Response to reviewer # 1 and #2 input on "A new metric for climate models that includes field and spatial dependencies using Gaussian Markov Random Fields" by Nosedal-Sanchez et al.

May 21, 2016

Response to Reviewer # 1 Reviewer comment given in blue.

1. Conditional independence is an assumption underlying Markov random fields. For three variables A, B, and C, the joint probability distribution of A and B conditioned on C, written as $p(A,B|C)$, can be factored into the product $p(A|C)$ times $p(B|C)$ for all values of C if A and B are conditionally independent of C. The authors should argue, or preferably demonstrate, that the necessary conditional independence properties approximately hold for the application of their method to climate model fields. The feedbacks across scales in the climate system and the coupled nature of the physical equations may serve as a basis for some degree of conditional independence, though I expect some cases where $p(A|C)$ is a poor approximation of $p(A|B,C)$ as implied by conditional independence.

   The assumption of conditional independence (for estimating precisions among points outside a neighborhood structure) does not need to be met exactly in order for GMRF to represent a useful step forward toward the goals outlined in the introduction. The assumption facilitates the sparse representation of the precision matrix and therefore is convenient. It enables us to capture some but perhaps not the full extent of the dependencies that exist across space and fields in the climate data. The witch-hat graphs provide a measure of how well GMRF captures observed covariances. One could enlarge the neighborhood structure indicating conditional dependencies of the precision matrix beyond nearest neighbors, but we felt that the present treatment was adequate.

2. The authors appear to neglect temporal relationships in the method and example, even though such relationships are prevalent in the climate system. A perturbation

in the pattern of sea surface temperature in the tropics, for example, may take months before the signal shows up in the spatial distribution of precipitation in the mid-latitudes. While introducing temporal correlations into their method is beyond the scope of the manuscript and not required at this stage, it would still be beneficial to readers if the authors described how their method could be extended is this way.

It is common for climate model evaluation to place most of its emphasis on long-term means and that is our target application. GMRFs may be extended to include temporal relationships (e.g. Cressie and Wikle, 2011), but we did not attempt to develop those ideas in the present manuscript. Note that the assumption that the distribution of errors are Gaussian does not hold as well on short(er) time scales. The text mentions that the effects of teleconnection patterns shape local covariances (through a set of processes that are influenced by anomalously low or high pressures). Space-field GMRFs would be sensitive to these effects since the teleconnections shape long term means. We will update the text to provide more information about the possibility to extend the analysis to include temporal relationships.

Statistics for Spatio-Temporal Data, by Noel Cressie and Christopher K. Wikle. Wiley, Hoboken, NJ, 2011 (588 pp.)

3. The opening paragraph states that there is skepticism in using a scalar metric to assess climate model performance. This gives the impression that everything gets boiled down to a single number, which isn't the case. Climate models are often assessed using a vector of scalar quantities (e.g. as in Gleckler et al), a scalar measure of a vector field, or combinations of these and other metrics. A single field projected on a Taylor diagram, for example, considers two orthogonal scalar quantities (centered rms and correlation). Please clarify the description.

   We agree with the reviewer's point that the scientific community makes use of many metrics to judge a model's credibility. The section explaining our point was poorly written. The issue is not that climate scientists already make use of many metrics in model selection. We needed to first say that formal methods for model calibration operate on a single scalar metric. The scientific community is skeptical that a scalar metric (and therefore formal calibration methods) could adequately capture all the scientific sensibilities that are needed for judging model acceptability. We will clarify this point.

4. The opening paragraph also describes the need to account for spatial and field dependencies. Field dependence is an essential feature of your methodology, so it would be useful to readers to provide a specific example of what you mean by field dependence early in the introduction.

Thanks for this suggestion. We will provide an example.

5. The first sentence in the second paragraph in the introduction is a little awkward and should be rewritten (lines 12-15, page 3). There is an observational record of climate, but not an observational record of a climate model. Moreover, this statement seems to suggest that data assimilation is primarily a data imputation method, which it really isn't. Data assimilation minimizes the differences between the model state and observations, while insuring that the state fields abide by conservation laws (mass, energy, and momentum) and other important physical dependencies. This paragraph overall seems to imply that the models don't do a very good job with the dependencies, which is arguable. I have confidence that the models are getting many of the large scale dependencies about right (e.g. equator to pole gradients, land-ocean contrasts, temperature dependence of water vapor through Clausius-Clapeyron, etc).

   In this paragraph we wish to make the point that there exists a very limited observational record on which to estimate space and field dependencies of climate phenomena. Likely the best synthesis of these dependencies are from reanalysis products for some of the reasons you state. However there is a catch. The models used for data assimilation rely on their own physics packages to predict cloud characteristics and their radiative effects. Moreover, the data assimilation strategies for generating these products do not attempt to conserve mass, energy, and momentum, particularly between analysis steps. So the products are both a reflection of the observations that go into them and the physics and fluid motions of the model. Figure 1 below provides an illustration of this point. It shows that a multivariate measure of the distance between NCEP and ERA40, which are two reanalysis products, and CAM3.1 were nearly as different from one another as CAM3.1 was to them. Seasonally and regionally, the two products contained upwards of 200 Wm^{-2} differences in shortwave radiation reaching the surface which is emblematic of the different parameterizations each model uses for estimating cloud distributions and their radiative properties.

6. It's a good idea to present the general idea behind the metric in equation (1) in the introduction, though I found myself flipping back and forth between the introduction and section 2 to make better sense of the information. To make it easier for readers to get through the introduction without getting hung up on details, perhaps you could introduce the concept in more general terms. Also, the symbol Z is used for the metric in this section, but it doesn't appear elsewhere in the manuscript and should be dropped. And the times symbol in 'n_{obs} x n_{pts}' on line 3, page 4 suggests that v is a matrix with n_{obs} rows and n_{pts} columns. I recommend changing it to $n_{obs}n_{pts}$.

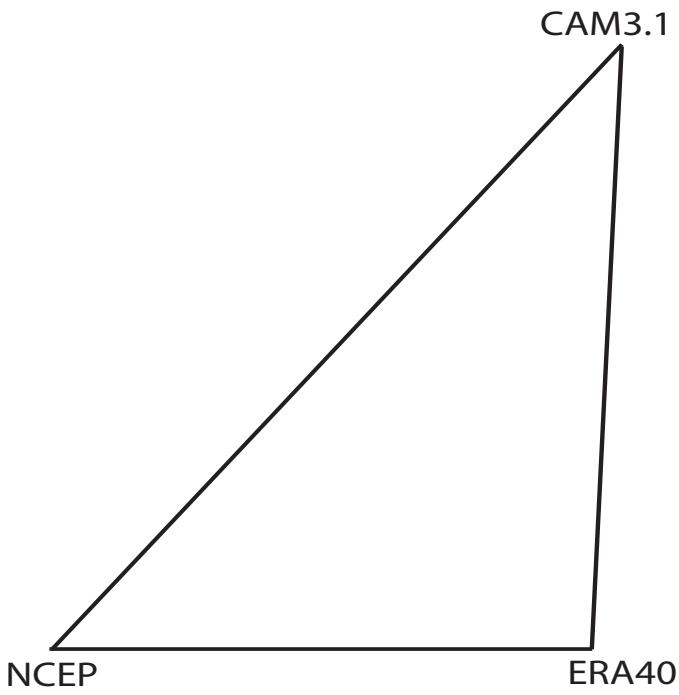

Figure 1: Average distance between two data assimilation products, NCEP (Kalnay et al.,1996); Kistler et al. 2001) and ERA40 (Uppala et al. 2005), is a good fraction of the distance to CAM3.1 (Community Atmosphere Model version 3.1). The length of each segment is based on the metric used in Jackson et al., (2008) and described by Mu et al., (2003) and includes shortwave radiation to surface, 2 m air temperature, surface sensible heat flux, relative humidity, air temperature, zonal winds, and sea level pressure from 1990 to 2001.

These are very good suggestions and clarifications. We will update the text.

7. There is a typo in the lower right element of the S inverse matrix on line 11, page 7. The sigma index should be 22, instead of 11. For consistency, use the same indices for off-diagonal terms (e.g. S12 is used for the lower left term on line 11, while S21 is used on line 13 page 7).

   Thanks for catching this error.

8. Regarding the alpha parameter, please provide references or further information about the statement that alpha depends only on the geometry of the neighborhood and not on the details of the fields. I have other questions about alpha. How much does it vary going from a first order neighborhood to a higher order neighborhood? Can alpha be extended from a scalar to a vector to optimize the covariances in

different regions in the neighborhood?

Equation (5) shows that α is only dependent on the eigenvalues of the Q matrix. The Q matrix itself is only a function of the domain (e.g. geometry and number of latitude and longitude grid points) and the neighborhood structure. Thus we would expect the value for α to be affected if we use a higher order neighborhood. We started the task of building a Q matrix with a higher order neighborhood structure, however implementing it correctly requires a lot of attention to detail to deal with how the stencil changes as one approaches a boundary and this task will require more time to complete than we have at the moment. Thus we don't know how much α would be affected. Because the higher order Q matrix will have more than one singular vector, we came to the realization that the concept of α may need to be expanded to accommodate all of the Q matrix singular vectors. Thus the answer to the question is not straight-forward and would require further consideration. In response to your last question, since α in the way we have been using it exists as an extension of the Q matrix, it does not make sense to use it to accommodate covariances for particular regions which may be field dependent.

9. The witch hat plots are convenient, but take some effort to get used to. It would be useful it you first stepped the readers through the concept with a simple example. How much does the shape of the witch hat depend on the selected indexing for the neighborhood? E.g. swapping x3 and x4 in figure 1 appears arbitrary, but results in a different Q. Does the averaging of the cells for a given distance from the diagonal hide information that could be important? Are there other simple ways to show the differences between the empirical and GMRF estimates (e.g. Hinton diagrams)?

   The results would not be altered by how boxes are indexed within a Q matrix that correctly identifies the neighborhood structure around each grid cell. The reason to present a summary of the covariance matrix in terms of a 'witch hat' graph is because there is not much variation in the estimate of the variances/covariances along any of the diagonals. There can be small deviations in the symmetry that occur because of how neighboring cells are indexed particularly as one approaches a boundary. An example of this deviation is provided below. However a fairly accurate estimate of the 'witch hat' could be constructed as an average of variances or covariances along the various diagonals relative to the main diagonal. However, as the example illustrates, this is not as simple as an evaluation of the distance to the diagonal. We construct our covariances for the 'witch hat' graph by explicitly identifying those cells that are a given distance from the diagonal (see example below). Hinton diagrams provide a graphic view of the size of values within a matrix. 'witch hat' graphs allow us to compare GMRF implied variances/covariances with those estimated empirically from data. The text explaining witch hats will be

further clarified.

Example of the construction of a 'witch hat'.

We will describe the construction of a witch hat graph for a 3×3 lattice, like the one shown below.

1	2	3
4	5	6
7	8	9

Suppose that variance estimates are available at each of the 9 grid points for one field, for example $S_{11}Q^*$. In this case, $S_{11}Q^*$ is a 9×9 matrix.

Like any other graph, a witch hat graph is formed by points. We will find the points that define a witch hat graph that makes comparisons in the N - S direction. From the figure shown above, it is clear that grid cells $1,2,3,4,5$, and 6, have one grid cell below them: 4, 5, 6, 7, 8, and 9, respectively. Now, we use these numbers to form pairs: $(4,1),(5,2),(6,3),(7,4),(8,5)$, and $(9,6)$. Then, using the corresponding elements of our matrix of estimates, we compute the following average:

$$w(-1) = \frac{\hat{\sigma}_{41} + \hat{\sigma}_{52} + \hat{\sigma}_{63} + \hat{\sigma}_{74} + \hat{\sigma}_{85} + \hat{\sigma}_{96}}{6}$$

(where $\hat{\sigma}_{ij} = \hat{\sigma}(i,j)$ = element located on ith row and jth column of matrix of estimates).

Thus, we define $(-1, w(-1))$ as one point of our witch hat graph. Let us find another point. Again, using the same figure, it is clear that grid cells $4,5,6,7,8$, and 9, have one grid cell above them: $1,2,3,4,5$, and 6, respectively. As we did before, we proceed to form pairs with these numbers: $(1,4),(2,5),(3,6),(4,7),(5,8)$, and $(6,9)$. Then, we use the corresponding elements of our matrix of estimates to compute another average:

$$w(1) = \frac{\hat{\sigma}_{14} + \hat{\sigma}_{25} + \hat{\sigma}_{36} + \hat{\sigma}_{47} + \hat{\sigma}_{58} + \hat{\sigma}_{69}}{6}$$

This couple of numbers, $(1, w(1))$, gives another point of the witch hat graph. Doing something similar, with grid cells that have neighbours located two rows down or up of themselves, we obtain:

$$w(-2) = \frac{\hat{\sigma}_{71} + \hat{\sigma}_{82} + \hat{\sigma}_{93}}{3}$$

and

$$w(2) = \frac{\hat{\sigma}_{17} + \hat{\sigma}_{28} + \hat{\sigma}_{39}}{3}.$$

Note that the number of cells from one grid to itself is zero. So, a fifth point for our graph is

$$w(0) = \frac{\hat{\sigma}_{11} + \hat{\sigma}_{22} + \ldots + \hat{\sigma}_{99}}{9}.$$

A witch hat graph is a graphical representation of these pairs of points: $(-2, w(-2))$, $(-1, w(-1))$, $(0, w(0))$, $(1, w(1))$, and $(2, w(2))$. By construction, $w(-1) = w(1)$ and $w(-2) = w(2)$ (recalling that the matrix of estimates is symmetric). Which is convenient for computational purposes. It is worth noting that we could define a graph in the E-W direction in a similar fashion. In general, a witch hat graph in the E-W direction will differ from the one constructed to make comparisons in the N-S direction.

Note. Making a graph of a 9×9 matrix of estimates would allow us to see that $w(1)$ = average of entries located **three** columns to the right of main diagonal. Similarly, $w(2)$ = average of entries located **six** columns to the right of main diagonal. However, these numbers (three and six) **depend on the number of columns of lattice in question**.

$\hat{\sigma}_{11}$	$\hat{\sigma}_{12}$	$\hat{\sigma}_{13}$	$\hat{\sigma}_{14}$	$\hat{\sigma}_{15}$	$\hat{\sigma}_{16}$	$\hat{\sigma}_{17}$	$\hat{\sigma}_{18}$	$\hat{\sigma}_{19}$
$\hat{\sigma}_{21}$	$\hat{\sigma}_{22}$	$\hat{\sigma}_{23}$	$\hat{\sigma}_{24}$	$\hat{\sigma}_{25}$	$\hat{\sigma}_{26}$	$\hat{\sigma}_{27}$	$\hat{\sigma}_{28}$	$\hat{\sigma}_{29}$
$\hat{\sigma}_{31}$	$\hat{\sigma}_{32}$	$\hat{\sigma}_{33}$	$\hat{\sigma}_{34}$	$\hat{\sigma}_{35}$	$\hat{\sigma}_{36}$	$\hat{\sigma}_{37}$	$\hat{\sigma}_{38}$	$\hat{\sigma}_{39}$
$\hat{\sigma}_{41}$	$\hat{\sigma}_{42}$	$\hat{\sigma}_{43}$	$\hat{\sigma}_{44}$	$\hat{\sigma}_{45}$	$\hat{\sigma}_{46}$	$\hat{\sigma}_{47}$	$\hat{\sigma}_{48}$	$\hat{\sigma}_{49}$
$\hat{\sigma}_{51}$	$\hat{\sigma}_{52}$	$\hat{\sigma}_{53}$	$\hat{\sigma}_{54}$	$\hat{\sigma}_{55}$	$\hat{\sigma}_{56}$	$\hat{\sigma}_{57}$	$\hat{\sigma}_{58}$	$\hat{\sigma}_{59}$
$\hat{\sigma}_{61}$	$\hat{\sigma}_{62}$	$\hat{\sigma}_{63}$	$\hat{\sigma}_{64}$	$\hat{\sigma}_{65}$	$\hat{\sigma}_{66}$	$\hat{\sigma}_{67}$	$\hat{\sigma}_{68}$	$\hat{\sigma}_{69}$
$\hat{\sigma}_{71}$	$\hat{\sigma}_{72}$	$\hat{\sigma}_{73}$	$\hat{\sigma}_{74}$	$\hat{\sigma}_{75}$	$\hat{\sigma}_{76}$	$\hat{\sigma}_{77}$	$\hat{\sigma}_{78}$	$\hat{\sigma}_{79}$
$\hat{\sigma}_{81}$	$\hat{\sigma}_{82}$	$\hat{\sigma}_{83}$	$\hat{\sigma}_{84}$	$\hat{\sigma}_{85}$	$\hat{\sigma}_{86}$	$\hat{\sigma}_{87}$	$\hat{\sigma}_{88}$	$\hat{\sigma}_{89}$
$\hat{\sigma}_{91}$	$\hat{\sigma}_{92}$	$\hat{\sigma}_{93}$	$\hat{\sigma}_{94}$	$\hat{\sigma}_{95}$	$\hat{\sigma}_{96}$	$\hat{\sigma}_{97}$	$\hat{\sigma}_{98}$	$\hat{\sigma}_{99}$

Estimates in blue represent numbers that would be used to make graph in the E-W direction. Estimates in red represent numbers that would be used to make graph in the S-N direction. This suggests that using the "second main diagonal" to plot witch hat graphs at $w(-1) = w(1)$ will result in a very different value.

10. Figures 3 and 4 are positioned before section 4 in the manuscript, but the figures rely on information about the climate model data from that section (e.g. the estimates are from 15 samples). Please cross reference the material from section 4 where needed to avoid confusion.

    Thank you for this suggestion.

11. In the last paragraph on page 11, the authors state that the only meaningful correlations are of TREFHT with PSL and PRECT with PSL. However, if TREFHT and PRECT are individually correlated with PSL, shouldn't TREFHT and PRECT also be correlated to each other? Moreover, there is a contradiction between the physical explanation on lines 23-25, page 11 and the sign of the correlation between PSL and PRECT (low pressure systems increase precipitation).

    To address this question we created maps of the grid point correlations between JJA mean 2m air temperature (TREFHT), sea level pressure (PSL), and precipitation (PRECT) with sea level pressure (PSL) (Figure 2). What is clear between all these figures is that there is a lot of structure to all these maps. The sign of the correlation is regionally dependent. Therefore providing a mechanistic explanation of the spatially averaged correlation is not going to be particularly meaningful. However it may be useful for readers to know that there is a lot of structure to these maps and that the reason that the spatially averaged correlation between PSL and PRECT is so small is not because local correlations are small. Rather the average includes regions of large negative correlations as well as regions with large positive correlations. Despite losing this regional information in the S matrix summary of field covariances, this does not affect GMRF estimated field covariances between these fields as can be seen within the 'witch hat' graphs.

12. The model simulations use prescribed sea surface temperatures, which strongly constrain the near surface air temperature, so it seems surprising that the biggest changes in cost are associated with the 2-m air temperature. Can the authors provide a physical explanation for their finding?

    By our read of manuscript Figure 6, cost changes related to 2m air temperature (TREFHT) are the smallest relative to the three other fields. It is true that specifying sea surface temperatures will limit the model's TREFHT response over the ocean, however the models response to changes in parameters can affect the atmospheric boundary layer over the ocean including TREFHT. Moreover TREFHT is less restricted over land which was an important fact explaining why TREFHT showed the biggest qualitative differences in cost when using the Q matrix to include space dependencies within the cost. The latter is due to the sensitivity of the Q operator to the sharp spatial structures that arise from model-observational differences in and around the poorly resolved Andes mountains.

13. The authors find that spatial dependencies are more important to capture than field dependencies for the four selected outputs (PSL, TREFHT, U, PRECT). Do they have any reason to suspect (or can they show) that the field dependencies will dominate over spatial dependencies for other fields? If not, then this suggests that it may not be critical to capture the field dependencies and that their method does not offer many clear benefits over standard model assessment techniques. From their example in figure 5 of optimizing model performance by changing two parameters (c0 and Ke), it even looks like adding the spatial dependence alone would not greatly affect the conclusions drawn from assuming spatial independence (i.e., that high values of c0 and low values of Ke are best).

    While the current results do not provide a strong case for why including field dependencies is important, we were only looking at four fields within the tropics in JJA. We don't yet know whether field dependencies become important for other fields, regions, or seasons. We elected to keep the scope of the present manuscript focused on the mathematical treatment of GMRF. It is helpful to know that the results look reasonable which would be hard to evaluate without this limited example. We are in the process of generating results for 11 fields, 3 regions, and 4 seasons which will be reported separately. It also could be that the importance of field dependencies may depend on what parameters are being varied.

14. Observational uncertainty does not appear to be taken into account in their method. Can the authors comment on and suggest ways to incorporate observational uncertainty into their test statistic?

    This is an excellent and important question. The most obvious place to include this information is within the S matrix. For instance if a grid point error variance is known, it could be added to the diagonal elements. However we have already run into a case where satellite observations of cloud fields include linear structures that are obviously related to the satellite tracks. We suspect that the Q operator within GMRF may be particularly sensitive to these artifacts in the data. We do not have a full answer to this question. There is not much experience in the community as a whole for quantifying these uncertainties and representing them within metrics of climate model performance.

Response to Reviewer # 2 Reviewer comment given in blue.
Correlations in space and across variables were handled by Gaussian Markov Random Fields (GMRFs). I had a hard time understanding whether this is an appropriate technique or whether it was implemented correctly. In particular:

1. Equation 1 is introduced in the introduction and is said to be the culmination of the subsequent derivations but is never fully explained. Better explanation is needed. In particular, I dont think it makes sense to provide this equation in the introduction.

We will remove the specific details about GMRF, including equation (1) to a subsequent section.

2. I think Eq. 1 is a log-likelihood function derived from assuming model errors follow a multivariate Gaussian distribution (eq. 2) with the inverse covariance matrix Σ^{-1} replaced by GMRF precision matrix. These points need clarification and the reasonableness of assuming a multivariate Gaussian distribution for model output and for approximating the covariance matrix with a GMRF precision matrix both require further justification.

   Currently climate model evaluation makes use of relatively short, few year model integrations for testing the effects of uncertain parameters. When compared to the 10 to 30-year climatologies of observations the distribution of errors is approximately Gaussian (see Figure 3). The distribution of climate fields on very short, hourly to daily time scales can be decidedly non-Gaussian which is not the case for longer term means. We include here a few examples of monthly mean climate model output that show an approximately Gaussian distribution. The reasonableness for using GMRF to estimate the inverse covariance matrix is provided by the 'witch-hat' graphs.

3. I think the log-likelihood function in eq 1 is missing the following term: $ln((2\pi)^{-n/2}tr(\Sigma)^{-1/2})$. Is this true? In any case, this derivation needs to be more clear.

   Yes this is true, although it would include the determinant of the covariance matrix not its trace. Within the statistics community the argument of the likelihood function is referred to as the log-likelihood since the factor you are referring to is a constant offset. However this can be confusing especially since this community often also neglects the factor of $\frac{1}{2}$ that should be included within the exponential argument for a Gaussian distribution. We will make our statement more clear.

4. The precision matrix is only described for the 2x2 case. Are the "rules" on page 6 followed only for the 2x2 case, or are they followed for all cases?

   Yes, the rules apply to all cases.

5. Because Q seems to be defined independently of the spatial autocorrelation in the actual data, I find it hard to believe that it can be a good approximation for Σ^{-1} except by chance. In particular, I bet the "witch hat graph" for surface precipitation alone (which has short autocorrelation length scales) looks very different than that for surface temperature (which has long autocorrelation length scales) and that Fig. 4 only looks reasonable for quantities which happen to have the autocorrelation structure matching the precision matrix assumptions. I would like to see the comparison between Q^{-1} and Σ (note Im asking for things in correlation-matrix

space rather than precision matrix space because the former is easier to interpret physically) for several different output fields to gain confidence in the method. The fact that Q is defined independently of autocorrelation in the actual data is my single biggest concern with this paper.

You are correct that the Q matrix is defined independently of field information. This matrix is a differential operator which 'senses' how much fields change within the neighborhood structure. The units come from scaling the Q matrix with S^{-1} using the Kronecker product. Together the GMRF provides a decent approximation to the inverse covariance matrix Σ^{-1}. 'Witch-hat' graphs show observed variances and covariances with the inverse precisions (implied variances/covariances) and are being used to test how well GMRF capture observed space and field dependencies. Figures 4 and 5 show several additional 'witch-hat' graphs to provide a more complete evaluation of GMRFs representation of observed space and field dependencies.

6. The fact that the precision matrix has a zero eigenvalue seems to be an obvious result of the fact that Q indicates the neighbors of each cell and neighbors of the last cell can be predicted from the others (because the cells which are its neighbors have already tagged it as being their neighbor). I am surprised and alarmed that your solution to this problem is to add a small perturbation to make your singular matrix merely nearly-singular. It seems like this nearly-singular matrix will at best have numerical issues and at worst isn't actually solving the system you meant to solve. Wouldnt it make more sense to replace the system with a matrix of 1 lower dimension?

   The issue is that Q matrix needs to have the same dimension as each field so changing its dimension is not the solution. Note that we only need to take an inverse of the Q matrix for testing GMRF predictions of observed covariances within the 'witch-hat' graphs. Even then the R codes provided robust results.

Other Comments:

1. In the title and elsewhere, you call your method a "metric". I think the benefit of your approach is that it allows you to evaluate a log-likelihood function in order to choose the best parameter settings to an uncertainty-quantification problem. While the log-likelihood function does give you a scalar value for a particular set of parameters and is therefore a metric of sorts, I think emphasizing that youre defining a metric is kind of missing the main point of what you're doing. In particular, you have to specify exactly what output you want to use to define a metric and I think a benefit of your method is that it should work on a wide variety of output data choices. In short, Id suggest changing "metric" to "method" throughout the text.

We appreciate your thoughts on this matter and agree with your point. The term "metric" is often used for climate model evaluation although that term does not capture the fact that we are evaluating a signal to noise ratio for testing the null hypothesis of whether changes in a climate model are significant. In other publications we sometimes refer to this normalized metric as a "test statistic". We therefor prefer using that term over "method". We thank you for this suggestion.

2. using CAM3.1 is odd and detracts from the publication-worthiness of the paper because it is an ancient model which nobody cares about anymore. Can you really not find data from more recent model runs? It would be worth the effort.

   CAM3.1 output is more than adequate for purposes of examining how GMRF would be applied to climate model evaluation.

3. In eq. 2, you need to indicate that $|x|$ is the determinant of x.

   We can indicate this.

4. p. 6 line 8: "fuller" should be "more full"

   Thanks for catching this.

5. You should define what the Kronecker product is for climate people, who may not know off the top of their heads.

   We can provide the following example of how the Kronecker product works: Consider the following 2×2 matrices

$$A = \begin{pmatrix} 1 & 4 \\ 2 & 5 \end{pmatrix} \text{ and } B = \begin{pmatrix} 1 & 3 \\ 0 & 4 \end{pmatrix}.$$

   The Kronecker product of A and B, $A \otimes B$, is given by

$$A \otimes B = \begin{pmatrix} 1(B) & 4(B) \\ 2(B) & 5(B) \end{pmatrix} = \begin{pmatrix} 1 & 3 & 4 & 12 \\ 0 & 4 & 0 & 16 \\ 2 & 6 & 5 & 15 \\ 0 & 8 & 0 & 20 \end{pmatrix}$$

6. p. 7, line 18: youre missing a word between supplemental and carries.

   Thanks for catching this.

7. p. 11 line 25: low pressure cooling the underlying surface would be a *positive* correlation. Perhaps youre seeing a "thermal low" effect?

See also response to item 12 from reviewer # 1. Because covariances vary by region (e.g. Figure 2), we will back off from providing particular explanations for explaining covariances that may only apply to particular regions of the domain.

8. You show in Fig. 5 that using GMRF or not doesnt make a big difference. Is this the result of your particular choice of parameters and/or model version and/or output variables? Taking the time to test your method in other cases would at a minimum make your conclusions more robust and could potentially show that your method has an important impact in certain circumstances.

   While we agree that testing GMRF in all possible cases (with more fields, regions, seasons, and model parameters) would provide a more thorough examination of the question you raise of whether the GMRF can make an important difference, our purpose for the present manuscript was to develop the mathematical application of GMRF to climate model output. The effort represents several years of concerted effort. The testing of GMRF in more cases is being developed with more scientific goals in mind.

9. In the supplementary material, why assume x has means which are all zero?

   It was not a necessary assumption, but it did facilitate the derivation without complicating the expressions.

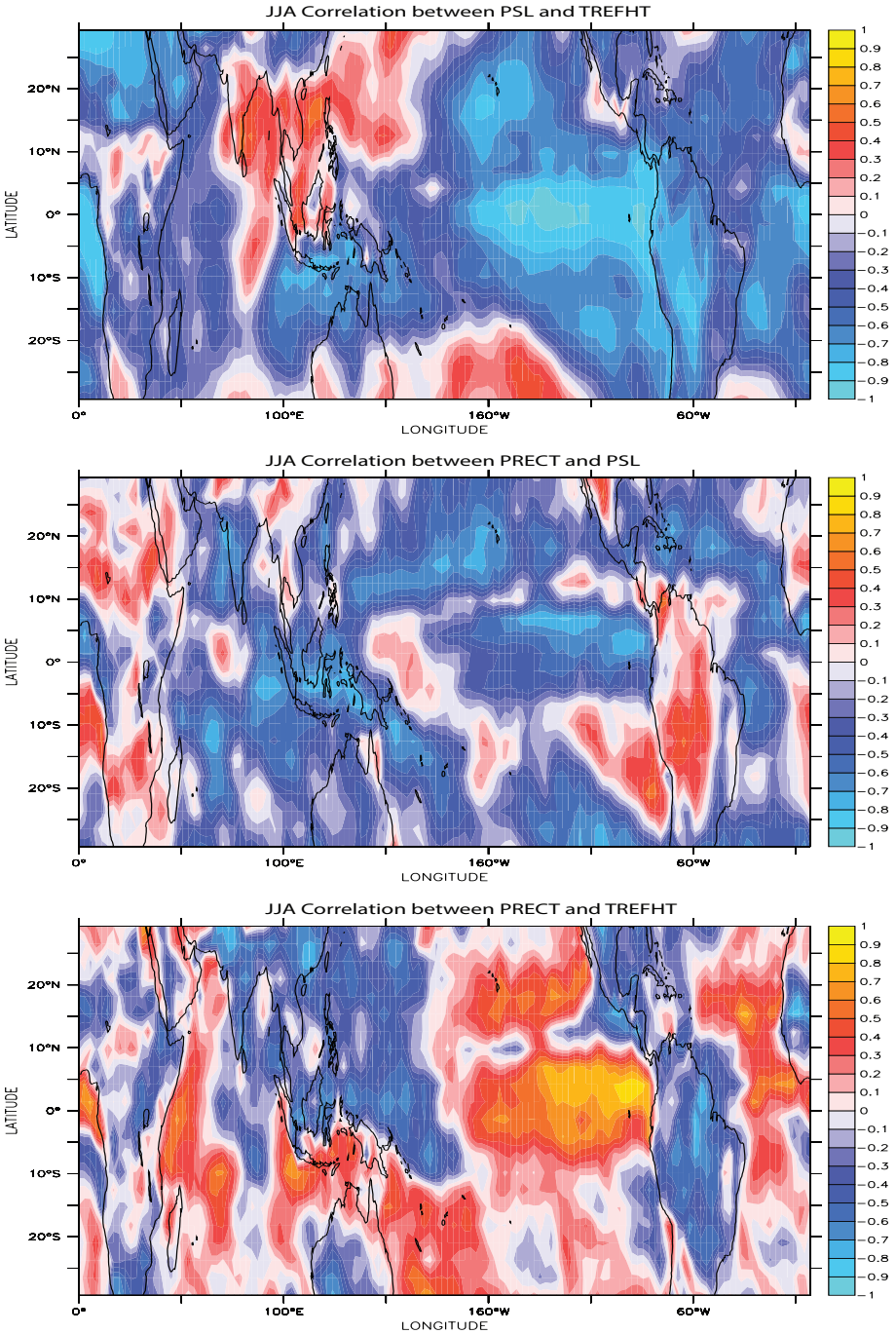

Figure 2: JJA correlations between 2m air temperature (TREFHT), sea level pressure (PSL), and precipitation (PRECT).

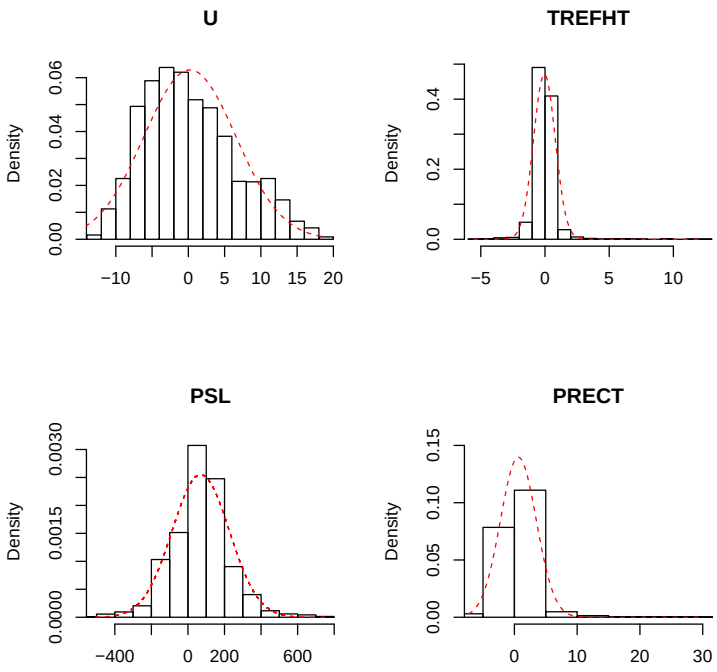

Figure 3: Histograms of differences between observations and model output for four fields (U, TREFHT, PSL, and PRECT) for an experiment that includes changes to cloud parameters C0 and ke.

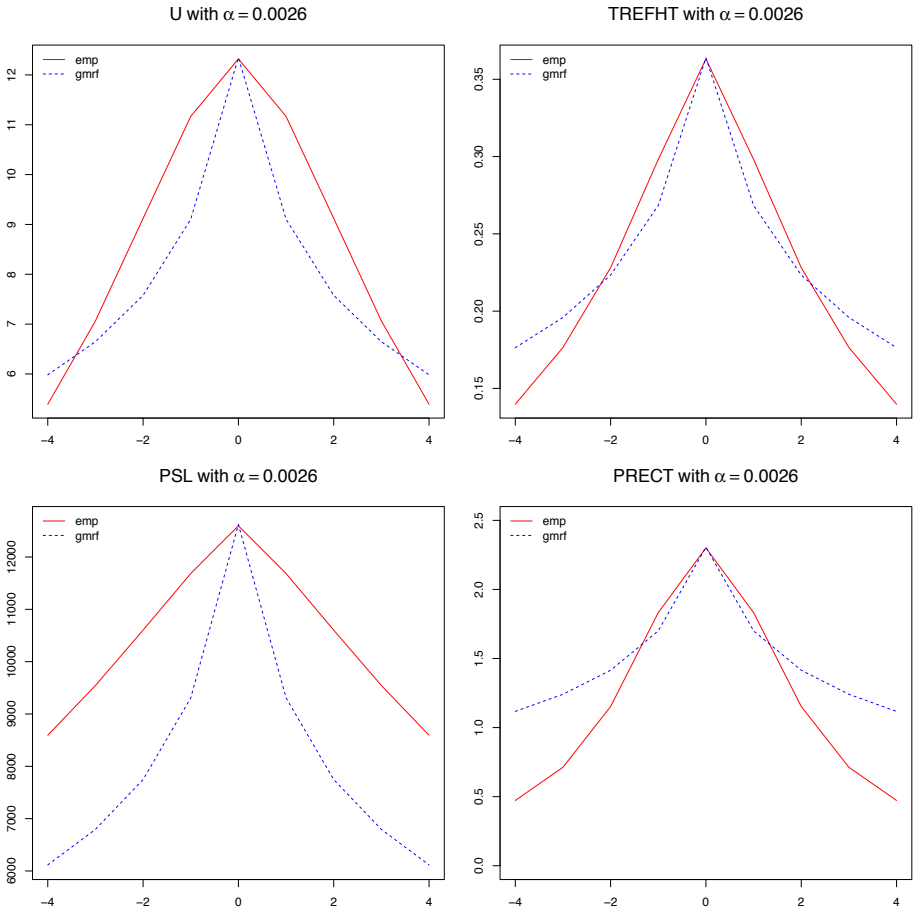

Figure 4: 'Witch hat' graphs testing GMRF approximations to empirical estimates of variances of U, PSL, TREFHT, and PRECT.

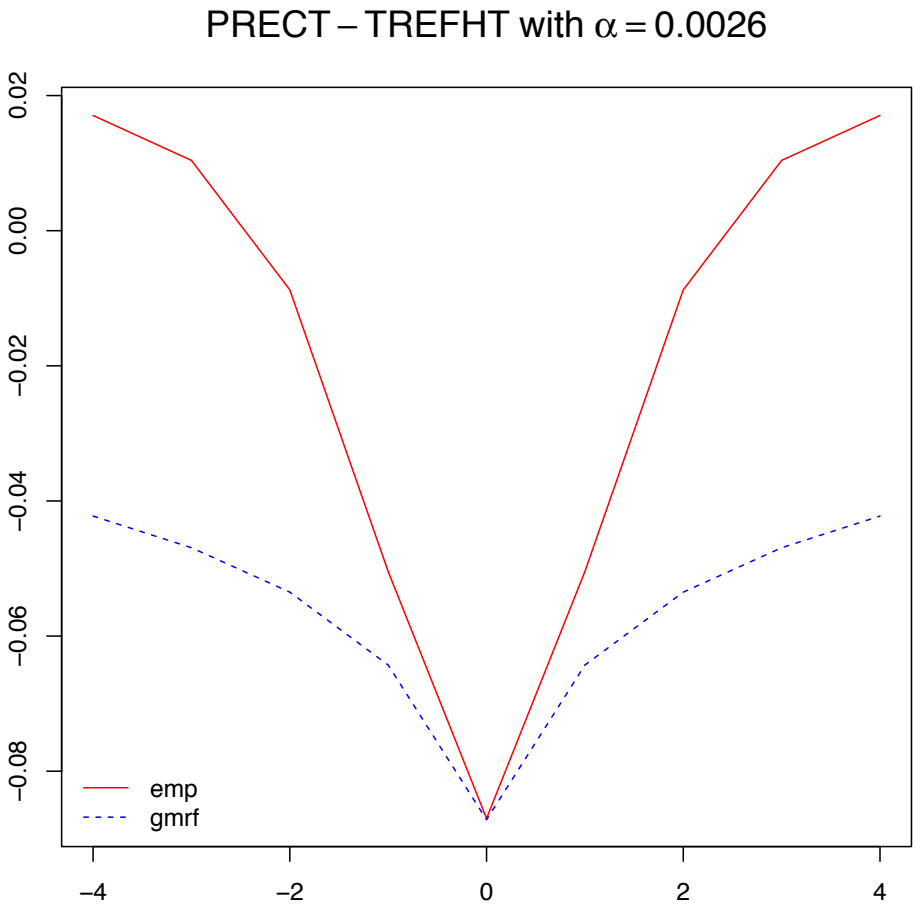

Figure 5: 'Witch hat' graphs testing GMRF approximations to empirical estimates of covariances between TREFHT and PRECT.

Manuscript prepared for Geosci. Model Dev.
with version 2015/04/24 7.83 Copernicus papers of the LaTeX class copernicus.cls.
Date: 15 May 2016

A new ~~metric~~ test statistic for climate models that includes field and spatial dependencies using Gaussian Markov Random Fields

Alvaro Nosedal-Sanchez[1,2], Charles S. Jackson[3], and Gabriel Huerta[1]

[1]Department of Mathematics and Statistics, The University of New Mexico, Albuquerque
[2]Department of Mathematical and Computational Sciences, University of Toronto, Mississauga
[3]Institute for Geophysics, The University of Texas at Austin, Austin

Correspondence to: Charles Jackson (charles@ig.utexas.edu)

Abstract. A new ~~metric~~ test statistic for climate model evaluation has been developed that potentially mitigates some of the limitations that exist for observing and representing field and space dependencies of climate phenomena. Traditionally such dependencies have been ignored when climate models have been evaluated against observational data, which makes it difficult to assess whether any given model is simulating observed climate for the right reasons. The new ~~metric~~ statistic uses Gaussian Markov Random Fields for estimating field and space dependencies within a first order grid point neighborhood structure. We illustrate the ability of Gaussian Markov Random Fields to represent empirical estimates of field and space covariances using 'witch hat' graphs. We further use the new ~~metric~~ statistic to evaluate the tropical response of a climate model (CAM3.1) to changes in two parameters important to its representation of cloud and precipitation physics. Overall, the inclusion of dependency information did not alter significantly the recognition of those regions of parameter space that best approximated observations. However there were some qualitative differences in the shape of the response surface that suggest how such a measure could affect estimates of model uncertainty.

1 Introduction

~~Within the climate assessment community, there is an interest to develop metrics of how well~~ Climate scientists are interested in developing new metrics for assessing how well climate simulations reproduce observed climate for purposes of comparing models, driving model development, and evaluating model prediction uncertainties ~~(Gleckler et al., 2008; Reichler and Kim, 2008; Santer et al., 2009; Knutti et al., 2010; Weigel et~~ ~~Nevertheless, a certain~~ (Gleckler et al., 2008; Reichler and Kim, 2008; Santer et al., 2009; Knutti et al., 2010; Weigel et al., 2010; B Formal methods for accomplishing these goals, such as Bayesian calibration, operate with a single test statistic[1] for determining likelihood measures of different model configurations. A level of skepticism exists ~~about whether a scalar metric can be sufficiently informative for these purposes~~within

[1] A test statistic is a metric that includes information about the significance of modeling errors.

the climate assessment community concerning the sufficiency of any one metric to judge a climate model's scientific credibility. Climate phenomena involve interactions of multiple ~~quantities~~ fields (observables) on a wide range of time and space scales from minutes to decades (and longer) and from meters to planetary scales. Thus ~~it can be challenging to summarize what is physically meaningful.~~ there are plenty of challenges that exist for synthesizing the many ways that a climate model can be tested against observational data.

The most common approach to climate model evaluation among climate scientists is to display maps of long-term means of well-known ~~quantities~~ fields (e.g. temperature, sea-level pressure, precipitation) whose distribution is familiar and well understood in order to identify sources of model error. ~~The Taylor metric that is~~ Taylor metrics that are often generated as part of model evaluation ~~is~~ are based on spatial means of squared grid point errors for individual ~~quantities~~ fields (Taylor, 2001). Such measures neglect field and space dependencies ~~and thus may be insensitive to mechanisms giving rise to model errors. There is a need to develop metrics that can evaluate whether a model is capturing observed spaceand field relationships sufficiently well (Braverman et al., 2011) . The hope is that by accounting for relationship information within climate model metrics, they will prove to be more useful for scientific evaluation~~that arise as a consequence of how the physics of the climate system correlate multiple quantities in space. Neglecting these dependencies therefore ignores additional information that can be used to test whether models are simulating observables for the right reasons.

~~Given that there is only a~~ Here we present a new test statistic based on Gaussian Markov Random Fields (GMRFs) that addresses some of the challenges that currently exist for estimating the significance of modeling errors across multiple fields that takes into account field and space dependencies that exist within observations. Perhaps one of the under-recognized challenges in this regard is the limited amount of observations available to quantify ~~field and space relationships of climate phenomena, data assimilation is the most common approach~~ dependencies. Data assimilation is commonly used to fill in gaps in the observational record ~~of a climate model~~ (Trenberth et al., 2008). While assimilation ~~data products help solve~~ products help address some aspects of the problem of how one compares point measurements to the scales resolved by climate models, these ~~data~~ products include the space and field dependencies of the model that was used to assimilate ~~the data. Here we introduce a new kind of metric based on~~ observations. The imprint of the reanalysis model is readily seen when comparing two or more assimilation products, particularly quantities that are directly related to parameterized physics such as precipitation and radiation. One of the advantages of Gaussian Markov Random Fields ~~that only needs limited~~ (GMRFs) is that it only needs a limited amount of data to decipher space and field dependencies of climate phenomena. This is because GMRFs summarize relationship information as it is expressed across fields of gridded data.

The present application of GMRFs operates on long-term means. While it may be possible to extend GMRFs to capture time dependencies (Cressie and Wikle, 2011) , the present application represents an advance over more traditional metrics.

We define a new Z-test statistic, alternatively referred to as a log-likelihood or cost for assessing the significance of a discrepancy between model output and observations. The statistic makes use of Gaussian Markov Random Fields to estimate field and space dependencies that exist within gridded climate model output that can be assessed against space and field dependent observational data. The matrix form of the test statistic is given by:

$$\mathbf{v}^T \mathbf{S}^{-1} \otimes (\alpha \mathbf{I} + (1-\alpha)\mathbf{Q})\mathbf{v}$$

where $\mathbf{v}$ is the vector of differences between model output and observations with a length given by the product of the number of observational fields and number of grid points, $n_{obs} \times n_{pts}$, α is a scalar with a value close to zero, $\mathbf{I}$ stands for an identity matrix (a diagonal matrix of ones) of a dimension corresponding to $\mathbf{v}$, and $\mathbf{Q}$ is a precision matrix of dimension $n_{pts} \times n_{pts}$ from a Gaussian Markov Random Field (GMRF) induced by a first order neighborhood structure. This cost function captures field dependenciesthrough $\mathbf{S}^{-1}$ which is a matrix of dimension $n_{obs} \times n_{obs}$ where each of its elements represents a spatial-average of grid point variances and covariances between fields. The spatial dependency between grids is approximated through $\mathbf{Q}$. The quantity α could be interpreted as a weight of the spatial relationship between grid cells. The Kronecker product $\otimes$ provides a means for associating the different matrix dimensions of the metric, essentially combining its field and space components.

The sections of this paper explain, test, and provide examples of how various components of equation (2) GMRF work. Section 2 gives a brief introduction to GMRFs . This section will allow us to understand how and the use of a neighborhood structure for estimating dependency information using a precision operator $\mathbf{Q}$is obtained and the information that it provides about spatial dependency between grid cells. In this section we also define and discuss Kronecker products, and how the Kronecker product and how it is used to generalize GMRFs to use this concept to generalize GMRF ideas to deal with more than one field. Section 3 introduces a graph for testing the extent to which equation (2) captures GMRFs represent observed variance-covariances of tropical temperature, precipitation, sea level pressure, and upper level winds. Finally, in Section 4, we consider the field and space dependencies that are captured by the GMRF-based metric within the response of an atmospheric general circulation model CAM3.1 to two model parameters important to cloud and precipitation physics. What we learned in general is that including the space and field dependencies provides some qualitatively different perspectives about which model configurations are more similar to what is observed. For the example we consider, the effects of space dependencies turn out to be more critical than field dependencies.

2 Gaussian Markov Random Fields (GMRFs)

A Gaussian Markov Random Field (GMRF) is a special case of a multivariate normal distribution~~,~~ ~~one that satisfies additional properties related to conditional independence~~. The density of a normal random vector $\mathbf{x} = (x_1, x_2, ..., x_n)^T$ (where T denotes the operation of transposing a column to a row), with mean μ ($n \times 1$ vector) and covariance matrix $\mathbf{\Sigma}$ ($n \times n$ matrix), is

$$f(\mathbf{x}) = (2\pi)^{-n/2} |\mathbf{\Sigma}|^{-\frac{1}{2}} exp\left\{ -\frac{1}{2} (\mathbf{x} - \mu)^T \mathbf{\Sigma}^{-1} (\mathbf{x} - \mu) \right\} \quad (1)$$

Here, $\mu_i = E(x_i)$, $\Sigma_{ij} = Cov(x_i, x_j)$, ~~and~~ $\Sigma_{ii} = Var(x_i) > 0$~~.~~, and $|\mathbf{\Sigma}|$ is the determinant of $\mathbf{\Sigma}$. Estimating $\mathbf{\Sigma}$ can be quite challenging in many contexts, especially for climate models where there is only limited data. All eigenvalues of $\mathbf{\Sigma}$ must be greater than zero, otherwise ~~$\mathbf{\Sigma}$~~ $\mathbf{\Sigma}^{-1}$ becomes a singular matrix and ~~does~~ it does not define a valid multivariate normal distribution. It can also be shown that if all eigenvalues of $\mathbf{\Sigma}$ are positive then all eigenvalues of $\mathbf{\Sigma}^{-1}$ are also greater than zero. ~~We define $\mathbf{Q} = \mathbf{\Sigma}^{-1}$ and refer to $\mathbf{Q}$ as~~ Rather than estimating $\mathbf{\Sigma}$ and ensuring all eigenvalues of $\mathbf{\Sigma}^{-1}$ are positive, GMRFs makes use of the precision matrix ~~, and denote $\mathbf{x} \sim \mathbf{N}(\mu, \mathbf{Q})$~~ $\mathbf{P} = \mathbf{\Sigma}^{-1}$. We denote $\mathbf{x} \sim \mathbf{N}(\mu, \mathbf{P})$ to represent $\mathbf{x}$ as a multivariate normal distribution with vector mean μ and precision matrix ~~$\mathbf{Q}$.~~ $\mathbf{P}$. GMRFs approximate $f(\mathbf{x})$ using a sparse representation for $\mathbf{P}$ by setting all precisions outside a neighborhood structure to zero. Thus GMRFs make the assumption that points outside a neighborhood structure are conditionally independent. As we shall show below, this limitation does not prevent GMRFs from capturing covariances outside the neighborhood structure used to define precisions.

2.1 ~~Precision matrix of a GMRF~~

~~The precision matrix $\mathbf{Q}$ is an operator for obtaining information about dependencies among neighboring grid~~ The GMRF-based expression that we have developed for quantifying the significance of differences between model output and observations is

$$\mathbf{v}^T \mathbf{S}^{-1} \otimes (\alpha \mathbf{I} + (1 - \alpha) \mathbf{Q}) \mathbf{v} \quad (2)$$

where $\mathbf{v}$ is the vector of differences between model output and observations with a length given by the product of the number of observational fields and number of grid points, $n_{obs} n_{pts}$, α is a scalar with a value close to zero, $\mathbf{I}$ stands for an identity matrix (a diagonal matrix of ones) of dimension n_{pts} corresponding to $\mathbf{v}$, and $\mathbf{Q}$ is a precision operator of dimension $n_{pts} \times n_{pts}$ from a Gaussian Markov Random Field (GMRF) induced by a first order neighborhood structure. This cost function captures field dependencies through $\mathbf{S}^{-1}$ which is a matrix of dimension $n_{obs} \times n_{obs}$ where each of its elements represents a spatial-average of grid point variances and covariances between fields. The spatial dependency between grids is approximated through $\mathbf{Q}$. The quantity α could be interpreted as a weight of the spatial relationship between grid cells. The Kronecker product $\otimes$ provides a means for associating the different matrix dimensions of the metric, essentially combining its field

and space components. Each of the following subsections provides additional information about the derivation and application of equation (2).

2.1 Precision operator of a GMRF

The precision operator of a GMRF $\mathbf{Q}$ provides a way to estimate dependencies among neighboring grid cells. ~~Although $\mathbf{Q}$ is sparse, its inverse, as a model for the covariance matrix $\mathbf{\Sigma}$, presumes all grid points are conditionally dependent.~~ $\mathbf{Q}$ needs to be constructed such that it:

- Reflects the kind of spatial dependency we assume our data has.
- Yields a legitimate covariance matrix, $\mathbf{\Sigma}$, i.e. symmetric and positive definite, so that it can be used to compute a likelihood function.

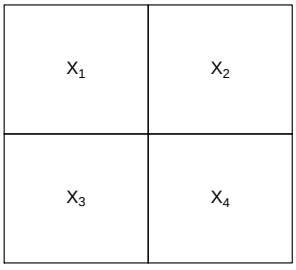

Figure 1. Graphical representation of 2×2 lattice and elements of $\mathbf{x}$.

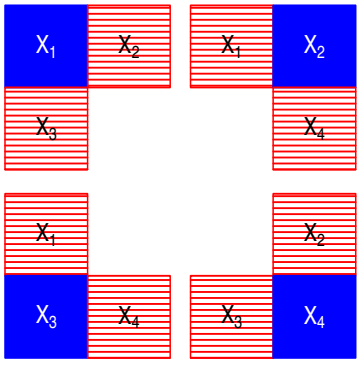

Figure 2. Neighbors of x_1, x_2, x_3 and x_4

Consider $\mathbf{x}$, a vector of measurements on a 2×2 lattice, as represented in Figure 1. Assume a neighborhood structure between the four elements of $\mathbf{x}$. In Figure 2, the neighbors for each element

of $\mathbf{x}$ are defined graphically. Given the neighborhood structure shown in Figure 2, the precision matrix that works for this problem is

$$\mathbf{Q} = \begin{pmatrix} 2 & -1 & -1 & 0 \\ -1 & 2 & 0 & -1 \\ -1 & 0 & 2 & -1 \\ 0 & -1 & -1 & 2 \end{pmatrix}$$

which follows these rules,

- $\mathbf{Q}_{ij} = -1$, if x_i and x_j are neighbors.
- $\mathbf{Q}_{ij} = 0$, if x_i and x_j are not neighbors.
- $\mathbf{Q}_{ii}$ gives the total number of neighbors of x_i.

While the implementation of GMRFs is simple, the theory and mathematics are rather involved. A ~~fuller~~ more full description of the mathematics of this example is provided in the supplemental material. It may also not be immediately clear to a physical scientist that such a simple specification, where only relationships among neighboring grid cells are taken into account, would be sufficient to quantify correlated quantities across large distances. The mathematics of working with precisions allows one to infer the net effect of long distance relationships through relationship information that exists among neighboring cells. While the GMRF approach does not include information about particular teleconnection structures such as ENSO, the approach is sensitive to how changes in large scale conditions induce local covariances across multiple fields within the entire domain. In this way teleconnections are represented through a conditional dependence.

A problem arises in that one of the eigenvalues of the $\mathbf{Q}$ matrix is 0, which implies that this definition of the precision matrix does not induce an invertible covariance matrix. This problem is solved by using $\alpha\mathbf{I} + (1-\alpha)\mathbf{Q}$, instead of $\mathbf{Q}$. If α is small, the neighborhood structure remains essentially unchanged. Section 3 describes our approach to ~~specifying~~ specify a value for α.

2.2 Generalizing concepts to deal with multiple fields

The generalization of $\mathbf{Q}$ to handle multiple fields ~~will be illustrated by an example using~~ involves a Kronecker product ($\otimes$) between $\mathbf{S}^{-1}$ and $\mathbf{Q}$. For reference, a Kronecker product of $A \otimes B$ where

$$A = \begin{pmatrix} 1 & 4 \\ 2 & 5 \end{pmatrix} \text{ and } B = \begin{pmatrix} 1 & 3 \\ 0 & 4 \end{pmatrix}$$

is given by

$$A \otimes B = \begin{pmatrix} 1(B) & 4(B) \\ 2(B) & 5(B) \end{pmatrix} = \begin{pmatrix} 1 & 3 & 4 & 12 \\ 0 & 4 & 0 & 16 \\ 2 & 6 & 5 & 15 \\ 0 & 8 & 0 & 20 \end{pmatrix}.$$

Consider $\mathbf{x}$ and $\mathbf{y}$ which represent observations for two different fields of interest ~~. These observations are taken~~ on a 2×2 lattice. First, $\mathbf{x}$ and $\mathbf{y}$ are combined to form one vector $\mathbf{v}$ as follows: $\mathbf{v}^T = (x_1, x_2, x_3, x_4, y_1, y_2, y_3, y_4)$. The average covariances among these observations can be represented by a 2×2 matrix between the first field, $\mathbf{x}$, and the second field, $\mathbf{y}$:

$$\mathbf{S} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

~~$\mathbf{S} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$~~

where $Var(\mathbf{x}) = \sigma_{11}$, $Var(\mathbf{y}) = \sigma_{22}$, and $Cov(\mathbf{x}, \mathbf{y}) = \sigma_{12}$. Recalling that the correlation between fields 1 and 2 is defined as: $\rho = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$, one ~~may~~ can show that the inverse of $\mathbf{S}$ is

$$\mathbf{S}^{-1} = \begin{pmatrix} \frac{1}{\sigma_{11}(1-\rho^2)} & \frac{-\rho}{(1-\rho^2)\sqrt{\sigma_{11}\sigma_{22}}} \\ \frac{-\rho}{(1-\rho^2)\sqrt{\sigma_{11}\sigma_{22}}} & \frac{1}{\sigma_{22}(1-\rho^2)} \end{pmatrix} = \begin{pmatrix} S_{11}^{-1} & S_{12}^{-1} \\ S_{21}^{-1} & S_{22}^{-1}. \end{pmatrix}$$

If we consider the Kronecker product in ~~Equation 1~~ equation (2) when $\alpha = 0$,

$$\mathbf{S}^{-1} \otimes \mathbf{Q} = \begin{pmatrix} S_{11}^{-1}\mathbf{Q} & S_{12}^{-1}\mathbf{Q} \\ S_{21}^{-1}\mathbf{Q} & S_{22}^{-1}\mathbf{Q} \end{pmatrix}$$

then

$$\mathbf{v}^T\mathbf{S}^{-1} \otimes \mathbf{Q}\mathbf{v} = S_{11}^{-1}\mathbf{x}^T\mathbf{Q}\mathbf{x} + S_{12}^{-1}\mathbf{y}^T\mathbf{Q}\mathbf{x} + S_{21}^{-1}\mathbf{x}^T\mathbf{Q}\mathbf{y} + S_{22}^{-1}\mathbf{y}^T\mathbf{Q}\mathbf{y}.$$

In this last expression, one can see that the inverse of $\mathbf{S}$ in combination with the Kronecker product with $\mathbf{Q}$ includes terms involving cross products between fields. The supplemental materials carries this expression one step further by estimating the conditional mean for the the first element of $\mathbf{v}$ to illustrate how this element is related to itself and its neighbors across multiple fields.

3 A test of GMRF estimates of variance

GMRFs provide a way to approximate field and space dependencies contained in the inverse covariance matrix $\mathbf{\Sigma}^{-1}$ of equation~~(1~~ (1) by its GMRF equivalent $\mathbf{S}^{-1} \otimes (\alpha\mathbf{I} + (1-\alpha)\mathbf{Q})$. In this section, we will test how well ~~GMRF~~ GMRFs are able to reproduce observed space and field dependencies.

This may be achieved by comparing field and spatial variance and covariance estimates obtained from the inverse of the GMRF ~~equation~~ estimate of the precision matrix with those obtained empirically from observational data. It turns out this comparison is sensitive to the value that is selected for α. ~~Fortunately~~By construction, the optimal choice of α depends only on geometric considerations of the neighborhood model that is used for GMRF and the number of grid points in the fields and not the properties of the field data. We introduce a 'witch hat' graph that provides a compact summary of variance-covariance information between these two methods in order to show that GMRFs do a reasonable job approximating observed field and space relationships.

3.1 Finding an appropriate value of α

In the effort to compare space and field dependencies approximated by GMRF with empirical estimates we need to determine an optimal value for α. In order to carry out this comparison, we need to find the inverse of $\mathbf{S}^{-1} \otimes (\alpha \mathbf{I} + (1-\alpha)\mathbf{Q})$, our proposed precision matrix based on GMRF. Using results of Kronecker products, we have that $\left[\mathbf{S}^{-1} \otimes (\alpha \mathbf{I} + (1-\alpha)\mathbf{Q})\right]^{-1} = \mathbf{S} \otimes (\alpha \mathbf{I} + (1-\alpha)\mathbf{Q})^{-1}$. Letting $\mathbf{Q}^* = (\alpha \mathbf{I} + (1-\alpha)\mathbf{Q})^{-1}$, then $\mathbf{S} \otimes \mathbf{Q}^*$ for two fields can be written as

$$\begin{pmatrix} S_{11}\mathbf{Q}^* & S_{12}\mathbf{Q}^* \\ S_{21}\mathbf{Q}^* & S_{22}\mathbf{Q}^* \end{pmatrix}.$$

If n is the total number of grid points of the lattice, $\mathbf{S} \otimes \mathbf{Q}^*$ is a ~~$(2 \times n) \times (2 \times n)$~~ $2n \times 2n$ covariance matrix. Note that each element of $diag(S_{ij}\mathbf{Q}^*)$ contains the estimated variance or covariance at each grid point for fields i and j using a GMRF where i can be equal to j. If we average these estimates across the whole lattice, we obtain G_{ij}, the GMRF estimate of the variance or covariance for fields i and j. Therefore,

$$G_{ij} = \frac{S_{ij} \sum_{k=1}^{n} Q^*_{kk}}{n} = \frac{S_{ij} tr(\mathbf{Q}^*)}{n} \tag{3}$$

where $tr(\mathbf{Q}^*)$ denotes the trace of $\mathbf{Q}^*$ and Q^*_{kk} are its diagonal elements. We will now select a value for α that allows the GMRF estimate for field variances and covariances to be equal, on average, to what has been calculated for $\mathbf{S}$. In order to achieve this, G_{ij} needs to equal S_{ij}. Satisfying this condition is equivalent to finding the solution for

$$\frac{tr(\mathbf{Q}^*)}{n} = 1. \tag{4}$$

It may not be so obvious what the diagonal elements of $\mathbf{Q}^*$ are. However, one can use the fact that for any matrix $\mathbf{A}$ that admits a Singular Value Decomposition, $tr(\mathbf{A})$ is equal to sum of its eigenvalues. In our case, if the eigenvalues of $\mathbf{Q}$ are $\lambda_1, \lambda_2, ..., \lambda_n$, the eigenvalues of $\alpha \mathbf{I} + (1-\alpha)\mathbf{Q}$ are $\alpha + (1-\alpha)\lambda_1, \alpha + (1-\alpha)\lambda_2, ..., \alpha + (1-\alpha)\lambda_n$. The eigenvalues of $\mathbf{Q}^* = (\alpha \mathbf{I} + (1-\alpha)\mathbf{Q})^{-1}$ are $(\alpha + (1-\alpha)\lambda_1)^{-1}, (\alpha + (1-\alpha)\lambda_2)^{-1}, ..., (\alpha + (1-\alpha)\lambda_n)^{-1}$. This implies that in order to satisfy

~~Equation 4~~equation (4), we need to find α from

$$f(\alpha) = \sum_{i=1}^{n} \frac{1}{n(\alpha + (1-\alpha)\lambda_i)} = 1. \tag{5}$$

Figure 3 shows the relationship between various values of α and $f(\alpha)$. The eigenvalues used to obtain this figure correspond to ~~a precision matrix~~the precision operator, $\mathbf{Q}$, for a GMRF induced by a first order neighborhood structure and considering a 128×22 lattice (which is the dimension of our data). From the figure we can see that the curve crosses the value of 1 when α is close to 0. By using linear interpolation, we determine that α is approximately 0.0026. Note that this value is independent of fields since equation (5) does not contain any field-specific information.

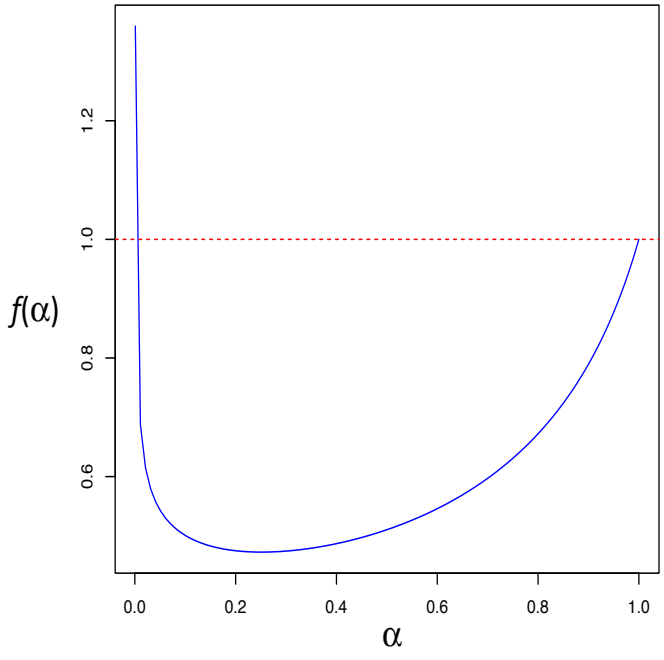

Figure 3. α vs $f(\alpha)$.

3.2 'Witch hat' comparison test

To illustrate any differences that may exist between empirical estimates of the covariance matrix $\boldsymbol{\Sigma}$ and its GMRF equivalent $\mathbf{S} \otimes (\alpha\mathbf{I} + (1-\alpha)\mathbf{Q})^{-1}$, we rely on a graph that shows the spatial average grid point variance and covariances as a function of distance for cells and their neighbors. We compute the average entries of the covariance matrix corresponding to each grid cell and the corresponding element to the north or east (for the positive distances) or to the south or west (for the negative distances) relative to the main diagonal of the matrix. The zero distance case is the

average of variances of the main diagonal. ~~Alternatively, we can produce a graph that considers the east and west directions~~The cells corresponding to one or more grid cells away are mostly on entries in parallel with the main diagonal. On average, covariances decrease with distance making the graph have the shape of a witch's hat. This graph is symmetric because covariance matrices are symmetric.

Figure 4 shows a ~~witch hat~~ 'witch hat' test of estimated variances for air temperatures simulated by the Community Atmosphere Model version 3.1 (CAM3.1). The variances are estimated from 15 samples of two year mean summertime temperatures. Setting $\alpha = 1$ provides a solution to equation (5), however, this will shut down the effect of $\mathbf{Q}$ and only the variances at the reference point (lag 0) will be well ~~estimated~~represented. On the other hand, when $\alpha = 0.0026$, we allow $\mathbf{Q}$ to play more of a role which results in a better representation of covariances at neighboring points (lags different of zero).

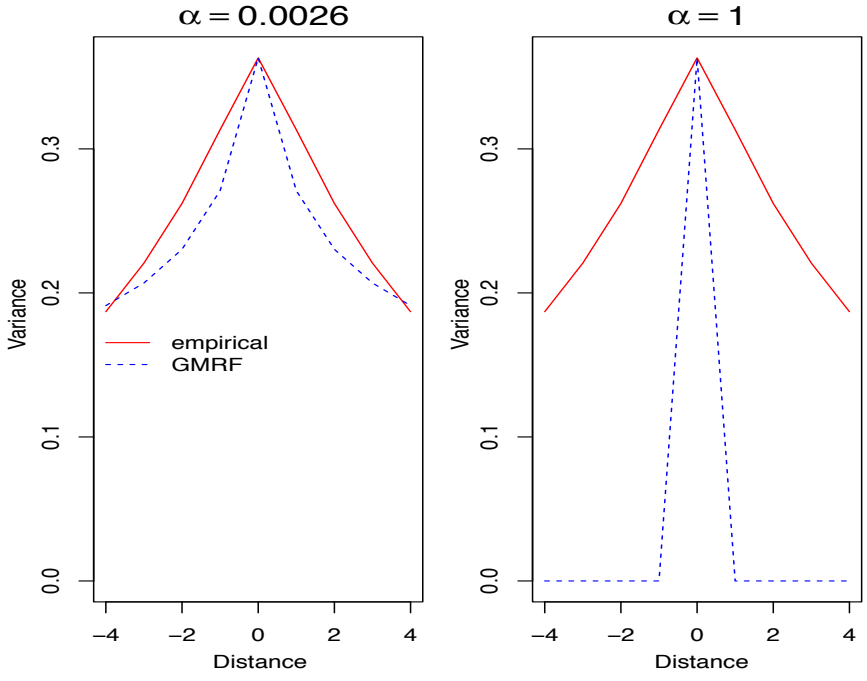

Figure 4. 'Witch hat' graphs for air temperature on a 128×22 lattice of the tropics from 30°S to 30°N. The empirical estimates are given by the solid red line. The GMRF estimate is given by the dashed blue line.

4 Climate response to uncertain parameters

In this section we show how inclusion of field and space dependencies using GMRF affect comparisons of the Community Atmosphere Model (CAM3.1) (Collins et al., 2006) with observations. We consider CAM3.1's response to ~~to~~ changes in parameter *ke*, which controls rain drop evaporation rates, and parameter *c*0, which controls precipitation efficiency through conversion of cloud

water to rain water. For this comparison we only consider the response for the June, July, and August (JJA) seasonal mean between 30°S to 30°N on four variables including 2 meter air temperature (TREFHT), 200-millibar zonal winds (U), sea level pressure (PSL), and precipitation (PRECT). Experiments with CAM3.1 use observed climatological sea surface temperatures and sea ice extents. Each experiment with CAM3.1 is 32-years in duration.

The observational data that is used to evaluate the model comes from a reanalysis product ECMWF-ERA interim (Uppala et al., 2005) for 2 m air temperature, 200-millibar zonal winds, and sea level pressure and GPCP (Adler et al., 2009) for precipitation. We make use of approximately 30 years of JJA mean fields between 1979 and 2009. For constructing S, we calculate variances from 2-year means (i.e. 15 samples).

A total of 64 experiments were completed, varying each of the two parameters within an 8×8 lattice. For each experiment we calculate three versions of ~~GMRF-based cost(equation~~ the GMRF test statistic which we refer to as a 'cost' (equation 2). The first version is the traditional cost based on the assumption of space and field independence ~~set here by setting~~ where the off diagonal components of $\mathbf{S}$ are set to zero and setting $\alpha = 1$. This approach is similar to what has been done previously for Taylor (2001). The second version of evaluating the cost takes field dependencies into account by including all components of $\mathbf{S}$ and setting $\alpha = 1$. The third version for the cost takes field and space dependencies into account by including all components of $\mathbf{S}$ and setting $\alpha = 0.0026$.

The correlation matrix, $\mathbf{R}$, corresponding to the $\mathbf{S}$ matrix of 2-year JJA seasonal mean variances and covariances, as estimated from 30 years of observations, is:

	PRECT	PSL	TREFHT	U
PRECT	1	-0.219	-0.047	0.015
PSL	-0.219	1	-0.313	-0.112
TREFHT	-0.047	-0.313	1	-0.145
U	0.015	-0.112	-0.145	1

The primary field correlations are the values of (-0.313) and (-0.219) occurring between sea level pressure (PSL) and 2 m air temperature (TREFHT), and precipitation (PRECT) and sea level pressure (PSL), respectively. ~~These correlations make physical sense in that precipitation mainly occurs within low pressure storm systems which tends to cool the underlying surface. The other correlations are minimal and there is not a good physical argument supporting their relationship.~~ Maps of the grid point correlations between these fields show a lot of structure with regions of both positive and negative correlations. Therefore, providing a mechanistic explanation of the spatially averaged correlation is not particularly meaningful. Despite losing regional information in the S matrix summary of field covariances, GMRF estimated field covariances as seen within 'witch hat' graphs are reasonable as compared to empirical estimates (see supplemental).

Figure 5 shows a comparison of the three versions of the GMRF-based cost for the 64 experiments within an 8×8 lattice. All versions of cost result in qualitatively similar results with high and low

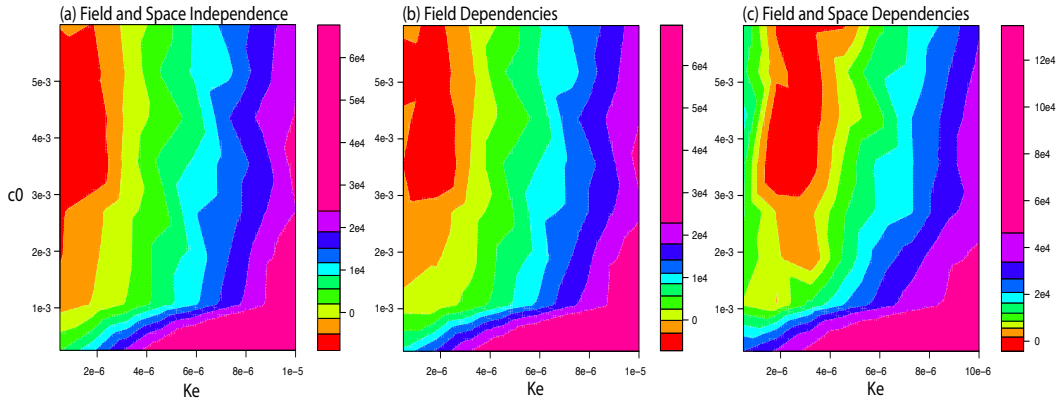

Figure 5. Three versions of the GMRF-based cost as a function of two CAM3.1 parameters *k*e and *c*0 that assumes the data has (a) field and space independence, (b) field dependencies, and (c) field and space dependencies. Each color represents ten percentiles of the cost distribution. The cost is shown relative to the value of the default model configuration.

cost values roughly in the same portions of parameter space. The main difference among the versions of cost comes from taking space dependencies into account within the field-space version. In this case, extremely low values of *k*e result in higher metric values. Figure 6 examines the reasons for this by graphing the different field contributions to the GMRF-based costs for a slice where $c0 = 0.0035$ which corresponds to one of the rows of the lattice. By plotting everything differenced from metric values at $ke = 3 \times 10^{-6}$, one can learn that the biggest qualitative difference comes from cost values associated with 2 m air temperature. Closer inspection of differences between model output and observations of 2 m air temperature (not shown) indicates that the traditional cost is likely reflecting large-scale differences over the southern hemisphere oceans. Inclusion of space dependencies places much greater significance on smaller-scale anomalies occurring over the continents, particularly over the Andes Mountains. This finding is a result of the mathematics of GMRF. It does not imply that the large-scale errors are of lesser scientific importance. It only means that GMRF is less sensitive to large-scale anomalies, perhaps because they are associated with fewer degrees of freedom than highly structured errors. Understanding whether and how these distinctions aid model assessment needs further study. We do find it reassuring that GMRF-based metrics of distance to observations are similar, at least in the example provided, to a traditional metric.

5 Summary

We have developed a new test statistic as a scalar measure of model skill or cost for evaluating the extent to which climate model output captures observed field and space relationships using Gaussian Markov Random Fields (GMRFs). The challenge has been that few observations exist for estab-

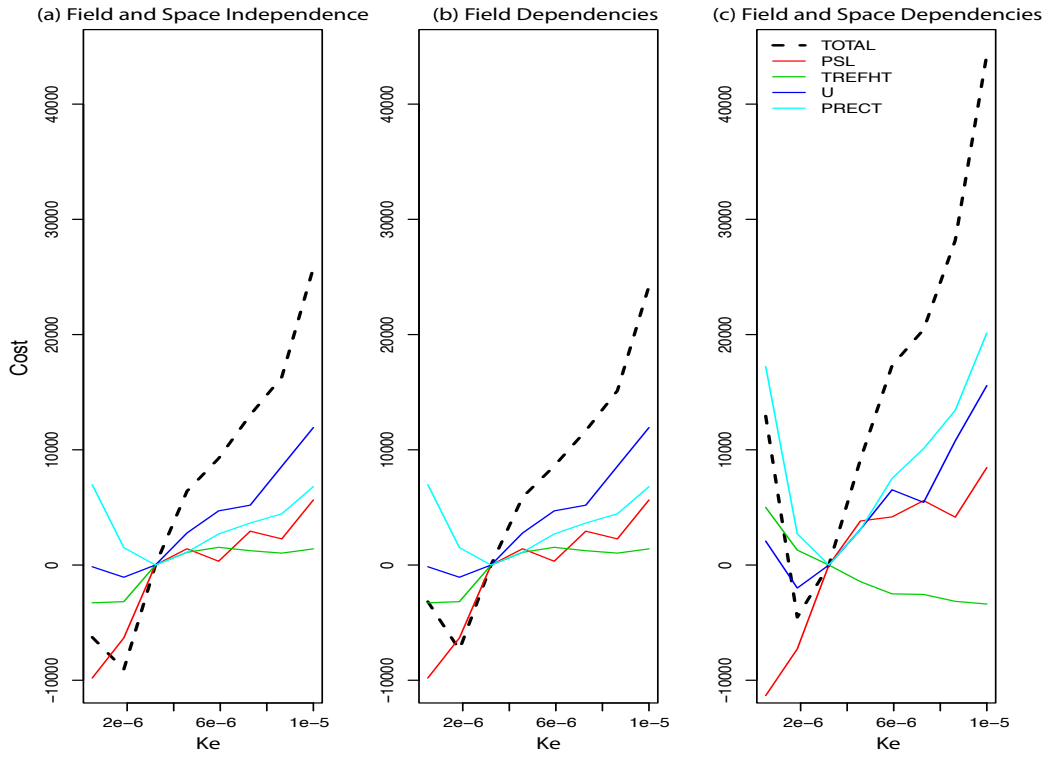

Figure 6. Different field contributions to the GMRF-based costs for a slice of Figure 5 where $c0 = 0.0035$. Cost values are relative to the default parameter setting for *k*e. Note that total cost (black dashed line) is a weighted sum of field contributions as given by $\mathbf{S}^{-1}$ with contributions from sea level pressure (PSL, red line), 2-m air temperature (TREFHT, green line), 200-millibar zonal winds (U, blue line), and total precipitation (PRECT, cyan line).

lishing a meaningful observational basis for quantifying field and space relationships of climate phenomena. Much of the data that is typically used for model evaluation is suspected of having its own relationship biases introduced by the numerical model that is used to synthesize measurements into gridded products. The GMRF-based metric overcomes some of these limitations by considering field and space variations within a neighborhood structure thereby lowering the metric's data requirements. The form of the metric separates space and field dependencies using a Kronecker product that, when multiplied out, has all the terms necessary to represent how different points in space are tied together across multiple field. We also include a scalar α that weights the importance of spatial relationships between grid cells. Its optimal value turns out to be independent of the data type which aids the use of GMRFs for comparing model output to data across multiple fields. Using 'witch hat' graphs, we show a first order (nearest neighborhood) structure does an excellent job of capturing empirical estimates of field and space relationships for various lag-windows or distances. We have applied three versions of cost that selectively turn on or off field and space dependencies

in a climate model (CAM3.1) output against observational products for tropical JJA climatologies for 2 m air temperature, sea level pressure, precipitation, and 200-millibar zonal winds. The results show subtle, but potentially important differences among these versions of the cost which may prove beneficial for selecting models that capture observed climate phenomena for the right reasons.

6 Code and data availability

R code and data for generating ~~figures~~ Figures 5 and 6 can be obtained through https://zenodo.org/record/33765, Nosedal-Sanchez et al. (2015)

Acknowledgements. This material is based upon work supported by the U. S. Department of Energy Office of Science, Biological and Environmental Research Regional & Global Climate Modeling Program under Award Numbers DE-SC0006985 and DE-SC0010843. Nosedal was partially supported by the National Council of Science and Technology of Mexico (CONACYT).

References

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P.-P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P., and Nelkin, E.: The Version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979–Present), http://dx.doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2, 4, 1147–1167, 2009.

Braverman, A., Cressie, N., and Teixeira, J.: A likelihood-based comparison of temporal models for physical processes, Statistical Analysis and Data Mining, 4, 247–258, 2011.

Collins, W. D., Rasch, P. J., Boville, B. A., Hack, J. J., McCaa, J. R., Williamson, D. L., and Briegleb, B. P.: The formulation and atmospheric simulation of the Community Atmosphere Model version 3 (CAM3), Journal of Climate, 19, 2144–2161, 2006.

Cressie, N. and Wikle, C. K.: Statistics for Spatio-Temporal Data, Wiley, Hoboken, NJ, 2011.

Gleckler, P. J., Taylor, K. E., and Doutriaux, C.: Performance metrics for climate models, Journal of Geophysical Research: Atmospheres (1984–2012), 113, 1–20, 2008.

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

Nosedal-Sanchez, A., Jackson, C. S., and Huerta, G.: Code for "A new metric for climate models that includes field and spatial dependencies using Gaussian Markov Random Fields", Zenodo, doi:10.5281/zenodo.33765, 2015.

Reichler, T. and Kim, J.: How Well Do Coupled Models Simulate Today's Climate?, Bulletin of the American Meteorological Society, 89, 303–311, 2008.

Santer, B. D., Taylor, K. E., Gleckler, P. J., Bonfils, C., Barnett, T. P., Pierce, D. W., Wigley, T. M. L., Mears, C., Wentz, F. J., Brüggemann, W., Gillett, N. P., Klein, S. A., Solomon, S., Stott, P. A., and Wehner, M. F.: Incorporating model quality information in climate change detection and attribution studies, Proceedings of the National Academy of Sciences, 106, 14 778–14 783, 2009.

Taylor, K.: Summarizing multiple aspects of model performance in a single diagram, Journal Of Geophysical Research-Atmospheres, 106, 7183–7192, 2001.

Trenberth, K. E., Koike, T., and Onogi, K.: Progress and Prospects for Reanalysis for Weather and Climate, Eos, Transactions American Geophysical Union, 89, 234–235, 2008.

Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Berg, L. V. D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., Mcnally, A. P., Mahfouf, J. F., Morcrette, J. J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, Quarterly Journal Of The Royal Meteorological Society, 131, 2961–3012, 2005.

Weigel, A. P., Knutti, R., Liniger, M. A., and Appenzeller, C.: Risks of Model Weighting in Multimodel Climate Projections, Journal Of Climate, 23, 4175–4191, 2010.

Supplemental Material: A new metric for climate models that includes field and spatial dependencies using Gaussian Markov Random Fields.

Alvaro Nosedal, Charles S. Jackson and Gabriel Huerta.

May 15, 2016

Mathematical details to find Q for a 2×2 lattice.

Consider $\mathbf{x} \sim N(\mathbf{0}_{4\times 1}, \mathbf{Q}_{4\times 4})$, a vector of measurements on a 2×2 lattice, as represented in Figure 1 of the main manuscript. Assume a neighborhood structure between the four elements of $\mathbf{x}$. In Figure 2 of the main manuscript, the neighbors for each element of $\mathbf{x}$ are defined graphically. Given this structure, one can write expressions for the conditional means that reflect how information at each grid point might be related to its neighbors. Therefore,

$$E(x_1|x_2, x_3, x_4) = \beta_{12}x_2 + \beta_{13}x_3, \tag{1}$$

$$E(x_2|x_1, x_3, x_4) = \beta_{21}x_1 + \beta_{24}x_4, \tag{2}$$

$$E(x_3|x_1, x_2, x_4) = \beta_{31}x_1 + \beta_{34}x_4, \tag{3}$$

$$E(x_4|x_1, x_2, x_3) = \beta_{42}x_2 + \beta_{43}x_3. \tag{4}$$

These expressions are used to find a relationship between the β coefficients and the elements of $\mathbf{Q}$. Since $\mathbf{x} \sim N(\mathbf{0}_{4\times 1}, \mathbf{Q}_{4\times 4})$, the joint probability distribution of $\mathbf{x}$ is given by,

$$\begin{aligned} f(x_1, x_2, x_3, x_4) \propto \quad exp(-\tfrac{1}{2}(&Q_{11}x_1^2 + Q_{22}x_2^2 + Q_{33}x_3^2 + Q_{44}x_4^2 + 2Q_{12}x_1x_2 + 2Q_{13}x_1x_3 + 2Q_{14}x_1x_4 \\ &+2Q_{23}x_2x_3 + 2Q_{24}x_2x_4 + 2Q_{34}x_3x_4)). \end{aligned}$$

Using this joint probability distribution, we derive the full conditional of x_1 given x_2, x_3, x_4,

$$f(x_1|x_2, x_3, x_4) \propto exp\left\{-\frac{1}{2}Q_{11}\left(x_1^2 - 2x_1\left(-\frac{Q_{12}}{Q_{11}}x_2 - \frac{Q_{13}}{Q_{11}}x_3 - \frac{Q_{14}}{Q_{11}}x_4\right)\right)\right\}. \tag{5}$$

This expression can be re-written as

$$f(x_1|x_2, x_3, x_4) \propto exp\left\{-\frac{1}{2}Q_{11}\left(x_1 - \left(-\frac{Q_{12}}{Q_{11}}x_2 - \frac{Q_{13}}{Q_{11}}x_3 - \frac{Q_{14}}{Q_{11}}x_4\right)\right)^2\right\}. \tag{6}$$

From matching (6) to the expression of a univariate normal distribution,

$$E(x_1|x_2,x_3,x_4) = -\frac{Q_{12}}{Q_{11}}x_2 - \frac{Q_{13}}{Q_{11}}x_3 - \frac{Q_{14}}{Q_{11}}x_4, \tag{7}$$

and

$$Prec(x_1|x_2,x_3,x_4) = Q_{11}. \tag{8}$$

By comparing equations (1) and (7), we obtain

$$\beta_{12} = -\frac{Q_{12}}{Q_{11}}, \qquad \beta_{13} = -\frac{Q_{13}}{Q_{11}}, \qquad \beta_{14} = -\frac{Q_{14}}{Q_{11}} = 0.$$

Considering the full conditionals for x_2, x_3 and x_4 and its conditional expectations respectively, yield similar relationships between the β coefficients and the elements of $\mathbf{Q}$:

$$\begin{aligned}
&\beta_{21} = -\frac{Q_{21}}{Q_{22}}, \qquad \beta_{23} = -\frac{Q_{23}}{Q_{22}} = 0, \qquad \beta_{24} = -\frac{Q_{24}}{Q_{44}} \\
&\beta_{31} = -\frac{Q_{31}}{Q_{33}}, \qquad \beta_{32} = -\frac{Q_{32}}{Q_{33}} = 0, \qquad \beta_{34} = -\frac{Q_{34}}{Q_{33}} \\
&\beta_{41} = -\frac{Q_{41}}{Q_{44}} = 0, \qquad \beta_{42} = -\frac{Q_{42}}{Q_{44}}, \qquad \beta_{43} = -\frac{Q_{43}}{Q_{44}}.
\end{aligned}$$

These relationships hold for an n dimensional distribution as established in Rue and Held [1]. If the conditional means and precisions can be written as

$$E(x_i|x_{-i}) = \mu_i + \sum_{j\neq i} \beta_{ij}(x_j - \mu_j) \quad \text{and} \tag{9}$$

$$Prec(x_i|x_{-i}) = k_i > 0, \tag{10}$$

then $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ follows a multivariate normal distribution with mean $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)$ and precision matrix $\mathbf{Q}$ of entries Q_{ij}, where

$$Q_{ij} = \begin{cases} -k_i\beta_{ij} & i \neq j \\ k_i & i = j \end{cases} \tag{11}$$

provided $k_i\beta_{ij} = k_j\beta_{ji}$, $i \neq j$.

If we let $Prec(x_i|x_{-i}) = 2$ $(i = 1,2,3,4)$, $\beta_{12} = \beta_{13} = \beta_{21} = \beta_{24} = \beta_{31} = \beta_{34} = \beta_{42} = \beta_{43} = 1/2$ and $\beta_{14} = \beta_{23} = \beta_{32} = \beta_{41} = 0$ using equations (9)-(11), $\mathbf{x} = (x_1, x_2, x_3, x_4)$ follows a multivariate normal distribution with mean $\mu = (0,0,0,0)^T$ and precision matrix

$$\mathbf{Q} = \begin{pmatrix} 2 & -1 & -1 & 0 \\ -1 & 2 & 0 & -1 \\ -1 & 0 & 2 & -1 \\ 0 & -1 & -1 & 2 \end{pmatrix}.$$

Generalizing Q to deal with multiple fields.

The generalization of **Q** to handle multiple fields is illustrated by a case with two fields, **x** and **y** which represent the difference between a model and observations for these fields. These observations are assumed to be on a 2×2 lattice, as shown in Figure 1.

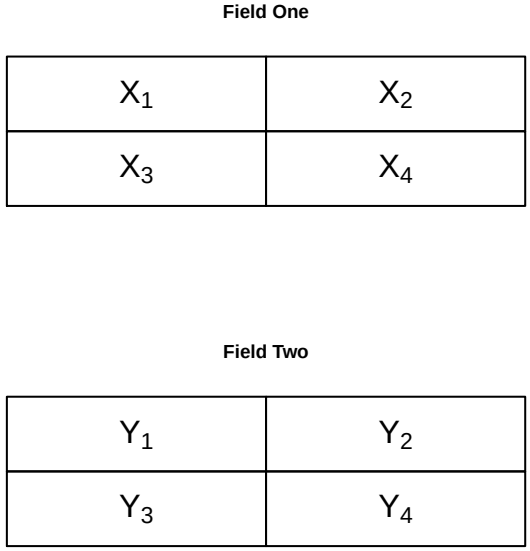

Figure 1: Two fields with observations **x**, **y** defined on a 2×2 lattice.

Firstly **x** and **y** are combined into one vector **v** so that $\mathbf{v} = (v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8)^T = (x_1, x_2, x_3, x_4, y_1, y_2, y_3, y_4)^T$. The covariance among these observations can be represented by a 2×2 matrix between the field 1, **x**, and the field 2, **y**,

$$\mathbf{S} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix},$$

where $Var(\mathbf{x}) = \sigma_{11}$, $Var(\mathbf{y}) = \sigma_{22}$, and $Cov(\mathbf{x}, \mathbf{y}) = \sigma_{12}$. Recalling that the correlation between fields 1 and 2 is defined as: $\rho = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$, it is easy to verify that the inverse of **S** is

$$\mathbf{S}^{-1} = \begin{pmatrix} \frac{1}{\sigma_{11}(1-\rho^2)} & \frac{-\rho}{(1-\rho^2)\sqrt{\sigma_{11}\sigma_{22}}} \\ \frac{-\rho}{(1-\rho^2)\sqrt{\sigma_{11}\sigma_{22}}} & \frac{1}{\sigma_{11}(1-\rho^2)} \end{pmatrix}.$$

Defining $\mathbf{Q}^*$ as $\mathbf{S}^{-1} \otimes \mathbf{Q}$, the Kronecker product of $\mathbf{S}^{-1}$ and **Q**, then,

$$\mathbf{Q}^* = \mathbf{S}^{-1} \otimes \mathbf{Q} = \begin{pmatrix} \frac{1}{\sigma_{11}(1-\rho^2)}\mathbf{Q} & \frac{-\rho}{(1-\rho^2)\sqrt{\sigma_{11}\sigma_{22}}}\mathbf{Q} \\ \frac{-\rho}{(1-\rho^2)\sqrt{\sigma_{11}\sigma_{22}}}\mathbf{Q} & \frac{1}{\sigma_{11}(1-\rho^2)}\mathbf{Q} \end{pmatrix}.$$

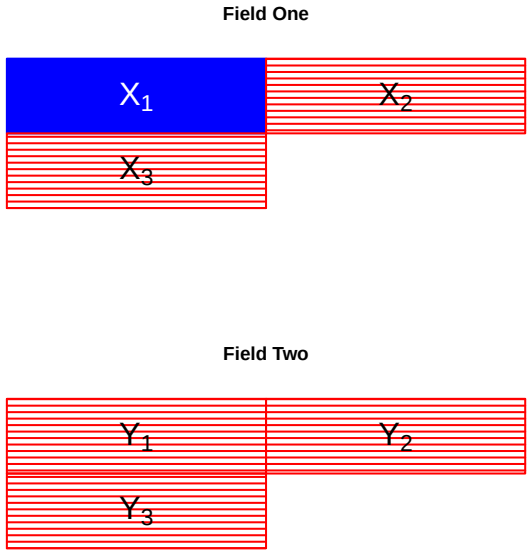

Figure 2: Neighbors of x_1 for a 2×2 lattice and two fields $\mathbf{x}$ and $\mathbf{y}$.

To see what type of relationships are imposed by $\mathbf{Q}^*$ on the elements of $\mathbf{v}$, consider the first element $v_1 = x_1$. Also notice that the first row of $\mathbf{Q}^*$ is,

$$\left(\frac{2}{\sigma_{11}(1-\rho^2)} \quad \frac{-1}{(1-\rho^2)\sigma_{11}} \quad \frac{-1}{(1-\rho^2)\sigma_{11}} \quad 0 \quad \frac{-2\rho}{\sqrt{\sigma_{11}\sigma_{22}}(1-\rho^2)} \quad \frac{\rho}{\sqrt{\sigma_{11}\sigma_{22}}(1-\rho^2)} \quad \frac{\rho}{\sqrt{\sigma_{11}\sigma_{22}}(1-\rho^2)} \quad 0 \right). \tag{12}$$

Using equations (9)-(11), it can be easily checked that the value for $\beta_{12} = \frac{-Q^*_{12}}{Q^*_{11}} = \frac{1}{2}$. The other β values can be determined in a similar fashion. Using these β coefficients, the equations for the conditional mean and precision of $v_1 = x_1$ given the rest of the elements of $\mathbf{v}$ are

$$E(v_1|\mathbf{v_{-1}}) = \frac{1}{2}x_2 + \frac{1}{2}x_3 + \frac{\rho\sigma_{11}}{\sqrt{\sigma_{11}\sigma_{22}}}y_1 - \frac{\rho\sigma_{11}}{2\sqrt{\sigma_{11}\sigma_{22}}}y_2 - \frac{\rho\sigma_{11}}{2\sqrt{\sigma_{11}\sigma_{22}}}y_3 \tag{13}$$

and

$$Prec(v_1|\mathbf{v_{-1}}) = \frac{1}{\sigma_{11}(1-\rho^2)}. \tag{14}$$

The expression for the conditional mean can be rewritten in terms of the slope b of the linear regression between $\mathbf{x}$ and $\mathbf{y}$, $b = \frac{\rho\sqrt{\sigma_{11}}}{\sqrt{\sigma_{22}}}$, with ρ equal to their correlation,

$$E(v_1|\mathbf{v_{-1}}) = \frac{1}{2}x_2 + \frac{1}{2}x_3 + by_1 - \frac{b}{2}y_2 - \frac{b}{2}y_3. \tag{15}$$

Equation (15) implies that the neighbors of x_1 are x_2, x_3, y_1, y_2 and y_3. Figure 2 shows a graphical display of all neighbors of x_1 in the context of the two fields $\mathbf{x}$, $\mathbf{y}$ and a 2×2 lattice.

Interpretation of S matrix

Reviewers raised the question about the physical interpretation of the correlation matrix **R**, corresponding to the **S** matrix of 2-year JJA seasonal mean variances and covariances. We noted that it is difficult to ascribe a particular interpretation to these numbers since taking a spatial average may result in a small correlation from fields that have large but opposing correlations. Figure 3 shows maps of the grid point correlations between JJA mean 2m air temperature (TREFHT), sea level pressure (PSL), and precipitation (PRECT) with sea level pressure (PSL). What is clear between all these figures is that there is a lot of structure to all these maps. The sign of the correlation is regionally dependent. This is the case for 2m air temperature (TREFHT) and precipitation (PRECT) which has a near zero correlation within the correlation matrix **R** but have regionally very high correlations. Figure 4 shows that the 'witch hat' test of the GMRF-based estimate of covariances between these two fields show that GMRFs are doing a reasonable job.

References Cited

[1] Rue, H. and Held, L.: Gaussian Markov random fields, vol. 104, Chapman & Hall/CRC, 2005.

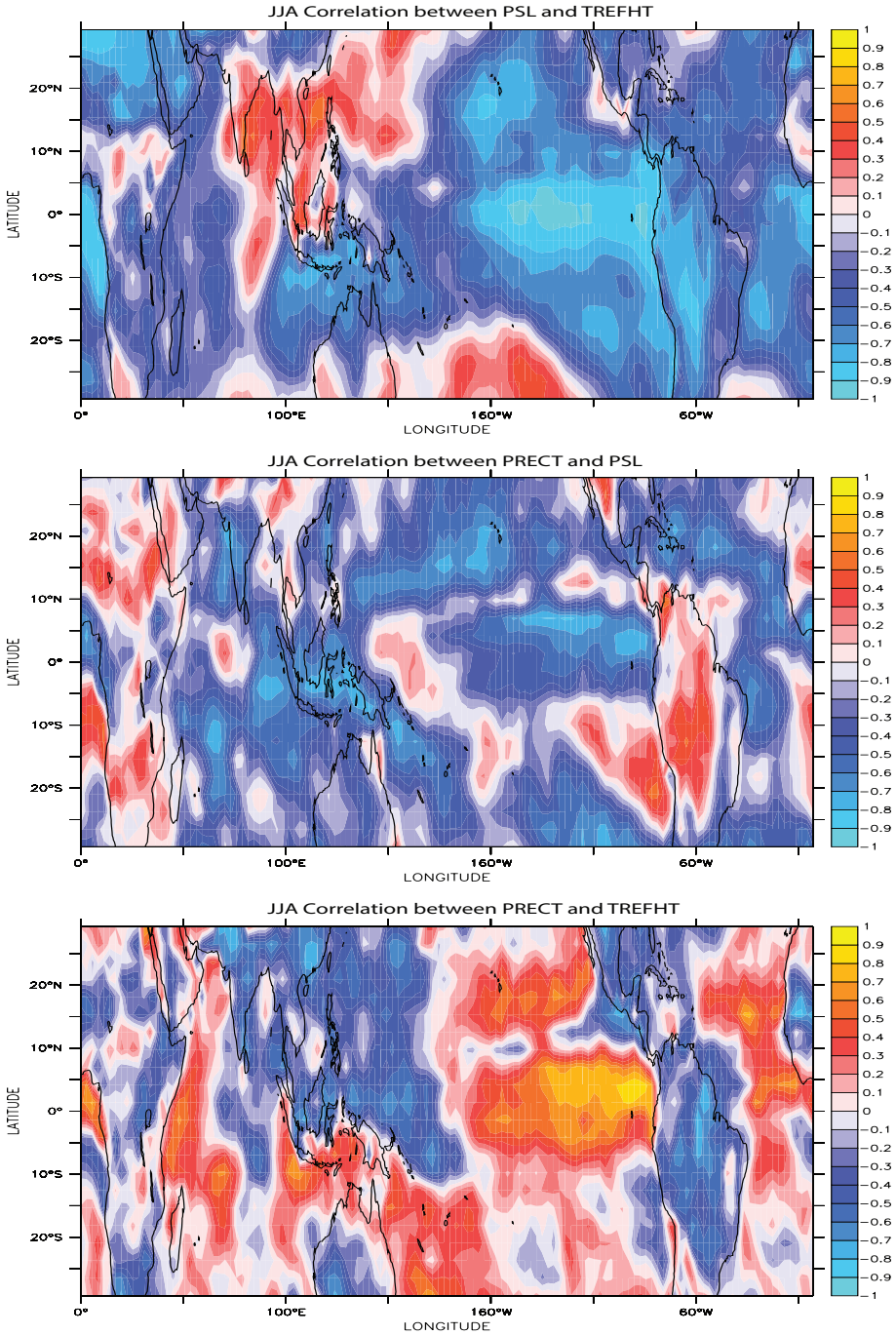

Figure 3: JJA correlations between 2m air temperature (TREFHT), sea level pressure (PSL), and precipitation (PRECT).

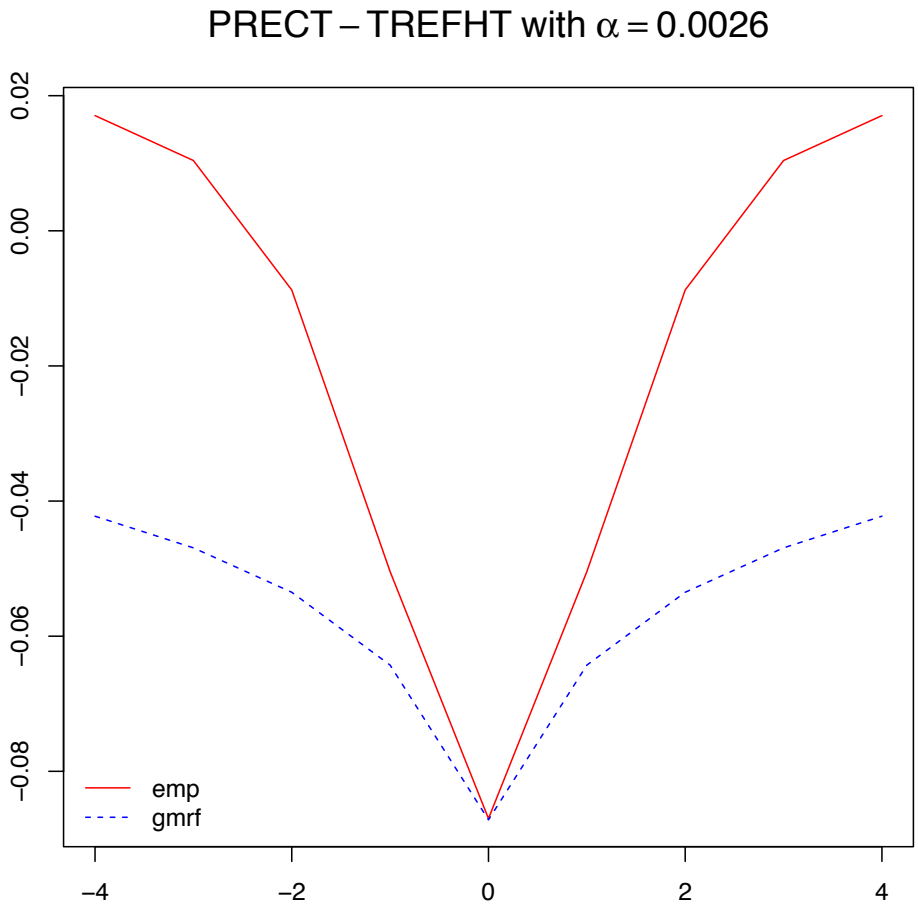

Figure 4: 'Witch hat' graphs testing GMRF approximations to empirical estimates of covariances between TREFHT and PRECT.