

This paper presents a metric for the evaluation of observing networks. Previous studies have used some scalar function of the posterior covariance. Most commonly this is also the uncertainty in some linear functional of the posterior estimate (such as the total flux over a given region) but metrics like the trace of the posterior covariance are also used. The metric proposed by this paper is the information content defined as the difference in entropy between prior and posterior. Under the common linear Gaussian assumption this turns out to be closely related to the log of the generalised variance or determinant of the posterior covariance.

The authors develop the necessary mathematics for their metric then present two examples, adding isotope measurements for methane or radiocarbon measurements for estimating fossil fuel CO₂ emissions. The mathematics is presented well and the examples are clear and pertinent. Furthermore the paper clearly lies within scope for the journal.

My concern with the paper is its lack of evaluation of the metric itself. There is a comparison with the independence metric but not with the covariance metrics. It is not clear to me that the information theory metric does the same job as the covariance metric or a better or worse job. There are two related problems:

First the generalised variance is one metric and probably not a very flexible one. It must include all potential sources. Normally this is not what we want. We have some target quantity like national emissions for which we are designing the network. Normally our target quantity will be a subset of pixels (e.g. pixels in one country). I'm not sure we can easily calculate the determinant of a submatrix of the Hessian.

Next there is the choice of uncertainty quantity to minimise. Rayner et al. (1996) pointed out that the preferred network depended on details of this quantity, such as the total ocean flux vs the average uncertainty for each ocean basin. The determinant is the volume of the hyperellipse described by the posterior covariance. It might be a good general choice but is likely to obscure these differences.

Finally I think the computational advantages of the new metric need a bit more justification. The authors claim that the covariance metric requires the inversion of a large matrix. Depending on the uncertainty metric we choose to minimise this might not be true. In general our target quantity is a linear functional of the posterior sources. Examples include the sum over some subregion and average over time. From what I learned to call Riesz's Representation Theorem (though there seem to be several of these) for any

linear functional f on R^n there is a vector \mathbf{v} such that $f(\mathbf{x}) = \mathbf{v} \cdot \mathbf{x}$ for all $\mathbf{x} \in R^n$. Thus for any target quantity t we can find some vector \mathbf{v}_t such that $t^a = \mathbf{v}_t \cdot \mathbf{x}^a$. The superscript a refers to the analysis or posterior. An example \mathbf{v}_t contains 1 for every pixel in a region and 0 otherwise. This will sum over the region of interest. By the Jacobian law of probabilities the uncertainty in t is given by $\mathbf{v}_t^T \cdot \mathbf{A} \cdot \mathbf{v}_t$ where the superscript T denotes transpose and \mathbf{A} is the posterior or analysis covariance. \mathbf{A} is the inverse of the Hessian \mathbf{G} so we need to calculate $\mathbf{v}_t^T \cdot \mathbf{G}^{-1} \cdot \mathbf{v}_t$. I believe this calculation can be efficiently accomplished by the Cholesky decomposition of \mathbf{G} . If we write $\mathbf{G} = \mathbf{L}\mathbf{L}^T$ (Cholesky decomposition) then I believe $\mathbf{G}^{-1} = \mathbf{L}^{-1,T}\mathbf{L}^{-1}$. Substituting this we see $t = \mathbf{y}^T \cdot \mathbf{y}$ where $\mathbf{y} = \mathbf{L}^{-1} \cdot \mathbf{v}_t$. Thus I think the target uncertainty can be performed with a Cholesky decomposition, a matrix-vector product and a dot-product. This may even be less costly than the determinant via the Cholesky decomposition.

I may just as easily be wrong here but think the comparison of the cost and generality of the new metric of the existing uncertainty metric does need more consideration than it gets here.

I only have two specific comments on the paper:

- L45** When citing early literature it is probably fair to cite the paper that gave rise to the field, Hardt and Scherbaum (1994).
- L65** Summing over the submatrix does indeed account for the covariance of uncertainty but that isn't it's most important property. This is that it calculates the uncertainty on the summed regional flux rather than the individual pixels.

References

- Hardt, M. and Scherbaum, F.: The Design of Optimum Networks for After-shock Recordings, *Geophys. J. Int.*, 117, 716–726, 1994.
- Rayner, P. J., Enting, I. G., and Trudinger, C. M.: Optimizing the CO₂ Observing Network for Constraining Sources and Sinks, *Tellus*, 48B, 433–444, 1996.