

# Variational quality control of non-Gaussian innovations and its parametric optimizations for the GRAPES m3DVAR system

Jie HE<sup>1</sup>, Yang SHI<sup>2</sup>, Boyang ZHOU<sup>3</sup>, Qiuping WANG<sup>4</sup>, Xulin MA (✉)<sup>4</sup>

<sup>1</sup> Guangzhou Institute of Tropical and Marine Meteorology, China Meteorological Administration, Guangzhou 510640, China

<sup>2</sup> Guangdong Meteorological Observatory, Guangzhou 510640, China

<sup>3</sup> Qingdao Air Traffic Management Station of Civil Aviation of China, Qingdao 266108, China

<sup>4</sup> Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Key Laboratory of Meteorological Disaster, Nanjing University of Information Science and Technology, Nanjing 210044, China

© Higher Education Press 2022

**Abstract** The magnitude and distribution of observation innovations, which have an important impact on the analyzed accuracy, are critical variables in data assimilation. Variational quality control (VarQC) based on the contaminated Gaussian distribution (CGD) of observation innovations is now widely used in data assimilation, owing to the more reasonable representation of the probability density function of innovations that can sufficiently absorb observations by assigning different weights iteratively. However, the inaccurate parameters prevent VarQC from showing the advantages it should have in the GRAPES (Global/Regional Assimilation and PrEdiction System) m3DVAR system. Consequently, the parameter optimization methods are considerable critical studies to improve VarQC. In this paper, we describe two probable CGDs to include the non-Gaussian distribution of actual observation errors, Gaussian plus flat distribution and Huber norm distribution. The potential optimization methods of the parameters are introduced in detail for different VarQCs. With different parameter configurations, the optimization analysis shows that the Gaussian plus flat distribution and the Huber norm distribution are more consistent with the long-tail distribution of actual innovations compared to the Gaussian distribution. The VarQC's cost and gradient functions with Huber norm distribution are more reasonable, while the VarQC's cost function with Gaussian plus flat distribution may converge on different minimums due to its non-concave properties. The weight functions of two VarQCs gradually decrease with the increase of innovation but show different shapes, and the VarQC with Huber norm distribution shows more elasticity to assimilate the

observations with a high contamination rate. Moreover, we reveal a general derivation relationship between the CGDs and VarQCs. A novel schematic interpretation that classifies the assimilated data into three categories in VarQC is presented. They are conducive to the development of a new VarQC method in the future.

**Keywords** data assimilation, variational quality control, contaminated Gaussian distribution, non-Gaussian distribution, innovation

## 1 Introduction

The density distribution function of observation errors and background errors play an important role in the variational data assimilation (VarDA). VarDA expects the background errors and observation errors to be Gaussian distribution. With this hypothesis in mind, the estimation of its posterior analysis will be the Best Linear Unbiased Estimation (Swinbank et al., 2003; Pires et al., 2010). In general, the assumption of Gaussian error distribution has been the basis of data assimilation (DA) theory and is widely used in the development of DA system (Parrish and Derber, 1992; Courtier et al., 1998; Houtekamer and Mitchell, 1998; Rabier et al., 2000).

Nevertheless, the actual errors are not necessarily fully Gaussian. For example, many observation errors are non-Gaussian distribution that shows longer tails compared to the pure Gaussian distribution (Tukey, 1960; Fernández and Steel, 1998). The outliers are well outside of the expected magnitude of observations, thus leading to a non-Gaussian innovation (observation minus background) distribution (Tavolato and Isaksen, 2015). The assimilation

of these outliers is believed to hamper posterior analysis. Therefore, the outliers' quality control will be a critically important process to improve posterior analysis.

As one quality control method, the variational quality control (VarQC) and VarDA are mutually compatible. In other words, VarQC known as online quality control was performed simultaneously with the VarDA (Lorenc and Hammon, 1988), rather than in a separate step before the assimilation as in traditional offline quality control. It is assumed that the non-Gaussian innovation distribution comes from the non-Gaussian observation (outlier) errors, while the background error is of a Gaussian distribution (He et al., 2021). This non-Gaussian VarQC idea (Lorenc, 1988) was performed for laser Doppler wind observations, which postulated the gross error with 50% probability (Dharsai et al., 1992). Here the gross errors, which are commonly identified as "bad" data or outliers, are assumed to have come from instrument errors, human origins, or during transmission of the observations (Kalnay, 2003). Considering the long tails caused by outliers, Ingleby and Lorenc (1993) derived a VarQC theory by using the probability density function (PDF) of Gaussian plus a flat distribution of the observation errors. Subsequently, this VarQC algorithm was applied at the European Center for Medium-Range Weather Forecast (ECMWF) in the four-dimensional variational (4DVAR) system (Anderson and Jarvinen, 1999). Besides, the Huber norm, which was a non-Gaussian PDF, was introduced to VarDA as an observation error distribution to develop another VarQC (Tavolato and Isaksen, 2015). This study also revealed that not all outliers are gross errors but instead are probably genuine or important observations, as described by Hampel (2001). For example, in the situation where a forecasted thunderstorm was substantially displaced from the observed thunderstorm, the wind site was located between the forecasted thunderstorm and the observed one. This wind, a correct observation, is likely to be treated as gross error due to the anomalous innovation. With the Huber-norm VarQC, the prediction system greatly improved the accuracy of mesoscale weather simulations, particularly for typhoons. Currently, VarQCs strategies with different non-Gaussian PDFs have been adopted in many DA systems. VarQC, in general, performs well in operational prediction systems (Gauthier et al., 2003; Su and Pursera, 2013; Wang and Lei, 2014; Storto, 2016) compared with the systems that only the conventional quality control methods were used, including the background quality check (BgQC; Jarvinen and Uden, 1997), extreme value check and other methods.

Two VarQCs were established in the GRAPES (Global and Regional Assimilation and Prediction) m3DVAR (the three-dimensional variational on model layer) system in our prior work (Ma et al., 2017; He et al., 2021). Despite the improvement of the mass field compared to the control experiment, the GRAPES VarQCs did not produce an apparent advantage for wind, specific humidity and the upper levels' mass field, possibly because of the inaccurate parameters setting at different observation

variables and levels. Consequently, this study aims to provide the potential parameter optimization methods to overcome the inaccurate parameters problem in the current GRAPES VarQC system. Meanwhile, a contaminated Gaussian distribution for observation errors is proposed to reveal a general derivation relationship with the present VarQCs. This derivation system will be utilized to develop other new VarQCs in the future.

The paper is organized as follows. The non-Gaussian CGD mode following the Gaussian errors mode is proposed in Section 2. The potential parameters optimization methods involved in VarQCs are provided for the GRAPES m3DVAR system in Section 3. Subsequently, the potential influences and sensitivities of different setting parameters for VarQCs are analyzed in Section 4, as well as a novel schematic interpretation of VarQC. Finally, the conclusions and discussions are given in Section 5.

## 2 Contaminated Gaussian distribution

The cost function in the GRAPES m3DVAR assimilation (Ma et al., 2009) is defined as

$$J = J_b + J_o. \quad (1)$$

On the right side are the background cost function ( $J_b$ ), which will not be changed in VarQC, and the observational cost function ( $J_o$ ), which will be updated in VarQC. The observation error distribution, in general, obeys the following Gaussian PDF:

$$P_G = \frac{1}{\sqrt{2\pi}\sigma_o} \exp\left[-\frac{1}{2}\left(\frac{y-Hx}{\sigma_o}\right)^2\right], \quad (2)$$

where the subscript  $G$  denotes "Gaussian".  $y$  and  $Hx$  represent, respectively, the observed value and its model equivalent.  $H$  is the observation operator.  $\sigma_o$  is the observational error standard deviation. Under the condition that observation error and background error are Gaussian distribution and mutually independent, the observational cost function can be derived by substituting  $P = P_G$  into the following formula:

$$J_o = -\ln P + C, \quad (3)$$

where  $C = -\ln(\sqrt{2\pi}\sigma_o)$ , and is a constant. Subsequently, the observation cost function of the pure 3DVAR and its gradient can be expressed as follow:

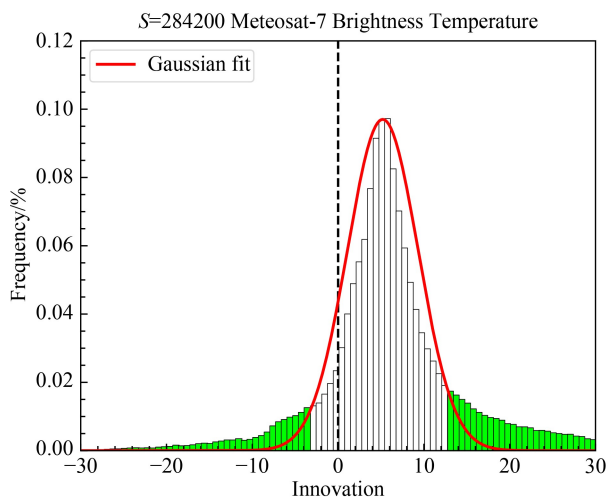
$$J_G = \frac{1}{2}\left(\frac{y-Hx}{\sigma_o}\right)^2, \quad (4)$$

$$\nabla J_G = -\frac{1}{\sigma_o}\left(\frac{y-Hx}{\sigma_o}\right). \quad (5)$$

VarDA also assumes that the probability of gross errors (outliers) must be zero (after "perfect" quality control). In

view of this, all the outliers in the observation data set would be removed by using a conventional quality control strategy. However, the outlier probability can never equal zero after an imperfect conventional quality control. This can be illustrated by the innovation statistics (Fig. 1) of the geostationary meteorological satellite (Meteosat-7) observations over a limit area and at a random time. Note that the Gaussian profile is fitted with a similar expectation-maximization algorithm (Sondergaard and Lermusiaux, 2013) which is performed by searching the ideal means and variances in a range near the real means and variances of innovations. The green boxes show a non-Gaussian distribution with two apparent long tails beyond the Gaussian distribution (red line), suggesting that the outlier probability is not zero. Hence, due to the existence of these outliers, the innovation (observation errors) distributions should always be non-Gaussian distributions with longer tails. It is well-known that the inconsistency between the assumed observational error distribution mode (OEDM: Zhu and Zeng, 1999) and the actual mode can greatly reduce the effectiveness of DA, or even result in wrong analysis (Fowler and van Leeuwen, 2013; Legrand et al., 2016).

To fit the distribution of the errors more realistically, a mixed-Gaussian distribution, coined as the contaminated Gaussian distribution (Tukey, 1960), is put forward aimed to reduce the bad impacts of outliers on the posterior analysis. Basically, a mixed-Gaussian distribution represents a “main” distribution slightly “contaminated” by a wider Gaussian distribution which has the same



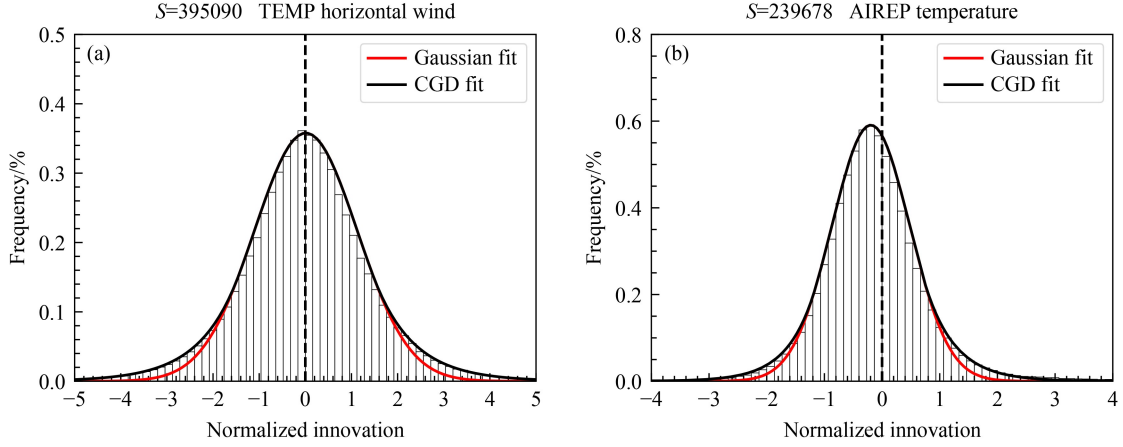
**Fig. 1** The fitted Gaussian distributions (histogram) of innovation of brightness temperature from geostationary meteorological satellite (Meteosat-7). The green boxes indicate the non-Gaussian long-tail histograms (distribution). The dashed line represents the approximation between observation and background (ERA5 reanalysis). The  $x$ -axis represents the size of the innovation; the  $y$ -axis represents frequency.  $S$  shows the total sample number of observations at 1200 UTC 10 June 2015.

mean as the “main” distribution but with a larger variance. In this way, the long tails of observation error distribution caused by outliers can be represented reasonably with this mixed-Gaussian CGD. In general, the Gaussian distribution as the “main” distribution plus some other distribution (including Gaussian distributions with different means and variances) as the “perturbing distribution” is defined as the CGD in subsequent studies. The PDF of a CGD is generally constructed with a combination of two distributions:

$$P_{CGD} = (1 - \varepsilon)P_G + \varepsilon P_c, \quad (6)$$

where  $\varepsilon$  is the contamination rate (the prior probability of outliers),  $P_G$  is the “main” (Gaussian) distribution, and  $P_c$  is the contaminating (perturbing) distribution. The subscript  $G$  and  $c$  denote “Gaussian” and “contaminating”, respectively. Unlike Gaussian distribution, the CGD can better fit the actual observation error distribution that is widely used to deal with the long tail problem in the field of surveying and mapping (Yang, 1991; Zhu, 1996).

To illustrate the representativeness of CGD for the actual observational error in meteorology, Fig. 2 shows the normalized innovation histograms and their fitted profiles for the Global Telecommunications System (GTS) observations that are assimilated in the GRAPES m3DVAR system after the BgQC. The fitted profiles come from two distributions’ PDF, including a Gaussian PDF (red lines) and a Huber norm PDF (black lines) which is one of the CGDs to be described in Section 3.2. The fitting method is the same as Fig. 1. As shown in Fig. 2(a), the CGD profile represents a more reasonable distribution for the innovations of the horizontal radiosonde (TEMP) wind compared to the pure Gaussian distribution. The difference between the two profiles is that the tails of the CGD are longer than those of the Gaussian distribution. The long tails fitted by CGD are also found in the normalized temperature innovation (Fig. 2(b)) from aircraft-reported observation (AIREP). It is visible that the CGD is close to the Gaussian distribution near the center of innovations since its “main” distribution is also a Gaussian distribution within the Gaussian-fitted domain (approximately within  $(-1.5, 1.5)$  in Fig. 2(a)). The CGD, however, can match the distribution of the actual innovations better than the pure Gaussian distribution beyond the Gaussian-fitted domain. Moreover, the skewness of the normalized innovation distribution is not equal to zero (not symmetry), particularly in Fig. 2(b), indicating that it must be a non-Gaussian distribution. Therefore, the CGD, a non-Gaussian OEDM, can give a better-fitted innovation distribution in the DA systems. The other observational innovations from the same period also have similar long-tails features (Zhao et al., 2015; He et al., 2021). It should be noted that the non-Gaussianity of innovation can come from observation errors, background errors, or both, as discussed by Pires et al. (2010). In this study, we only consider the non-Gaussianity of observation errors caused by outliers.



**Fig. 2** The distribution (histogram) statistics of normalized innovation by using the horizontal wind (a) and the temperature (b) from sounding and aircraft-reported observations are fitted with the pure Gaussian distribution (red line, Gaussian fit) and the CGD fit (black line). The x-axis represents the size of the innovation; the y-axis represents frequency.  $S$  shows the total sample number of observations from August 2013.

### 3 Parametric optimizations of VarQCs

#### 3.1 Flat-VarQC parameters

In the CGD, a flat distribution (box-car; [Lahoz and Schneider, 2014](#)) is often chosen as a perturbed distribution. The Gaussian plus flat distribution is designed to match the non-Gaussian innovation distribution to make VarDA absorb the long-tails outliers effectively. These outliers can be of use but would be removed by conventional quality control through gross check. This CGD considers the available outliers without gross error in the gray zone to have caused the flat distribution. In this way, the contamination rate should be the prior probability of outliers without gross error in the assimilated observations. According to this, the Gaussian plus flat distribution is constructed by choosing the flat distribution  $F$  as the perturbing distribution of outliers. The hypothesis of no correlation in observation errors promises that this approach can be calculated conveniently. The Gaussian plus a flat distribution's PDF ( $P_F$ ) is defined as follows:

$$P_F = (1 - \varepsilon)P_G + \varepsilon F, \quad (7)$$

$$F = \begin{cases} \frac{1}{2d\sigma_o} & |y - Hx| < d\sigma_o \\ 0 & |y - Hx| \geq d\sigma_o \end{cases}, \quad (8)$$

where  $d$  represents the maximum multiple of observational error standard deviation allowed for outliers (the width of  $F$ ). The observations falling into  $|y - Hx| \geq d\sigma_o$  ( $F = 0$ ) are discarded by the extreme check or background quality check of conventional quality control. After substituting  $P_F = P$  into Eq. (3), the observational cost function and gradient of this CGD in the 3DVAR method are derived as

$$J_F = -\ln\left(\frac{\gamma + \exp(-J_G)}{\gamma + 1}\right), \quad (9)$$

$$\nabla J_F = \nabla J_G W_F, \quad (10)$$

$$W_F = 1 - P_{\text{posterior}} = 1 - \frac{\gamma}{\gamma + \exp(-J_G)}, \quad (11)$$

where  $\gamma = \varepsilon \sqrt{2\pi}/2d(1 - \varepsilon)$ ,  $W_F$  and  $P_{\text{posterior}}$  are the analytical weighting function and the posterior probability of outliers, respectively. The other symbols are the same as in the previous definition. Eqs. (9)–(11) build the basis of using the Gaussian plus flat CGD in VarQC (referred to as “Flat-VarQC” hereafter) in VarDA.

The parameter  $\gamma = \varepsilon \sqrt{2\pi}/2d(1 - \varepsilon)$  in Flat-VarQC is a constant once known  $\varepsilon$  and  $d$ . But it can also be obtained approximately without known  $\varepsilon$  and  $d$  based on the statistics of the normalized analysis residual  $((y - Hx_a)/\sigma_o)$  between historical observation and analysis field. This method is similar to the derivation of rejection thresholds in BgQC. The only difference is that the rejection thresholds in BgQC are derived from the statistics of the normalized innovations  $((y - Hx_b)/\sigma_o)$  with historical data. Therefore, the analysis residual should satisfy the following formulation:

$$|y - Hx_a| \leq \alpha \sqrt{\sigma_o^2 + \sigma_a^2}, \quad (12)$$

where  $\alpha$  is the statistical rejection coefficient of BgQC. The GRAPES m3DVAR system has provided  $\alpha$  value in this study. For different observations and variables,  $\alpha$  takes different values to reflect the sensitivity. The variance of the analysis residual is expressed as the sum of observation error variance and analysis error variance, i.e.  $\sigma_o^2 + \sigma_a^2$ . [Figure 3\(a\)](#) shows a frequency distribution of AIREP temperature data from 1 to 14 August 2013. The variance of the analysis residual (the departure from GRAPES 3DVAR analysis) is computed by fitting a pair of straight lines based on a new histogram ([Fig. 3\(b\)](#)) of the analysis residual ([Hollingsworth, 1989; Anderson and](#)

Jarvinen, 1999). The new histogram transferred from the histogram of analysis residual is defined as

$$f_{\text{new}} = \sqrt{-2\ln\left[\frac{f}{\max(f)}\right]}, \quad (13)$$

where  $f$  is the number of observations corresponding to each box of the analysis residual histogram (Fig. 3(a)). In this manner, the symmetric straight lines in Fig. 3(b) define the standard deviation of the Gaussian profile shown in Fig. 3(a). The standard deviation of the analysis residual is indicated with the inverse of the positive slope of the straight lines ( $y = \text{abs}(x/\lambda)$ ), where

$$\lambda = \sqrt{\sigma_o^2 + \sigma_a^2}. \quad (14)$$

Unlike BgQC, VarQC does not have specific rejection thresholds and alleviates (promotes) the negative (positive) impacts from outliers by reducing the weighting of innovation (e.g.  $W_F < 1$ ) iteratively. The rejection intensity of VarQC is related to the posterior probability  $P_{\text{posterior}}$  which is the “rejected” probability of innovations, and their relationship is shown as follows based on Eqs. (5) and (11):

$$|y - Hx_b| = \sqrt{2\sigma_o^2 \ln \frac{P_{\text{posterior}}}{\gamma(1 - P_{\text{posterior}})}}. \quad (15)$$

The outliers that have caused the non-Gaussian CGD will result in larger innovations that are flagged to reduce weight using VarQC during each iteration. Suppose that the innovations with a weight less than 0.25 are represented to be “rejected” in VarQC, i.e., the “rejected” probability of this outlier  $P_{\text{posterior}}$  will equal to 0.75 based on Eq. (11), the rejection threshold ( $\rho$ ) concerning  $\gamma$  for this observation can be defined as the maximum of  $|y - Hx_b|$  based on Eq. (15):

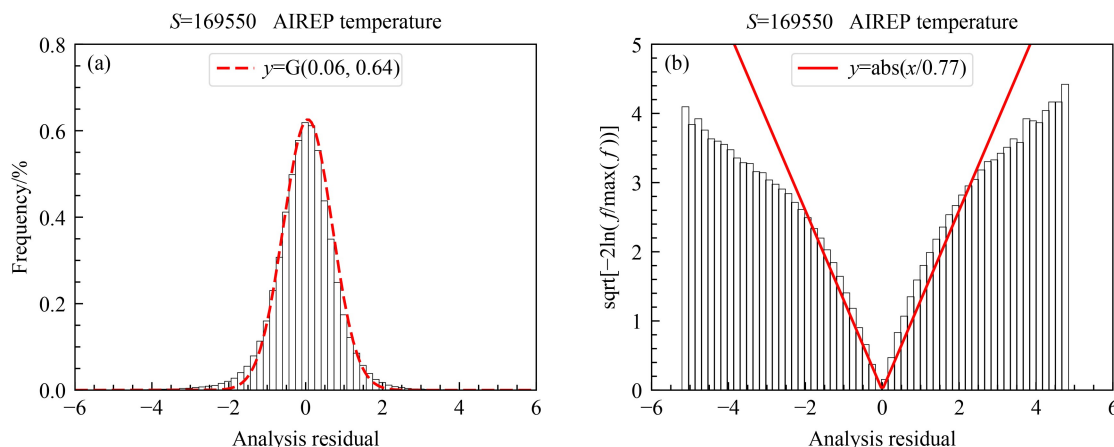
$$\rho = \sqrt{2\sigma_o^2 \ln \frac{P_{\text{posterior}}}{\gamma(1 - P_{\text{posterior}})}} = \sqrt{2\sigma_o^2 \ln\left(\frac{3}{\gamma}\right)}. \quad (16)$$

The rejection threshold from Eq. (16) can be defined approximately to equal the right-hand side of Eq. (12) based on Anderson and Jarvinen (1999), i.e.,  $\rho = \alpha\lambda$ . Then, the parameter  $\gamma$  can be obtained from the following equation:

$$\gamma = \exp\left(\ln 3 - \frac{\alpha^2 \lambda^2}{2\sigma_o^2}\right). \quad (17)$$

It is an approximate rejection limit when the innovation assigns a weight of 0.25. But VarQC does not require the specified threshold limits. Based on the statistic observations in Fig. 3, the rejection coefficient  $\alpha = 4.0$  (Table 1), the inverse of the slope  $\lambda = 0.77$  from Eq. (14) and the observation error variance  $\sigma_o^2 = 1$  reported by AIREP observations,  $\gamma \approx 0.026$  can be obtained for the case given. In this way, the contamination rate can be computed with a specified  $d$  which is similar to a rejection coefficient. The parameter  $\gamma$  of other observation types and variables are calculated similarly by Eq. (17).

As a case statistic, we retrieved the approximate rejection threshold of VarQC for other GTS observations. These historical observations statistics (Table 1) are spanning two weeks from 1 to 14 August 2013. SYNOP and SATOB represent surface and satellite wind observation, respectively. For the rejection threshold of BgQC in Table 1, the approximate range is due to that the observed variance is reported in a confidence interval. The parameters  $\lambda$  and  $\alpha$  are crucial to determine the Flat-VarQC parameters. Table 1 also shows that the rejection limit of Flat-VarQC is generally smaller than that of BgQC, indicating that Flat-VarQC will repeat the quality control for observations with a large deviation after BgQC. In this way, the outliers with very large



**Fig. 3** Histogram (a) and transformed histogram (b) of temperature departure from analysis from the AIREP observations collected from 1 to 14 August 2013. The slope of the straight lines fitted in (b) defines the standard deviation of the Gaussian curve (dashed line) drawn in (a). The red dashed label represents a Gaussian curve with a mean of 0.06 and a variance of 0.64. The solid red lines are displayed by an expression in  $y = \text{abs}(x/0.77)$ . The top  $S$  shows the total sample number of observations.

**Table 1** Statistics of rejection threshold for Flat-VarQC and BgQC in the GRAPES m3DVAR system. The horizontal wind (u and v), relative humidity (rh), pressure, and temperature (temp) of different observations are obtained from the standard GTS observations over the Chinese mainland (10°N – 60° N, 70°E – 140° E). These observations are collected four times every day spanning the period from 1 to 14 August 2013

Obs.	Variables	Sample	max( <i>f</i> )	$\lambda$	$\alpha$	Approximate rejection threshold	
						Flat-VarQC	BgQC
TEMP	wind u	177089	16162	1.81	4.0	7.24 m/s	7.2–12 m/s
	wind v	177089	15064	1.81	4.0	7.24 m/s	7.2–12 m/s
	rh	101827	7712	9.13	3.5	32.0%	59.5%
	pressure	148670	18451	0.30	2.5	0.75 hPa	0.4–2.3 hPa
AIREP	wind u	170138	18296	1.75	4.0	7.00 m/s	10–16 m/s
	wind v	170138	17416	1.75	4.0	7.00 m/s	10–16 m/s
	temp	169550	17565	0.77	4.0	3.08 K	4.8–5.6 K
SYNOP	rh	64767	5066	9.21	4.0	36.8%	52–92%
	pressure	26345	1930	0.60	4.0	2.40 hPa	2.9–3.2 hPa
SATOB	wind u	8984	795	2.27	4.0	9.08 m/s	8–20 m/s
	wind v	8984	949	2.27	4.0	9.08 m/s	8–20 m/s

innovations will be discarded (referred to the weight less than 0.25) to ensure that the analysis is free from “bad” observations. Note that the weight of valid observations will not be reduced too much to be approximately 1 due to the assumption of Gaussian distribution over the Gaussian domain. However, the parameter of short-term case statistics cannot fully represent that of the current real-time assimilated observation. Consequently, the parameter  $\gamma$  from the long-period statistics is necessary, and so are the further sensitive experiments to Flat-VarQC parameters.

### 3.2 Huber-VarQC parameters

As a powerful representation in robust statistics (Huber, 2011), the Laplace distribution is used frequently as the perturbed distribution to deal with outliers. The PDF expression of “Gaussian plus Laplace distribution” (Tavolato and Isaksen, 2015) is defined as

$$P_L = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma_o} \exp(-\frac{1}{2}\delta^2) & |\delta| \leq c \\ \frac{1}{\sqrt{2\pi}\sigma_o} \exp(\frac{c^2}{2} - c|\delta|) & |\delta| > c \end{cases}, \quad (18)$$

where  $\delta = (y - Hx)/\sigma_o$ . Equation (18) is the PDF of Huber (norm) distribution, which is a commonly used distribution in the robust statistical theory. The robustness of the Huber norm, which refers to the insensitivity of small differences between the actual and the specified OEDM, is regularly used to respond to the impacts of outliers. After substituting  $P_L = P$  into Eq. (3), the cost function and gradient for this CGD are derived as follows:

$$J_L = \begin{cases} J_G & |\delta| \leq c \\ c\sqrt{2J_G} - \frac{c^2}{2} & |\delta| > c \end{cases}, \quad (19)$$

$$\nabla J_L = \begin{cases} \nabla J_G & |\delta| \leq c \\ \nabla J_G \frac{c}{\sqrt{2J_G}} & |\delta| > c \end{cases}, \quad (20)$$

$$W_L = \begin{cases} 1 & |\delta| \leq c \\ \frac{c}{\sqrt{2J_G}} & |\delta| > c \end{cases}, \quad (21)$$

where  $c$  is the transition point,  $W_L$  refers to the analytical weight function. As also applied in other fields (Guitton and Symes, 2003; Huber, 2011), the Huber norm had shown robustness compared to the posterior analysis with the assumption of Gaussian error distribution. This indicates that it can alleviate (promote) the negative (positive) impacts of outliers on the optimization analyses. Therefore, the VarQC of Huber CGD (referred to as “Huber-VarQC” hereafter) can potentially produce more promising applications with respect to Flat-VarQC.

As the discussion in Huber (1972), Eq. (18) has a special formulation defined as

$$P_{L, \text{special}} = \begin{cases} \frac{1-\varepsilon}{\sqrt{2\pi}\sigma_o} \exp(-\frac{1}{2}\delta^2) & |\delta| \leq c \\ \frac{1-\varepsilon}{\sqrt{2\pi}\sigma_o} \exp(\frac{c^2}{2} - c|\delta|) & |\delta| > c \end{cases}. \quad (22)$$

Because the integration of Huber norm PDF in Eq. (22) from positive infinity to negative infinity equals 1, the relationship between the contamination rate  $\varepsilon$  and the transition point  $c$  can be derived easily as follows:

$$2\Phi(c) - 1 + \frac{2\phi(c)}{c} = \frac{1}{1-\varepsilon}, \quad (23)$$

where  $\phi(*)$ , a standard Gaussian distribution, is equal to the first derivative of  $\Phi(*)$ , e.g.,  $\phi(\delta) = \Phi'(\delta) = 1/\sqrt{2\pi}\exp(-\delta^2/2)$  with  $\sigma_o = 1$ . Based on Eq. (23), there is a corresponding relationship between the contamination rate  $\varepsilon$  and the transition point  $c$  in Table 2. One of the contamination rate and the transition point  $c$  in Huber-VarQC is determined when another one is known. The larger the contamination rate is, the smaller the transition point  $c$  can be used, and thus allowing Eq. (18) with longer tails. Although Eq. (23) requires the contamination rate and variance to be known, it objectively has presented a new way to improve the accuracy of the parameter  $c$ . Moreover, Tavolato and Isaksen (2015) and Duan et al. (2017) evaluated the optimal parameter  $c$  in their studies by fitting the best Huber norm profile for the innovational histograms. In this way, the contamination rate, which is a reference for the Flat-VarQCs, will be estimated while known  $c$  and  $\sigma_o$ . With these in mind, the Huber-VarQC (Eqs. (19)–(21)) can perform very well while known transition point  $c$  but does not need to compute the contamination rate. In general, it is reasonable to have  $c \geq 1.14$  with a contamination rate below 0.1 for conventional observations (Hampel, 1977).

Due to different observational quality over different regions, the contamination rates are different. Also, different assimilation systems show a distinct sensitivity to the parameter  $c$ . Hence, it should be best fitted for the GRAPES m3DVAR system. In general, it is an unbiased estimation when the transition point ( $\pm c$ ) is symmetric that is an ideal situation. The asymmetrical left and right transition points, however, can also obtain the least biased estimation (Pires et al., 2010) due to the robustness of Huber-VarQC. That means the Huber-VarQC promises the different left and right transition points. Despite all this, the reasonable transition points also should be determined after further sensitivity tests in the GRAPES m3DVAR system.

## 4 Optimization analysis of the VarQC

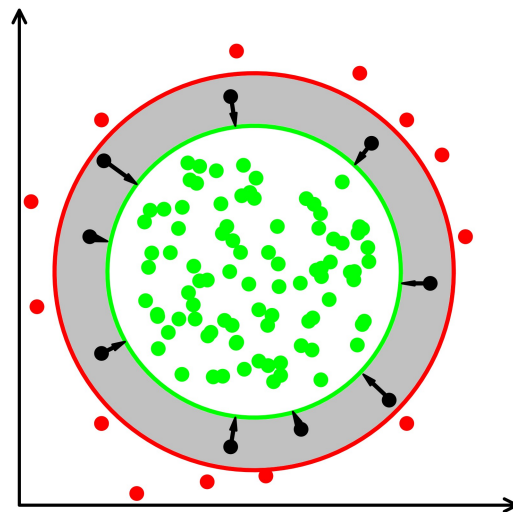
### 4.1 Observation classification of VarQC

Derived from the Bayesian theorem under the assumption that observation error follows a non-Gaussian CGD, VarQC can effectively alleviate (promote) the negative (positive) impact on posterior analysis from outliers or larger innovations. A schematic diagram of observation classification for VarQC is shown in Fig. 4. In VarQCs,

**Table 2** The corresponding relationship between the contamination rate  $\varepsilon$  and the transition point  $c$

Parameter names	Parameter values
$\varepsilon$	0.002 0.005 0.010 0.020 0.050 0.100 0.200 0.300 0.500
$c$	2.435 2.160 1.945 1.717 1.399 1.140 0.862 0.685 0.436

the observations can be classified into three categories depending on the magnitude of deviation: valid observation (VO: green dots), available observation (AO: black dots), and deleterious observation (DO: red dots). The valid observations can be assimilated as correct observations with Gaussian error distribution. The available observations, which cause long tails in the observational error distribution and contain useful information in the gray zone, are outliers without gross errors. The deleterious observations represent outliers with gross errors as faulty observations whose larger departure in innovations will lead to irreversible destruction in the posterior analysis. Generally, as the outliers with gross errors, both the AOs and DOs are excluded by the gross check in conventional quality control of VarDA system, such as the extreme value check, background quality check, and so on. However, these AOs may be correct or contain at least partially useful information, which can sometimes improve the quality of analysis, especially in extreme weather events, but its errors do not fit into the Gaussian distribution. With the use of VarQC, it tries to extend the rejection threshold (from the green circle to the red circle in Fig. 4) of the conventional quality control for assimilating more available observations. VarQC can handle the AOs by assigning relatively smaller (larger) weight to the observation in a gray zone based on larger (smaller) innovations. The weights of observations in the gray zone are adjusted (shown with black arrows in Fig. 4) through iterative analysis during the VarDA so that a more



**Fig. 4** A schematic diagram of observation classification for variational quality control. The green dots represent the valid observations (VOs); the black dots represent the available observations (AOs) shown as outliers but are not gross errors; the red dots represent the deleterious observations (DOs) shown as outliers but are gross errors. The green circle denotes the perfect threshold for a pure Gaussian error domain, the red circle represents the empirical (extended) rejection threshold in conventional quality control for gross errors. The VarQC pushes the observations (the black dots) in the gray zone iteratively into the green circle to be ingested in the VarQC.

accurate solution can be obtained. In this way, the VarQC is effective in absorbing the available information of outliers and less sensitive to the rejection threshold of the conventional quality control. It should be noted that the gray zone data, which causes the long tails in Fig. 2, had existed in current assimilated observations before extending the rejection threshold. This is probably due to that the empirical rejection threshold in conventional quality control is overestimated. It also implies that the actual observational error is non-Gaussian distribution. In view of this, it is quite necessary to perform VarQC for better posterior analysis in the current GRAPES m3DVAR system.

In summary, observational errors used in DA usually display a non-Gaussian distribution that can be fitted more reasonably by the CGDs. A combination of multiple profiles as the PDF allows us to include more outliers, and more available observation information can be used that will benefit the posterior analysis. By using VarQC along with the optimization iteration in the VarDA, more useful observations can be effectively assimilated.

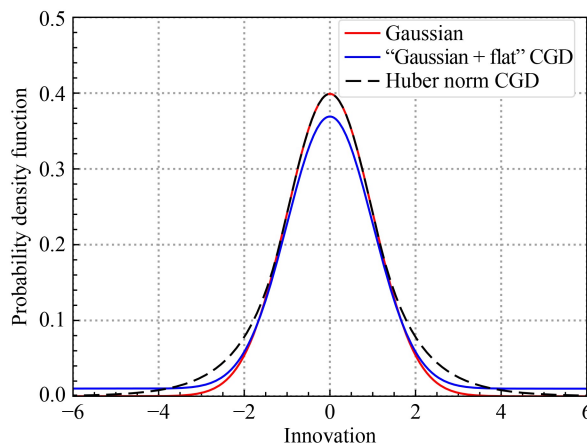
#### 4.2 The CGD modes of observational errors

As described in Section 3, the VarQC takes the Gaussian and the contaminated distribution characteristics of observation error in innovation into consideration. With the VarQC, the original VarDA now has an extra term defined as analysis weighting function  $W$  ( $W_F$  and  $W_L$ ). The  $W$  is only a function of the normalized innovation or the posterior probability  $P_{\text{posterior}}$  of the outliers. The VarQC adjusts the magnitude of  $W$ , which follows the variation of innovation size in the minimization process, to capture more accurate analysis value and synchronizes with VarDA. To explore the validity and reasonability of the CGDs in the VarQC, and the ability of the CGDs to describe the characteristics of actual innovation distribution is first investigated by comparing with the pure Gaussian distribution.

It is an idealized assumption that observation errors follow Gaussian distribution. The innovation statistics proved that only some of the observation errors strictly follow the Gaussian distribution. Often the actual innovation distributions are closer to the contaminated Gaussian distribution with long tails led by the small-sample outliers in observations. With the given parameters, two CGDs (Fig. 5) show longer tails than those of the Gaussian distribution, which can better match the actual observational error feature, as well-documented in Fig. 2. The Huber distribution has longer (heavier) tails than those of the Gaussian distribution and the Gaussian plus flat distribution. Moreover, the heavier tails are generally at critical positions, where the rejection thresholds of background quality check among the gray zone are relatively ambiguous, as shown in Fig. 4. This implies that the Huber distribution probably can better process the outliers located in the gray zone. Overall, the two non-Gaussian CGDs are able to fit the long tails of the actual error distribution well. However, the optimal CGD, which is in good agreement with the characteristics of the actual observation error, needs to be further studied with real observations.

#### 4.3 Optimization analysis of cost, gradient and weight function

Figure 6 shows the curves of the observational cost function, gradient function and weighting function of the two VarQCs, as well as the observational cost function ( $J_G$ ) of the original VarDA. The cost function curves of Flat-VarQC and Huber-VarQC approximately overlap with  $J_G$  within the Gaussian domain (within  $(-2, 2)$  in Fig. 6(a) and  $[-1.14, 1.14]$  in Fig. 6(b)) located on both sides of zero, because the main distribution mode of two CGDs is the same as the pure Gaussian distribution. With the growth of the absolute value of innovations beyond the Gaussian domain, the  $J_F$  curve gradually flattens to be



**Fig. 5** The profiles of the probability density distribution of observation error for the pure Gaussian and two CGDs (Gaussian + flat and Huber norm). The parameters used are  $\varepsilon = 0.1$ ,  $d = 5$ ,  $\sigma_o = 1$ ,  $c = 1.14$ .

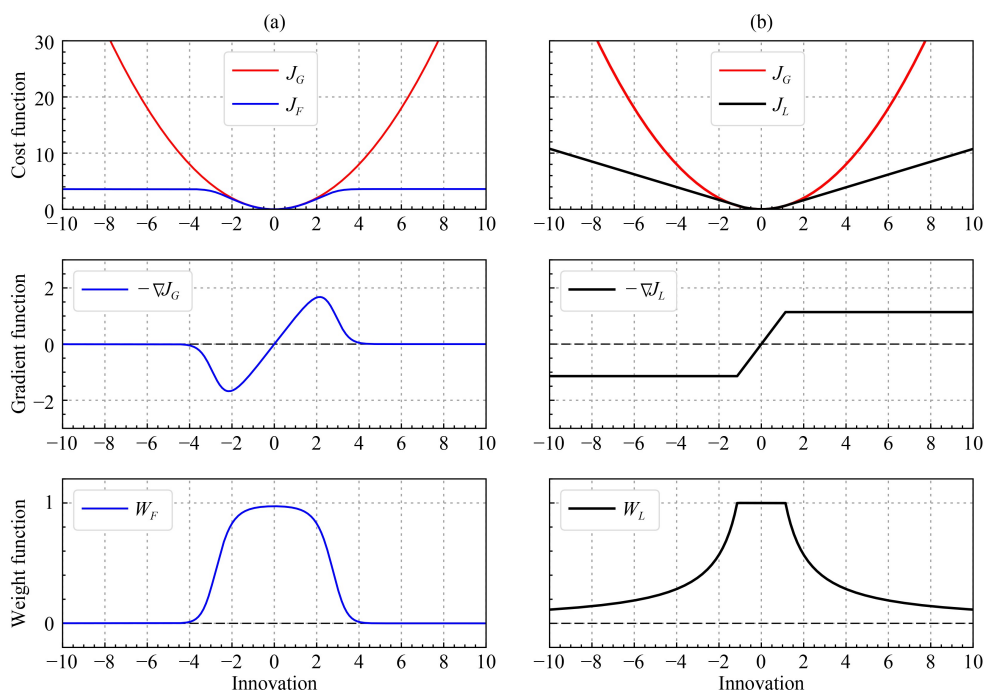
a flat distribution. Meanwhile, its gradient function curve drops to zero from a peak value and subsequently keeps unchanged. It indicates that this (strictly) non-concave function ( $J_F$ ) probably generates multi-minimum values or unstable convergence during the iteration of the minimization process. As for the weight function of Flat-VarQC,  $W_F$  rapidly decreases from  $\sim 1$  to 0 across the Gaussian domain. The function variations of Huber-VarQC (Fig. 6(b)) are different from those of the Flat-VarQC. The cost function, gradient function and analysis weight function show a piecewise feature due to the partitioned CGD mode but keep continuity in the range of innovations. Over the Gaussian domain, the impact of the Gaussian distribution on the cost function of the Huber norm makes it the same as the  $J_G$  curve. Beyond the Gaussian domain with larger innovations, whereas the cost function displays symmetric straight lines under the influence of Laplace distribution. The gradient curve flattens to be a constant with the increase of innovations. Moreover, the analysis weight maintains full weight in the Gaussian domain before approaching around zero smoothly.

Overall, the gradients of two VarQCs near the Gaussian domain kept pace with  $\nabla J_G$  are monotonic increasing, which is beneficial for convergence. But the cost functions of Flat-VarQC are non-concave functions that may give multiple-minimums during the minimization process, although its CGD can represent the long-tail feature. As discussed by Anderson and Jarvinen (1999), however, if the minimization starts with an intermediate

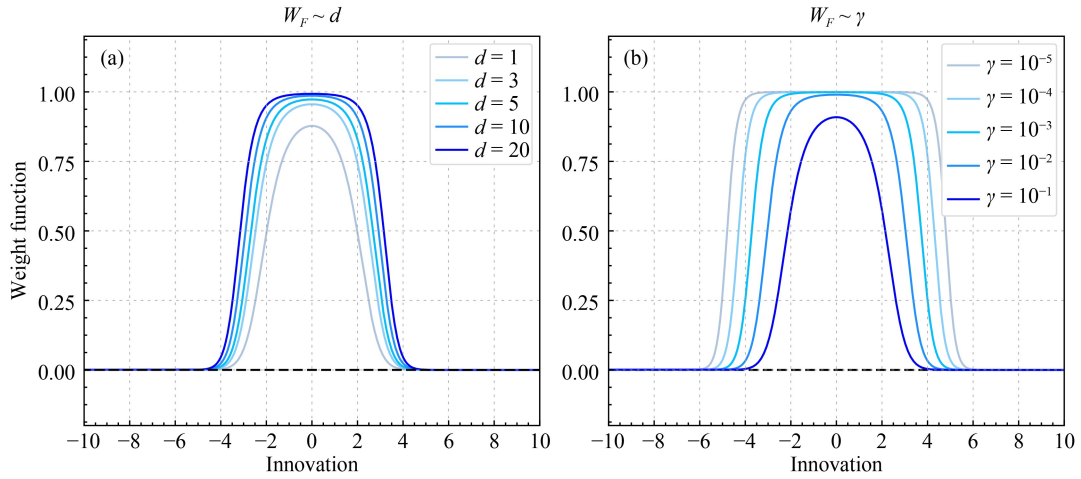
analysis that has a small departure from the background, the VarQC will guarantee a correct convergence in the DA system. The weighting functions of two VarQCs are (approximately) equal to full 1 in the Gaussian domain where the innovations are adequately small due to that VarQC flags them as the VOs. With the innovation's increase beyond the Gaussian domain, the weight should decrease to zero, as shown in Fig. 6. In contrast to the Flat-VarQC, the Huber-VarQC is reasonable for minimization analysis due to its strict concave cost function. The Huber-VarQC puts more weight on large innovations beyond triple variance ( $3\sigma_0^2$ ) and allows more influences from the outliers compared to the Flat-VarQC scheme. As for two VarQCs, further parameter optimization experiments are needed to verify the performance, including their convergence issues.

#### 4.4 VarQCs' weight sensitivity

The Flat-VarQC weight  $W_F$  is obtained by calculating the innovation and the parameter  $\gamma$  based on Eq. (11). The parameter  $\gamma$  is a single constant that is defined by  $\varepsilon$  and  $d$ . Figure 7 shows the weight sensitivity to  $d$  and  $\gamma$  in the Flat-VarQC scheme. The weight curves in Fig. 7(a) show a slight difference while the width  $d$  equals 3, 5, 10, 20. However, the weight function with  $d = 1$  shows a thinner curve with a narrow Gaussian domain. The thinner and narrower the weight curve is, the tighter the VarQC would be performed, and the more observations would be "rejected." Overall, the Flat-VarQC weights are not too sensitive to  $d$  unless the value chosen is rather small.



**Fig. 6** The cost function, gradient function, and analysis weight function of (a) Flat-VarQC (blue line) and (b) Huber-VarQC (black line). The red lines represent the observational cost function of the pure 3DVAR under the assumption of Gaussian distribution. The parameters are used as shown in Fig. 5.

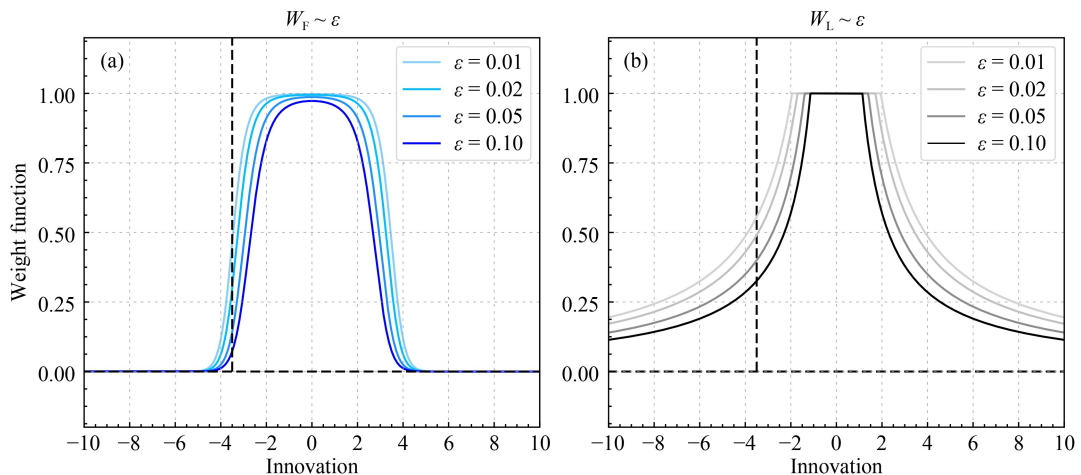


**Fig. 7** Same as the analysis weight functions in Fig. 6(a) but for the sensitivity to  $d$  (the width of flat distribution) and  $\gamma$  ( $\gamma = \varepsilon \sqrt{2\pi}/2d(1-\varepsilon)$ ) in Flat-VarQC. The other parameters are (a)  $\varepsilon = 0.1$  and (b)  $d = 5$ . The same parameter is  $\sigma_o = 1$ .

With a small value of  $10^{-5}$  for  $\gamma$  (Fig. 7(b)), the Flat-VarQC weight exhibits a steeper reduction of weights compared to the weights of 3DVAR without VarQC. The larger the parameter  $\gamma$  is, the narrower the Gaussian domain in Flat-VarQC will be used. With the decrease of  $\gamma$  by the magnitude of 10, the curves show a wider weight of Gaussian domain, meaning that more observations will sign larger weights but less than 1. The largest weight will reduce to  $\sim 0.9$  while  $\gamma = 0.1$ , suggesting that the parameter  $\gamma$  should not be too large. Therefore, the accuracy of  $\gamma$  can determine the performance of Flat-VarQC. With the convergence issue in Flat-VarQC, however, the smaller  $\gamma$  is benefit for a successful minimization iteration.

To a considerable degree, the magnitude of contamination rate determines the shape of the weight function that would impact the accuracy of posterior analysis in VarQC. Figure 8 shows the sensitivities to the different

contamination rates for different VarQC weight functions. The weight curves, which reveal the “n” shape in Flat-VarQC (Fig. 8(a)), are significant differences from the “ $\pi$ ” shape of Huber-VarQC (Fig. 8(b)). With the increasing contamination rate, the width of the Gaussian domain in the weight curves is getting narrower for two VarQCs. In other words, VarQCs will reduce more observational weight to impact posterior analysis. Furthermore, the weights over the Gaussian domain in Flat-VarQC are approximately 1 but equal to 1 for those of Huber-VarQC. Beyond the Gaussian domain, the Flat-VarQC decreases the weights to 0 steeply whereas the Huber-VarQC decreases its weight to 0 smoothly. For a fixed innovation far from the Gaussian domain, the higher the contamination rate is, the less weight will be gotten. Moreover, the Flat-VarQC assigns the weights in the range of [0.05, 0.5] to a fixed innovation value (e.g. the black vertical dashed line) while Huber-VarQC assigns the larger weights in the range of [0.3, 0.55].



**Fig. 8** Same as the analysis weight functions in Fig. 6 but for the sensitivity to the contamination rate ( $\varepsilon$ ) in different VarQCs. The parameter  $c$  corresponding to  $\varepsilon$  is based on Table 2. The other parameters are the same as those of Fig. 5.

Compared with Flat-VarQC, therefore, Huber-VarQC's weight is not very sensitive to larger contamination rates, implying that it probably gives more inclusiveness to outliers.

## 5 Discussion and conclusions

The observation error in innovation is a key factor to impact the quality of VarDA analysis as its distribution determines the observational cost function. In this study, it is found that the previous OEDMs, which are essentially the CGDs, can fit the actual observational error distribution more accurately based on the non-Gaussian-distributed characteristics of innovations from the GRAPES m3DVAR system. Furthermore, we present a novel interpretation of VarQC in detail, i.e., the valid, available and deleterious observations classified by VarQC. Two VarQCs are formulated by using two CGDs including the Gaussian plus flat distribution and Huber norm distribution based on interdisciplinary theories of DA and robust statistics. We also present detailed methods to optimize the different parameters in the two VarQCs. Subsequently, the CGDs are verified to be more consistent with the actual observational error distribution compared to the pure Gaussian distribution. The evaluation of the two VarQCs with different setting parameters demonstrated that all two VarQCs showed reasonable weight variations with the increase of innovations. But the Flat-VarQC probably posts convergence issues when the innovations are quite large at the beginning of the variational minimization process. This convergence issue may be alleviated by switching the VarQC during an intermediate analysis step or giving a value of  $\gamma$  as small as possible. Flat-VarQC weights are not too sensitive to  $d$  unless the value chosen is rather small. The Laplace distribution in the Huber norm distribution shows better robustness based on the robust statistics, making the Huber-VarQC more attractive to study the impacts for outliers. Huber-VarQC shows a more elastic weight function for different parameters against outliers compared with that of Flat-VarQC.

This paper reveals a general relationship between the CGD and VarQC. The parameter optimization and analysis results will promote the further development and improvement of VarQCs in the GRAPES m3DVAR system. Further real-data parameter experiments are also important to improve the posterior analysis by utilizing the VarQCs. For example, in order to obtain the optimal analysis step that can ensure the correct convergence when the Flat-VarQC switches on, sensitive experiments are necessary to determine the optimal initial iteration steps of VarQC. The relaxed coefficient in rejection thresholds for background quality check is also required to be determined based on the robust intensity of VarQC which allows more observations to be assimilated by

promising a larger rejection threshold in the GRAPES m3DVAR system.

**Acknowledgments** The authors thank Yinghui Lu for his helpful advice and grammar correction. Jie He is supported by the China Scholarship Council. This work is primarily sponsored by the National Key R&D Program of China (Nos. 2018YFC1506702 and 2017YFC1502000). We acknowledge the High Performance Computing Center of Nanjing University of Information Science & Technology (NUIST) for their support of this work. The data sets in this paper are archived and accessible on the super computer of NUIST.

## References

- Anderson E, Jarvinen H (1999). Variational quality control. *Q J R Meteorol Soc*, 125(554): 697–722
- Courtier P, Andersson E, Heckley W, Vasiljevic D, Hamrud M, Hollingsworth A, Rabier F, Fisher M, Pailleux J (1998). The ECMWF implementation of three-dimensional variational assimilation (3D-Var). Part I: formulation. *Q J R Meteorol Soc*, 124(550): 1783–1807
- Dharssi I, Lorenc A C, Ingleby N B (1992). Treatment of gross errors using maximum probability theory. *Q J R Meteorol Soc*, 118(507): 1017–1036
- Duan B H, Zhang W M, Yang X F, Dai H J, Yu Y (2017). Assimilation of typhoon wind field retrieved from scatterometer and SAR based on the Huber norm quality control. *Remote Sens (Basel)*, 9(10): 987
- Fernández C, Steel M F (1998). On Bayesian modeling of fat tails and skewness. *J Am Stat Assoc*, 93(441): 359–371
- Fowler A, van Leeuwen P J (2013). Observation impact in data assimilation: the effect of non-Gaussian observation error. *Tellus*, 65(1): 20035
- Gauthier P, Chouinard C, Brasnett B (2003). Quality control: methodology and applications. In: Swinbank R, Shutyaev V, Lahoz W A, eds. *Data Assimilation for the Earth System*. Dordrecht: Springer-Verlag, 177–187
- Guitton A, Symes W W (2003). Robust inversion of seismic data using the Huber norm. *Geophysics*, 68(4): 1310–1319
- Hampel F R (1977). Rejection rules and robust estimates of location: an analysis of some Monte Carlo results. In: *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, and Random Processes*. Dordrecht, Hingham, MA, 187–194
- Hampel F R (2001). Robust statistics: a brief introduction and overview. In: *Research report/Seminar für Statistik, Eidgenössische Technische Hochschule. Zurich, Switzerland*, 1–5
- He J, Ma X L, Ge X Y, Liu J J, Cheng W, Chan M Y, Xiao Z N (2021). Variational quality control of non-Gaussian innovations in the GRAPES m3DVAR system: mass field evaluation of assimilation experiments. *Adv Atmos Sci*, 38(9): 1510–1524
- Hollingsworth A (1989). The role of real-time four-dimensional data assimilation in the quality control, interpretation, and synthesis of climate data. In: Anderson D L T, and Willebrand J, eds. *Oceanic Circulation Models: Combining Data and Dynamics*. Dordrecht: Springer-Verlag, 304–339
- Houtekamer P L, Mitchell H L (1998). Data assimilation using an ensemble Kalman filter technique. *Mon Weather Rev*, 126(3): 796–811
- Huber P J (1972). The 1972 wald lecture robust statistics: a review.

- Ann Math Stat, 43(4): 1041–1067
- Huber P J (2011). Robust statistics. In: Lovric M, ed. International Encyclopedia of Statistical Science. Heidelberg: Springer-Verlag, 1248–1251
- Ingleby N B, Lorenc A C (1993). Bayesian quality control using multivariate normal distributions. *Q J R Meteorol Soc*, 119(513): 1195–1225
- Jarvinen H, Uuden P (1997). Observation screening and first guess quality control in the ECMWF 3D-Var data assimilation system. In: ECMWF Technical Memoranda(236): 1–33
- Kalnay E (2003). Atmospheric Modeling, Data Assimilation and Predictability. New York: Cambridge University Press: 198–204
- Lahoz W A, Schneider P (2014). Data assimilation: making sense of earth observation. *Front Env Sci—Switz*, 2: 16
- Legrand R, Michel Y, Montmerle T (2016). Diagnosing non-Gaussianity of forecast and analysis errors in a convective scale model. *Nonlinear Process Geophys*, 23: 1–12
- Lorenc A C (1988). Optimal nonlinear objective analysis. *Q J R Meteorol Soc*, 114(479): 205–240
- Lorenc A C, Hammon O (1988). Objective quality control of observations using Bayesian methods: theory, and a practical implementation. *Q J R Meteorol Soc*, 114(480): 515–543
- Ma X L, Zhuang Z R, Xue J S, Lu W S (2009). Development of 3-D variational data assimilation system for the nonhydrostatic numerical weather prediction model-GRAPES. *Acta Meteorol Sin*, 67(1): 50–60
- Ma X L, He J, Zhou B Y, Li L, Ji Y, Guo H (2017). Effect of variational quality control of Non-Gaussian distribution observation error on heavy rainfall prediction. *Trans Atmos Sci*, 40(2): 170–180 (in Chinese)
- Parrish D F, Derber J C (1992). The national meteorological center spectral statistical interpolation analysis system. *Mon Weather Rev*, 120(8): 1747–1763
- Pires C A, Talagrand O, Bocquet M (2010). Diagnosis and impacts of non-Gaussianity of innovations in data assimilation. *Physica D*, 239(17): 1701–1717
- Rabier F, Järvinen H, Klinker E, Mahfouf J F, Simmons A (2000). The ECMWF operational implementation of four-dimensional variational assimilation. I: experimental results with simplified physics. *Q J R Meteorol Soc*, 126(564): 1143–1170
- Sondergaard T, Lermusiaux P F (2013). Data assimilation with Gaussian mixture models using the dynamically orthogonal field equations. Part I: theory and scheme. *Mon Weather Rev*, 141(6): 1737–1760
- Storto A (2016). Variational quality control of hydrographic profile data with non-Gaussian errors for global ocean variational data assimilation systems. *Ocean Model*, 104: 226–241
- Su X J, Pursera R J (2013). A new observation error probability model for nonlinear variational quality control and applications within the NCEP gridpoint statistical interpolation. In: Sixth WMO Symposium on Data Assimilation. Riggs Alumni Center, College Park, MD, 33–34
- Swinbank R, Shutyaev V, Lahoz W A (2003). Data assimilation for the earth system. *Maratea: Springer*, 31–35
- Tavolato C, Isaksen L (2015). On the use of a Huber norm for observation quality control in the ECMWF 4D-Var. *Q J R Meteorol Soc*, 141(690): 1514–1527
- Tukey J W (1960). A survey of sampling from contaminated distributions. In: Olkin I, ed. Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling. Stanford, CA: Stanford University Press, 448–485
- Wang X G, Lei T (2014). GSI-based four-dimensional ensemble-variational (4DEnsVar) data assimilation: formulation and single-resolution experiments with real data for NCEP global forecast system. *Mon Weather Rev*, 142(9): 3303–3325
- Yang Y X (1991). Robust bayesian estimation. *J Geod*, 65(3): 145–150
- Zhao H, Zou X L, Qin Z K (2015). Quality control of specific humidity from surface stations based on EOF and FFT—case study. *Front Earth Sci*, 9(3): 381–393
- Zhu J J (1996). Robustness and the robust estimate. *J Geod*, 70(9): 586–590
- Zhu J J, Zeng Z Q (1999). The theory of surveying adjustment under contaminated error model. *Acta Geodaetica et Cartographica Sinica*, 28(3): 91–91(in Chinese)

---

## AUTHOR BIOGRAPHIES

Jie He received his Ph.D. Degree in meteorology from Nanjing University of Information Science and Technology (NUIST), Nanjing, China in 2021, mainly focusing on the research of variational quality control and ensemble reanalysis. He continues as a postdoc to study data assimilation and ensemble forecast at the Guangzhou Institute of Tropical and Marine Meteorology.  
E-mail: hej@gd121.cn.

Yang Shi received her M.S. Degree in meteorology from NUIST, Nanjing, China in 2015. She is an Engineer at the Guangdong Meteorological Observatory and her current research includes ensemble forecasts and postprocessing of NWP.  
E-mail: shiyang@gd121.cn.

Boyang Zhou received her M.S. Degree in meteorology from NUIST, Nanjing, China in 2017. She is an engineer at the Qingdao Air Traffic Management Station of Civil Aviation of China and her current research includes data assimilation and ensemble perturbation schemes.  
E-mail: zby12357@126.com.

Qiuping Wang is a graduate student in meteorology from the School of Atmospheric Sciences at NUIST, Nanjing, China. Her research interest focuses on ensemble forecast and data assimilation.  
E-mail: wangqp@nuist.edu.cn.

Xulin Ma received his Ph.D. Degree in meteorology from NUIST, China in 2008. His major field is data assimilation, ensemble forecast, and meso- and micro-scale numerical simulation. He works as a Professor at the School of Atmospheric Sciences, NUIST, Nanjing, China since 2008. He is devoted to studying ensemble forecast theory, hybrid data assimilation, and adaptive or targeting observation.  
E-mail: xulinma@nuist.edu.cn.