Global and regional variability in marine surface temperatures

T. Laepple,[1] P. Huybers[2]

1. Key points

1. Methods are introduced to compare instrumental and model SST variability
2. Regional SST variability is underestimated by the CMIP5 models at decadal timescales
3. Lack of intrinsic variability may explain the difficulty in simulating recent global trends

Corresponding author: T. Laepple, AWI (thomas.laepple@awi.de)

[1]Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Potsdam, Germany

[2]Earth and Planetary Sciences, Harvard University, Cambridge, MA, USA.

The temperature variability simulated by climate models is generally consistent with that observed in instrumental records at the scale of global averages, but further insight can also be obtained from regional analysis of the marine temperature record. A protocol is developed for comparing model simulations to observations that accounts for observational noise and missing data. General consistency between CMIP5 model simulations and regional sea surface temperature variability is demonstrated at interannual timescales. At interdecadal timescales, however, the variability diagnosed from observations is significantly greater. Discrepancies are greatest at low-latitudes, with none of the 41 models showing equal or greater interdecadal variability. The pattern of suppressed variability at longer timescales and smaller spatial scales appears consistent with models generally being too diffusive. Suppressed variability of low-latitude marine temperatures points to underestimation of intrinsic variability and may help explain why few models reproduce the observed temperature trends during the last fifteen years.

1. Introduction

Accurate representation of the spread in predictions of future climate is, arguably, as important as correctly predicting a central value. Comparison against observed variability is one means of evaluating the skill of general circulation models (GCMs) in simulating the spread of plausible temperatures. At the global scale, the observed temperature variability is generally consistent with that produced by GCMs both in terms of overall magnitude and spectral distribution [*Solomon et al.*, 2007; *Jones et al.*, 2013]. Although regional model-data consistency has also generally been found at synoptic to interannual timescales [*Collins et al.*, 2001; *Min et al.*, 2005], discrepancies have been noted in regional variability at longer timescales. *Stott and Tett* [1998] found that simulations from a climate model underestimate surface temperature variability at scales less than 2000 km. *Davey et al.* [2002] and *DelSole* [2006] also suggested that collections of models underestimate regional low-frequency variability at decadal timescales relative to observations, and *Santer et al.* [2006] found a similar mismatch for Eastern Tropical Atlantic SST.

There are two classes of explanation for model-data discrepancies in regional SST variability. The first is for model simulations to inadequately simulate variability. The second class of explanation is for observational errors, data inhomogeneities, or interpolation artefacts to bias instrumental estimates of variability. These data issues were not systematically treated in foregoing studies, raising the question of whether discrepancies arise from model or data short-comings.

To address these possibilities we extend upon foregoing model-data comparison studies in three respects. First, analysis of the CMIP5 archive [*Taylor et al.*, 2012] offers a more recent set of 163 historical simulations to compare against observations. Second, recently developed corrections for data inhomogeneities along with more complete estimates of uncertainty [*Kennedy et al.*, 2011a, b] permit for more accurate assessment of observational variability. Finally, we introduce and apply a new technique to correct for the effects of data gaps upon variance and spectral estimates. Such accounting for variance contributions to the observed SST variability permits for less biased model-data comparison.

2. Simulations and data

For simulations we rely on the CMIP5 collection of coupled atmosphere-ocean model runs. Analysis is of the SST fields of historical simulations covering 1861-2005 (CMIP5) that are forced by reconstructed natural and anthropogenic radiative forcing from solar variations, greenhouse gas concentrations, and volcanic and anthropogenic aerosols. In all, there are 163 simulations from 41 models. Simulations are placed onto the $5 \times 5°$ grid of the HadSST3 dataset by first interpolating to a uniform $0.25 \times 0.25°$ grid and then averaging to $5 \times 5°$ boxes. This high-resolution interpolation followed by averaging avoids spatial aliasing that would otherwise lead to biases in estimated variability. SST anomalies are then computed by removing the monthly climatology calculated between 1960-1990.

Instrumental observations are from the HADSST3 compilation of sea surface temperatures (SST) [*Kennedy et al.*, 2011a, b]. This dataset consists of binned SST observations from ships and buoys on a 5° by 5° grid, where averaging is conducted after excluding outliers. The time series are bias corrected for spurious trends caused by changes in measurement techniques but are not interpolated or variance adjusted, as is appropriate for our purposes. Uncertainty estimates associated with observational noise, binning, and bias correction are all provided [*Kennedy et al.*, 2011a, b].

SST records are primarily from ship measurements that, outside of certain heavily trafficked routes, tend to contain observational gaps. Annual mean SST estimates are only computed when at least ten observations are present within the year. Analysed time-series are the longest possible at each grid box for which no more than 10% of years are missing and for which data is present during the first and last years. Missing years are linearly interpolated for. The last year is always fixed at 2005 in order to overlap with the time span covered by the historical CMIP5 simulations. Further, as our focus is on multidecadal variations in SSTs, time-series must cover at least 100 years after interpolation in order to be included.

To provide for an equivalent basis for model-data comparison, missing months in the observations are censored in the simulation results. Interpolation will typically alter spectral estimates [*Wilson et al.*, 2003; *Rhines and Huybers*, 2011], but because equivalent months and years are missing from both the simulations and observations, comparisons between the two are not biased, excepting for certain issues involving correcting for noise components in the observational dataset that are addressed shortly.

3. Spectral estimation and noise correction

Timescale dependent variance is estimated in both the instrumental observations and model simulations by summing spectral energy estimates between frequencies of 1/2-1/5 years^{-1} for interannual variations and 1/20-1/50 years^{-1} for interdecadal variations. For the variance estimate, we sum across the relevant frequencies of a periodogram [e.g. *Bloom-*

field, 1976], whereas the multitaper method with three windows [*Percival and Walden*, 1993] is used for visually presenting results. The periodogram is used for timescale dependent variance estimates because the multitaper methods is slightly biased at the lowest frequencies [*McCoy et al.*, 1998]. All spectral analyses are performed after linearly detrending the SST time series.

Instrumental SST records contain substantial noise, with the average monthly observation having a one-standard-deviation uncertainty of 0.48°C [*Kennedy et al.*, 2011a]. Noise estimates are available for each month and grid box and are calculated taking into account random measurement errors, errors stemming from incomplete spatial coverage of the 5° by 5° grid-box, and incomplete temporal coverage of the observed month. For regional variance estimates, we treat these sources of noise as independent between months because measurements from ships are unlikely to correlate in a single location over different months, and measurements from buoys have relatively small uncertainties (pers. comm. Kennedy 2012). For the global mean SST estimate, we use measurement and sampling error estimates that account for spatial and temporal correlations [*Kennedy et al.*, 2011a].

Independent realization of normally distributed noise is expected to have a uniform spectral distribution in the case of uniform sampling, but the presence of gaps in regional observational records leads to a variable noise influence with frequency. Essentially, interpolation between noisy values introduces autocorrelated noise. To correct for these noise contributions, we generate annually resolved time-series from draws of a normal distribution having time-variable standard deviation consistent with the reported error. Years with missing observations are linearly interpolated for, and the spectral estimate of the realized noise sequence is computed. This process is repeated 10,000 times, and the average across noise spectra is calculated and removed from the corresponding instrumental SST spectral estimate. This technique shares some similarities with that introduced by *Laepple and Huybers* [2013] for correcting the spectral estimates associated with paleoclimate records, and it is applied to the time-series associated with each grid-box included in the analysis. The correction for excess variance has the largest proportional effects at interannual timescales, rather than decadal ones, because spectral magnitudes are smaller at higher frequencies. The correction at the global level is more simple, having a uniform distribution across frequency, because there are no data gaps.

Prior to correction, the variance ratio between the observed and simulated temperatures has a cross-correlation with the average number of observations per year across grid boxes of r=-0.38. This negative correlation is significant at the 95% confidence level, assuming at least 28 degrees of freedom, and is expected on the basis of fewer observations leading to greater noise in the annual temperature estimates. After correction, the magnitude of the correlation is reduced to a value that is statistically indistinguishable from zero, r=0.03, indicating that the correction is successful in removing excess noise. Also important is that, after correction, the variance ratio shows no dependence on what time interval is analyzed nor upon what data coverage criteria are applied for admitting annual temperature estimates (Table 1). Note that variance adjusted products were provided in earlier versions of the HadSST dataset, but are not used here because variance adjustment is accomplished through exclusively rescaling the amplitude of high-frequency variability in order to homogenize variance given differences in expected signal-to-noise ratios [*Brohan* *et al.*, 2006]. We have no expectation for noise to be band-limited and apply a correction across the entirety of spectrum, which partially reduces model-data differences at low frequencies.

Uncertainties reported in Table 1 include those usually associated with finite data as well as the uncertainties associated with removal of the noise component. In addition, there also exist uncertainties in the instrumental SST dataset stemming from corrections applied for systematic changes in measurement techniques [*Kennedy et al.*, 2011b]. To account for these systematic uncertainties, we analyse the 100 available realizations of the HadSST3 field that seek to cover the range of instrumental biases, and include the resulting spread in the estimated temperature spectra in our final uncertainty estimate. Uncertainties associated with the mean of the regional spectral estimates are computed assuming ten spatial degrees of freedom [*Jones et al.*, 1997], except for those associated with measurement changes, which are treated as systematic across records.

Available ensemble members associated with each model range from 1 to 23. In order to achieve uniform model weighting when calculating multimodel means, spectral analysis results associated with each ensemble member are inversely weighted according to the total number of ensemble members. This gives equal weighting across models, which is appropriate because ensemble members are generally tightly clustered relative to inter-model spread. Note that the spread of the ensemble provides a description of the CMIP5 collection but is only a lower bound on total model uncertainty [*Knutti et al.*, 2010]. The results that we present from our analysis are robust to using either nearest neighbor or linear interpolation techniques, various filters to isolate variance at a particular timescale, and for the allowance of 2%, 10%, or 20% of missing data in choosing what records to include.

4. Model-data comparison

Spectral estimates associated with regional SST variability are much greater in magnitude than those associated with global average SST variability (Fig. 1). The difference in variability is about two orders of magnitudes at interannual timescales and decreases to less than an order of magnitude on multidecadal timescales. The global-regional differences reflect cancellation of variability in the global mean, and the weaker cancellation toward lower frequencies is consistent with findings that temperature anomalies have greater spatial autocorrelation toward longer timescales [*Jones et al.*, 1997].

For the global average, instrumental and model spectral estimates are generally consistent to within uncertainties across frequencies, as also reported elsewhere [*Solomon et al.*, 2007; *Crowley*, 2000; *Jones et al.*, 2013], excepting near the frequencies associated with the El Niño Southern Oscillation between 1/2-1/7years, which is more strongly expressed in the observations than in most simulations. The mean of the regional spectra agree at once per decade and higher frequencies, but at lower frequencies the observations show significantly greater spectral energy. Agreement for global-average spectral estimates but disagreement at the regional level demonstrates that model temperature variability has, on average, greater positive spatial covariance than the observations at decadal timescales.

More insight into the mismatch between models and data can be gained from considering the ratio of spectral energies as a function of space (Fig. 2). At interannual timescales, between 1/2-1/5 year^{-1}, the data-model ratio of spectral energy is near one when taking the zonal mean at most latitudes. Regionally, it is around half in the Northern North Atlantic, Northwestern Pacific, and Northern Indian Ocean, and 1.5 in the remainder of the Atlantic and Eastern Pacific (Table 1).

The data-model ratio at decadal timescales, between 1/20-1/50 years^{-1}, is larger than at interannual timescales (Fig. 2 and Fig. 3). At middle and higher latitudes ($\geq$30°) the average data-model ratio is 1.3, with portions of the North Atlantic and Northwestern Pacific showing values less than one in a pattern similar to that seen at interannual timescales. At lower latitudes ($\leq$30°) the data-model ratio is 1.9, with only 4 out of 163 ensemble members showing greater variability than the observations: 2 of 10 ensemble members from GFDL-CM2 and 2 of 10 members from HadCM3. It is also worth emphasizing that the correction for instrumental noise sources reduces the data-model ratio by as much as 100% at interannual timescales but by less than 30% at decadal timescales (Table 1). Temperature variations are of larger amplitude toward lower frequencies and are associated with a greater signal-to-noise ratio and are, therefore, less sensitive to noise correction. The noise correction would have to be more than a factor of three too small at decadal timescales, while being unchanged at interannual timescales, for the data and simulations to be consistent.

Our results thus confirm and update foregoing indications that regional model variability is weak relative to the observations at low latitudes and at decadal timescales [*Stott and Tett*, 1998; *Davey et al.*, 2002; *DelSole*, 2006]. It is also relevant to address the fact that other studies found general consistency when comparing the variability in average Eastern Tropical Pacific SSTs against the CMIP3 [*Santer et al.*, 2006] and CMIP5 [*Fyfe* *and Gillett*, 2014] model ensembles. These results can be understood in that averaging over the Eastern Equatorial Pacific reduces the apparent model-data inconsistency in the multidecadal band from a ratio of 2 to 1.6. This result follows from greater suppression of variability in the observations than in the models, consistent with our hypothesis of the models being too diffusive. Furthermore, analysis of average temperature produces a spread in variance ratios that is 24% larger than when the average is taken across the ratios computed for each grid box. Thus, analysis of average temperature reduces both discrepancies and detectability of discrepancies.

5. Discussion and conclusion

These results raise the question of why model simulations do not generate greater low-frequency SST variability at regional scales. It could be that models are too weakly forced at multidecadal time-scales or contain insufficient positive feedback to amplify such forcing, but such a scenario seems unlikely to be a complete explanation because externally forced variability only accounts for a small fraction of regional model variance [*Goosse et al.*, 2005]. Comparing unforced simulations to an ensemble of forced simulations of the ECHAM5/MPIOM AOGCM, [*Jungclaus et al.*, 2010] show that externally forced variability accounts for only 20% of the multidecadal tropical variability at $5 \times 5°$ scales and even smaller fractions when including the extratropics. Assuming linearity, it can be inferred that doubling regional variability at $5 \times 5°$ scales would require at least a five-fold increase in the externally forced contribution. Furthermore, interannual consistency at the regional level and across all timescales at the global level suggests that a marked increase in external variability would lead to other model-data mismatches.

More consistent with our findings is for the models to underestimate internal variability. This structure of the model-data mismatch suggests that model effective horizontal diffusivity may be too large, as this would lead to suppression of regional variability at low-frequencies. Diffusivity would become important for the grid scale size that we analyze at approximately 8 years, where the square of the 500 km domain is divided by an effective horizontal diffusivity of 1000 m^2/s. This timescale is consistent with the appearance of divergence between regional data and model spectra beginning in the vicinity of 1/8 $years^{-1}$ and increasing toward lower frequencies (Fig. 1). Also of note is that *Stammer* [2005] showed that an initial specification of a uniform 1000 m^2/s horizontal diffusivity in the MIT-GCM was generally revised downward through a formal data-fitting procedure.

Further insight can be gained by separating the multimodel ensemble according to resolution. Models are grouped into quartiles according to horizontal ocean resolution at the equator, and results are consistent with the diffusion hypothesis in the sense that lower resolution quartiles show less variability and a larger discrepancy with the observations. Specifically, the low resolution quartile of models has an average ratio of observed versus model variability of 2.8 in the tropics and 2.2 globally, whereas the quartile of highest-resolution models has analogous ratios of 1.7 and 1.4. Resolution is at best only a partial determinant of variability, however, as indicated by a 0.2 cross-correlation between resolution and multidecadal variability across models.

Recent trends in global average temperature largely fall below those simulated by general circulation models [*Fyfe et al.*, 2013], and observed trends in Eastern Equatorial Pacific SSTs are even more anomalously low relative to the models [*Fyfe and Gillett*, 2014]. These trends in EEP and global temperature appear related [*Rahmstorf et al.*, 2012; *Kosaka and Xie*, 2013; *Fyfe et al.*, 2013; *Fyfe and Gillett*, 2014]. We speculate that some of the model-data trend difference comes from simulations having too small internal variability. Greater internal variability in the models would widen the spread in the ensemble of temperature trends and increase the likelihood of including the observed trends, especially if the greater variability is in regions having strong global teleconnections, such as in the EEP. Note that our results are largely independent of the interval in question because all records span at least 100 years and end by 2005.

Although our results agree with earlier studies and are stable with respect to the time interval considered and various correction choices, there is some complication inherent to inferring variability during an interval containing substantial trends in global temperature. Spectral estimation and filtering assume quasi-stationarity over the interval of the record that cannot be entirely ensured through detrending. Distinguishing natural variability from forced variations that project onto natural modes of variability is also difficult. The use of paleodata to extend model-data comparisons and to include intervals prior to this last century seems a logical next step. Insomuch as the hypothesis that excessive horizontal diffusion damps regional model variability holds, we expect even greater data-model discrepancies in variability toward lower frequencies.

Acknowledgments. The Program for Climate Model Diagnosis and Intercomparison and the World Climate Research Programme Working Group on Coupled Modeling made the WCRP CMIP5 simulations available. R. Ferrari, B. Fox-Kemper, and M. Miller provided helpful suggestions with regard to model diffusivity and J. Kennedy with regard to the SST data. We thank the twon anonymous reviewers for their constructive comments. TL was supported by the Initiative and Networking Fund of the Helmholtz Association and the Daimler and Benz foundation. PH acknowledges NSF grant 1304309.

References

Bloomfield, P. (1976), *Fourier Decomposition of Time Series: An Introduction*, John Wiley, New York.

Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones (2006), Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, *Journal of Geophysical Research: Atmospheres*, *111*(D12), doi:10.1029/2005JD006548.

Collins, M., S. F. B. Tett, and C. Cooper (2001), The internal climate variability of HadCM3, a version of the Hadley Centre coupled model without flux adjustments, *Climate Dynamics*, *17*(1), 61–81.

Crowley, T. J. (2000), Causes of Climate Change Over the Past 1000 Years, *Science*, *289*(5477), 270–277, doi:10.1126/science.289.5477.270.

Davey, M.,et al., (2002), STOIC: a study of coupled model climatology and variability in tropical ocean regions, *Climate Dynamics*, *18*(5), 403–420, doi:10.1007/s00382-001-0188-6.

DelSole, T. (2006), Low-frequency variations of surface temperature in observations and simulations, *Journal of Climate*, *19*(18), 4487–4507.

Fyfe, J. C., and N. P. Gillett (2014), Recent observed and simulated warming, *Nature Climate Change*, *4*(3), 150–151, doi:10.1038/nclimate2111.

Fyfe, J. C., N. P. Gillett, and F. W. Zwiers (2013), Overestimated global warming over the past 20 years, *Nature Climate Change*, *3*(9), 767–769.

Goosse, H., H. Renssen, A. Timmermann, and R. S. Bradley (2005), Internal and forced climate variability during the last millennium: a model-data comparison using ensemble simulations, *Quaternary Science Reviews*, *24*(12-13), 1345–1360.

Jones, G. S., P. A. Stott, and N. Christidis (2013), Attribution of observed historical near-surface temperature variations to anthropogenic and natural causes using CMIP5 simulations, *Journal of Geophysical Research*, *118*(10), 4001–4024, doi:10.1002/jgrd.50239.

Jones, P. D., T. J. Osborn, and K. R. Briffa (1997), Estimating sampling errors in large-scale temperature averages, *Journal of Climate*, *10*(10), 2548–2568.

Jungclaus, J. H., S. J. Lorenz, C. Timmreck, C. H. Reick, V. Brovkin, K. Six, J. Segschneider, M. A. Giorgetta, T. J. Crowley, J. Pongratz, N. A. Krivova, L. E. Vieira, S. K. Solanki, D. Klocke, M. Botzet, M. Esch, V. Gayler, H. Haak, T. J. Raddatz, E. Roeckner, R. Schnur, H. Widmann, M. Claussen, B. Stevens, and J. Marotzke (2010), Climate and carbon-cycle variability over the last millennium, *Clim. Past*, *6*(5), 723–737.

Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby (2011a), Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. measurement and sampling uncertainties, *Journal of Geophysical Research*, *116*, doi:201110.1029/2010JD015218.

Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby (2011b), Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. biases and homogenization, *Journal of Geophysical Research*, *116*(D14), doi:10.1029/2010JD015220.

Knutti, R., R. Furrer, C. Tebaldi, J. Cermak, and G. A. Meehl (2010), Challenges in combining projections from multiple climate models, *Journal of Climate*, *23*(10), 2739–2758.

Kosaka, Y., and S.-P. Xie (2013), Recent global-warming hiatus tied to equatorial pacific surface cooling, *Nature*, *advance online publication*, doi:10.1038/nature12534.

Laepple, T., and P. Huybers (2013), Reconciling discrepancies between uk37 and Mg/Ca reconstructions of holocene marine temperature variability, *Earth and Planetary Science Letters*, *375*, 418–429, doi:10.1016/j.epsl.2013.06.006.

Lean, J. L., and D. H. Rind (2008), How natural and anthropogenic influences alter global and regional surface temperatures: 1889 to 2006, *Geophysical Research Letters*, *35*(18), doi:10.1029/2008GL034864.

McCoy, E., A. Walden, and D. Percival (1998), Multitaper spectral estimation of power law processes, *Signal Processing, IEEE Transactions on*, *46*(3), 655–668.

Min, S. K., S. Legutke, A. Hense, and W. T. Kwon (2005), Internal variability in a 1000-yr control simulation with the coupled climate model ECHO-G I. Near-surface temperature, precipitation and mean sea level pressure, *Tellus Series A*, *57*, 605.

Percival, D. B., and A. T. Walden (1993), *Spectral analysis for physical applications: multitaper and conventional univariate techniques*, Cambridge Univ Press.

Rahmstorf, S., G. Foster, and A. Cazenave (2012), Comparing climate projections to observations up to 2011, *Environmental Research Letters*, *7*(4), 044,035, doi:10.1088/1748-9326/7/4/044035.

Rhines, A., and P. Huybers (2011), Estimation of spectral power laws in time uncertain series of data with application to the Greenland Ice Sheet Project 2 d18O record, *Journal of Geophysical Research*, *116*(D1), D01,103.

Santer, B. D., T. M. L. Wigley, P. J. Gleckler, C. Bonfils, M. F. Wehner, K. AchutaRao, T. P. Barnett, J. S. Boyle, W. Brüggemann, M. Fiorino, N. Gillett, J. E. Hansen, P. D. Jones, S. A. Klein, G. A. Meehl, S. C. B. Raper, R. W. Reynolds, K. E. Taylor, and W. M. Washington (2006), Forced and unforced ocean temperature changes in Atlantic and Pacific tropical cyclogenesis regions, *Proceedings of the National Academy of Sciences*, *103*(38), 13,905–13,910, doi:10.1073/pnas.0602861103.

Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller (Eds.) (2007), *Climate Change 2007 - The Physical Science Basis: Working Group I Contribution to the Fourth Assessment Report of the IPCC*, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Stammer, D. (2005), Adjusting internal model errors through ocean state estimation, *Journal of Physical Oceanography*, *35*(6), 1143–1153.

Stott, P. A., and S. F. B. Tett (1998), Scale-Dependent Detection of Climate Change, *Journal of Climate*, *11*(12), 3282–3294.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2012), An overview of CMIP5 and the experiment design, *Bulletin of the American Meteorological Society*, *93*(4), 485–498, doi:10.1175/BAMS-D-11-00094.1.

Wilson, P. S., A. C. Tomsett, and R. Toumi (2003), Long-memory analysis of time series with missing values, *Physical Review E*, *68*(1), 017,103, doi:

10.1103/PhysRevE.68.017103.

Figure 1. Regional vs. global SST variability. At top is the average of local spectral estimates from instrumental observations and model simulations, and at bottom are the spectra estimated of global mean SST. Also shown are the 66% and 90% quantiles of the models (light and dark grey) and the 90% quantiles of the different realizations of the bias-corrected instrumental SSTs (light blue). Correction for the excess variance in SST observations caused by sampling and measurement error (dashed blue line vs. blue line) has the strongest relative effect at interannual timescales.

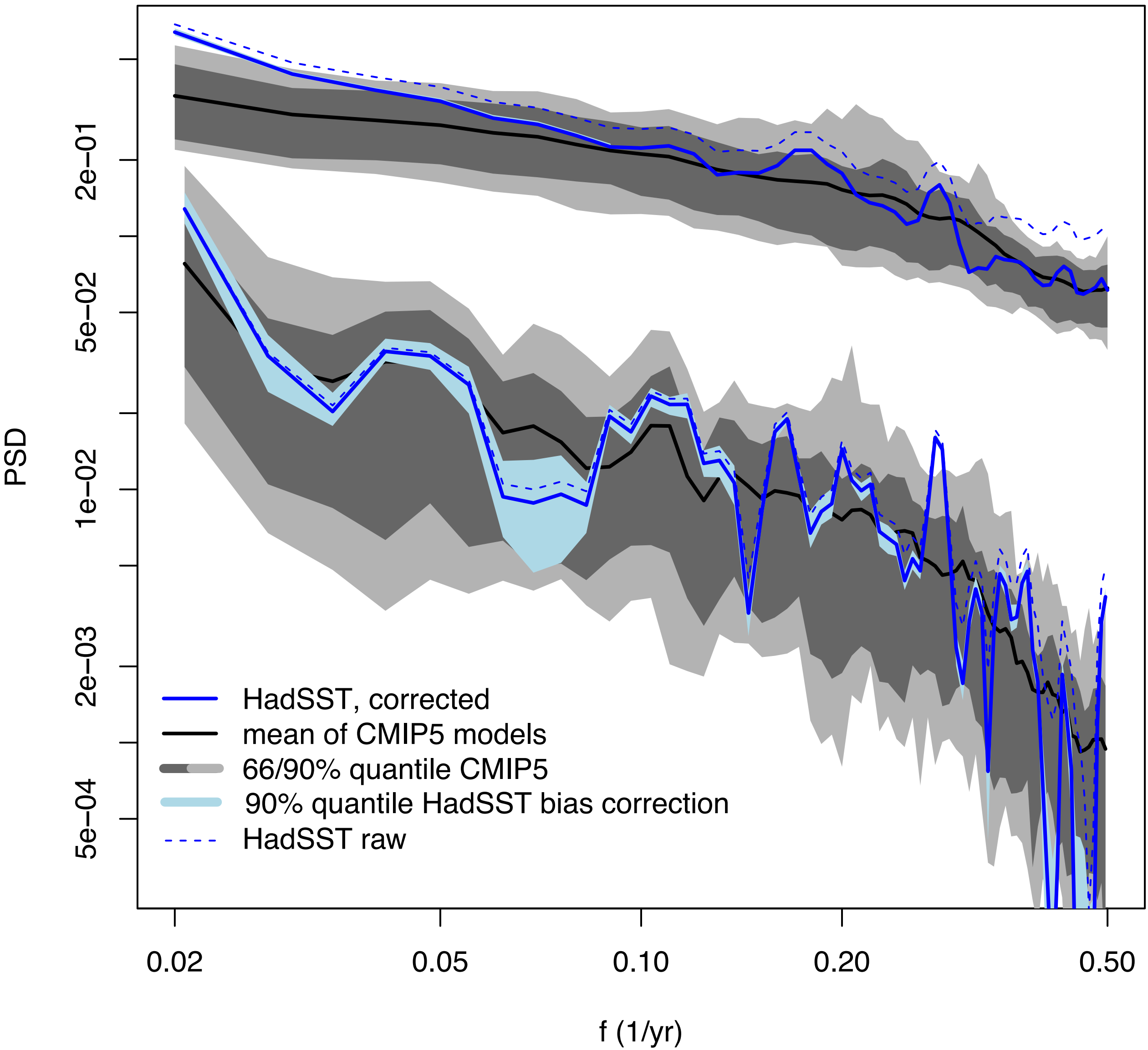
Figure 2. Variance ratio between the observed and simulated SSTs for interannual (2-5yr, a.) and multidecadal (20-50yr, b.) timescales. Simulated variance is the mean variance of all CMIP5 simulations. Observed variance is corrected for sampling and instrumental errors (see methods). Also shown is the zonal mean variance ratio between observed and simulated SSTs.

Figure 3. Distribution of the ratio between average instrumental and model SST variance for individual simulations. Shown are 2-5yr timescales (blue) and 20-50yr timescales (black) at middle to high latitudes (>30N and >30S) and low-latitude region (>30S <30N).

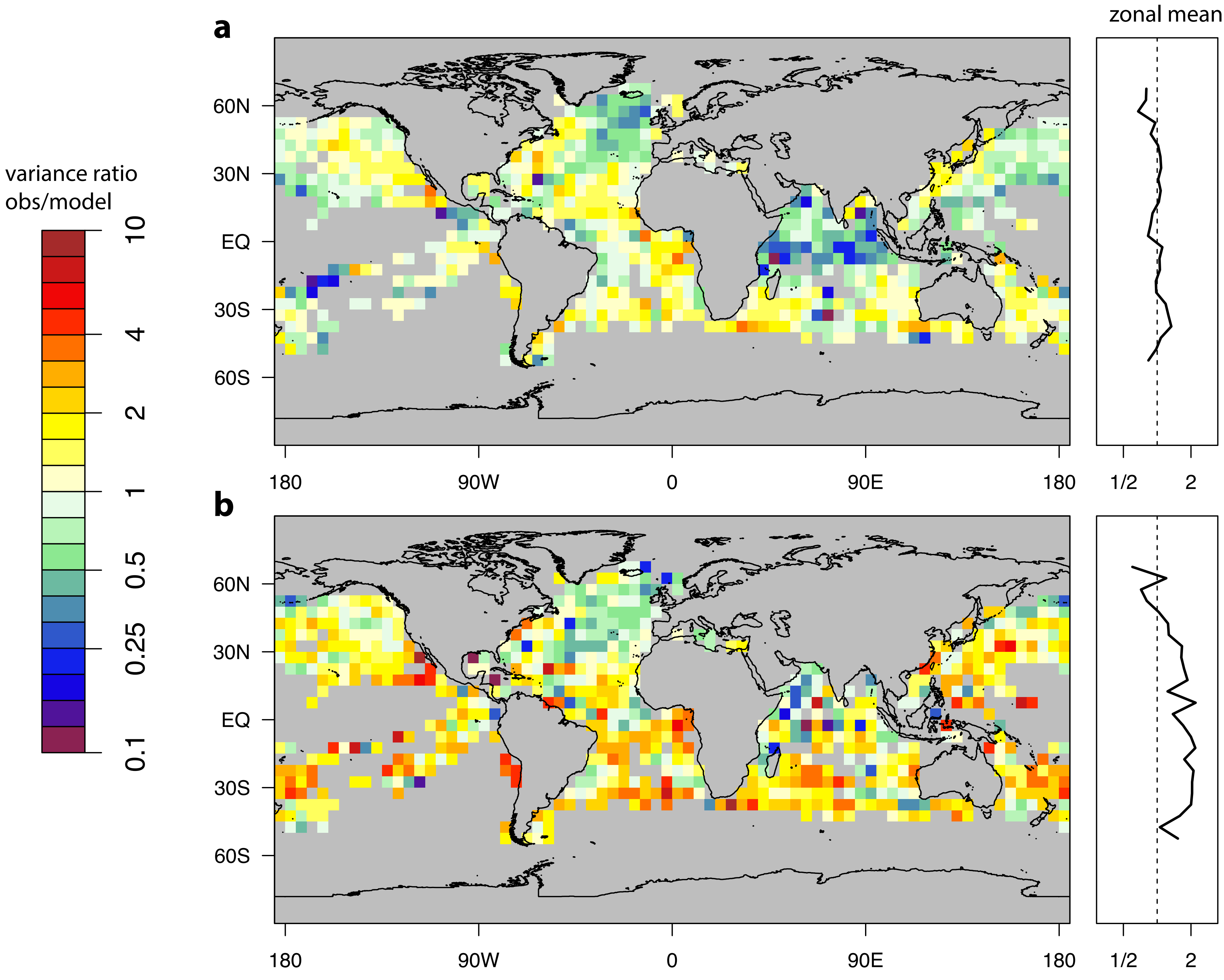
Table 1. Variance ratios of instrumental and simulated SSTs and their dependence on correction choices and data restriction criteria.

	time period	data restriction	mid-high latitudes >30S >30N		tropics and sub-tropics 30S-30N	
			2-5yr	20-50yr	2-5yr	20-50yr
uncorrected	1861-2005	≥1 obs/year	2.04 (1.85-2.23)	1.8 (1.33-2.34)	2.11 (1.92-2.31)	2.86 (2.11-3.72)
	1861-2005	≥10 obs/year	1.44 (1.3-1.57)	1.43 (1.06-1.87)	1.63 (1.48-1.78)	2.24 (1.65-2.92)
	1900-2005	≥10 obs/year	1.25 (1.12-1.39)	1.37 (0.97-1.83)	1.48 (1.32-1.65)	2.12 (1.51-2.84)
	1900-1960	≥10 obs/year	1.39 (1.18-1.61)	1.31 (0.87-1.84)	1.6 (1.36-1.85)	2.64 (1.76-3.7)
	1961-2005	≥10 obs/year	1.43 (1.21-1.68)	1.33 (0.81-1.98)	1.47 (1.24-1.73)	1.82 (1.11-2.7)
corrected	1861-2005	≥1 obs/year	1.19 (1.08-1.3)	1.55 (1.14-2.02)	1.02 (0.93-1.12)	2.19 (1.62-2.86)
	1861-2005	≥10 obs/year	1.04 (0.94-1.14)	1.32 (0.98-1.72)	1.06 (0.97-1.16)	1.92 (1.42-2.51)
	1900-2005	≥10 obs/year	0.99 (0.89-1.1)	1.3 (0.93-1.74)	1.09 (0.97-1.21)	1.93 (1.37-2.58)
	1900-1960	≥10 obs/year	1.07 (0.91-1.24)	1.23 (0.82-1.72)	1.01 (0.86-1.17)	2.28 (1.52-3.2)
	1961-2005	≥10 obs/year	0.98 (0.82-1.15)	1.19 (0.72-1.76)	1.08 (0.91-1.27)	1.51 (0.92-2.24)

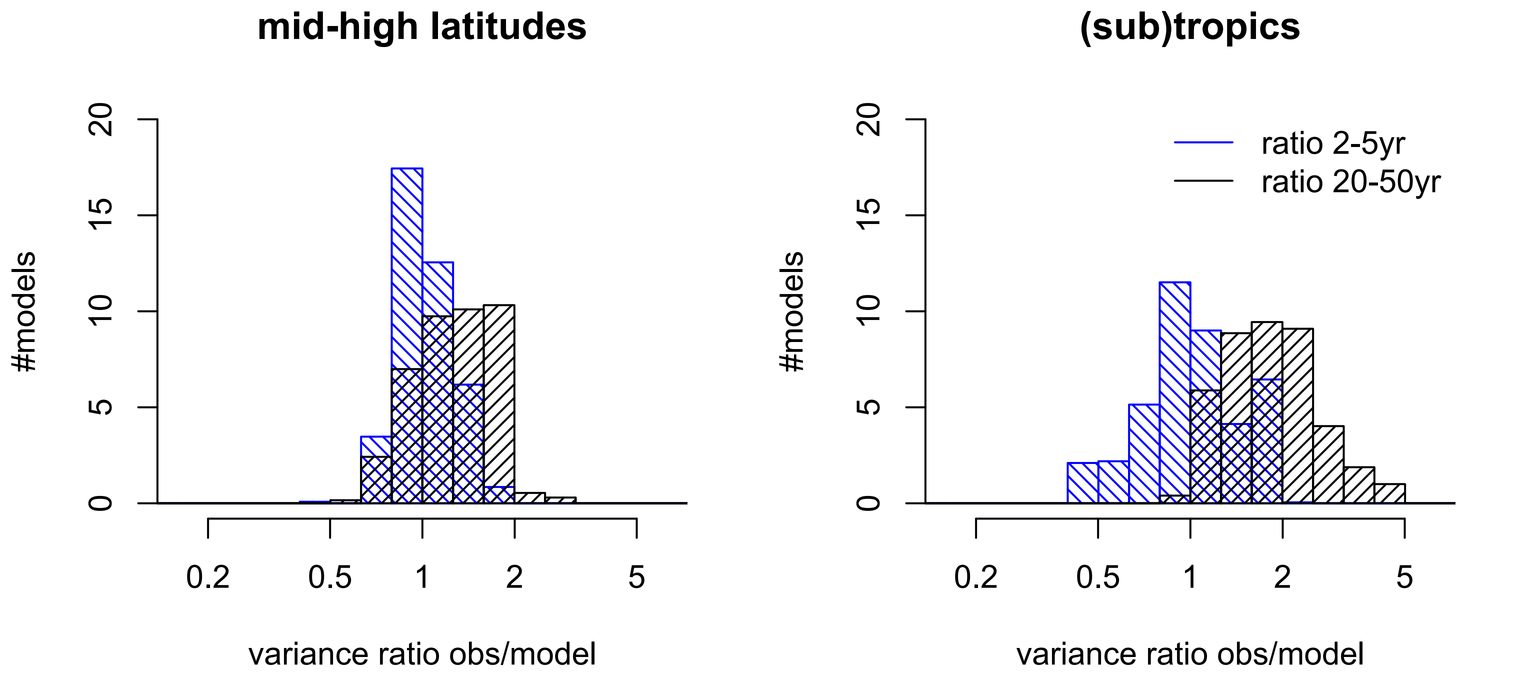
Note that variance ratios are independent of the data restriction criteria after correction for noise sources, whereas the inclusion of sparsely sampled grid-boxes otherwise leads to greater variance. 95% confidence intervals are calculated assuming ten spatial degrees of freedom and one degree of freedom per model simulation.



PSD
2e−01
5e−02
1e−02
2e−03
5e−04
0.02
0.05
0.10
0.20
0.50
f (1/yr)
HadSST, corrected
mean of CMIP5 models
66/90% quantile CMIP5
90% quantile HadSST bias correction
HadSST raw



variance ratio
obs/model
10
4
2
1
0.5
0.25
0.1
a
b
zonal mean
60N
30N
EQ
30S
60S
180
90W
0
90E
180
1/2
2


mid-high latitudes
(sub)tropics
#models
20
15
10
5
0
0.2
0.5
1
2
5
variance ratio obs/model
ratio 2-5yr
ratio 20-50yr