Modified Relative Humidity Based on the Johnson's S_B Distribution Function

Yasutaka Wakazuki

Center for Research in Isotopes and Environment Dynamics, *University of Tsukuba*, *Tsukuba*, *Japan*
Japan Agency for Marine-Earth Science and Technology, *Yokohama*, *Japan*

Abstract

A new atmospheric humidity variable called the modified relative humidity (MRH) is proposed based on properties of the Johnson's S_B distribution function. The frequency distribution of MRH can be roughly approximated by the normal distribution, while other variables such as relative humidity and the water vapor mixing ratio cannot. This characteristic suggests that MRH is convenient for statistical variable controls such as data assimilation and climatological grid data controls. Super-saturation and negative water vapor states induced by positive and negative humidity increments are avoided by using MRH. Three types of MRH, Types-I, -II, and -III, were examined. Type-III, with three fixed parameters, was the best function for approximating to the normal distribution. However, Types-I and -II, each with two fixed parameters, were beneficial for stable statistical humidity variable estimations.

(**Citation:** Wakazuki, Y., 2013: Modified relative humidity based on the Johnson's S_B distribution function. *SOLA*, **9**, 111−114, doi:10.2151/sola.2013-025)

1. Introduction

Statistical controls of atmospheric variables are often performed in data assimilation and climatological data controls that include bias corrections. In data assimilation, numerical simulation results predicted from a prior time are statistically combined with observed variables, and the combined grid data are used as the initial conditions of the next simulation cycle. In the pseudo-global-warming (PGW) method (Sato et al. 2006; Kimura and Kitoh 2007), which is a method of climatological data control, mean climatological differences estimated by general circulation model (GCM) simulations are added to an objective analysis dataset. The combined data are used as the boundary conditions for regional climate model (RCM) simulations of the future climate.

Humidity variables such as specific humidity, relative humidity (RH), the water vapor mixing ratio (q_v), and dew point depression (TTD; temperature minus dew-point temperature) are also included among the controlled variables. For example, a positive increment of the humidity in the lower atmosphere induces unstable stratification, which is able to cause intense precipitation. In a physical sense, q_v seems to be appropriate as a humidity variable because q_v is used as a variable accompanied by the conservation laws of atmospheric models.

However, statistical humidity variable controls have a serious problem caused by their non-Gaussian nature. In addition, the range of humidity is restricted, i.e., between dry (no water vapor) and saturation states. For example, RH cannot be negative, and RH in excess of 100% is not allowed in general. When RH is used as the humidity variable, super-saturation is often induced by the statistical variable controls and unnatural intense rainfall may be predicted.

Several humidity variables have been proposed for data assimilation, including pseudo relative humidity (PRH; Dee and Da Silva 2002) and its extensions (Gustafssona et al. 2011). If a humidity variable is divided into a base state and perturbation, data assimilation mainly controls the perturbation part. Meanwhile, the base state is mainly controlled by climatological variable controls, for which the control techniques have not been studied sufficiently. In the PGW method, for example, four worlds of climatological states, namely objective analysis (a), present climate (p), future climate (f), and pseudo future climate (w), are present. Variables (V) in the PGW method are denoted by

$$V_w = V_a + \delta V, \quad \delta V = \overline{V_f} - \overline{V_p}, \tag{1}$$

where δV is the mean climatological difference and is a single value. On the other hand, V_a and V_w are time series data with specific time intervals. Namely, V_a has a specific probability distribution and the distribution is modified by δV to create the probability distribution of V_w. The probability distribution of V_w becomes unnatural for traditional humidity variables, and unrealistic values of V_w often generated. Therefore, humidity variables are not independently controlled in the PGW method, and the RH is assumed not to change between the present and future climate states (e.g., Hara et al. 2008).

We propose a new humidity variable that is able to avoid range restrictions. In Section 2, the Gaussianity of controlled variables is discussed. In Section 3, the statistical properties of traditional humidity variables are examined. In Section 4, the new humidity variable is introduced and its properties are investigated. The summary and discussion are presented in Section 5.

2. The Gaussianity of probability distributions

Most data assimilation systems require Gaussian variables. For climatological variable controls, Gaussianity is a sufficient but not a necessary condition. In general, distribution function parameters are divided into those for shape and those for location. For climatological variable controls, only the location parameters of the distribution functions are modified, while the shapes of the distributions do not change. Here, non-Gaussian variables are considered. Shapes with large skewness are often associated with physical properties and restrictions. In particular, there are variables that do not permit shifts of the peak or restriction points to avoid unnatural distributions. For instance, shifting the restriction points of relative humidity is not permitted. In another example, the probability density function (PDF) for the amount of short-term precipitation is an exponential-type function (e.g., Wakazuki 2011). The peak point of the PDF is located at around 0 mm. The climatological variable control shifts the location of a peak point, but the peak point not located at around 0 mm is unnatural. The physical meanings of the peak and restriction points are complicated and differ depending on the variables. A normal distribution is often observed in the statistics of natural random values of a variable. When the distribution of a variable is Gaussian, the value of the variable is regarded as generated by natural random phenomena without a specific physical forcing. In addition, it is expected that unrealistic values of the variable will not be generated by climatological variable controls. Therefore, it is desirable for the variables in climatological variable controls to be Gaussian.

3. Traditional humidity variables

Here we examine the statistical properties of traditional humidity variables. Humidity data were extracted from the

Corresponding author: Yasutaka Wakazuki,Center for Research in Isotopes and Environment Dynamics, University of Tsukuba, 1-1-1, Tsukuba, Ibaraki 305-8577, Japan. E-mail: ywakazki@gmail.com.

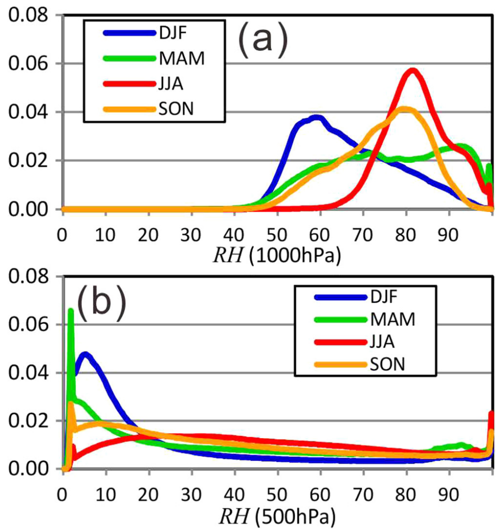

Fig. 1. Probability density distributions of the relative humidity (RH; unit is %) at (a) 1000 hPa and (b) 500 hPa.

initial-time data of the Japan Meteorological Agency (JMA) Meso-Scale Model (MSM). Grid point values within a domain [130°E−140°E, 25°N−30°N] on the southern side of the main island of Japan were used to examine the statistical properties. The analysis period is from March 2006 to February 2013. Figure 1 shows the probability density distributions of the RH at 1000 and 500 hPa in DJF (December, January, and February), MAM (March, April, and May), JJA (June, July, and August), and SON (September, October, and November). Some parts of the distributions at 1000 hPa resemble a normal distribution. However, many of these show a large skewness and are clearly different from a normal distribution. In addition, if positive relative humidity increments are given to the distributions, for example, many samples around 100% would become super-saturated. Similar problems are induced when other traditional humidity variables are used. Unrealistic humidity states would be induced in some cases, such as negative TTD with negative increments, negative q_v with negative increments, and super-saturation of q_v with positive increments. The shapes of the distributions of q_v and TTD are also clearly different from a normal distribution (not shown). Therefore, these humidity variables are not appropriate as statistical humidity controls.

4. A new humidity variable

Humidity has range restrictions from dry (no water vapor) to saturation states. The RH is limited to the range of 0 to 100%. Therefore, a new humidity variable with a normal distribution must be found. In general, variables with a specific two-sided restriction range often follow the Johnson's S_B distribution (Johnson 1949), which is a distribution function of the Johnson's distribution system. The log-normal distribution is also included in the Johnson's distribution system. The PDF of the Johnson's S_B distribution as a function of x is expressed by

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\left(\frac{\lambda}{(x-\xi)(\lambda+\xi-x)}\right)\times\exp\left[-\frac{1}{2}\left(\frac{\ln\left(\frac{x-\xi}{\lambda+\xi-x}\right)-\mu}{\sigma}\right)^2\right], \tag{2}$$

where ξ, λ, σ, and μ are the shape and location parameters of the function. A humidity variable x such as RH is assumed to be approximated by the Johnson's S_B distribution. If a variable z is introduced as

$$z=\ln\left(\frac{x-\xi}{\lambda+\xi-x}\right), \tag{3}$$

then Eq. 2 can be rewritten as the PDF of z as

$$f(z)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2\right], \tag{4}$$

where

$$f(x)=f(z)\cdot\frac{dz}{dx}.$$

Equation 4 is the normal distribution of z, where parameters μ and σ are the average and standard deviation of z, respectively, and ξ and $\lambda+\xi$ are the universal minimum and maximum values of x, respectively. Values of x smaller than or equal to ξ and x greater than or equal to $\lambda+\xi$ are regarded as out of range. It is preferable for x to be a ratio variable such as RH because x is restricted by the two specific constant values ξ and λ. When x is the RH with the unit of percent, ξ and λ can be fixed at values of 0 and 100, which correspond to dry and saturated conditions, respectively. Figure 2 shows the relation between RH and z in the case where ξ and λ are fixed to be 0 and 100, respectively. Any value of z is convertible to the RH value within the restriction range.

Next we discuss the merits of using z with the increment equation (Eq. 1), in which V is a humidity variable. An increment δV is added to V_a to estimate V_w, and V_a and V_w are time series data. The relative humidity values of V_a, V_w, and δV are RH_a, RH_w, and δRH, respectively. Table 1 shows RH_w for the cases where (a) V is equal to RH, (b) V is q_v, and (c) V is z. For the estimations of RH_w in (b) and (c), RH_a and δRH are converted to V_a and δV, respectively. Values of V_a are modified by δV to estimate V_w, and V_w values are converted to RH_w. In the case where V is RH or q_v, large positive (negative) humidity increments (δRH) often induce super-saturation (a negative water vapor state) when RH_a is very large (small). On the other hand, when V is z, unrealistic humidity states are not found.

When fitting RH to the distribution functions by using the maximum likelihood estimation (MLE) method, the sample summation of logarithms of likelihood functions (L) is maximized to estimate values of the parameters, where large values of L correspond to better fitting results. Comparison of values of L (not shown) shows that the Johnson's S_B distribution is a better function for RH than the normal distribution. Thus, we use a new humidity variable z (Eq. 3) assuming the Johnson's S_B distribution for x with the fixed parameters ξ and λ. The new variable z is named the "modified relative humidity" (MRH), where x is assumed to be RH.

There is, however, a problem in the estimation of z. Which type of x is most appropriate? Therefore, we examined three types of x (Table 2). In Type-I, RH is used as x, where ξ and λ are fixed to be 0 and 100, respectively. In Type-II, ξ and λ are not fixed, and are estimated by the MLE method. Here, ξ and λ are searched

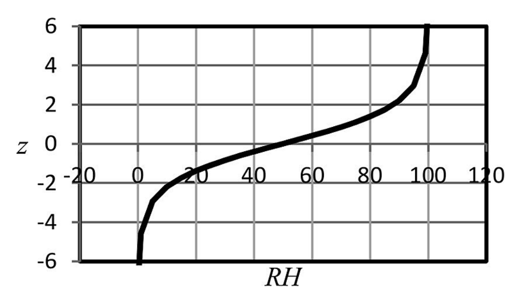

Fig. 2. The relationship between RH and z.

Table 1. Relative humidity RH_w using Eq. 1 for various values of RH_a and δRH when (a) V is RH, (b) V is q_v, and (c) V is z. δRH is assumed to be the difference from 50%. In (b), the pressure and temperature used in the calculation of relative humidity are fixed to be 1000 hPa and 293.15 K, respectively. In (c), ξ and λ are 0 and 100, respectively. The gray shaded boxes show unrealistic values.

(a)

RH_a \ δRH	−30	−10	−5	−1	0	1	5	10	30
1	−29	−9	−4	0	1	2	6	11	31
5	−25	−5	0	4	5	6	10	15	35
10	−20	0	5	9	10	11	15	20	40
30	0	20	25	29	30	31	35	40	60
50	20	40	45	49	50	51	55	60	80
70	40	60	65	69	70	71	75	80	100
90	60	80	85	89	90	91	95	100	120
95	65	85	90	94	95	96	100	105	125
99	69	89	94	98	99	100	104	109	129

(b)

RH_a \ δRH	−30	−10	−5	−1	0	1	5	10	30
1	−29.7	−9.2	−4.1	0.0	1.0	2.0	6.1	11.2	31.7
5	−25.6	−5.2	−0.1	4.0	5.0	6.0	10.1	15.2	35.6
10	−20.6	−0.2	4.9	9.0	10.0	11.0	15.1	20.2	40.6
30	−0.3	19.9	25.0	29.0	30.0	31.0	35.0	40.1	60.3
50	20.0	40.0	45.0	49.0	50.0	51.0	55.0	60.0	80.0
70	40.3	60.1	65.0	69.0	70.0	71.0	75.0	79.9	99.7
90	60.6	80.2	85.1	89.0	90.0	91.0	94.9	99.8	119.4
95	65.6	85.2	90.1	94.0	95.0	96.0	99.9	104.8	124.4
99	69.7	89.2	94.1	98.0	99.0	100.0	103.9	108.8	128.3

(c)

RH_a \ δRH	−30	−10	−5	−1	0	1	5	10	30
1	0.3	0.7	0.8	1.0	1.0	1.0	1.2	1.5	3.9
5	1.3	3.4	4.1	4.8	5.0	5.2	6.0	7.3	17.4
10	2.7	6.9	8.3	9.6	10.0	10.4	12.0	14.3	30.8
30	9.7	22.2	26.0	29.2	30.0	30.8	34.4	39.1	63.2
50	20.0	40.0	45.0	49.0	50.0	51.0	55.0	60.0	80.0
70	36.8	60.9	65.6	69.2	70.0	70.8	74.0	77.8	90.3
90	69.2	85.7	88.0	89.6	90.0	90.4	91.7	93.1	97.3
95	82.6	92.7	94.0	94.8	95.0	95.2	95.9	96.6	98.7
99	96.1	98.5	98.8	99.0	99.0	99.0	99.2	99.3	99.7

Table 2. The three types of x and their parameters ξ, λ, and α used to estimate z.

type	x	ξ	λ	α
I	RH	0	100	
II	RH	not fixed	not fixed	
III	RH^{α}	not fixed	not fixed	not fixed

under the following restrictions to avoid unrealistic relative humidity values:

$$\xi \geq 0, \quad \xi + \lambda \leq 100, \quad \lambda > 0.$$

Figure 3 shows the probability density distributions of MRH Types-I and -II. The shapes of the functions of MRH are obviously more similar to a normal distribution than those of RH (Fig. 1). These results show that MRH is more appropriate for statistical humidity variable controls. Figure 4 shows the skewness and kurtosis values of the PDFs of the RH and the three types of MRH in MAM. Skewness and kurtosis should both be 0 for a normal distribution. In the lower atmosphere, the magnitudes of the skewness and kurtosis of MRH Types-I and -II are smaller than those of the RH. This also shows that the shapes of the functions of MRH Types-I and -II are obviously more similar to a normal distribution than those of the RH. In addition, Type-II is better than Type-I, but the difference is small.

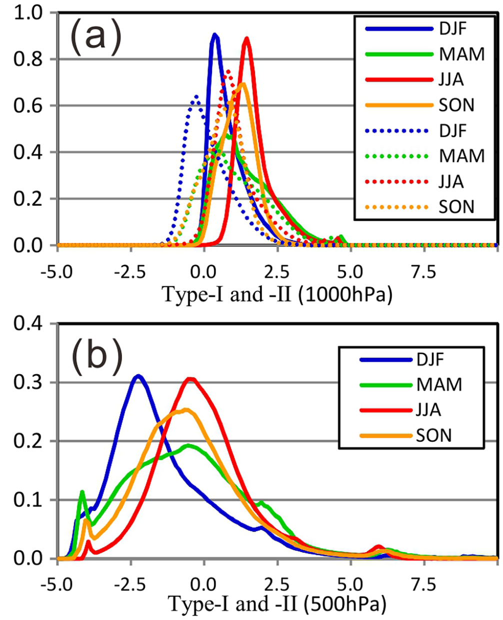

Fig. 3. Probability density distributions of the modified relative humidity (MRH Types-I and -II) at (a) 1000 hPa and (b) 500 hPa. The solid and dashed lines are Types-I and -II, respectively. The distributions of Type-II at 500 hPa are the same as those of Type-I because ξ and $\xi + \lambda$ estimated by the MLE method exceed the 0 and 100% range restrictions, respectively.

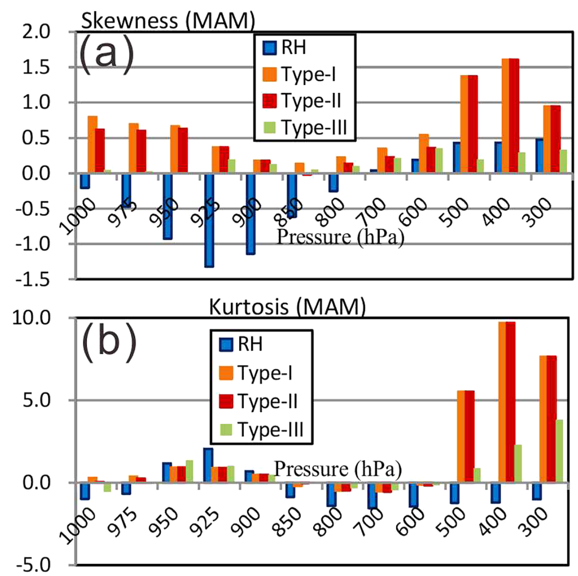

Fig. 4. (a) Skewness and (b) kurtosis of RH and MRH Types-I, -II, and -III at each pressure level in MAM.

In Fig. 4, the skewness values of MRH Types-I and -II are still large (positive) especially in the uppers atmosphere. This characteristic is also found in Fig. 3 (the right-side tails of the PDFs are long). To correct for the positive skewness, we propose MRH Type-III (Table 2). In Type-III, x is not RH but RH^{α}, and parameters ξ, λ, and α are estimated by the MLE method, where α is restricted to be positive. A value of α larger (smaller) than 1 is able to correct a positive (negative) value of the skewness. The distribution function is modified into the following formula from the Johnson's S_B distribution, where x is RH:

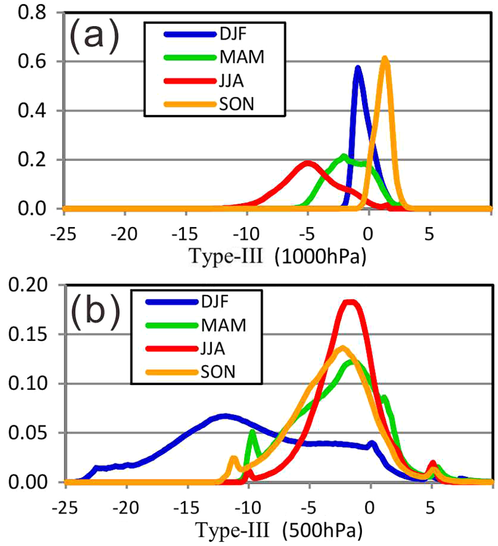

Fig. 5. Probability density distributions of the modified relative humidity (MRH Type-III) at (a) 1000 hPa and (b) 500 hPa.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\left(\frac{\lambda\alpha x^{\alpha-1}}{(x^\alpha-\xi)(\lambda+\xi-x^\alpha)}\right) \times \exp\left[-\frac{1}{2}\left(\frac{\ln\left(\frac{x^\alpha-\xi}{\lambda+\xi-x^\alpha}\right)-\mu}{\sigma}\right)^2\right]. \quad (5)$$

This formula is regarded as the extended function of the Johnson's S_B distribution. Figure 5 shows the probability density distributions of MRH Type-III. The shapes of the PDFs are more similar to a normal distribution than those in Fig. 3. The skewness and kurtosis values of MRH Type-III are obviously smaller than those of Types-I and -II (Fig. 4). Therefore, the best-fit function of RH (x) is regarded as Eq. 5, and the humidity variable MRH Type-III is expressed as

$$z = \ln\left(\frac{x^\alpha-\xi}{\lambda+\xi-x^\alpha}\right). \quad (6)$$

The MRH denoted by Eqs. 3 and 6 is accompanied by constant parameters ξ, λ, and α. The MRH Type-I is determined only by the natural humidity restriction range. Meanwhile, ξ, λ, and α of MRH Types-II and -III are the climatological parameters estimated by the function fitting processes. The parameters would vary depending on the season and location, and the climatological parameters must be universal ones.

5. Summary

For applications to statistical humidity variable controls such as climatological variable controls and data assimilations, we have proposed a new humidity variable called MRH. The MRH is constructed to have a normal distribution by using the Johnson's S_B distribution function. We have proposed and examined three types of MRH. Among them, MRH Type-III is the most appropriate variable. Three constant parameters, ξ, λ, and α, must be determined. The values of ξ and $\lambda + \xi$ are considered to be associated with the climatological minimum and maximum of the RH. In Type-I, they are fixed to be 0 and 100%, respectively. In Type-II, they are estimated by the MLE method, but it is not clear whether the parameters estimated by the fitting processes are universal. Values of ξ and λ should be estimated at each grid point and season. The difference in shape between Types-I and -II is small. In Type-III, another constant parameter α must be estimated by the MLE method to correct for the large skewness values of the distributions. The parameter α is also considered to be a climatological universal parameter that varies with location and season. However, the physical meaning of α cannot be explained. In addition, estimation errors on the exponentiation number (α) often induce a large estimation error on MRH. Therefore, we believe that MRH Types-I and -II are adequate variables for the statistical control of humidity at the present stage of study because humidity in the lower atmosphere, which is associated with the triggers of convection and rainfall, is well controlled by MRH Types-I and -II in comparison with the RH. The proposed MRH must be further studied because there are many distribution functions, and the fitting results are better but not perfect.

Acknowledgments

This study is supported by the Tenure Track System at the University of Tsukuba (Yasutaka Wakazuki) (the Japan Science and Technology Agency); and the Research Program on Climate Change Adaptation (RECCA) "An Integrated Water Resources Management System Enhancing Adaptability to Climate Change" (the Ministry of Education, Culture, Sports, Science & Technology in Japan). We express appreciation for beneficial comments from Dr. F. Kimura (Japan Agency for Marine-Earth Science and Technology (JAMSTEC)), Mr. M. Hara (JAMSTEC), the editor Dr. K. Saito (Meteorological Research Institute of JMA) and two anonymous reviewers.

References

Dee, D. P., and A. M. Da Silva, 2003: The choice of variable for atmospheric moisture analysis. *Mon. Wea. Rev.*, **131**, 155−171.

Gustafssona, N., S. Thorsteinssonb, M. Stengela, and E. Holm, 2011: Use of a nonlinear pseudo-relative humidity variable in a multivariate formulation of moisture analysis. *Q. J. R. Meteorol. Soc.*, **137**, 1004−1018.

Hara, M., T. Yoshikane, H. Kawase, and F. Kimura, 2008: Estimation of the impact of global warming on snow depth in Japan by the pseudo-global-warming method. *Hydrological Research Letters*, **2**, 61−64.

Johnson, N. L., 1949: Systems of frequency curves generated by methods of translation. *Biometrika*, **36**, 149−176.

Kimura, F., and A. Kitoh, 2007: Downscaling by pseudo global warning method. *Final report of ICCAP*.

Sato, T., F. Kimura, and A. Kitoh, 2006: Projection of global warming onto regional precipitation over Mongolia using a regional climate model. *Journal of Hydrology*, **333**, 144−154.

Wakazuki, Y., 2011: New distribution functions for hourly and daily precipitation intensities during the snowless season in Japan. *J. Meteor. Soc. Japan*, **89**, 29−45.

Manuscript received 4 April 2013, accepted 2 July 2013
SOLA: http://www.jstage.jst.go.jp/browse/sola/