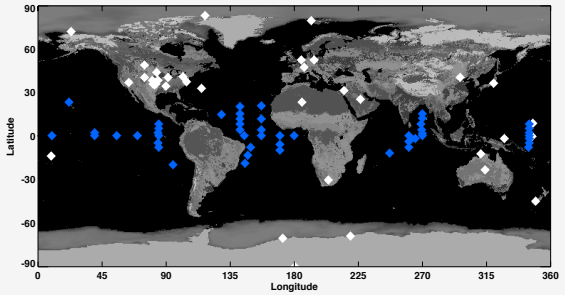
Sampling Error In Model and Observation Comparisons

David Rutan[2], Seiji Kato[1], Fred Rose[2]

1. NASA Langley Research Ctr., Science Directorate, Hampton, VA. 2. Science Systems and Applications Inc, Hampton, VA

It is now common to use surface observations (such as those provided by the ARM community) to assess the quality of radiative transfer calculations from GCM's, re-analyses, and satellite data products that produce surface irradiance estimates.

Sample of Surface Observation Sites



How many surface sites are "enough" and what is the impact on statistics as sites come and go?

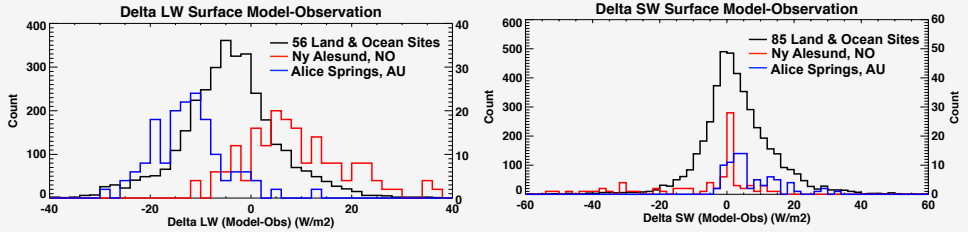
Comparing Observation to Calculation

Surface Longwave Irradiance (Observed Mean 330 Wm^{-2})				
Model	Monthly Mean Bias Wm^{-2}(%)	Standard Deviation Wm^{-2}(%)		
		3-hr	Day	Month
SYN	-4.1 (-1.2)	20.9 (6.3)	15.8 (4.8)	9.8 (3.0)
ISCCP	8.3 (2.5)	36.0 (10.9)	30.6 (9.3)	20.6 (6.2)
MERRA	-17.4 (-5.3)	28.2 (8.5)	24.1 (7.3)	20.6 (6.2)
SRB	-0.3 (0.0)	30.0 (9.1)	20.6 (6.2)	10.6 (3.2)
ERA-interim	-6.2 (-1.9)			11.7 (3.5)
EBAF	0.4 (0.1)			10.2 (3.1)

Surface Shortwave Irradiance (Observed Mean 201 Wm^{-2})				
Model	Monthly Mean Bias Wm^{-2}(%)	Standard Deviation Wm^{-2}(%)		
		3-hr	Day	Month
SYN	2.9 (1.4)	55.0 (28.4)	26.4 (13.1)	11.4 (5.7)
ISCCP	-10.0 (-5.0)	81.9 (40.7)	38.0 (18.9)	19.3 (9.6)
MERRA	14.0 (7.0)	88.1 (43.8)	46.2 (23.0)	22.9 (11.4)
SRB	-10.9 (5.4)	80.9 (40.2)	35.9 (17.9)	20.3 (10.1)
ERA-interim	9.7 (4.8)			18.1 (9.0)
EBAF	-0.1 (0.0)			11.8 (5.9)

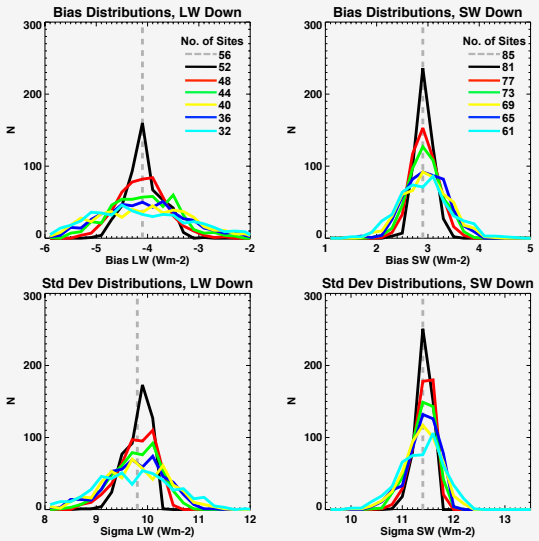
March 2000 to Dec 2007 (85 land & ocean surface sites)

However, the statistics resulting from such comparisons are not always as 'normal' as they appear. *Individual surface locations often have unique bias characteristics with respect to the model calculations.* Thus, should a site come or go, its effect is not a simple random sampling of the overall bias distribution.

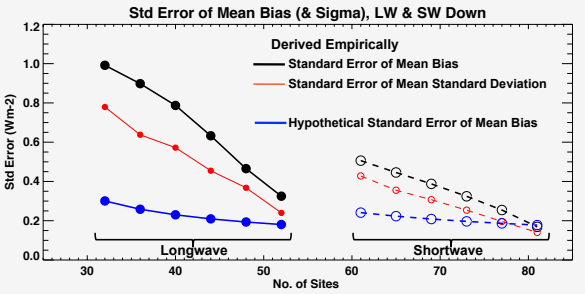


Removing or adding an individual site impacts the bias distribution in a non-random manner. General statistical theory gives that as a population is ***randomly*** sub-sampled the standard error of the mean is represented as σ/√n. But, removing a site is not randomly sampling the original distribution so this theory is not directly applicable for this type of comparison.

The standard error of the mean is instead derived empirically. To do this we progressively remove more sites re-sampling the original distribution 500 times with the new number of sites.



Distributions of bias & σ as more and more surface sites are removed from the population. Each of the 500 samples has a different "n" (sites*months_available*years) since each site has a distinct number of monthly mean values. Hence the breakdown of standard statistical theory.



The hypothetical blue lines are based on the idea that if all site/months were available and if one could sample the bias distribution randomly then these lines represent the σ/√n values. The black lines show the empirical results approaching the hypothetical as the number of sites increases. The current trend among institutions is to remove sites, not add them. So "how many surface sites are are enough?" We don't know, but this indicates reducing the number of sites takes our results in the wrong direction.

Corresponding Author: David.A.Rutan@NASA.gov