A New Metric to Quantify the Added Value of Regional Models

Masao Kanamitsu and Laurel DeHaan

Scripps Institution of Oceanography, University of California, San Diego

January 5, 2011

Submitted to the Journal of Geophysical Research-Atmospheres

Corresponding author: Dr. Masao Kanamitsu, Mail Code 0224; CASPO/SIO/UCSD; 9500 Gilman Drive; La Jolla, CA 92093-0224

E-mail: mkanamitsu@ucsd.edu

Abstract

A metric to quantify the value added by high resolution models is introduced. It is based on a characteristic spatial distribution of skill rather than the averages of skill values. Normal distribution functions are fit to the model skill distribution of coarse and fine resolution models and a new metric (Added Value Index, AVI) is defined as the area enclosed by the two distribution functions, with information on the way the two curves cross each other. The AVI is computed for a case of downscaling seasonal forecasts and is shown to properly provide a different degree of added value by high resolution models.

1. Introduction

Regional models have been used extensively for the purpose of short range forecasting and dynamical downscaling. The merits and demerits of regional models have been discussed by many (Anthes et al, 1989; de Elía and Laprise, 2003; Castro, 2005; Feser, 2006; Rockel et al, 2008; Prömmel et al, 2009; Winterfeldt and Weisse, 2009). In those studies, one of the most difficult subjects is to quantitatively represent the magnitude of value added by high resolution models in comparison to the coarse resolution models used as large scale forcing. Many studies simply show the high resolution spatial distribution maps and argue that the model output provides more realistic small scale features without any quantitative measures. Such studies are misleading to those who intend to use the high resolution results for quantitative applications. For a quantitative measure, most studies calculate bias errors, root mean square errors or temporal correlations with available station observations, or with a high resolution gridded analysis, and their individual values or area means are compared. Some studies drive application models (such as river routing, hydrology, agricultural, forest fire, etc.) and compare the output with observations, while others examine the spatial and temporal spectrums of the model results (Castro et al. 2005). For some idealized cases, it can be possible to quantitatively measure the errors of different spatial scales when the truth is given as a control experiment. In reality, such validation is generally very difficult for regional forecasts and downscalings since good high resolution analyses are hard to obtain. In the idealized framework, de Elia and Laprise (2003) used a distribution-oriented approach that makes it possible to measure the skill depending on the value of the field itself. This provides a useful tool for understanding the predictability based on the spatial spectrum of the field and measuring the skill of rare events. More studies have been done to measure the skill in probabilistic forecasts, but this paper will not address them.

For a quantitative measure of the value added by regional models, some estimate of the degree of fit of the simulation to observation must be used. In the next subsections, we will point out two independent problems with applying the most commonly used geographical distribution of temporal correlation or root mean square fit between simulations and observations as a measure for added value. The first problem concerns the difference in model resolution, which complicates the relation between fit to observation and model error. The second problem addresses the geographical distribution of the skill. We will then propose a new metric, the Added Value Index (AVI) which addresses the second problem.

1.1. Representativeness error

The above measures are based on how the model output deviates from the validating observations. Such measures may not be appropriate to quantify the added value of a regional model since the difference in the output spatial resolution itself may make the comparison problematical. In other words, it becomes difficult to differentiate the cause of the added value, whether it is from model resolution or model error. The inference of this differentiation between resolution and model error will be explained below.

The model grid point value is considered as a mean of the field represented by a grid point, which is a function of model grid size. Since the value is the most likely estimate at the grid, there is an error associated with it. This error may be named the representativeness error (ε_R), as it is commonly called in objective analysis. ε_R varies with model resolution as well as with the spatial variability of the field. For example, for near surface fields ε_R will be large over complex terrain and small over smooth land or over ocean. ε_R will be smaller for a smooth field, such as 500 hPa height, but larger for noisier vorticity, divergence and precipitation. When a model simulation is verified against station observations or fine resolution analysis on a grid, the model grid point values are interpolated to the station (or fine resolution grid) locations and the difference from the observation is used as a measure of the fit of the model to observation. Considering ε_R as introduced above, this process will be expressed as:

$$F^M(x_{obs}) = [\, F^T(x_{grid}) + \varepsilon_M + \varepsilon_R],$$

where $F(x_{grid})$ is a field examined at grid points x_{grid}, ε_M is a model error, and the bracket indicates a spatial interpolation operator. Subscript 'obs' indicates the observation at the location. Superscript T indicates truth and M indicates model. The interpolation introduces an additional error ε_I from the interpolation of $F^T(x_{grid})$, ε_M and ε_R, which leads to the following relation:

$$F^M(x_{obs}) = [\, F^T(x_{grid})\,] + [\varepsilon_M] + [\varepsilon_R] + \varepsilon_I$$

Thus, the model grid point values interpolated to the observation point have three types of errors, $[\varepsilon_M]$, $[\varepsilon_R]$ and ε_I. It is important to note that among these errors, ε_R and ε_I are not dependent on the model and may be estimated separately from observations or historical forecasts. In the following argument, we assume that ε_I is small compared to ε_M and ε_R . In addition to the previously mentioned errors, an observation at a location has its own error ε_{obs} which consists of instrument, retrieval, and representativeness errors. In addition, when we use a grid point analysis of observations for the skill calculation, we need to consider the additional error due to interpolation of irregularly spaced observations onto fine resolution regular analysis grid points. The error of the model at an observation location will be written as:

$$F^M(x_{obs}) - F^O(x_{obs}) = [\, F^T(x_{grid})] + [\varepsilon_M] + [\varepsilon_R] + \varepsilon_I - F^T(x_{obs}) + \varepsilon_{obs}$$

Since $[\, F^T(x_{grid})]$ is equal to $F^T(x_{obs})$,

$$F^M(x_{obs}) - F^O(x_{obs}) = [\varepsilon_M] + [\varepsilon_R] + \varepsilon_I + \varepsilon_{obs}$$

Therefore, the difference between model and observation at an observation location is a combination of four errors: model error, model grid point representativeness error, interpolation error, and the error in the observation itself consisting of instrument, retrieval representativeness and observation interpolation errors.

The model representativeness error ε_R can be estimated from the method proposed by Tustison et al (2001), which interpolates a field from a fine resolution analysis grid to a lower resolution model grid by area averaging (field A), and then interpolating back to the analysis grid (field B). The difference between the two (A-B) provides an estimate of the representativeness error.

Figure 1 shows the ε_R and ε_M for two model resolutions (a global model at 200 km and a Regional model 'b', hereafter referred to as Model-b, at 35 km resolution). The representativeness error (ε_R) is computed from the North American Regional Reanalysis (Mesinger et al. 2010) for the global model and CaRD10 (California Reanalysis Downscaling at 10 km, Kanamitsu & Kanamaru, 2007) for Model-b. Generally, ε_R decreases with decreasing grid distance, as expected. The ε_R is larger over the complex topography, where the small scale features dominate. Compared to ε_M , ε_R is smaller but still significant for the coarse resolution model, while it is much smaller than the ε_M for the fine resolution model. This indicates that the skill of the coarse resolution model is simply penalized by the $\varepsilon_{R,}$ and does not represent the true skill of the model.

The key point of this argument is that when we discuss the added value of the regional model, conventional skill comparisons provide a combination of different types of errors, which makes it difficult to understand the true meaning of the "valued added." For example, if the ε_M of the regional model is greater than that of the coarse resolution model, but ε_R is smaller due simply to the increased resolution, the fit to observations becomes better. Do we conclude that the regional model added value? For the model product users, the answer is probably yes, but for the modelers, the answer will probably be no. For the case of Figure 1, the magnitude of the fit of the simulations to analysis is about the same or slightly worse for Model-b, indicating that the high resolution model error is much larger than that of the coarse resolution CFS model.

The above discussion shows there are conceptual differences in interpreting a simple fit of model grid point values interpolated to observation as a metric for added value, depending whether one approaches the issue from a modeling or application point of view.

1.2. Spatial distribution of skill

Recognizing the limitation of the simple fit of model grid point values to observation as noted above, there is an additional weakness in utilizing the skill improvements, particularly their area average, as a measure of the value added. In Figure 2, we show a comparison of correlation skill against PRISM (Precipitation Elevation Regression Independent Slopes Model, Daly, et al, 2002) observations of January mean precipitation for two different resolution models. One is the NCEP/NCAR Reanalysis (Kalnay et al, 1996) and the other is the downscaling of the NCEP/NCAR Reanalysis using the 10 km resolution RSM model. As we would expect, we see much smaller-scale detail in the skill for RSM. When we computed the area average, the skill of NCEP/NCAR happened to have a small advantage. The regional model's disadvantage is coming from areas of large negative skill over the eastern slope of the Sierra Nevada Mountains, but at the same time we see enhanced skill over the western side of the mountain range. The figure clearly indicates that the regional model's maximum skill is larger but the area of high skill is much narrower. This implies that the regional model will be much more useful than the coarse resolution model over these high skill score areas. Since users will eventually look at areas where the model has useful skill, the regional model will apparently be adding value to those areas. Simply using the skill average over the regional domain does not allow such increases in local areas to be highlighted. In order to quantify this spatial distribution of high skill regions, we developed a new metric that compares the spatial distributions of high skill areas rather than the fit of the model simulation to observations.

In Section 2, we introduce the new metric. Section 3 describes some details of the new metric calculation. Section 4 presents results as applied to several cases, and in Section 5 we conclude the paper.

2. Added Value Index (AVI)

Figure 3 is an idealized example of probability distribution functions (PDF) generated from a geographical map of temporal correlations of an arbitrary variable, often called a skill map. Each curve is constructed by counting the number of grid points with the skill between skill values of S and S+ΔS, normalized by the total number of grid points in the domain of verification. When the skill is computed as a correlation, the S value ranges from -1 to 1, but in order to fit the curve to a Gaussian distribution, we need to apply a transformation of S such that the transformed S^* ranges from $-\infty$ to $+\infty$. The choice of the transformation and further details of the computational method are described in the next section. The thin vertical line marked S_c^* is the transformed *critical useful skill* of S_c using the above transformation. In the example given and throughout the paper, we chose the S_c for seasonal prediction of 0.3. This number is somewhat arbitrary and may have to be modified depending on the type of simulation (short-range forecast, seasonal forecast downscaling or climate downscaling) and the user's objective. In Figure 3, the solid line is assumed to be the PDF of the skill of a coarse resolution model, while the dashed line is that of the fine resolution regional model over the same domain. It is easily shown that the two curves cross each other at two points, except for the case when the variances of the skill of the two models are equal.

We can see four situations as shown in Figure 3. Panel (A) is the case when the average skill of the regional model is less than that of the coarse resolution model, and the higher skill tail of the distribution is lower than that of the coarse resolution model. In this case, the regional model is inferior to the coarse resolution model in all skill ranges above the critical useful skill. The area shaded by horizontal lines indicates the number of grid points (or areas) for which the regional model is inferior. When the regional model is inferior in all skill ranges, one of the cross points of the two PDFs is located to the left of the critically useful skill. The other point is located to the far right of the skill axis, but this point is regarded as an artifact due to its very small area between the two curves (an example will be shown in Section 4).

The second case (B) is when the mean skill is lower for the regional model, but the two curves cross due to a larger variance of the skill for the regional model. In this case, the regional model is inferior to the coarse resolution model up to the skill at the cross point (indicated by XP in the figure) but superior at higher skill. The area shaded by the horizontal lines is the area where the regional model is less skillful, while the cross-hatched area is where the regional model skill is higher. We may interpret this case as the useable skill redistributed from low to high skills. The example shown in Figure 2 corresponds to this case. The third case (C) is a complete opposite of (B). The fourth case (D) is the case when the mean skill of the regional model is higher, and also the area of high skill exceeds that of the global model.

In terms of added value by the regional model, case (D) is when the regional model performance is better than the coarse resolution model at all skill ranges. This is often the primary goal of regional modelers. But case (B) is also apparently adding value compared to the coarse resolution model, since the regional model has areas of much higher skill at higher skill values. This case is important for application since the regional model has higher utility over the areas of high skill. Note that in this case the mean skill over the regional domain is less than that of the coarse resolution model, thus one may falsely conclude that the regional model does not add value. Case (C) is when the regional model behavior is unreasonable, since the high resolution model provides smaller high skill areas than the coarse resolution model. Case (A) is a catastrophic case when the regional model has no advantage over the coarse resolution model at any skill level.

What we propose here is an index comprised of one number representing the area where the regional model skill is greater than that of the coarse resolution model and a symbol indicating the existence of the cross point. When the cross point does not exist or the point is far to the right of the x-axis, we show the area between the PDFs of the two models from the critical useful skill to infinity. When the cross point exists, we simply show the area to the right of the cross point. The difference between the cross-hatched and horizontal-line-hatched areas in Figure 3 may be of interest for overall performance, but it is already indicated by the domain mean skill. Choosing only the area of higher skill will augment the added value of the regional model. If we look at this index together with the mean skill, we will be able to see how the skill values are distributed in space and whether the regional model is adding value. We will call this index, Added Value Index (AVI).

Not surprisingly, the newly defined index, AVI varies with the nature of the field used to compute the skill, the size of the domain and, of course, the model used to make the simulations. In Section 4, we will show several examples of the AVI and demonstrate its usefulness. We will also briefly discuss the errors in AVI depending on the size of the area.

3. Computational details

The fit of the geographical distribution of skill to the Gaussian distribution requires some caution. The distribution may not necessarily follow the Gaussian distribution, requiring a transformation of the skill values, S. We examined the fit of the geographical distribution of skill of various variables to a normal distribution by constructing several Normal Test Plots. These plots are scatter diagrams of a theoretical normal distribution vs. observed skill data. Depending on the shape of the curve, we can identify the skewness and short/long tail distribution. If the skewness is found to be significant, we can reconstruct the skill data by assuming that the data less than the mean are symmetric to the mean. This will remove the skewness without affecting our results since we are only interested in the positive useable skill.

The shorter and longer tails can be adjusted by varying the transformation functions used to convert the skill value from -1 to +1 to $-\infty$ to $+\infty$. The function we used in our study is:

$$S^* = S/(1-ABS(S)^n) \quad (1)$$

By changing the value of n (larger n will shorten the tail and fractional n will lengthen the tail), we can improve the fit of the sample to the Gaussian distribution. When we applied this method with various 'n' to precipitation, near surface temperature, and 500 hPa height, we found that the data transformed with n=8 had the best fit. Figure 4 shows the Normal Test Plot before the transformation and after the n=8 transformation. For all three variables, the data fit the normal distribution very well. There is little skewness in the distribution, however there was enough that we adjusted for skewness in our study. The n=8 transformation slightly improved the fit of 500 hPa height to a normal distribution, but there was almost no difference for near surface temperature and precipitation. Table 1 shows the slope of the fitted line for the three variables with no transformation, n=4 and n=8 transformations. When we measure the goodness of the fit as a slope equal to one, we see improvement for 500 hPa height with n=8.

4. Examples

We applied the above definition of AVI to several cases of downscaled seasonal forecasts. The data used in this calculation are from the MRED (Multi-RCM Ensemble Downscaling of Seasonal Forecasts, details available from the data archive at https://docs.google.com/viewer?url=http://www.eol.ucar.edu/projects/cppa/meetings/200809/presentations/Tuesday/T0930_Arritt.pdf. Two regional models were chosen from the archive, Model-a and Model-b, both downscaled from the NCEP Climate Forecast System (Saha et al., 2006)). CFS has a horizontal resolution of about 200 km while both regional models are run using a 35 km resolution over the contiguous United States. An ensemble mean of 10 members for the period 1983 to 2008 was used for all three models. Only the downscaling for the January-February-March seasonal average is utilized.

Figure 5 shows an example of the difference of skill PDFs between CFS and the two regional models for near surface temperature over the southern Texas region. The cross point between the two PDFs occurs when the skill is near 0.5 for both regional models. Both models' skills are reduced up to the cross point and the skill greater than the cross point is increased. The rate of decrease and increase is larger for Model-a than for Model-b. This figure indicates that the geographical distribution of the low resolution CFS skill is redistributed to higher skills in the downscaled regional models, but the area mean skill is reduced slightly. This is a demonstration of small areas of higher skill in the regional simulations adding value compared to the CFS (example B described in Section 2).

Table 2 presents a summary of the cases examined for this paper. The AVI is obtained for surface temperature, precipitation, the u- and v-components of near surface winds, and 500 hPa geopotential height. Two areas, one over Texas and Mexico (110 to 96^{o} West, 25 to 36^{o} North, square area shown in Figure 6 left panel) and the other over the entire contiguous United States and northern Mexico are chosen to examine the variability of AVI with domain size. In addition, two regional models, a and b are validated.

The second and third columns of Table 2 (downscale mean and CFS mean) compare the area mean skill from a regional model and the coarse resolution CFS model. If we use this as a measure of value added, regional models have higher skill than the CFS only 8 times, while the CFS is better 9 times. Apparently, high resolution models do not add value to the CFS forecasts if we simply compare the domain average skill. The fifth and sixth columns of the table show the area between the two PDF curves, first from the critical skill level (.3) to the cross point, and then from the cross point to infinity. In the case where there is no cross point between the critical skill level and the far right of the x-axis, the sixth column shows the difference from 0.3

to infinity. This sixth column gives the AVI. The character 'x' attached at the end of the number indicates the presence of a cross point. When we examine the AVI, the regional model improves over the CFS 14 times, showing the added value of the regional downscaling clearly. For individual models, Model-a added value 4 times out of 10, while the domain mean skill is better only one time. For Model-b, value is added 10 out of 10 times, while the domain mean skill is better 5 times out of 10. Thus, among the two models, Model-b seems to be better than Model-a, always adding value to the CFS forecasts, while Model-a fails to add value in several cases. This table nicely demonstrates the value added by the regional models compared to the simple use of area mean skill

Comparing the near surface temperature skill maps of the three models (Figure 6), it is clear that Model-a and Model-b have larger areas of higher correlation over the Pacific Northwest coast. Model-a has an area of good skill to the south of Lake Superior but Model-b has a larger area of skill greater than 0.5 over Mexico. Thus, we expect AVI to be positive for Model-a and -b, but Model-b should have a slight edge over Model-a due to larger areas of higher skill over the Northwest. Over Texas/Mexico, the area of skill higher than 0.5 is greater for Model-b than Model-a. From these subjective observations, we expect AVI to be positive and larger for Model-a than for Model-b for the U.S. area while we expect the opposite for the Texas/Mexico area, which agrees with the AVI table discussed above. Thus, AVI is able to differentiate subtle differences in the high skill areas between the three models. Table 2 also gives AVI for 500 hPa height. This field is selected to highlight the different behavior of the model performance due to the spatial variability of the field. Interestingly, AVI did not show any different behavior, except much higher mean skill over the large US domain for both models and Model-a's unrealistic behavior of negative AVI with cross point, which is also seen in the skill of u-component of the wind.

These examples demonstrate clearly that the AVI can quantitatively present added value that cannot be shown by the area mean skill alone. The current example also was successful in differentiating skill of two models very well.

The error in the computation of the AVI can be estimated from the estimated error of mean and variance of skills that depends on sample size and variance. It was found that the estimated error of cross point and AVI are very small due to the large number of grid points used in our calculations.

5. Conclusions

A new metric to quantitatively measure the value added by regional models was introduced. The motivation comes from comparing the geographical patterns of temporal correlation skill maps between low and high resolution models. The high resolution model tends to give very high skill over small scale regions, while the low resolution model tends to give relatively lower skill over a larger domain. At the same time, the high resolution model often produces small regions of large negative correlation, and thus a simple area average skill cannot differentiate this important difference in the characteristics of the geographical distribution. In other words, it is necessary to provide not only the mean skill but also a measure of the geographical distribution of skill. The proposed method focuses on the probability distribution of the geographical distribution of temporal correlation in the regional model domain or its sub-domain. We first fit the skill distributions of two models to normal distributions, then overlay them and compute the cross points of the two PDFs. We define the Added Value Index (AVI) as the area beyond critical useful skill where the regional model skill is greater than that of the coarse resolution model. Here the critical useful skill is a predetermined skill beyond which the simulation is considered to be useful for the user's objectives. When the cross point between the two PDFs is far to the right of the skill axis, we assume that there is no cross point, and the AVI becomes the area between the two curves from critical useful skill to 1. The AVI will thus be expressed as one number and a symbol expressing the existence of a cross point.

This definition of the AVI was applied to several cases, and shown to satisfactorily characterize the model performance for different variables over different areas. Although our example uses a seasonal forecast downscaling, this result will also apply to short range forecasts.

We used temporal correlation as a skill map in the current example, but normalized RMS can also be used to calculate AVI. In addition, the AVI proposed in this study may be extended to a time series of pattern correlations. In this case, the AVI indicates the high resolution model's ability to represent high time frequency phenomena, or occasional high skill cases.

In a separate publication, we plan to apply the AVI to a much large number of regional downscaling simulations and present its usefulness.

Acknowledgements

This study was supported by NOAA ECPC and NOAA MRED. The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA. We thank Ms. Diane Boomer for the proof reading.

References

Anthes, R., Y. -H. Kuo, E. -Y. Hsie, S. Low-Nam and T. W. Bettege (1989), Estimation of skill and uncertainty in regional numerical models, *Quart. J. Royal Met. Soc.*, *115*, 763-806.

Castro, C. L. (2005), Dynamical downscaling: Assessment of value retained and added using the Regional Atmospheric Modeling System (RAMS), *J. Geophys. Res., 110*, 1-21.

Daly, C., W. P. Gibson, G. H. Taylor, G. L. Johnson, and P. Pasteris (2002), A knowledge-based approach to the statistical mapping of climate, *Climate Res., 22*, 99-113.

de Elía, R. & R. Laprise (2003), Distribution-Oriented Verification of Limited-Area Model Forecasts in a Perfect-Model Framework, *Mon. Wea. Rev.*, *131*, 2492-2509.

Feser, F. (2006), Enhanced Detectability of Added Value in Limited-Area Model Results Separated into Different Spatial Scales, *Mon. Wea. Rev.*, *134*, 2180-2190.

Kalnay, E., M. Kanamitsu, R. E. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, R. Jenne. and D. Joseph (1996), The NCEP/NCAR 40-Year Reanalysis Project, *Bull. Amer. Met. Soc.*, *77*, 437-471.

Kanamitsu, M. & H. Kanamaru (2007), Fifty-Seven-Year California Reanalysis Downscaling at 10 km (CaRD10). Part I: System Detail and Validation with Observations, *J. Climate*, *20*, 5553-5571.

Mesinger, F., G. DiMego, E. Kalnay, K. Mitchell, P. C. Shafran, W. Ebisuzaki, D. Jović, J. Woollen, E. Rogers, E. H. Berbery, M. B. Ek, Y. Fan, R. Grumbine, W. Higgins, H. Li, Y. Lin, G. Manikin, D. Parrish and W. Shi (2010), North American Regional Reanalysis, *Bull. Amer. Met. Soc.*, *87*, 343-360.

Prömmel, K., B. Geyer, J. M. Jones and M. Widmann (2009), Evaluation of the skill and added value of a reanalysis-driven regional simulation for Alpine temperature, *Int. J. Climatol.*, *30*, 760-773.

Rockel, B., C. L. Castro, R. A. Pielke, H. von Storch and L. Giovanni (2008), Dynamical downscaling: Assessment of model system dependent retained and added variability for two different regional climate models, *J. Geophys. Res.*, *113*, D21107, doi:10.1029/2007JD009461.

S. Saha, S. Nadiga, C. Thiaw, J. Wang, W. Wang, Q. Zhang, H. M. Van den Dool, H. -L. Pan, S. Moorthi, D. Behringer, D. Stokes, M. Pena, S. Lord, G. White, W. Ebisuzaki, P. Peng, and P. Xie (2006), The NCEP Climate Forecast System, *J. Climate*, **19**, 3483-3517.

Tustison, B., D. Harris, and E. Foufoula-Georgiou (2001), Scale issues in verification of precipitation forecasts, *J. Geophys. Res.*, *106*, 11775-11784.

Winterfeldt, J. and R. Weisse (2009), Assessment of Value Added for Surface Marine Wind Speed Obtained from Two Regional Climate Models, *Mon. Wea. Rev.*, *137*, 2955-2965.

Figure Captions

Figure 1. Model grid representativeness error (left panels) equivalent to CFS resolution (upper panel) and Model-b resolution (lower panel) compared with model error (left panels) for CFS (upper panel) and Model-b (lower panel). The variable is seasonally averaged precipitation root mean square error against NARR analysis.

Figure 2. Correlation skill of January mean precipitation for CaRD10 (left) and NCEP/NCAR Reanalysis (right) verified against PRISM gridded observation. Computation is made using 1950-1997 data. Figure taken from Kanamitsu and Kanamaru (2007) Figure 10.

Figure 3. Idealized distribution functions of correlation skill over the model domain for two different models. See text for more detail. The hatched area with horizontal lines indicates where the dashed line model has lower skill, while the cross hatched area indicates otherwise.

Figure 4. Normal test plot of near surface temperature (top), 500 hPa height (middle) and precipitation (bottom) with no transformation (left) and n=8 transformation (right).

Figure 5. An example of the PDF differences between Model-a and CFS (dark grey line) and Model-b and CFS (light grey line). Vertical axis is the normalized area (or number of grid points) and horizontal axis is skill.

Figure 6. An example of the geographic distribution of surface temperature skill for CFS (left),Model-a (middle), and the difference between the two (right).

Table 1. Slope of the linear fitted line between observed and normal distribution fitted skill for three variables with n=4 and 8 transformations.

	2m temperature	Precipitation	500 hPa height
No scaling	0.987	0.970	0.963
n=4 scaling	1.090	1.147	1.240
n=8 scaling	0.997	0.997	1.077

Table 2. Area mean skill, cross point, difference between the two PDFs and AVI, computed from downscaling of Model-b CFS over the TX/Mex area and the contiguous United States.

scaled with x/(1-x^8)							
	Down Scale Mean	CFS Mean	X pt	Diff .3 to X pt	Diff > X pt	AVI	Adde d value
T2m TX/Mex Model-a	**0.35**	0.34	0.41	-0.03	0.03	0.03x	yes
T2m TX/Mex Model-b	**0.35**	0.34	0.49	-0.02	0.04	0.04x	yes
T2m US Model-a	**0.16**	0.14	No X	0.00	0.02	0.02	yes
T2m US Model-b	0.13	**0.14**	0.47	-0.01	0.01	0.01x	yes
Precip Tx/Mex Model-a	0.22	**0.23**	No X	0.00	-0.04	-0.04	no
Precip Tx/Mex Model-b	**0.24**	0.23	No X	0.02	0.02	0.02	yes
Precip US Model-a	0.18	**0.23**	No X	0.00	-0.07	-0.07	no
Precip US Model-b	**0.24**	0.23	No X	0.00	0.03	0.03	yes
Usfc TX/Mex Model-a	0.24	**0.27**	0.55	-0.06	0.02	0.02x	yes
Usfc TX/Mex Model-b	0.25	**0.27**	0.50	-0.07	0.06	0.06x	yes
Usfc US MODEL-a	0.32	**0.33**	0.33	0.00	-0.03	-0.03x	no
Usfc US Model-b	0.33	0.33	0.56	-0.03	0.02	0.02x	yes
Vsfc TX/Mex Model-a	0.07	**0.13**	No X	0.00	-0.07	-0.07	no
Vsfc TX/Mex Model-b	**0.22**	0.13	No X	0.00	0.16	0.16	yes
Vsfc US Model-a	0.10	**0.12**	No X	0.00	-0.05	-0.05	no
Vsfc US Model-b	**0.13**	0.12	No X	0.00	0.02	0.02	yes
500 ht Tx/Mex Model-a	0.63	**0.64**	0.65	0.04	-0.04	-0.04x	no
500 ht Tx/Mex Model-b	**0.65**	0.64	0.63	-0.08	0.08	0.08x	yes
500 ht US Model-a	0.38	0.38	0.51	-0.01	0.02	0.02x	yes
500 ht US Model-b	0.38	0.38	0.46	-0.01	0.02	0.02x	yes

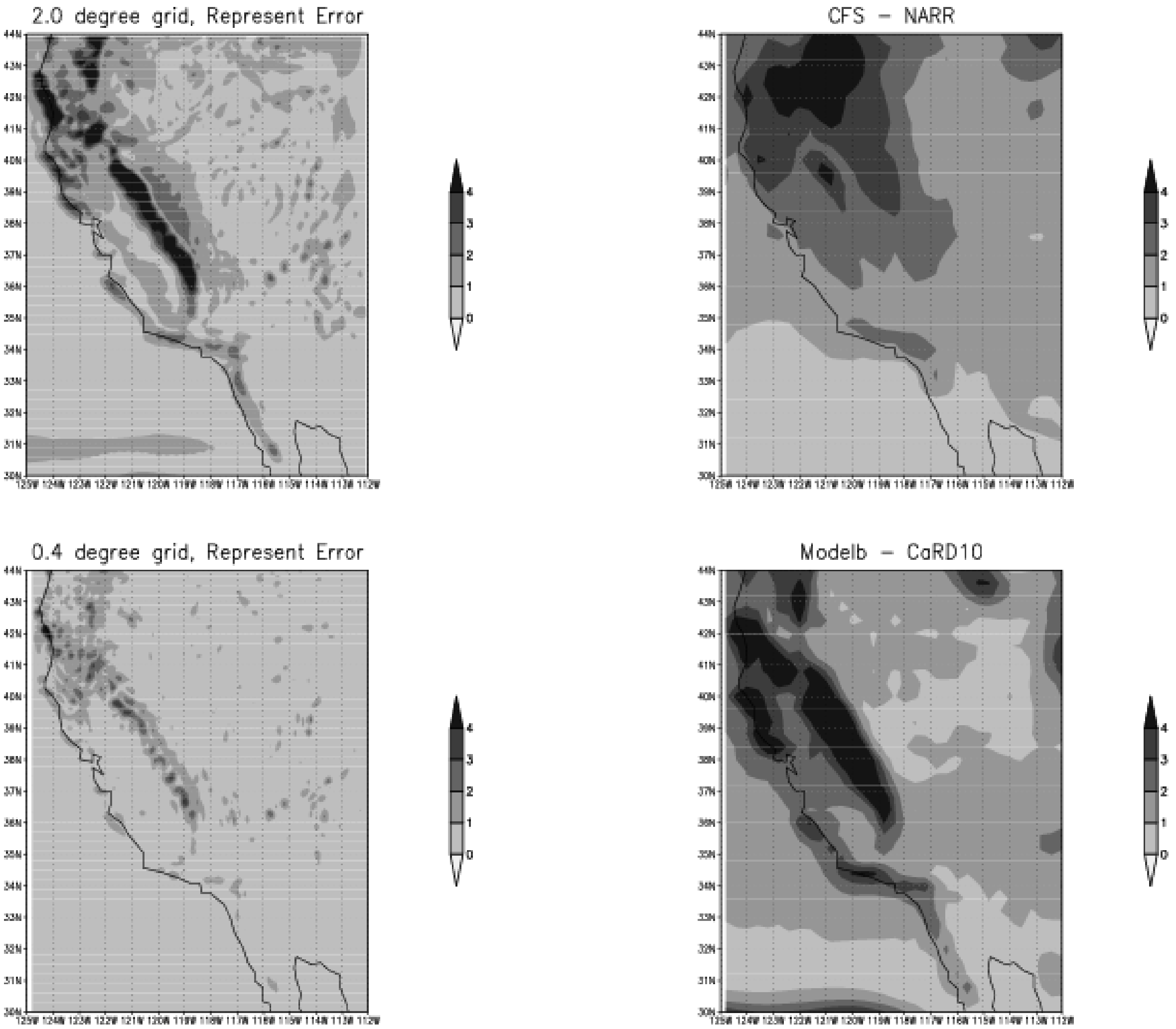

Figure 1. Model grid representativeness error (left panels) equivalent to CFS resolution (upper panel) and Model-b resolution (lower panel) compared with model error (left panels) for CFS (upper panel) and Model-b (lower panel). The variable is seasonally averaged precipitation root mean square error against NARR analysis.

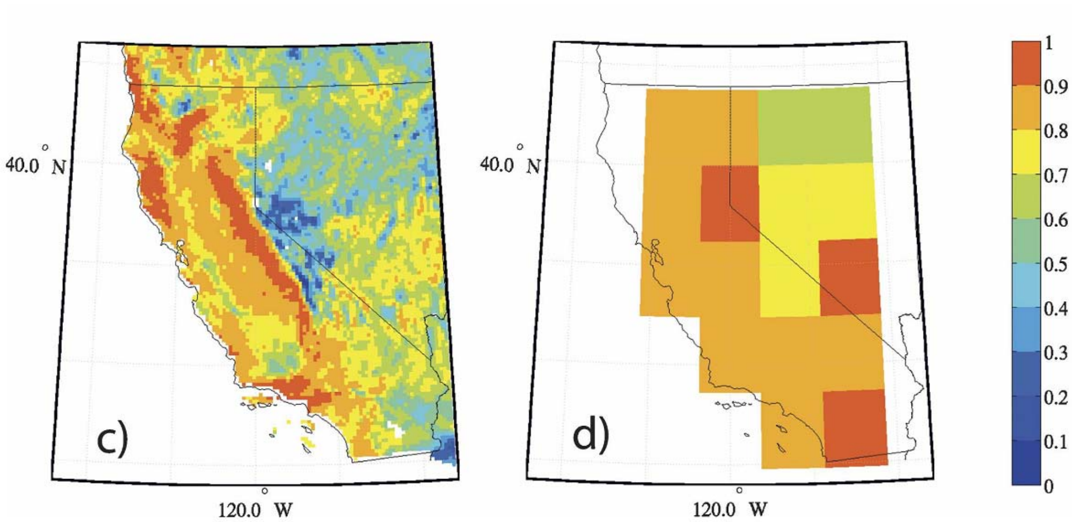

Figure 2. Correlation skill of January mean precipitation for CaRD10 (left) and NCEP/NCAR Reanalysis (right) verified against PRISM gridded observation. Computation is made using 1950-1997 data. Figure taken from Kanamitsu and Kanamaru (2007) Figure 10.

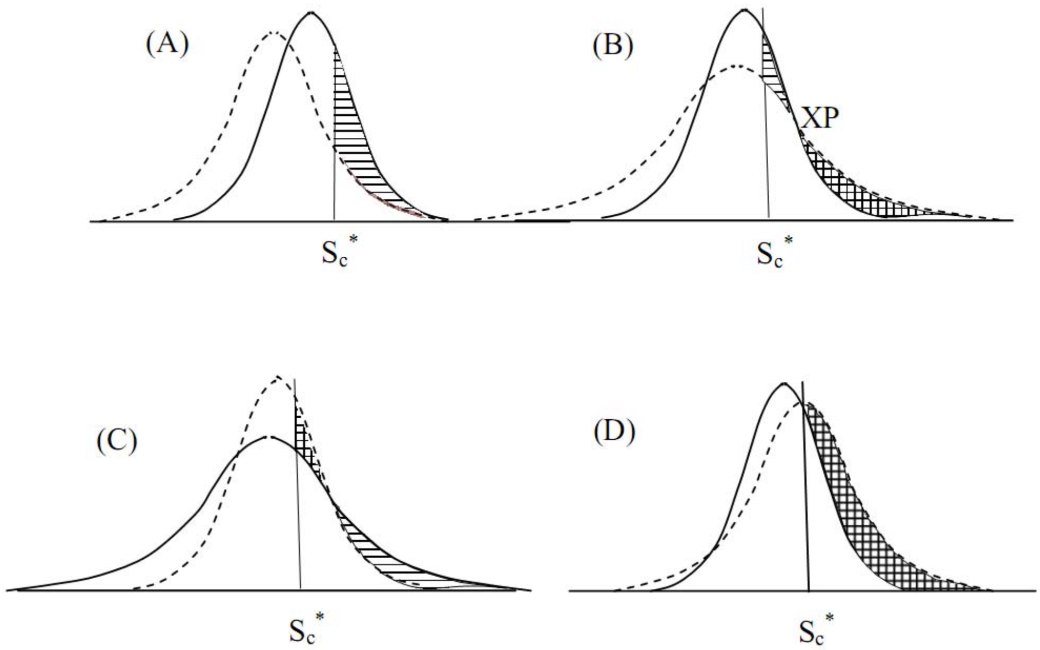

Figure 3. Idealized distribution functions of correlation skill over the model domain for two different models. See text for more detail. The hatched area with horizontal lines indicates where the dashed line model has lower skill, while the cross hatched area indicates otherwise.

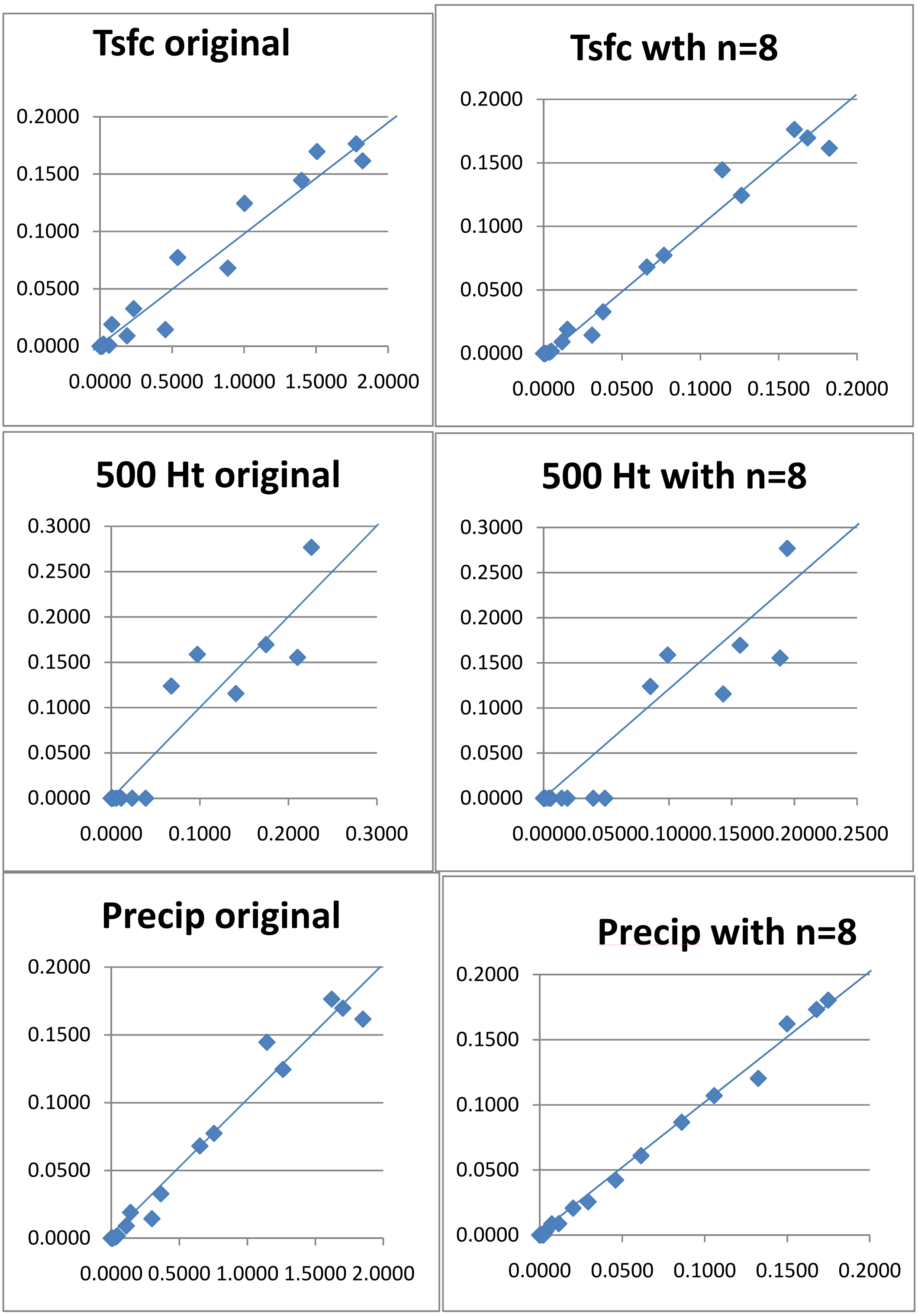

Figure 4. Normal test plot of near surface temperature (top), 500 hPa height (middle) and precipitation (bottom) with no transformation (left) and transformed with n=8 (right).

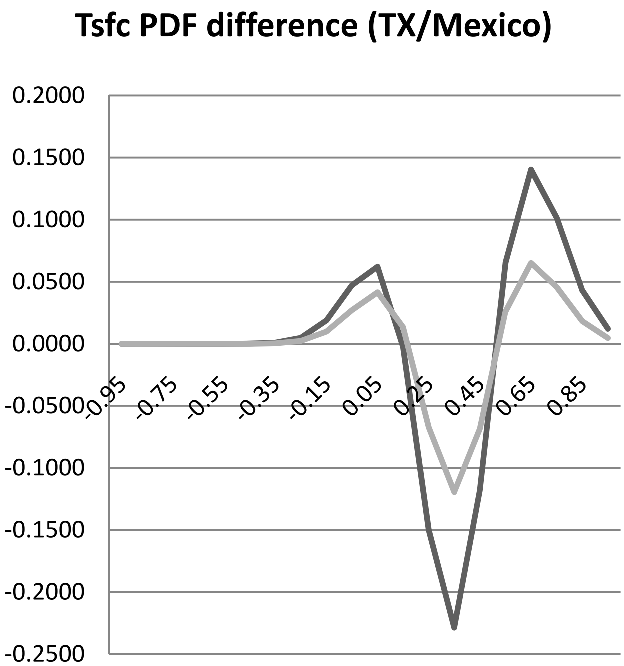

Figure 5. An example of the differences between Model-a and CFS (dark grey line) and Model-b and CFS (light grey line). Vertical axis is the normalized area (or number of grid points) and horizontal axis is skill.

439

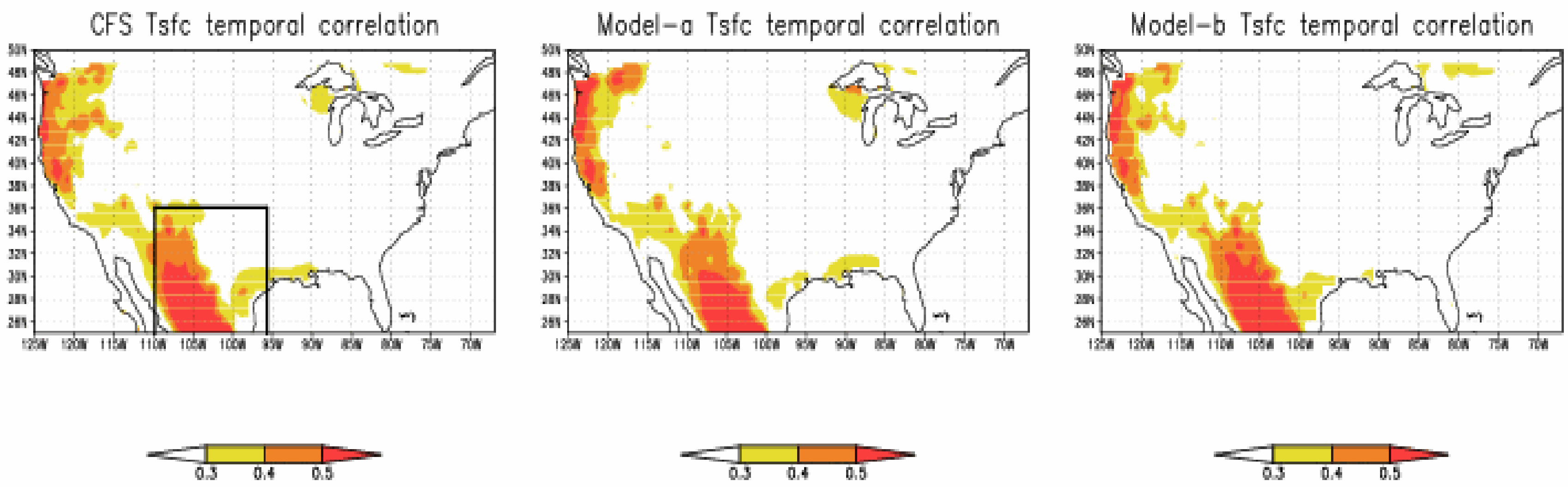

Figure 6. An example of the geographic distribution of near surface temperature skill for CFS (left), Model-a (middle), and Model-b (right).