Clim Dyn
DOI 10.1007/s00382-011-1260-5

Reducing biases in regional climate downscaling by applying Bayesian model averaging on large-scale forcing

Hongwei Yang · Bin Wang · Bin Wang

Received: 24 August 2011 / Accepted: 30 November 2011

Abstract Reduction of uncertainty in large-scale lateral-boundary forcing in regional climate modeling is a critical issue for improving the performance of regional climate downscaling. Numerical simulations of 1998 East Asian summer monsoon were conducted using the Weather Research and Forecast model forced by four different reanalysis datasets, their equal-weight ensemble, and Bayesian model averaging (BMA) ensemble means. Large discrepancies were found among experiments forced by the four individual reanalysis datasets mainly due to the uncertainties in the moisture field of large-scale forcing over ocean. We used satellite water–vapor-path data as observed truth-and-training data to determine the posterior probability (weight) for each forcing dataset using the BMA method. The experiment forced by the equal-weight ensemble reduced the circulation biases significantly but reduced the precipitation biases only moderately. However, the experiment forced by the BMA ensemble outperformed not only the experiments forced by individual reanalysis datasets but also the equal-weight ensemble experiment in simulating the seasonal mean circulation and precipitation. These results suggest that the BMA ensemble method is an effective method for reducing the uncertainties in lateral-boundary forcing and improving model performance in regional climate downscaling.

Keywords Regional climate simulation · Lateral boundary forcing · Bayesian model averaging

H. Yang (✉)
APEC Climate Center, Busan 612020, South Korea
e-mail: hwyang@apcc21.net

B. Wang
Department of Meteorology,
University of Hawaii at Manoa, Honolulu, HI 96822, USA

B. Wang
International Pacific Research Center,
University of Hawaii at Manoa, Honolulu, HI 96822, USA

B. Wang
LASG, Institute of Atmospheric Physics,
Chinese Academy of Sciences, Beijing 100029, China

1 Introduction

Since the first demonstration of regional climate model (RCM) by Dickinson et al. (1989) and Giorgi and Bates (1989), applications of the RCM have surged to meet variety of societal demands. Evaluating uncertainties in regional climate modeling is becoming increasingly important. A number of intercomparison projects of regional model climate simulations have been organized to examine the performance of RCMs (e.g., Christensen et al. 1997; Takle et al. 1999; Leung and Ghan 1999; Mearns et al. 2009; Fu et al. 2005).

Numerous studies demonstrated that the uncertainties in a nested RCM are mainly coming from the lateral boundary (LB) forcing that is used to drive the simulation (e.g., Giorgi and Bi 2000; Liang et al. 2001; Diaconescu et al. 2007; Gong and Wang 2000; Dimitrijevic and Laprise 2005). The uncertainties induced by LB forcing persistently influence the RCM simulation through entire integration (Anthes et al. 1985) and dominate the biases in simulation after spin-up time (Giorgi and Bi 2000). Thus, the uncertainties induced by LB forcing are generally larger than those induced by initial condition, model resolution, physical parameterizations (Vukicevic and Errico 1990; Elía et al. 2008), and even the RCM used (Arritt and Rummukainen 2011).

Published online: 23 December 2011

Moreover, various responses of the RCM to the uncertainties in LB forcing were observed. Wu et al. (2005) found the effect of perturbation in LB forcing shows no clear trend, whereas Christensen et al. (1997) found the RCM amplifies the biases in LB forcing, however, Hong and Leetmaa (1999) found the RCM reduces the biases.

Some studies try to lesson the influence of the uncertainties in LB forcing on the RCM simulation. Qian and Liu (2001) found that choosing the LB location where the LB forcing has rather small biases improves the performance of the RCM. Xue et al. (2007) found that the RCM is sensitive to its southern LB condition when the buffer zone is extended to the tropics and is able to obtain the reasonable simulations only with appropriate LB location. Similarly, Liang et al. (2001) found there is large discrepancy in humidity between the National Centers for Environmental Prediction–National Center for Atmospheric Research global reanalysis (Kalnay et al. 1996) and the 40-year reanalysis data from the European Centre for Medium-Range Weather Forecasts (ERA-40) over low latitude, and suggested that choosing the LB location should avoid this area.

However, none of the studies tried to directly decrease the uncertainties in LB forcing. In simulations of East Asian summer monsoon (EASM), Wang and Yang (2008) and Yang et al. (2011) found that the Weather Research and Forecast (WRF) model driven by (1) the 40-year reanalysis data from the European Centre for Medium-Range Weather Forecasts (ERA-40), (2) National Centers for Environmental Prediction-Department of Energy reanalysis data (NCEP-R2), and (3) the Japanese 25-year reanalysis (JRA-25) yields significantly different results. All experiments have significant deficiencies in reproducing the observed rainfall and large biases in simulating atmospheric circulation. They found that the large vapor uncertainties among reanalysis datasets over the Bay of Bengal and the Philippine Sea are the key factors that induce the largely different realizations. They for the first time proposed to use the equal-weight ensemble forcing (i.e., algebra average of different reanalysis datasets) which decreases the uncertainties in LB forcing and therefore improves model performance.

However, the experiments performed by Wang and Yang (2008) and Yang et al. (2011), driven by the equal-weight ensemble forcing, did not consider the specificity of different reanalysis datasets, especially the humidity fields that induce large model biases. An optimized ensemble scheme might be more competent than the equal-weight ensemble method when the models are very different (Arritt and Rummukainen 2011). Recently, the optimized ensemble method of Bayesian model averaging (BMA) (Hoeting et al. 1999) has been increasingly used in ensemble weather prediction. Raftery et al. (2005) successfully applied BMA to a 48-h regional weather forecast on surface temperature and sea level pressure. Min et al. (2007) performed their climate change study based on the BMA method. Duan et al. (2007) applied BMA to hydrologic prediction and found that BMA prediction is generally superior to the best individual prediction. Sloughter et al. (2007, 2010) applied BMA to precipitation and wind-speed forecasting. BMA is a statistical approach to generating a weighted average of the ensemble members that outperforms any single ensemble member. The weights are estimated according to the performance of individual members on simulating the given training data.

Thus, we were motivated to apply the BMA method to the humidity field, which has the largest uncertainty among the variables in LB forcing, to investigate efficient ways to reduce model biases. We sought to address the following questions: Does the BMA ensemble forcing effectively reduce model biases? Can BMA ensemble forcing improve RCM performance more than equal-weight ensemble forcing?

To answer these questions, we simulated the 1998 EASM with WRF model driven by NCEP-R2, ERA-40, JRA-25, the third-generation ECMWF reanalysis product INTERIM (ERA-IN), their equal-weight ensemble mean, and the same forcing of the previous one except the humidity field whose ensemble scheme is the BMA method. The model and data are described in Sect. 2. The experiments are described in Sect. 3. Section 4 presents the biases in the experiments driven by individual reanalysis datasets. The sources of the model biases are discussed in Sect. 5. In Sect. 6, we describe the application of the BMA ensemble method. Section 7 shows the performance of experiments forced by the equal-weight ensemble forcing and the BMA ensemble forcing. The final section presents the summary and discussion.

2 Model and data

The WRF model (Skamarock et al. 2005), a primitive equation model using sigma coordinates, was used in this study. The main physical options we used included Lin cloud microphysics (Lin et al. 1983; Chen and Sun 2002), the Yongsei University planetary boundary-layer scheme (Noh et al. 2003), the Betts-Miller-Janjic cumulus parameterization (Janjic 2000), the Dudhia (1989) scheme for shortwave radiation, the rapid radiative transfer model for long-wave radiation (Mlawer et al. 1997), and Noah's land-surface model (Chen and Dudhia 2001).

Four reanalysis datasets were used as the large-scale forcing fields: NCEP-R2 (Kanamitsu et al. 2002), ERA-40 (Uppala et al. 2005), ERA-IN (Dee et al. 2011), and JRA-25 (Onogi et al. 2007). The LB forcing fields were

geopotential height, air temperature, specific humidity, and horizontal winds. The width of the buffer zone in our model was set to be 10 grid points, where the prognostic variables of WRF model were nudged toward the reanalysis fields following the method of Davies and Turner (1977) with Newtonian nudging and horizontal diffusion within the RCM buffer zones. The initial state consisted of surface and sea-level pressures, 2-m-height moisture and temperature, 10-m-height horizontal winds, soil moisture and temperature, and skin temperature. The skin temperature over the ocean was considered the sea-surface temperature. The land-use and soil-type data in the land-surface component were obtained from the US Geological Survey and the Food and Agriculture Organization of the United Nations, respectively.

The water vapor path (WVP) of the Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data (HOAPS-3) (Andersson et al. 2007) was used as the observed training data to validate the humidity fields of the reanalysis datasets in the BMA method. The Special Sensor Microwave Imager (SSM/I)-based WVP of HOAPS-3 is a twice-daily gridded data product (1° × 1°). The gauge-based Monsoon Asia Analysis precipitation (resolution 0.5° × 0.5°) dataset, constructed by the Asian Precipitation–Highly–Resolved Observational Data Integration Towards Evaluation of the Water Resources project (Yatagai et al. 2009), was used to verify the simulated precipitation over land. The Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP) dataset (Xie and Arkin 1996) was used as the observation over ocean. Because the spatial resolution of the monthly CMAP dataset is 2.5° × 2.5°, we interpolated model precipitation to the grid of CMAP through bilinear scheme. The simulated circulation was verified with the corresponding large-scale forcing reanalysis datasets.

3 Experimental designs

The model domain covered the EASM region centered on 25°N and 115°E, from the middle of the Tibetan Plateau in the west to the western North Pacific in the east (Fig. 1). The low-level jet stream crossing the Bay of Bengal supplies abundant moisture to the EASM region. The western North Pacific subtropical high residing in the southeast dominates the evolution of monsoon precipitation variability. The model grid had an ~50-km horizontal resolution with 101 west–east grid points, 92 south–north grid points, and 31 vertical sigma levels up to 50 hPa.

Four sensitivity experiments named Exp-R2, Exp-40, Exp-IN, and Exp-25 were initially designed and respectively forced by the NCEP-R2, ERA-40, ERA-IN, and JRA-25 datasets. The control experiment (CTL) was driven

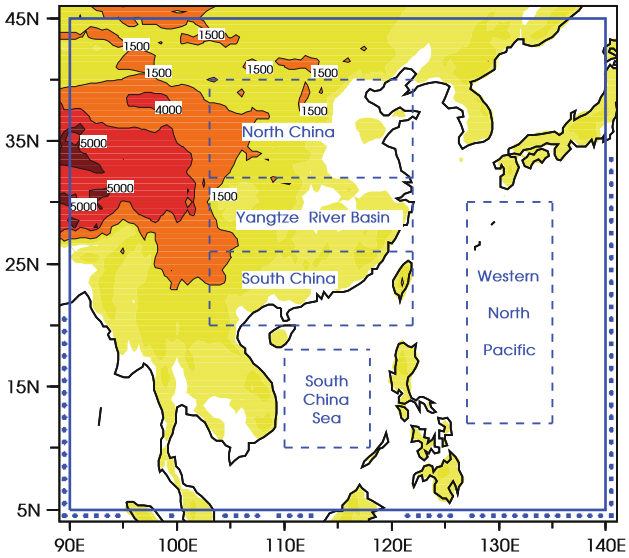

Fig. 1 Simulation domain (enclosed by *solid blue lines*: 5–45°N, 90–140°E) and topography (color shading in units of meters). Topographic contours of 1,500, 4,000, and 5,000 m are highlighted. The satellite-observed training data HOAPS-3 are located at the *blue points* outside the simulation domain. Analysis areas over land enclosed by *dashed lines* refer to South China (20–26°N, 103–122°E), the Yangtze River Basin (26–32°N, 103–122°E), and North China (32–40°N, 103–122°E). Analysis areas over ocean enclosed by *dashed lines* refer to the South China Sea (10–18°N, 110–118°E) and the western North Pacific (12–30°N, 127–135°E)

by the equal-weight ensemble mean of the four reanalysis datasets. To investigate the BMA method, we designed another sensitivity experiment the same as the CTL except that the ensemble mean of the humidity fields was based on the BMA method; we referred to it as Exp-BMA.

Generally, the reanalysis datasets NCEP-R2, ERA-40, and JRA-25 are available at 6-h intervals with a horizontal resolution of 2.5° on an identical grid. To keep all sensitivity experiments and the CTL in the same framework, we interpolated the ERA-IN from its original 1.5° grid to the 2.5° grid of the other reanalysis datasets using a bilinear scheme.

Because the NCEP-R2 and ERA-40 datasets have only 17 pressure levels, we used the same levels from the ERA-IN and JRA-25 to maintain the same number of vertical levels for all large-scale forcing. Different from the ERA-40 and JRA-25, the surface data of ERA-IN and NCEP-R2 are 1.5° and 1.85° respectively, and these datasets were interpolated to a 2.5° grid using a bilinear scheme. The initial conditions of experiments Exp-R2, Exp-40, Exp-IN, and Exp-25 were taken from the NCEP-R2, ERA-40, ERA-IN, and JRA-25 reanalysis datasets, respectively. The same initial conditions of CTL and Exp-BMA were from the equal-weight ensemble mean of the four reanalysis datasets. The physical parameterizations and model configuration were uniform for all experiments.

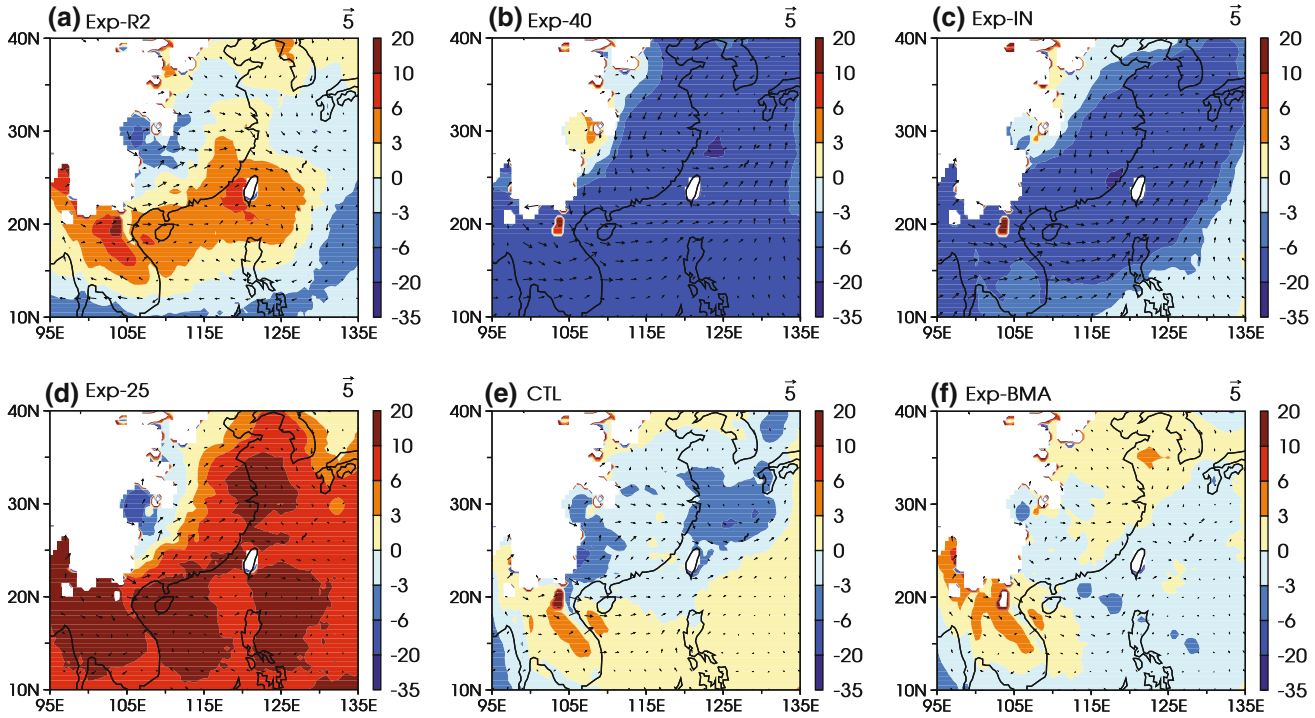

Fig. 2 The model biases of geopotential height (shading in units of meter) and horizontal winds (vector in units of 5 ms^{-1}) at 850 hPa in summer (JJA) mean. Respectively, **a**–**f** are the biases simulated with lateral boundary conditions derived from **a** NCEP/DOE reanalysis 2 (NCEP-R2), **b** ECMWF 40-year reanalysis (ERA-40), **c** ECMWF reanalysis INTERIM (ERA-IN), **d** the Japanese 25-year reanalysis (JRA-25), **e** the equal-weight ensemble mean of the four reanalysis datasets, and **f** the Bayesian model averaging (BMA) ensemble mean of the four reanalysis datasets. All the biases are defined by the departure of RCM simulations from the corresponding large-scale forcing. *White zones* are the missing value masked by the topography

All experiments were started from 00Z on 22 April and ended at 18Z on 31 August 1998. The days before 1 May were considered as a spin-up period (Giorgi and Mearns 1999), and the model output from 1 May to 31 August was analyzed.

4 Systematic biases forced by individual reanalysis datasets

The four sensitivity experiments forced by individual reanalysis datasets yielded different biases in simulations of the June–July–August (JJA) mean low-level circulations. Figure 2a–d show the biases in 850-hPa circulations produced by Exp-R2, Exp-40, Exp-IN, and Exp-25, respectively. The biases were defined by the deviation of the RCM simulations from each corresponding forcing. The bias in Exp-R2 (Fig. 2a) was a notable anticyclonic high-pressure bias over southeastern China and the adjoining oceanic region. Exp-40 produced a bias (Fig. 2b) much larger than that in Exp-R2 but with an opposite sign: a cyclonic low-pressure bias occupied almost the entire 850-hPa level. Exp-IN (Fig. 2c) had a bias pattern similar to Exp-40, but the range was smaller. Exp-25 (Fig. 2d) generated a high-pressure bias mainly over the Philippine Sea, the South China Sea, the Indochina Peninsula, and its associated circulations. These different biases reflected the large uncertainties in experiments driven by individual reanalysis datasets. The location and strength of the western North Pacific subtropical high directly influences the EASM. Therefore, these discrepancies would have marked effects on the monsoon precipitation in regional dynamic downscaling of the EASM.

The differences among simulations were observed not only in low-level circulation but also in seasonal mean precipitation. Compared with Exp-R2 and Exp-25, Exp-40 and Exp-IN produced much more precipitation over the Indo-China peninsula and the oceanic region in simulation domain. Exp-40 produced more (less) precipitation in the eastern (western) Yangtze River Basin than Exp-IN. Exp-25 simulated more precipitation over the Philippines and adjoining ocean than Exp-R2 (figures not shown).

Model skills in simulating seasonal precipitation are shown in Table 1, which presents the spatial correlation coefficient (R) and spatial root mean square error (RMSE) between the simulation and the rain-gauge observation in three regions over land in China (i.e., North China, the Yangtze River Basin, and South China in Fig. 1), and in

Table 1 Spatial correlation coefficients (R) and spatial root mean square error (RMSE, in unit of mm d^{-1}) of seasonal precipitation between the observation and Exp-R2, Exp-40, Exp-IN, Exp-25, CTL, and Exp-BMA in land of North China (NC), the Yangtze River Basin (YZ), South China (SC), and over ocean of the South China Sea (SCS) and the western North Pacific (WNP). R_t is the temporal correlation coefficient of daily precipitation averaged over land of NC, YZ, and SC between the observation and the simulations

R/RMSE/R_t	NC	YZ	SC	SCS	WNP
Exp-R2	0.71/1.65/0.49	0.26/4.04/0.38	0.42/4.97/0.67	−0.05/2.21	0.65/2.88
Exp-40	0.60/1.47/0.25	0.18/4.38/0.31	0.20/6.42/0.30	0.84/4.82	0.30/3.32
Exp-IN	0.64/1.37/0.45	0.08/4.54/0.11	0.12/7.79/0.39	0.32/6.83	0.45/3.61
Exp-25	0.72/1.34/0.63	0.33/4.14/0.57	0.41/5.28/0.49	0.55/2.08	0.56/2.84
CTL	0.68/1.53/0.40	0.10/4.96/0.39	0.43/6.70/0.62	0.30/3.06	0.55/3.62
Exp-BMA	0.78/1.07/0.42	0.02/3.76/0.44	0.48/4.63/0.52	0.64/2.16	0.51/2.65

two regions over ocean (i.e., the South China Sea and the western North Pacific). Among the sensitivity experiments forced by individual reanalysis datasets, Exp-25 had the largest R of 0.72, while Exp-40 had the smallest R of 0.60 over North China. Over the Yangtze River Basin, all sensitivity experiments forced by individual reanalysis datasets showed different Rs, but generally the Rs were small. Over South China, Exp-R2 had the largest R of 0.42, whereas Exp-IN had the smallest R of 0.12. Over the South China Sea, Exp-40 had the largest R of 0.84, whereas Exp-R2 had a small negative R of −0.05. Over the western North Pacific, Exp-R2 had the largest R of 0.65, while Exp-40 had the smallest R of 0.30.

In terms of RMSE, the results of sensitivity experiments forced by individual reanalysis datasets were also different. Exp-25 had the smallest RMSE, while Exp-R2 had the largest RMSE over North China. Over both the Yangtze River Basin and South China, the smallest (largest) RMSEs were observed from Exp-R2 (Exp-IN). But generally, the RMSEs over the Yangtze River Basin and South China were larger than those over North China, which means that tropical rainfall is difficult to simulate. Over both the South China Sea and the western North Pacific, Exp-25 had the smallest RMSEs, while Exp-IN showed the largest RMSEs.

Table 1 also shows the temporal correlation coefficient (R_t) of daily precipitation averaged over North China, the Yangtze River Basin, and South China between the simulation and the observation. Exp-R2 performed best in South China. Exp-25 performed best in both North China and the Yangtze River Basin.

With the same configuration, the aforementioned differences among the sensitivity experiments forced by individual reanalysis datasets could only be caused by the different initial and LB conditions. Our extra sensitivity experiments, in which the initial conditions of Exp-R2, Exp-40, Exp-IN, and Exp-JRA were replaced by the averaged initial conditions of themself, showed that compared with the biases induced by both initial and LB conditions, the biases that was solely induced by the initial condition were minor, which indicates that the biases were mainly determined by the LB forcing. As in the findings in Wang and Yang (2008) and Yang et al. (2011), here, the uncertainties in humidity fields of the LB forcing might also be the most important cause of the discrepancies. An analysis in detail is presented in the next section.

5 The source of the biases

To determine the forcing field most likely to result in the large discrepancies in the last section, we investigated the relative differences among the forcing fields: horizontal wind speed, geopotential height, absolute humidity, and temperature of the four reanalysis datasets. We used the spatial average of noise-to-signal ratio (NSR; the spatial average of the inverse of signal-to-noise ratio), on each side of the LBs from 1,000 mb up to 100 mb in JJA to present the relative differences of a variable among the four reanalysis datasets. For instance, the temperature mean of the four reanalysis datasets is a signal that indicates temperature quantity. Temperature deviation from the mean is a noise that indicates spread among the individual reanalysis dataset from their mean. The relative differences can be expressed as the ratio of noise to signal.

Table 2 shows that absolute humidity had the largest NSRs on all LBs, which means that the differences in absolute humidity were most likely to be the key factor resulting in the large discrepancies in the experiments forced by individual reanalysis datasets. Different from the western, eastern, and southern LBs, although the NSR value (0.17) was large at the northern boundary, the relative difference was not necessarily large, because the spatial average of the mean absolute humidity (i.e., an approximation of the signal in NSR) was minimal (i.e., 2.66). Thus, the uncertainties in the humidity field at the western, eastern, and southern boundaries were the major sources of the large discrepancies.

To determine where the humidity field had the largest differences, the standard deviation of the JJA mean absolute humidity of the four reanalysis datasets at the western,

Table 2 Spatial average of noise-signal ratio (NSR) of large-scale wind speed (WS), geopotential height (z), absolute humidity (a), and air temperature (T) at the western, eastern, southern, and northern lateral boundaries

NSR/mean	WS	z	a	T
West	0.07	0.00	0.17/5.53	0.00
East	0.04	0.00	0.12/5.94	0.00
South	0.06	0.01	0.14/7.00	0.00
North	0.04	0.00	0.17/2.66	0.00

The spatial average of the signal (mean, in unit of g m^{-3}) of absolute humidity is also shown

southern, and eastern LBs below 300 hPa were examined (Fig. 3). The large humidity differences at the western boundary were observed at the boundary layer and the middle atmosphere over the Bay of Bengal. At the southern boundary, the large humidity differences existed on the east side of the boundary. At the eastern boundary, large humidity differences existed in the boundary layer and the low-level atmosphere over the Philippine Sea. Although the vertical distribution of the humidity differences varied from place to place, it was clear that almost all of the large differences of humidity were pronounced over the ocean, which agrees with the findings in Wang and Yang (2008) and Yang et al. (2011). We reduced the large uncertainties in humidity field of the LB forcing over ocean using the BMA method as described in the next section.

6 Application of BMA ensemble method

In this study, we applied the BMA method to the humidity field of the LB forcing in dynamic downscaling of the EASM with WRF model. The approach is briefly described as follows.

Because the models used for the reanalysis datasets depict the atmospheric behaviors, the WVPs (the vertical integration of humidity) of the four reanalysis datasets are the estimations of the true WVP. The posterior probability $p(M_k^w|O^w)$ is the conditional probability of the WVP simulated by the k-th reanalysis model on the observed WVP, which reflects how well the reanalysis dataset mimics the observed WVP in the given places (locations indicated by the blue points in Fig. 1). The sum of those posterior probabilities is equal to one, thus they can be used as a set of weights to measure the skills of reanalysis models. Because WVP is the vertical integration of humidity, the weights obtained from two-dimensional WVP could be used for weighting humidity in three dimensions. Then the probability of BMA ensemble mean of the humidity near the whole LB region was the weighted sum of the probability of the humidity from the four reanalysis datasets:

$$p(h) = \sum_{k=1}^{4} w_k p(h|M_k),$$

where h is the BMA ensemble mean of humidity field, $p(h|M_k)$ is the probability of simulated humidity based on the k-th reanalysis model, the weight w_k is the posterior probability $p(M_k^w|O^w)$.

The observed training data was the WVP of HOAPS-3 twice daily from 00Z on 22 April through 18Z on 31 August 1998 at the locations indicated by the blue points in Fig. 1. The data captured WVP near the LB over the Bay of Bengal, the South China Sea, the Philippine Sea, and the western North Pacific, where the humidity fields differed greatly among reanalysis datasets and led to the large discrepancies in dynamic downscaling of EASM.

To calculate the weights (posterior probabilities), the humidity fields of reanalysis datasets were horizontally interpolated to the points in 17 levels that geographically overlaid the blue points in Fig. 1. Then the interpolated humidity fields were vertically integrated to obtain the WVP that was comparable to the observation. We calculated the weights using the Expectation–Maximization algorithm (details in Raftery et al. 2005 and Duan et al. 2007). To ensure the distribution of the WVP approximated

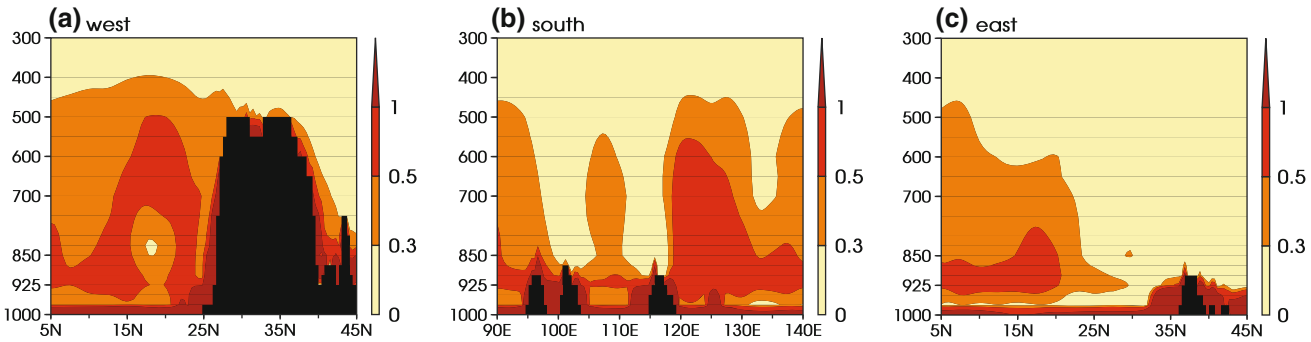

Fig. 3 The standard deviation (g m^{-3}) of JJA mean absolute humidity of the large-scale reanalysis forcing of NCEP-R2, ERA-40, ERA-IN, and JRA-25 at the western, southern, and eastern lateral boundaries below 300 hPa. *Black color* represents topography

normality, a Box–Cox transformation (Box and Cox 1964) was used to preprocess the WVP of the observation and the four reanalysis datasets before utilizing the Expectation–Maximization algorithm. This power transformation takes the following form:

$$z = \begin{cases} \frac{y^b - 1}{b}, & b \neq 0 \\ \log(y), & b = 0 \end{cases},$$

where y is the observation or reanalysis datasets and z is the transformed variable. The same transformation parameter b for the observation and reanalysis datasets was inferred from the maximum likelihood method.

Figure 4a shows the weights of the four reanalysis datasets with the same temporal resolution of HOAPS-3 in 12 h at 00z and 12z. The weights for one reanalysis dataset on all the LB points at the same time are equal. The weights at 06z (18z) took their former values at 00z (12z) because of the necessity of the 6-h interval of the LB forcing. The weights of ERA-IN had the greatest chances of being the largest, which means that the WVP of ERA-IN were closest to that of HOAPS-3 among the four reanalysis datasets, while the weights of NCEP-R2 had the greatest chances of being the smallest.

Figure 4b and c show the comparison of the WVP between the equal-weight ensemble mean and the BMA ensemble mean of the four reanalysis datasets based on the observation. Relative to HOAPS-3, the Rs of WVP of the BMA ensemble mean are higher than that of the equal-weight ensemble mean at almost all the times, and the RMSEs of WVP of the BMA ensemble mean are significantly reduced compared with those of the equal-weight ensemble mean at almost all the times. Forced by the BMA ensemble mean whose WVP were closer to HOAPS-3 than those of the equal-weight ensemble mean, what happened in the WRF model is analyzed in the next section.

7 Improved simulation with BMA ensemble method

It is well known that the ensemble climate prediction of global models from different centers yield better results than any individual prediction (e.g., Kalnay and Ham 1989; Krishnamurti et al. 1999; Shukla et al. 2000; Wang et al. 2004), because the multi-model ensemble (MME) reduces the uncertainties in the models's physical parameterization. Similarly, using multi-reanalysis ensemble mean as the LB forcing, RCM produces better results than most experiments forced by individual reanalysis dataset (Wang and Yang 2008 and Yang et al. 2011).

In this study, the equal-weight ensemble means of horizontal winds, geopotential height, temperature, and humidity from the NCEP-R2, ERA-40, ERA-IN, and JRA-25 reanalysis datasets were used as the LB forcing in the CTL experiment. Another sensitivity experiment Exp-BMA was similar to the CTL, but the ensemble mean of humidity field was based on the BMA method. The setup of these experiments was the same as the former.

The superiorities of CTL and Exp-BMA experiments were observed in low-level circulation. Figure 2e and f

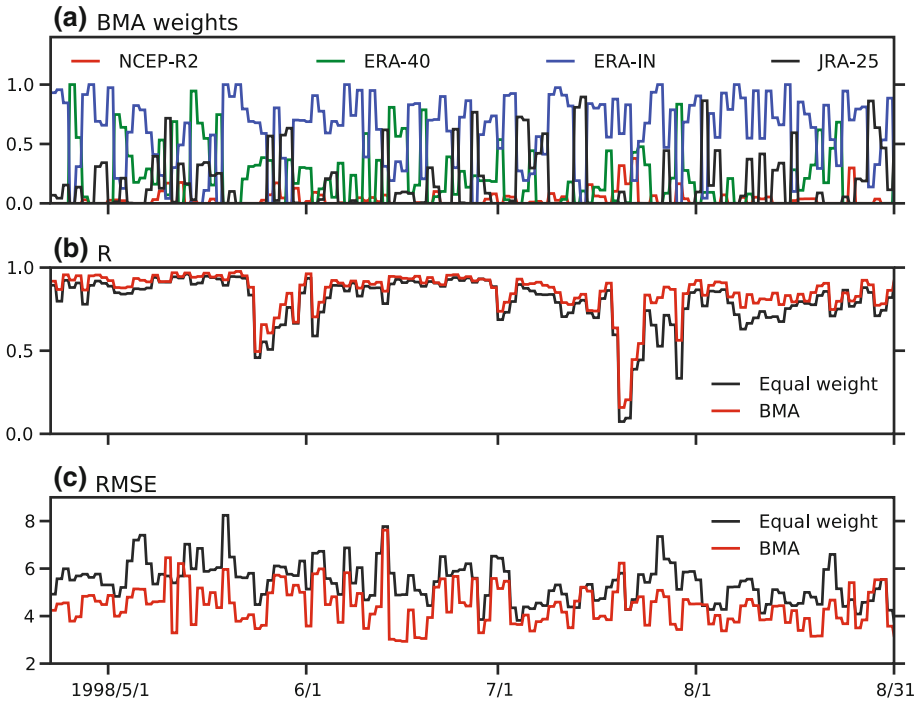

Fig. 4 **a** The Bayesian model averaging (BMA) weights of the water vapor path (WVP) of NCEP-R2 (*red*), ERA-40 (*green*), ERA-IN (*blue*), and JRA-25 (*black*) calculated from satellite-observed training data HOAPS-3 at the locations indicated by the *blue points* in Fig. 1**b** and **c** are the correlation coefficients (R) and root mean square error (RMSE, in unit of kg m^{-2}) of WVP of equal-weight ensemble mean (*black*) and BMA ensemble mean (*red*), respectively, of the four reanalysis datasets relative to HOAPS-3

shows that compared with the experiments forced by individual reanalysis datasets in Fig. 2a–d, the CTL experiment (Fig. 2e) dramatically reduced circulation biases. The results of the CTL experiment only slightly underestimated the geopotential height over the East China Sea (with a cyclone wind bias) and western South China, and they only overestimated the geopotential height in Laos. The improvement in the CTL forced by the equal-weight ensemble forcing agrees with the findings in Wang and Yang (2008) and Yang et al. (2011). Surprisingly, the Exp-BMA (Fig. 2f) significantly decreased the underestimated biases of geopotential height in the CTL even further. The underestimated geopotential height (<-3 m) exists only over several small areas without organized patterns of wind biases, although the area of overestimated geopotential height slightly increased over the Indo-China peninsula. Compared with the CTL, the further improvement in Exp-BMA was the result of the benefit from the BMA ensemble method applied to the LB humidity field. Thus, constrained by the observed humidity even in 2 dimensions over the ocean, the LB forcing forced the model results with much less bias.

Improvements were also observed in seasonal precipitation. Table 1 shows that the CTL had a larger R ($R = 0.68$) than those of Exp-40 ($R = 0.60$) and Exp-IN ($R = 0.64$) in North China. In South China, the CTL had larger R ($R = 0.43$) than those of all experiments forced by individual reanalysis datasets. Exp-BMA had the largest Rs in all experiments in both North China ($R = 0.78$) and South China ($R = 0.48$). In the Yangtze River Basin, almost all experiments had little accuracy in terms of R, which was probably caused by the pattern shift of precipitation. Over the South China Sea, the CTL did not have a good skill in R, while Exp-BMA was better than most experiments forced by individual reanalysis datasets except Exp-40. Over the western North Pacific, although both the CTL and Exp-BMA did not achieve the highest score in terms of R, they were better than the averaged R (0.49) of experiments forced by individual reanalysis datasets. Over all, the CTL was not as good as we expected, but the Exp-BMA performed better than expected.

In terms of RMSE, the CTL had relatively lower accuracy than most experiments forced by individual reanalysis datasets in both North China and South China, and it performed worse than all experiments forced by individual reanalysis datasets in the Yangtze River Basin and the western North Pacific, while Exp-BMA had the best performance in terms of RMSE in all experiments in four regions: North China, RMSE = 1.07; Yangtze River Basin, RMSE = 3.76; South China, RMSE = 4.63; and western North Pacific, RMSE = 2.65. Exp-BMA was only slightly worse than Exp-25 in the South China Sea. Relative to the CTL, the better performance of Exp-BMA in simulating seasonal precipitation can be attributed to the BMA ensemble mean of humidity in the LB forcing.

Table 1 also shows the temporal correlation coefficient (R_t) of daily precipitation between simulation and observation averaged over North China, the Yangtze River Basin, and South China. The CTL had better performance than Exp-40 in North China, and only performed worse than Exp-25 in the Yangtze River Basin and Exp-R2 in South China. Although Exp-BMA did not perform well compared with some experiments, it performed better than the CTL in both North China and the Yangtze River Basin.

The results in precipitation simulation suggest that the CTL forced by the equal-weight ensemble forcing might not reduce the biases in some aspects of the simulations; sometimes it was worse than the experiments forced by individual reanalysis datasets. But the BMA ensemble forcing was able to further reduce the biases in RCM simulations: over land, Exp-BMA outperformed not only the experiments forced by individual reanalysis datasets but also the CTL, except the daily precipitation in which the Exp-BMA was comparable to the CTL; over ocean, Exp-BMA outperformed the CTL and the most experiments forced by individual reanalysis datasets, except in the western North Pacific where its correlation coefficient was slightly lower than that of CTL but higher than the averaged correlation coefficient of the experiments forced by individual reanalysis datasets.

8 Summary

Simulations of the 1998 East Asian summer monsoon (EASM) were carried out using the WRF model forced by four reanalysis datasets (i.e., NCEP-R2, ERA-40, ERA-IN, and JRA-25), their equal-weight ensemble mean, and their BMA ensemble mean. There were large discrepancies in seasonal circulation at 850 hPa, seasonal precipitation, and daily precipitation (Fig. 2 and Table 1) in the experiments forced by individual reanalysis datasets. Diagnostic analysis of the spatial average of noise-to-signal ratio revealed that the differences in moisture fields of the large-scale forcing at the eastern, western, and southern boundaries over the ocean were responsible for the large discrepancies. The largest moisture uncertainties were found over the Bay of Bengal, which agrees with the findings of Wang and Yang (2008) and Yang et al. (2011).

As expected, the experiment (control) forced by the equal-weight ensemble forcing significantly reduced the biases in the simulated low-level circulation, while it only reduced the biases in simulated precipitation compared with some experiments forced by individual reanalysis datasets. These results suggest that the use of equal-weight ensemble forcing may not reduce the biases in all cases,

which agrees with the findings in Yang et al. (2011) that the use of equal-weight ensemble forcing systematically reduces the biases in RCM simulations in most of the years, but not in all years. The reason might be that the quality of the equal-weight ensemble forcing highly relied on the choice of the ensemble members when ensemble size is small, and low-performance members might make the ensemble mean unreliable.

The experiment (Exp-BMA) forced by the BMA ensemble forcing reduced the simulation biases not only in low-level circulation field but also in most cases in precipitation compared to the CTL and the experiments forced by individual reanalysis datasets. In the LB forcing of Exp-BMA, the weights of humidity field were determined by how well the reanalysis datasets mimicked the observed water vapor path (WVP) over ocean where the uncertainties in humidity field of the LB forcing were the leading factor that induced the large discrepancies in experiments forced by individual reanalysis datasets. Those weights changed with time according to the temporal variations of the quality of individual reanalysis dataset based on the observation. Thus, the BMA ensemble forcing was able to minimize the biases across time; therefore its quality was guaranteed (Fig. 4b, c).

Limited by availability of three-dimensional observation, this preliminary study was only carried out for simulations of the 1998 EASM and only relied on the two-dimensional observed WVP data to determine the weights using the BMA method. This type of study should be repeated under different circulation regimes with varying LB locations and different dynamic core and physical parameterizations of the WRF model to validate the BMA ensemble forcing in reducing the biases in regional climate downscaling. However, as long as the humidity fields of LB forcing have large uncertainty, RCM solutions would have large uncertainties. Our additional studies show that a climatological large uncertainty pertaining to moisture among the four reanalysis datasets existed in a large area around the LB location of our model domain (figure not shown). Thus, changing the LB location or the model setup would only change our results quantitatively. Even changing the simulation year would uncommonly change our results qualitatively. The biases produced there are expected to be further reduced by using three-dimensional satellite data including not only humidity but also other fields as the training base of the BMA method.

Acknowledgment The first author Hongwei Yang acknowledges support from APEC Climate Center (APCC), Chinese 973 Project (2010CB428403), and Chinese 863 Project (2010AA012301). The second author BW acknowledges support from NOAA Climate Test Bed project and the Global Research Laboratory (GRL) program, which is sponsored by the National Research Foundation of Korea (grant # 2100-0021927). The authors would also like to thank two anonymous reviewers of this manuscript for their useful and insightful comments.

References

Andersson A, Bakan S et al (2007) Hamburg ocean atmosphere parameters and fluxes from satellite data—HOAPS-3—twice daily composite. World Data Center for Climate doi:10.1594/WDCC/HOAPS3_DAILY

Anthes RA, Kuo YH, Baumhefner DP, Errico RM, Bettge TW (1985) Predictability of mesoscale atmospheric motions. Contribution to ''issues in atmospheric and oceanic modeling''. Adv geophys, vol 28B. Academic Press, pp 159–202

Arritt RW, Rummukainen M (2011) Challenges in regional-scale climate modeling. Bull Am Meteor Soc 92:365–368

Box GEP, Cox DR (1964) An analysis of transformations. J R Stat Soc B 26:211–252

Chen F, Dudhia J (2001) Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: model implementation and sensitivity. Mon Weather Rev 129:569–585

Chen S-H, Sun W-Y (2002) A one-dimensional time dependent cloud model. J Meteor Soc Jpn 80:99–118

Christensen JH, Machenhauer B, Jones RG, Schar C, Ruti PM, Castro M, Visconti G (1997) Validation of present-day regional climate simulations over Europe: LAM simulations with observed boundary conditions. Clim Dyn 13:489–506

Davies HC, Turner RE (1977) Updating prediction models by dynamical relaxation: an examination of the technique. Quart J R Meteor Soc 103:225–245

Dee DP, Uppala SM et al (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Quart J R Meteorol Soc 137:553–597

Diaconescu EP, Laprise R, Sushama L (2007) The impact of lateral boundary data errors on the simulated climate of a nested regional climate model. Clim Dyn 28:333–350

Dickinson RE, Errico RM, Giorgi F, Bates GT (1989) A regional climate model for western United States. Clim Change 15:383–422

Dimitrijevic M, Laprise R (2005) Validation of the nesting technique in a regional climate model and sensitivity tests to the resolution of the lateral boundary conditions during summer. Clim Dyn 25:555–580

Duan Q, Ajami NK, Gao X, Sorooshian S (2007) Multi-Model ensemble hydrologic prediction using Bayesian model averaging. Adv Water Resour 30:1371–1386

Dudhia J (1989) Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. J Atmos Sci 46:3077–3107

Elía R et al (2008) Evaluation of uncertainties in the CRCM-simulated North American climate. Clim Dyn 30:113–132

Fu C-B, Wang S-Y et al (2005) Regional climate model intercomparison project for Asia. Bull Am Meteor Soc 86:257–266

Giorgi F, Bates GT (1989) The climatological skill of a regional model over complex terrain. Mon Weather Rev 117:2325–2347

Giorgi F, Bi X (2000) A study of internal variability of a regional climate model. J Geophys Res 105:29503–29521

Giorgi F, Mearns LO (1999) Introduction to special section: regional climate modeling revisited. J Geophys Res 104:6335–6352

Gong W, Wang W-C (2000) A regional model simulation of the 1991 severe precipitation event over the Yangtze–Huai river Valley, Part II: Model Bias. J Clim 13:93–108

Hoeting JA, Madigan D, Raftery AE, Volinsky CT (1999) Bayesian model averaging: a tutorial. Stat Sci 14:382–417

Hong S-Y, Leetmaa A (1999) An evaluation of the NCEP RSM for regional climate modeling. J Clim 12:592–609

Janjic ZI (2000) Comments on ''development and evaluation of a convection scheme for use in climate models''. J Atmos Sci 57:3686

Kalnay E, Ham M (1989) Forecasting forecast skill in the Southern Hemisphere. Extended Abstracts, Third international conference on southern hemisphere meteorology and oceanography, Buenos Aires, Argentina, Am Meteor Soc 24–27

Kalnay E et al (1996) The NCEP/NCAR 40-Year Reanalysis Project. Bull Am Meteor Soc 77:437–471

Kanamitsu M, Ebisuzaki W et al (2002) NCEP-DOE AMIP-II reanalysis (R-2). Bull Am Meteorol Soc 83:1631–1643

Krishnamurti TN, Kishtawal CM, LaRow TE, Bachiochi DR, Zhang Z, Williford CE, Gadgil S, Surendran S (1999) Improved weather and seasonal climate forecasts from multi-model superensemble. Science 285:1548–1550

Leung LR, Ghan SJ (1999) Pacific Northwest climate sensitivity simulated by a regional climate model driven by a GCM. Part I: control simulations. J Clim 12:2010–2030

Liang X-Z, Kunkel KE, Samel AN (2001) Development of a regional climate model for US midwest applications. Part I—sensitivity to buffer zone treatment. J Clim 14:4363–4378

Lin Y-L, Farley RD, Orville HD (1983) Bulk parameterization of the snow field in a cloud model. J Clim Appl Meteorol 22:1065–1092

Mearns LO, Gutowski WJ, Jones R et al (2009) A regional climate change assessment program for North America. EOS 90:311–312

Min S-K, Daniel S, Andreas H (2007) Probabilistic climate change predictions applying Bayesian model averaging. Philos Trans R Soc A 365:2103–2116

Mlawer EJ, Taubman SJ, Brown PD et al (1997) Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated-k model for the longwave. J Geophys Res 102:16663–16682

Noh Y, Cheon WG, Hong SY, Raasch S (2003) Improvement of the K-profile model for the planetary boundary layer based on large eddy simulation data. Bound Layer Meteor 107:401–427

Onogi K, Tsutsui J, Koide H et al (2007) The JRA-25 reanalysis. J Meteorol Soc Jpn 85:369–432

Qian Y-F, Liu H-Q (2001) On nesting area selection of regional climate models coupled to a global climate model (in Chinese). Chin J Atmos Sci 25:492–502

Raftery AE, Gneiting T, Balabdaoui F, Polakowski M (2005) Using Bayesian model averaging to calibrate forecast ensembles. Mon Weather Rev 133:1155–1174

Shukla J, Anderson J, Baumhefner D et al (2000) Dynamical seasonal prediction. Bull Am Meteorol Soc 81:2593–2606

Skamarock WC, Klemp JB, Dudhia J et al (2005) A description of the advanced research WRF version 2. NCAR tech notes-468 + STR

Sloughter JML, Raftery AE, Gneiting T, Fraley C (2007) Probabilistic quantitative precipitation forecasting using bayesian model averaging. Mon Weather Rev 135:3209–3220

Sloughter JML, Gneiting T, Raftery AE (2010) Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. J Am Stat Assoc 105:25–35

Takle ES, Gutowski WJ Jr, Arritt RW et al (1999) Project to intercompare regional climate simulations (PIRCS): description and initial results. J Geophys Res 104:19443–19461

Uppala SM, Kallberg PW, Simmons AJ et al (2005) The ERA-40 re-analysis. Q J Roy Meteorol Soc 131:2961–3012

Vukicevic T, Errico RM (1990) The influence of artificial and physical factors upon predictability estimates using a complex limited-area model. Mon Weather Rev 118:1460–1482

Wang B, Yang H-W (2008) Hydrological issues in lateral boundary conditions for regional climate modeling: simulation of east asian summer monsoon in 1998. Clim Dyn 31:477–490

Wang B, Kang I-S, Lee J-Y (2004) Ensemble simulations of Asian–Australian monsoon variability by 11 AGCMs. J Clim 17:803–818

Wu W, Lynch AH, Rivers A (2005) Estimating the uncertainty in a regional climate model related to initial and lateral boundary conditions. J Clim 18:917–933

Xie P, Arkin PA (1996) Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. Bull Am Meteor Soc 78:2539–2558

Xue Y, Vasic R, Janjic Z, Mesinger F, Mitchell KE (2007) Assessment of dynamic downscaling of the continental U.S. regional climate using the eta/ssib regional climate model. J Clim 20:4172–4193

Yang H-W, Wang B, Wang B (2011) Reduction of systematic biases in regional climate downscaling through ensemble forcing. Clim Dyn doi:10.1007/s00382-011-1006-4

Yatagai A, Arakawa O, Kamiguchi K, Kawamoto H, Nodzu MI, Hamada A (2009) A 44-year daily gridded precipitation dataset for Asia based on a dense network of rain gauges. SOLA 5:137–140. doi:10.2151/sola.2009-035