Efficiently Constraining Climate Sensitivity with Ensembles of Paleoclimate Simulations

J. D. Annan[1], J. C. Hargreaves[1], R. Ohgaito[1], A. Abe-Ouchi[2,1] and S. Emori[3,1]
[1]*Frontier Research Center for Global Change, Yokohama, Japan*
[2]*Center for Climate System Research, The University of Tokyo, Japan*
[3]*National Institute for Environmental Studies, Tsukuba, Japan*

Abstract

We use a recently-developed efficient probabilistic estimation technique to estimate the sensitivity of the Earth's temperature to a doubling of atmospheric carbon dioxide. The method is based on the ensemble Kalman filter which we apply to the CCSR/NIES/FRCGC AGCM (the atmospheric component of MIROC 3.2) at T21L20 resolution coupled to a slab ocean. The method combines prior beliefs about the model, with observational data, to simultaneously estimate 25 model parameters in an efficient and objective manner. We perform a sensitivity analysis to investigate the effect of different assumptions regarding model error, since this is a necessarily subjective input which has not yet been well characterised. We attempt to validate the resulting ensembles against out-of-sample data by comparing their hindcasts of the Last Glacial Maximum (LGM) to paleoclimate proxy data, and demonstrate through this that our ensembles of simulations are probably biased towards too high a sensitivity. Within the framework of our single-model ensemble experiment, we show that climate sensitivity of much greater than 6°C is hard to reconcile with the paleoclimate record, and that of greater than 8°C seems virtually impossible. Our estimate for the most likely climate sensitivity is in the region of 4.5°C. Although these results are reasonably consistent with the most widely accepted estimates of climate sensitivity, they disagree with some recent research which has suggested a significant probability of sensitivities well in excess of these values. These results suggest that paleoclimatic evidence could provide a useful, albeit imprecise, constraint on ensemble forecasts of future climate change.

1. Introduction

Estimates of the future response of climate to anthropogenic forcing form an important input into the policy-making process for mitigation and adaptation. Evaluating the uncertainty in model forecasts of anthropogenically-forced climate change was identified as a high priority in the IPCC TAR (Houghton et al. 2001) but despite much effort, substantial uncertainty remains. Analyses of recent climate change data appear to indicate that climate sensitivity (the equilibrium temperature response to a doubling of atmospheric carbon dioxide) has a substantial probability of being well in excess of 6°C (Andronova and Schlesinger 2001; Gregory et al. 2002; Knutti et al. 2002), and a recent massive ensemble of GCM simulations illustrated that such extreme behaviour can also be exhibited by state-of-the-art models which simulate the present annually-averaged climate state reasonably well (Stainforth et al. 2005). The behaviour of a coupled ocean-atmosphere model on climatological time scales is highly dependent on the details of parameterisations which cannot be accurately determined from theory or direct observations. In order to constrain this source of uncertainty, we have recently developed a computationally efficient multivariate parameter estimation scheme, based on the ensemble Kalman filter (Evensen 2003), which has been tested on a variety of highly nonlinear models (Annan et al. 2005, and references therein). The method constructs samples according to the posterior joint probability distribution defined by a likelihood function, rather than drawing samples from a diffuse prior and then discarding (or substantially down-weighting) poor samples as has been previously performed (Knutti et al. 2002; Murphy et al. 2004; Stainforth et al. 2005). In this paper, we describe and present results from the first full-scale application of this method, in which it has been applied to simultaneously estimating 25 model parameters in a state-of-the-art AGCM coupled to a slab ocean. One critical input into the process is the estimate of uncertainty of the model inadequacy (also called discrepancy) (Kennedy and O'Hagan 2001) which has so far received surprisingly little attention given its important rôle in applications of this type (Annan and Hargreaves 2005; Rougier 2005). Since this term is as yet poorly characterised, we have performed a range of experiments using assumptions that we expect to cover the plausible range of model inadequacy (both doubling and halving our best prior estimate). This introduces an unavoidably subjective element into the estimation process, so we test the forecast skill of the resulting ensembles by simulating out-of-sample data (i.e., data independent of the modern data which were used to tune the model). In particular, we simulate the Last Glacial Maximum, a time when the climate was substantially different to today. Although direct calculations of climate sensitivity have been performed based on the LGM state (e.g., Lea 2004), we do not believe that such an approach to validation has been previously attempted in ensemble-based climate prediction with a state of the art GCM. From our LGM experiments, we can draw some admittedly tentative conclusions as to how skillfully our ensembles may be expected to predict the climate under conditions of doubled carbon dioxide.

2. Methods

We use the CCSR/NIES/FRCGC AGCM at T21L20 resolution, coupled to a slab ocean on the same horizontal grid. The AGCM is the atmospheric part of the MIROC3.2 model (Hasumi and Emori 2004). Heat fluxes in the slab ocean are first calculated via a 'nudging' run in which sea surface temperatures are strongly nudged to climatology, with the implied heat flux divergence calculated and stored for later use, and then a simulation is performed in which this previously computed heat flux is applied, with ocean temperatures allowed to evolve freely. The same heat flux divergence field is used for the present-day, doubled carbon dioxide and LGM integrations, and so these model runs all ignore changes that could arise in ocean heat transport due to circulation changes. We allowed 25 parameters (which describe the major uncertainties in the model, especially sub-grid-scale cloud parameterisations) to vary simultaneously, with prior estimates chosen to be as broad as reasonable.

We chose 15 diverse data types which together

Corresponding author: J. D. Annan, FRCGC/JAMSTEC, Yokohama, 236-0001, Japan. E-mail: jdannan@jamstec.go.jp.

describe the main features of the climate system, including temperature, moisture/precipitation, radiation balance and wind fields from the ERA-40 (Simmons and Gibson 2000), CMAP (Xie and Arkin 1997) and ERBE (Harrison et al. 1990) projects. All data consist of 2D fields with near-complete global coverage, and DJF and JJA seasons are considered separately, so our experimental design has more in common with Murphy et al. (2004) than Stainforth et al. (2005) who used annually averaged data.

The discrepancy between climate models and data is generally dominated by the model error, and therefore the perfect model assumption is not tenable. That is, the uncertainty of the observations is much smaller than the uncertainty of the model error itself (where the "model error" is defined as the difference between reality, and the model evaluated at the "best" set of parameters for the problem in question) (Annan and Hargreaves 2005). The theory and implications of these issues are also explored in more detail in Kennedy and O'Hagan (2001) and Rougier (2005). The practical effect of accounting for model error is to increase the denominator of the cost function for model fitness, thereby broadening the posterior distribution and increasing the spread of the results compared to when a perfect model assumption is made. The cost function under the perfect model assumption can be written as

$$J = \frac{1}{2}(\mathbf{m}-\mathbf{o})^T(\mathbf{R})^{-1}(\mathbf{m}-\mathbf{o}), \qquad (1)$$

where $\mathbf{m}$ is the vector of model variables corresponding to observations $\mathbf{o}$, and $\mathbf{R}$ is the covariance matrix of observational errors. To account for model error, we use

$$J = \frac{1}{2}(\mathbf{m}-\mathbf{o})^T(\mathbf{R}+\mathbf{T})^{-1}(\mathbf{m}-\mathbf{o}), \qquad (2)$$

where $\mathbf{T}$ is the covariance matrix of the uncertainty in model error. However, in the climate sciences, the magnitude and form of model error is not well known, certainly at present, if not fundamentally so, and its specification therefore relies on a substantially subjective judgement. We use our control run as a baseline from which to estimate model error, while noting that this is not an optimally tuned run but seems plausible on all measures. Our basic assumption, made partly on pragmatic grounds, is that the *uncertainty* in model error for each data type can be assumed spatially and seasonally invariant, and is proportional to the *actual* RMS error of that data type in the control run, with the constant of proportionality being a tunable factor. More research is needed into the specification of the uncertainty in model error, but we think that our assumptions are a reasonable starting point, and note that any explicit consideration of this point represents an advance over most previous work in which it has generally been ignored.

Estimates of model error for each of the 15 data types were based on 20 years of a control run using parameters from the T42 version of this model (with the exception of the highly resolution-dependent gravity wave drag parameter, where a default T21 value was used). Interannual variability of the model makes an insignificant contribution to the model-data discrepancy. For the central ('medium') estimate for model error, we weighted each data type equally to contribute a total of 0.5 to the log-likelihood cost function (giving the control run a cost of 7.5, and equivalent to assuming that the uncertainty in RMS model error is *equal* to the RMS error of the control run), and two more experiments were performed with this uncertainty estimate scaled by a factor of 2 in each direction. We also evaluated the prior directly with another 40-member ensemble. The costs (evaluating the original data constraint alone) for the resulting ensemble members for the present day simulations are shown in Table 1.

The ensemble size of 40 was chosen for computational convenience. This is too small for accurate sampling of the tails of the distributions, but the consistent trends in the bulk statistics of costs and sensitivities appear reliable. The estimation was performed using an iterative 4 year integration and analysis cycle. After roughly 25 cycles all ensembles had converged to statistically steady solutions in parameter space, and a final 16 year nudging run was performed to generate the ocean heat flux field for the present day, doubled carbon dioxide and Last Glacial Maximum (LGM) simulations. The present day simulations were integrated for a further 16 years from the final analysis states to generate mean climatologies, and had a typical drift of around 0.1°C or less over this interval, thus justifying the computational choices made. The doubled carbon dioxide simulations were integrated for 40 years and exponential curves fitted to the time series, which had not all fully converged. The LGM simulations were performed under standard PMIP2 boundary conditions (http://www-lsce.cea.fr/pmip2/), also for 40 years, by which time the drift was negligible so a fit was not performed and results were instead directly averaged over the final 4 years. Although PMIP2 does not include slab ocean models, our experimental approach is essentially that of the original PMIP.

Table 1. Statistics for the four ensembles: Tight, Medium and Loose are the three tuned ensembles in order of increasing assumed model error; Prior has parameters chosen from the prior estimates. The cost is the sum of normalised RMS errors of the model outputs compared to the present day data. Uncertainties are all one standard deviation.

	Costs	Climate Sensitivity (°C)	LGM Cooling (°C)
Tight	7.6±0.4	6.00±0.89	3.23±0.47
Medium	9.9±1.3	5.65±1.10	2.74±0.52
Loose	11.4±1.9	5.44±1.05	2.72±0.48
Prior	17.2±8.3	5.00±1.37	2.52±0.59

3. Results

The climate sensitivity results from our experiments are shown in Table 1 and Fig. 1. Sampling directly from the prior finds climate sensitivities of 3.7–6.5°C at the 90% confidence level, with one outlier from the 40-member ensemble exhibiting a runaway greenhouse (the exponential fit to 40 years gives the displayed value of close to 12°C of warming, but this run was integrated further and exceeded 16°C of warming after 60 years). As we tune the models more aggressively towards the modern climate, the ensemble estimates for climate sensitivity clearly increase. For our ensemble of highest quality models, which have comparable skill to the control run, more than half of the members have sensitivity greater than 6°C. Although our ensembles cover a substantial range of sensitivities, the median values contrast strongly with the bulk of previous work, which suggests a most likely value for climate sensitivity of around or below 4°C (Andronova and Schlesinger 2001; Knutti et al. 2002; Murphy et al. 2004).

In previous research, the value of the models and methods in generating useful estimates of climate sensitivity has generally been assumed and not directly tested. Testing a model against independent data is widely acknowledged to be a more stringent test of predictive skill than merely fitting a model to existing observations (Lipton 2005). To this end we have performed LGM simulations to investigate how well our ensembles can estimate a climate very different to the modern state. There are no good estimates of globally averaged temperature change for the LGM compared to the present day, so we focus on the tropical oceans (30 °S to 30°N) where there is a concentration of proxy data.

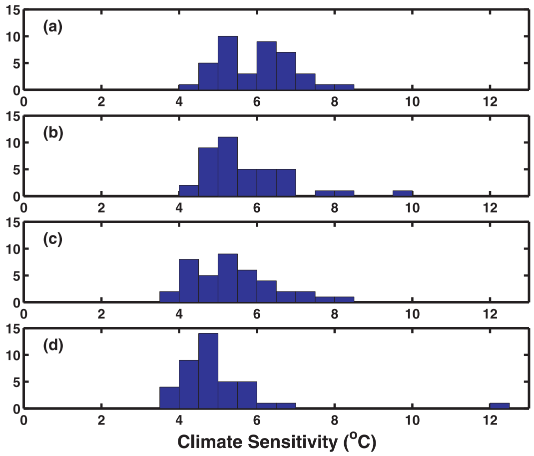

Fig. 1. Climate sensitivity distributions for four ensemble experiments: (a) tight fit to observations (b) medium fit (c) loose fit (d) sampled from prior parameter distributions with no tuning. See Section 2 for details.

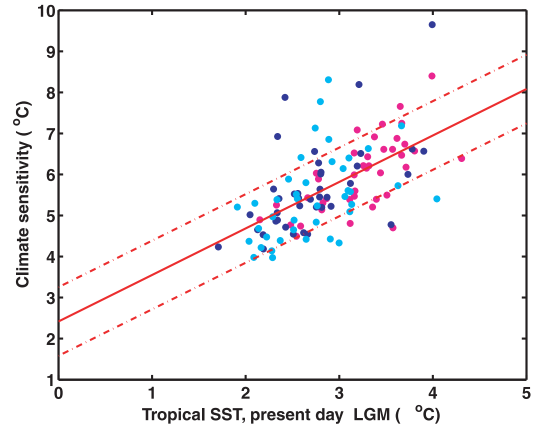

Fig. 2. Climate sensitivity versus Tropical SST temperature change at the LGM. Cyan, blue and magenta dots indicate loose, medium and tight ensembles respectively. Solid line indicates least-squares fit, and dot-dashed lines show RMS scatter of ensemble members about this line. The correlation is statistically significant at the 3 sigma level.

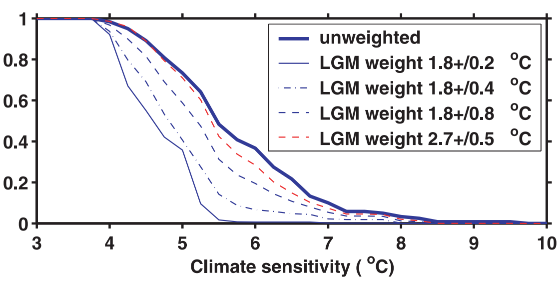

Fig. 3. Cumulative distribution functions for meta-ensemble with various constraints on LGM tropical SST.

Even here, uncertainty remains. The CLIMAP (1981) project produced a gridded SST data set which indicated an average cooling of 0.8°C in the tropics, a value which is now widely regarded as too small. The alkenone data set of Harrison (2000) has been widely used in recent years for comparison with models (Houghton et al. 2001), and so we focus primarily on this data set here. Averaging these data first by ocean basin, and then forming an area-weighted global average in the tropics, gives us an estimate of 1.8±0.2°C (at 1 sigma) cooler than the present day. Given that the large Pacific basin shows substantially less cooling (in both models and data) than the Atlantic Ocean, this figure is not inconsistent with the Atlantic average value of 3°C cooling from the GLAMAP project (Schäfer-Neth and Paul 2003). More recently, Ballantyne et al. (2005) presented an estimate of 2.7±0.5°C cooling, and the potential impact of this substantially colder value is considered below.

As with the present-day simulations, the presence of model error implies that we cannot expect our ensemble members to simulate the LGM data to within its observational errors, and applying such a "perfect model" constraint would result in an implausibly narrow posterior ensemble. The accuracy of our temperature estimate is limited by the representativity of the rather sparse data set used, and correlations between the observations which have not been accounted for. Given uncertainties in this estimate, the boundary conditions at the LGM, and the influence of model error (especially, but not solely, due to the use of modern heat fluxes), the sample uncertainty only provides a lower bound on the range of acceptable models. Therefore, we doubled and redoubled the error estimate to 0.4°C and 0.8°C and tried all three values in turn. Applying these constraints as a posterior weighting is equivalent to repeating the ensemble assimilation method with the extra data point included in the set of observations.

As shown in Fig. 2, our LGM simulations are generally biased towards the upper end of the range of this estimate of cooling and does not cover the full range of uncertainty. Since there remains a substantial likelihood that our ensembles have failed to simulate the glacial state adequately we cannot justify a probabilistic interpretation of our ensemble distributions for climate sensitivity. For our three tuned ensembles (but not the samples from the prior) there is a significant and very similar correlation between the cooling of the LGM tropical SST, and climate sensitivity, and this suggests that our climate sensitivity estimates are generally also too high. There is no guarantee that our modelled correlation is correct, due to uncertainties in both the forcing at the LGM and the model behaviour, and we would welcome results using a range of different models. However, a linear least-squares fit indicates that the most likely tropical SST cooling of 1.8°C corresponds to a climate sensitivity of around 4.5°C, with the scatter of the ensemble around that value adding 0.8°C of uncertainty at the 1 sigma level. This estimate is slightly higher than, but broadly consistent with, most previous research. Our conclusion is that this AGCM has a structurally high sensitivity and that greater changes than parameter estimation alone can provide would be required to produce a version of the model that is both realistic and has low climate sensitivity.

Since our ensembles have a considerable number of samples at high sensitivity we now investigate whether we can use the LGM data to constrain the upper end of climate sensitivity. For this we combine the three tuned experiments to form a single 120 member meta-ensemble, and apply the LGM data constraints as a posterior weighting factor. In Fig. 3 (blue lines) we see that the high sensitivity ensemble members are substantially down-weighted when the LGM constraint is used. Sensitivities of greater than 6°C are effectively eliminated unless we use the weakest of our three constraints (which does not even rule out at the 1% level a modelled LGM tropical ocean which is warmer than the present day). Even in this case, there is very low likelihood for sensitivity above 8°C. Although our biased ensemble cannot directly give probabilistic estimates, we can see that for the intermediate LGM constraint, only about 7% of our ensemble are above 6°C sensitivity, and virtually nothing above 8°C. Due to the strong bias towards high sensitivity in our ensembles, true climate sensitivity is likely to be lower than these values would imply. The impact of the colder temperature estimate of Ballantyne et al. (2005) is indicated by the dashed red line. Even in this case, our ensemble appears slightly too cold and the more sensitive models are generally downweighted, but the effect is less marked. Clearly, it is essential to establish robust estimates of the glacial state if approaches such as this are to be useful in constraining climate sensitivity.

4. Summary and discussions

Although it is difficult to place direct probabilistic estimates on climate sensitivity, due to both the presence of model error and the apparent bias of our results towards high sensitivities, our results contrast somewhat with the recent work of Stainforth et al. (2005). They found a significant proportion of their ensemble to have extremely high climate sensitivities of around 10°C. However, for their observational constraint, they only used annually-averaged rather than seasonally-differentiated data. Therefore, their constraint did not include any direct controls on the behaviour of the model in response to changes in radiative forcing. This may go some way towards explaining the differences between their results and those presented by Murphy et al. (2004) who, while using the same model, also used a constraint that includes seasonal variation, and found a moderate upper bound of 5.4°C at the 95% level.

Clearly, structural decisions (and to some extent also our prior distributions for the parameters) have pushed the MIROC3.2 model towards high sensitivities, even while it retains a reasonable spatial and seasonal pattern of behaviour. Our ensemble of climate sensitivity is substantially higher than that of Murphy et al. (2004). They have not yet checked their results with any out-of-sample data, but our LGM simulations suggest that our ensembles are biased towards excessively high sensitivity. Using the LGM as an additional constraint, we find a most likely value for climate sensitivity of around 4.5°C, although of course having used this LGM data, we are left with no further independent data to validate this estimate.

The substantial scatter in Fig. 2 implies that even if the LGM state was much better characterised, it would not be able to provide a very tight estimate of climate sensitivity, due to uncertainties in the model response to the different forcings. Our approach does, however, demonstrate a method for validation of model-based estimates which necessarily have a substantial subjective basis. Our model appears to have a strong structural bias towards high sensitivity, but by using the LGM we can discount the extreme end of our ensembles and therefore anticipate that these results may prove to be robust across different models. Any research which focusses on a single model will run a risk of overestimating the confidence of the results, by underestimating the extent of model error. We have taken some steps towards addressing this problem, by using independent data as validation. The main obstacle to improved estimation appears to be the influence of model error, and multi-model approaches may be very helpful in making more progress in addressing this issue.

Acknowledgments

We thank the K-1 Japan project members for support and discussion, and the reviewers for helpful comments. This work was partially supported by the Research Revolution 2002 of the Ministry of Education, Sports, Culture, Science and Technology. The model calculations were made on the Earth Simulator of JAMSTEC.

References

Annan, J., D. and J. C. Hargreaves, 2005: Climate prediction with imperfect models. *Quart. J. Roy. Meteor. Soc.*, (submitted).

Annan, J. D., D. J. Lunt, J. C. Hargreaves, and P. J. Valdes, 2005: Parameter estimation in an atmospheric GCM. *Nonlinear processes in geophysics*, **12**, 363–371.

Andronova, N. G., and M. E. Schlesinger, 2001: Objective estimation of the probability density function for climate sensitivity. *J. Geophys. Res.*, **108(D8)**, 22,605–22,611.

Ballantyne, A. P., M. Lavine, T. J. Crowley, J. Liu and P. B. Baker, 2005: Meta-analysis of tropical surface temperatures during the Last Glacial Maximum. *Geophys. Res. Lett.*, **32**, L05712.

CLIMAP Project Members, 1981: Seasonal reconstructions of the Earth's surface at the Last Glacial Maximum. *Geological Soc. of America*, Map and Chart Series MC-36.

Evensen, G. 2003: The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*, **53**, 343–367.

Gregory, J. M., R. J. Stouffer, S. C. B. Raper, P. A. Stott, and N. A. Rayner, 2002: An observationally based estimate of the climate sensitivity. *J. Climate*, **15(22)**, 3117–3121.

Harrison, E. F., P. Minnis, B. R. Barkstrom, V. Ramanathan, R. D. Cess, and G. G. Gibson, 1990: Seasonal variation of cloud radiative forcing derived from the Earth Radiation Budget Experiment. *J. Geophys. Res.*, **95(18)**, 18687–18703.

Harrison, S. P. 2000: Palaeoenvironmental data sets and model evaluation in PMIP. In P. Braconnot, editor, *Proceedings of the third PMIP workshop. WCRP-111; WMO/TD-No. 1007*, pages 9–25, La Huardire, Canada.

Hasumi, H. and S. Emori (eds), 2004: K-1 coupled model (MIROC) description, K-1 technical report 1. Technical report, CCSR, The University of Tokyo.

Houghton, J. T., Y. Ding, D. J. Griggs, M. Noguer, P. J. Van Der Linden, X. Dai, K. Maskell, and C. A. Johnson, 2001: *Climate Change 2001: Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change.* CUP.

Kennedy, M. C. and A. O'Hagan, 2001: Bayesian calibration of computer models. *J. Roy. Stat. Soc. Ser. B*, **3**, 425–450.

Knutti, R., T. F. Stocker, F. Joos, and G.-K. Plattner, 2002: Constraints on radiative forcing and future climate change from observations and climate model ensembles. *Nature*, **416**, 719–723.

Lea, D. W., 2004: The 100,000 year cycle in tropical SST, greenhouse forcing, and climate sensitivity. *J. Climate*, **17**, 2170–2179.

Lipton, P. 2005: Testing hypotheses: prediction and prejudice. *Science*, **307**, 219–221.

Murphy, J. M., D. M.~H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth, 2004: Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, **430**, 768–772.

Rougier, J. C. 2005: Probabilistic inference for future climate using an ensemble of simulator evaluations. *Climatic Change*, (submitted).

Schäfer-Neth, C. and A. Paul, 2003: The Atlantic Ocean at the last glacial maximum: 1. Objective mapping of the GLAMAP sea-surface conditions. *Wefer G., Mulitza S., and Ratmeyer V. (eds) The South Atlantic in the Late Quaternary: Material Budget and Current Systems*, Springer, Berlin, 531–548.

Simmons, A. J., and J. K. Gibson, 2000: The ERA-40 Project Plan, ERA-40 Project Report Series. TR 1, ECMWF.

Stainforth, D. A., T. Aina, C. Christensen, M. Collins, N. Faull, D. J. Frame, J. A. Kettleborough, S. Knight, A. Martin, J. M. Murphy, C. Piani, D. Sexton, L. A. Smith, R. A. Spicer, A. J. Thorpe, and M. R. Allen, 2005: Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, **433**, 403–406.

Xie, P., and P. A. Arkin, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.*, **78**, 2539–2558.

Manuscript received 29 July 2005, accepted 26 September 2005
SOLA: http://www.jstage.jst.go.jp/browse/sola/