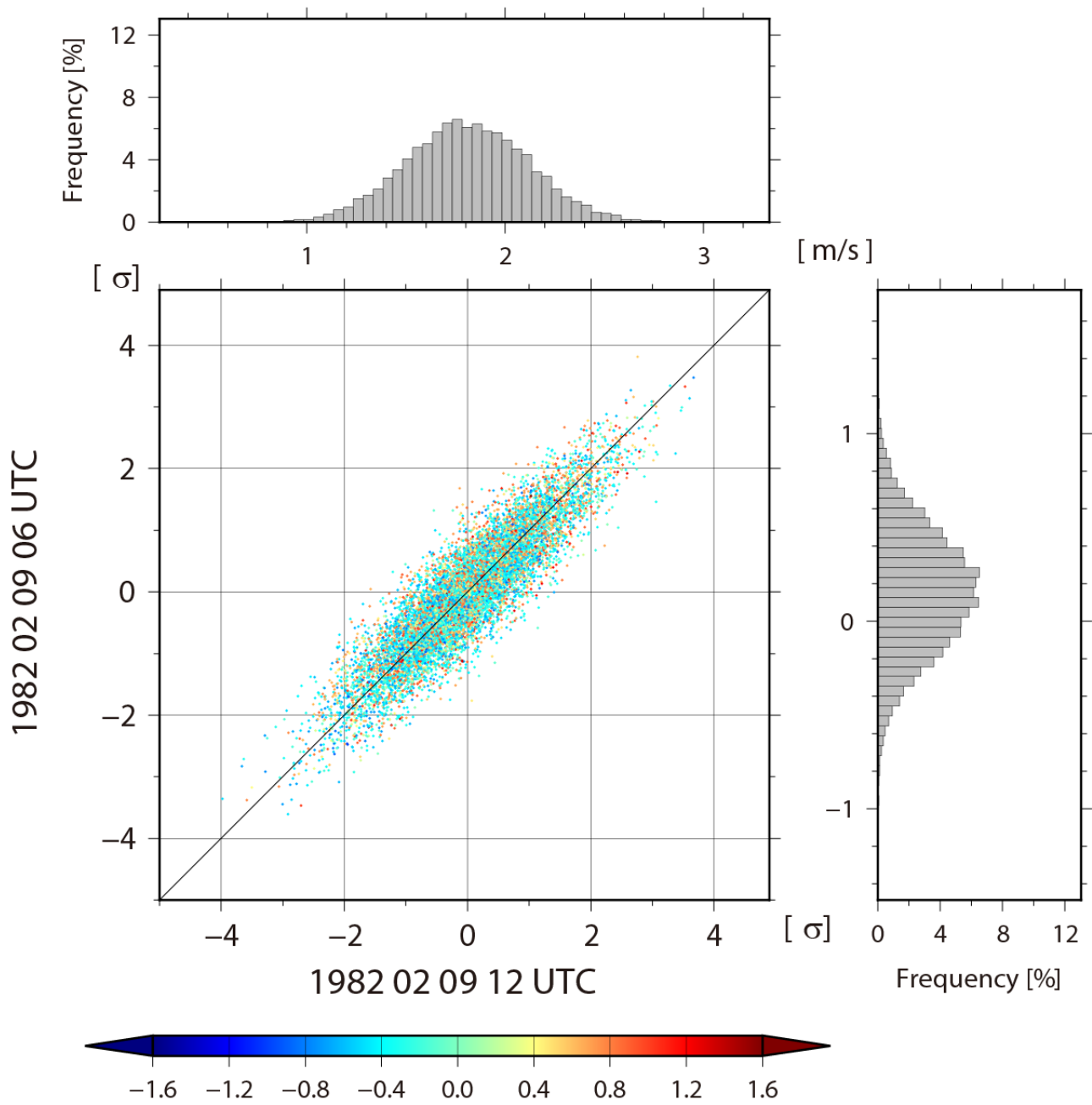



Figure 13: Similar to Fig. 11, but for 0600 UTC versus 1200 UTC 9 February for background zonal

wind at the fourth model level (~500 hPa).

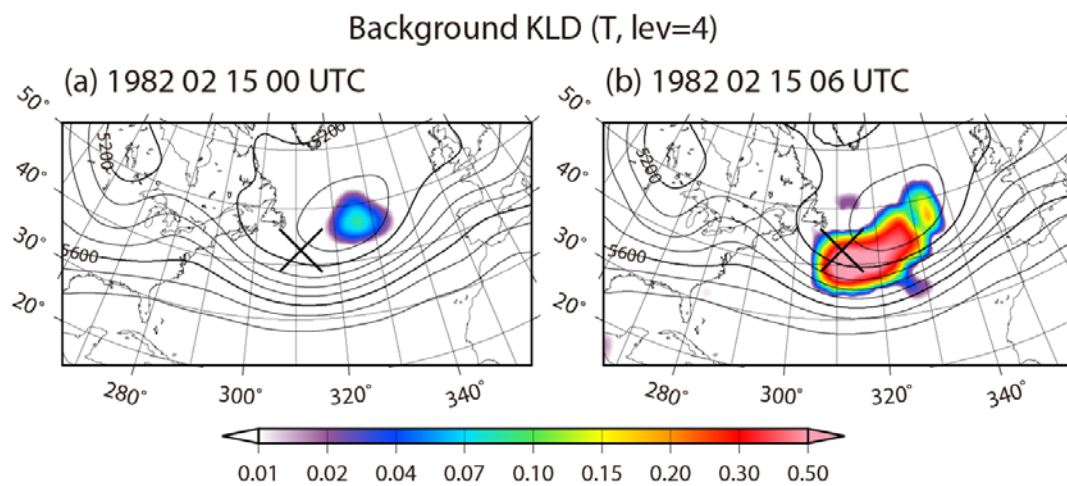



Figure 14: Spatial distributions of the KL divergence for background temperature at the fourth

model level (~500 hPa) (a) at 0000 UTC 15 February and (b) at 0600 UTC 15 February. Contours

show geopotential height of the ensemble mean at the 500 hPa level.

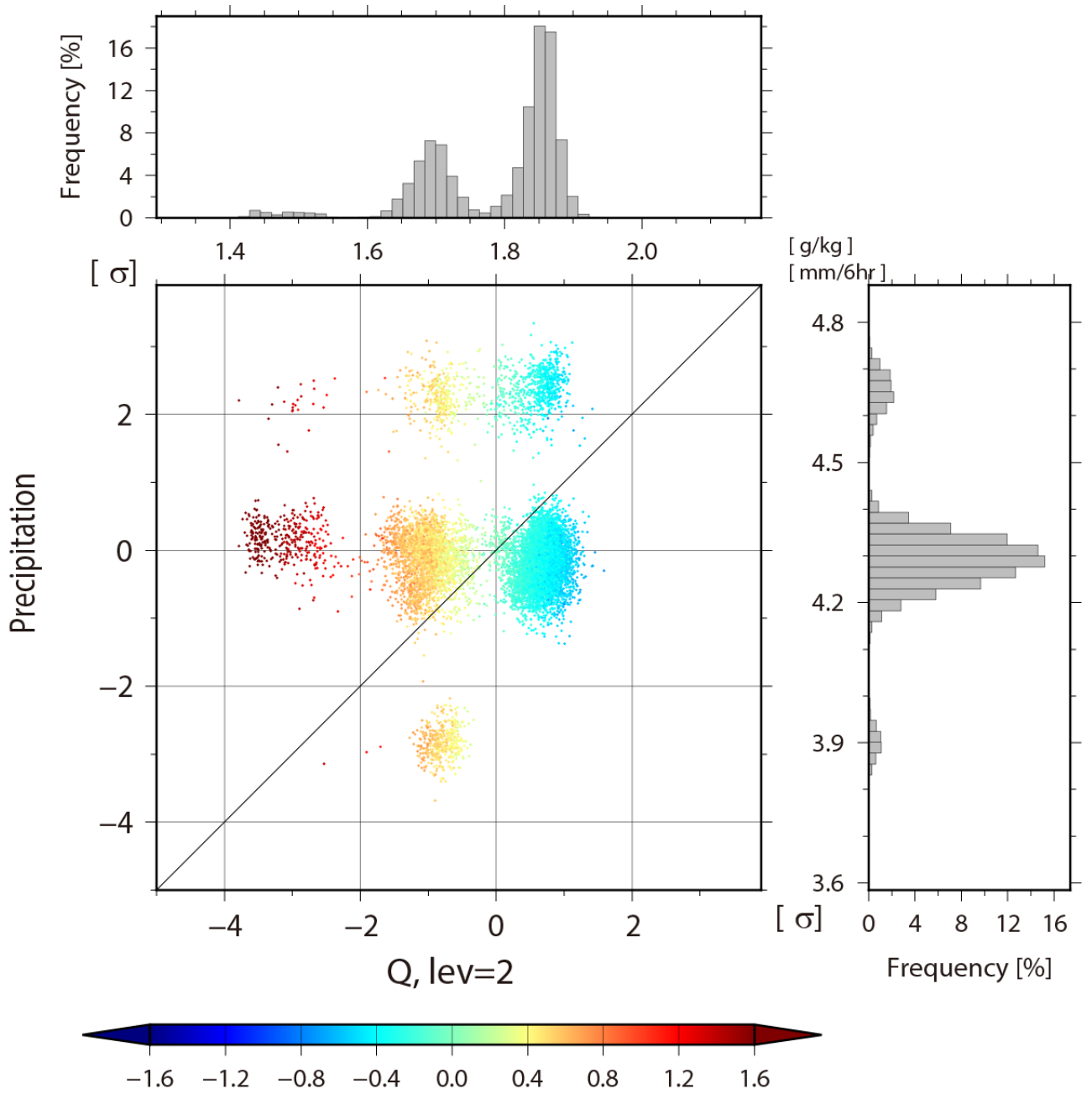



Figure 15: Scatter diagram of background specific humidity at the second model level (~850 hPa)

versus background precipitation at 42.7°N, 48.8°W (311.3°E) at 0600 UTC 15 February. The colors

show $d\theta'_e = (d\theta_{e\ 0600\ UTC} - d\theta_{e\ 0000\ UTC}) - (d\bar{\theta}_{e\ 0600\ UTC} - d\bar{\theta}_{e\ 0000\ UTC})$. The histograms on the

right side and on top show background precipitation and temperature at the same grid point,

respectively.

637

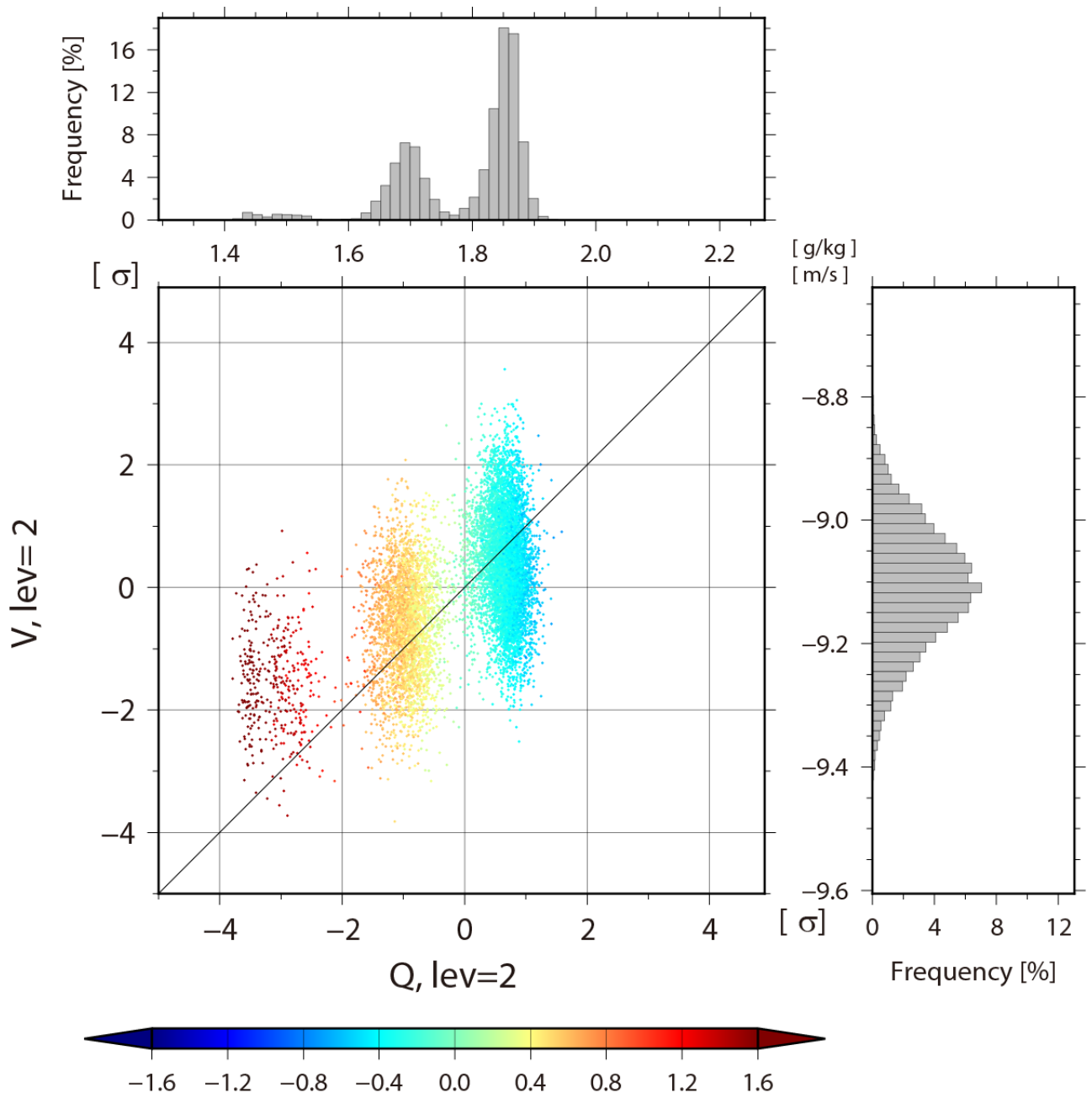



Figure 16: Similar to Fig. 14, but for background specific humidity versus meridional wind

639 background at the second level (~850 hPa).

640

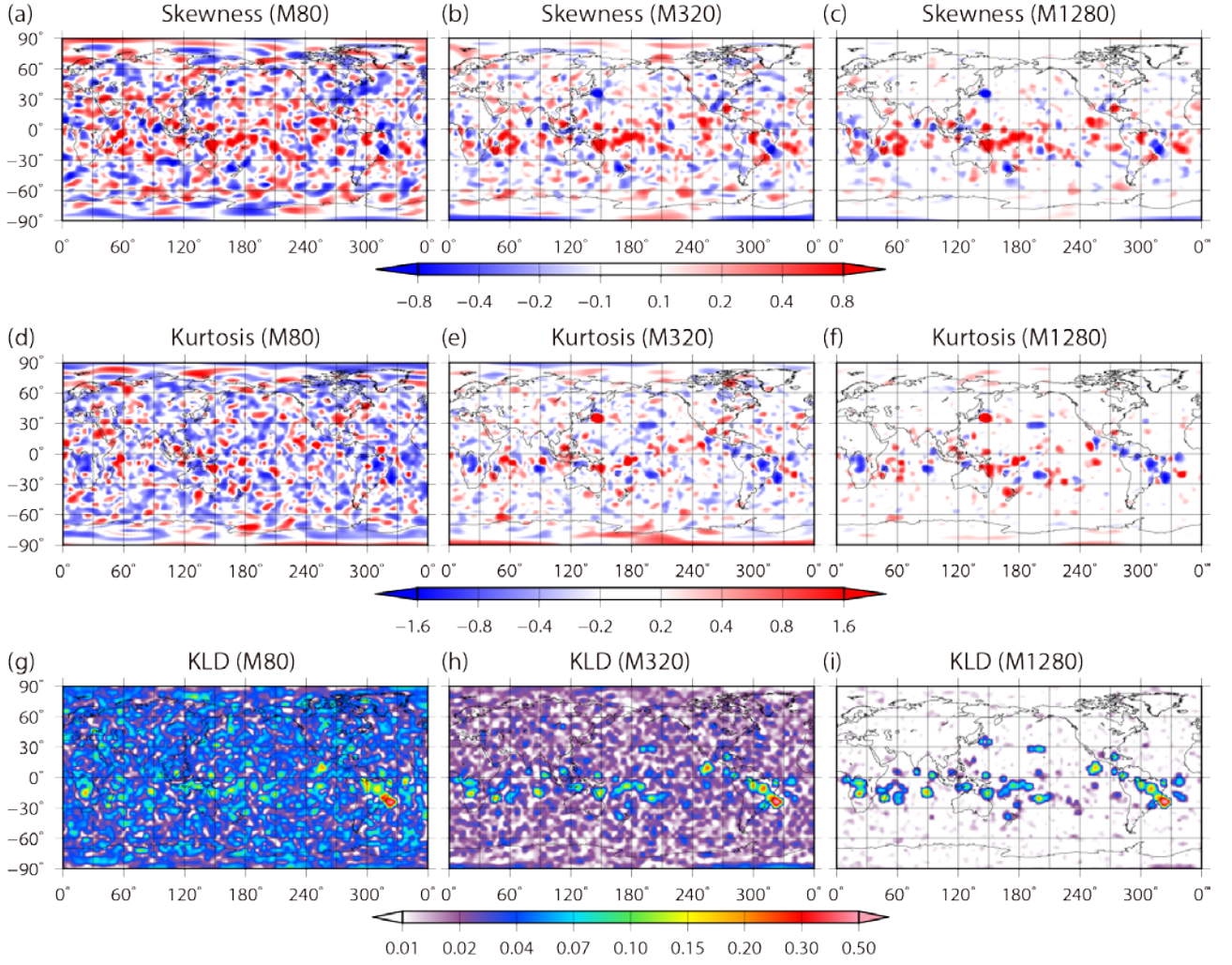



Figure 17: Spatial distributions of (a-c) skewness, (d-f) kurtosis, and (g-i) KL divergence for

temperature at the fourth model level (~500 hPa) at 0600 UTC 22 February. The left, center, and

right columns show 80, 320, and 1280 subsamples from 10240 members, respectively.

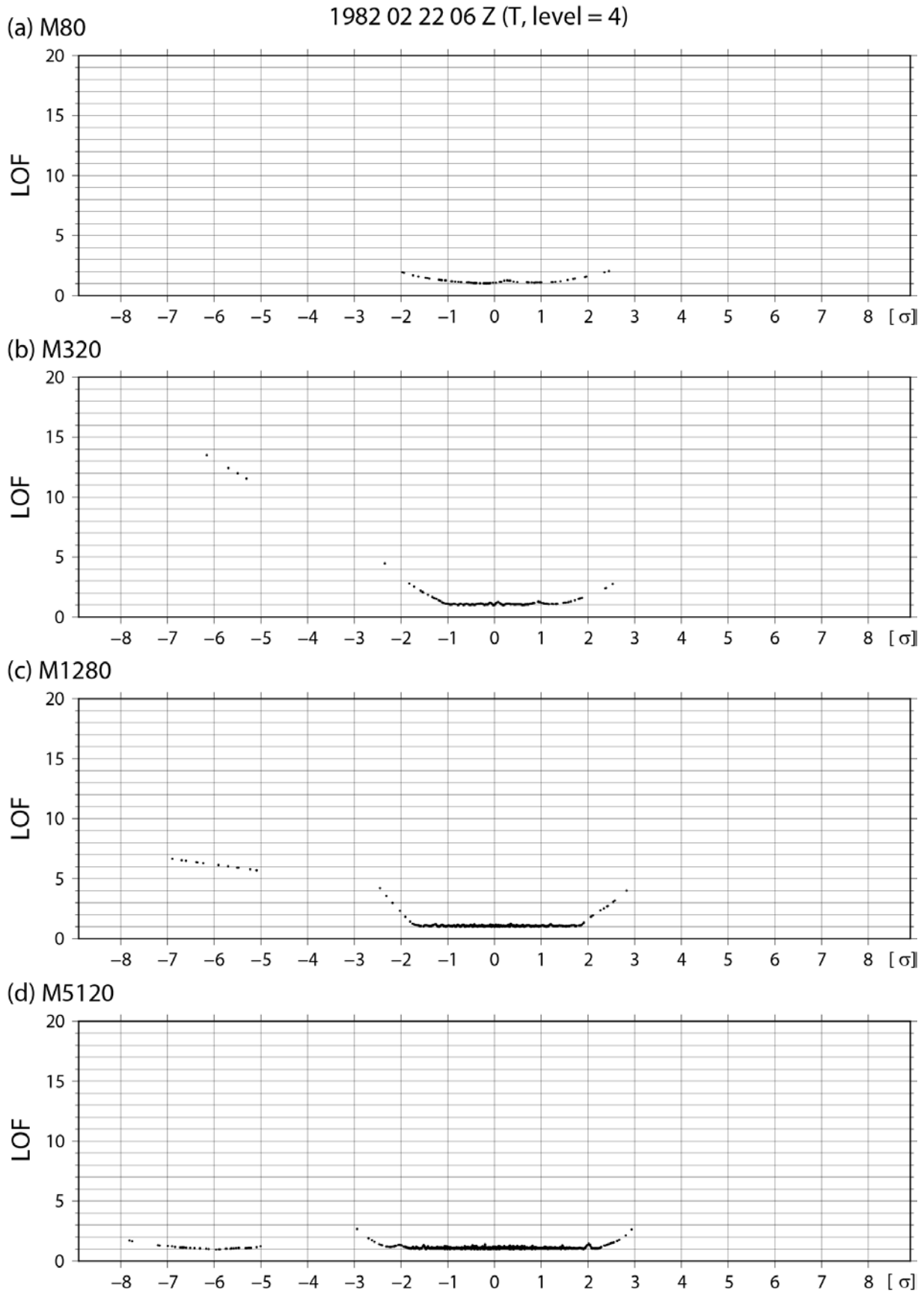
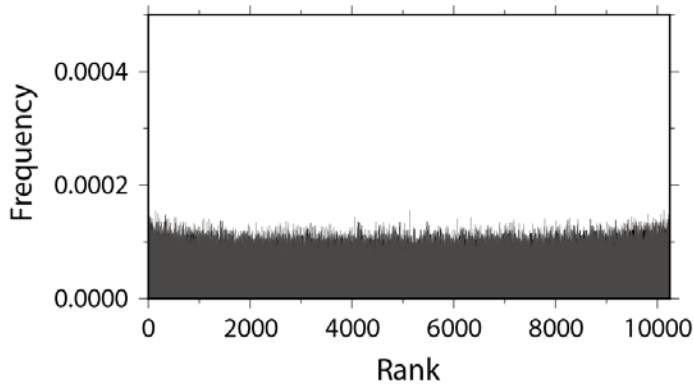
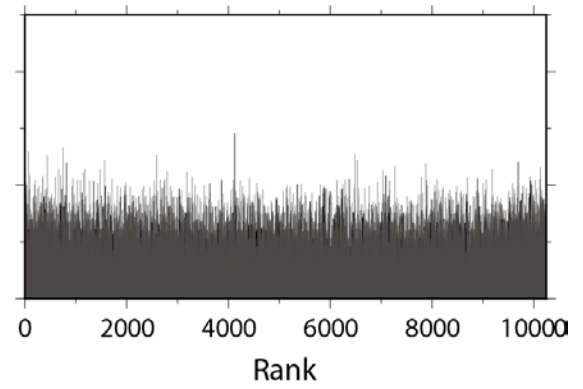



Figure 18: Similar to Fig. 5b, but for the ensemble sizes (a) 80, (b) 320, (c) 1280, and (d) 5120.

649

650 Figure 19: Rank histograms verified against truth for background specific humidity at the lowest

651 model level (~925 hPa) at (a) all grid points and (b) the grid points with non-Gaussian PDF from

652 0000 UTC 25 January to 1800 UTC 1 March.

653

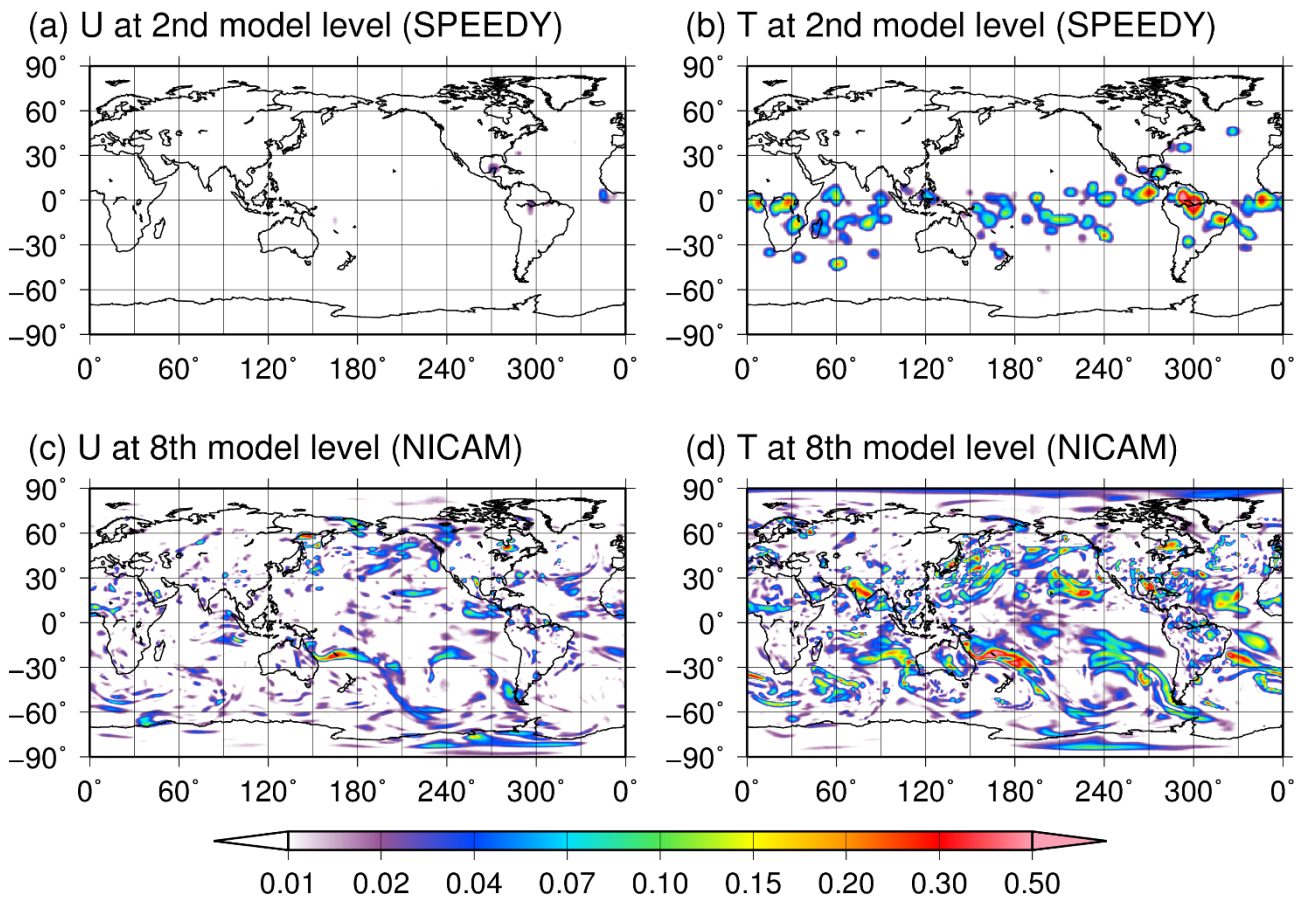



Figure 20: Spatial distributions of background KL divergence for SPEEDY model and NICAM.

Upper panels show (a) zonal wind and (b) temperature at the second model level (~850 hPa) for the

SPEEDY model at 0000 UTC 1 March. Bottom panels show (c) zonal wind and (d) temperature at

the eighth model level (~850 hPa) for NICAM at 0000 UTC 8 November 2011.

660 Table. 1: CRPS and its three components (reliability, resolution and uncertainty) for background

661 specific humidity at the lowest model level (~925 hPa) from 0000 UTC 25 January to 1800 UTC 1

662 March.

	CRPS [g kg^{-1}]	Reli [g kg^{-1}]	Resol [g kg^{-1}]	U [g kg^{-1}]
All grid points	0.0214	0.0000101	0.525	0.547
Grid points with non-Gaussian PDF	0.0475	0.0000244	0.030	0.077

663