Rougier, J., & Zammit Mangion, A. (2016). Visualization for Large-scale Gaussian Updates. *Scandinavian Journal of Statistics*, *43*. https://doi.org/10.1111/sjos.12234

# Visualisation for large-scale Gaussian updates

Jonathan Rougier

School of Mathematics

University of Bristol, UK

Andrew Zammit Mangion

National Institute for Applied Statistics Research Australia (NIASRA)

School of Mathematics and Applied Statistics

University of Wollongong, Australia

**Abstract**

In geostatistics, and also in other applications in science and engineering, it is now common to perform updates on Gaussian process models with many thousands or even millions of components. These large-scale inferences involve modelling, representational, and computational challenges. We describe a visualisation tool for large-scale Gaussian updates, the 'medal plot'. The medal plot shows the updated uncertainty at each observation location, and also summarises the sharing of information across observations, as a proxy for the sharing of information across the state vector (or latent process). As such, it reflects characteristics of both the observations and the statistical model. We illustrate with an application to assess mass trends in the Antarctic Ice Sheet, for which there are strong constraints from the observations and the physics.

KEYWORDS: VARIANCE UPDATE, VARIANCE BOUND, MEDAL PLOT, SPATIAL STATISTICS

# 1 Introduction

Statisticians are now attempting inferences of a scale and complexity that were unthinkable even a few years ago. This is for a number of reasons:

1. Computers are more powerful, and have larger memories,

2. New statistical techniques are available to represent judgements on large collections of random quantities, and to compute on those judgements,

3. Large new datasets, including from remote sensing, are becoming available, and

4. There is a political need, and research funding, to address inference for complex systems, notably environmental systems.

Similar assessments have been given by Kalnay (2002, ch. 1, concerning meteorology) and Smith (2010, ch. 1, concerning decision support). In our application, outlined below, the state vector has about $10^5$ components, and there are $3.5 \times 10^5$ observations. Statistical inferences of this scale are most easily handled using a Gaussian process prior, and the linearisation of the observation operator; or else the use of an optimisation approach that comes to very much the same thing (e.g., as in data assimilation for meteorology, see Apte *et al.*, 2008).

There are two major challenges in this type of inference. The first is modelling: constructing a statistical model over observables and predictands which embodies, in its structure and its values, the beliefs of the domain experts (e.g., glaciologists, in our application in Section 4). At the root of this challenge is the partial and somewhat qualitative nature of expert belief, in large and complex systems. It often falls to statisticians to implement the model in all of its details, the output of which is then discussed with the experts. At this point an expert may assert that something looks wrong. Together, the statisticians and the experts trace this 'wrongness' back to a modelling choice made by the statisticians or, surprisingly often, unrealistic values supplied as data or beliefs (measurement error covariances, for example). In this process, which can be iterated several times, it is very helpful

to have simple visualisations of model behaviour, which both experts and statisticians can interpret.

The second challenge is computational. There is a 'book-keeping' problem, of representing the joint distribution of observables and predictands which may have very different spatial and/or temporal scales, and encompass multiple interacting processes. This representation has to be compliant with efficient computation which means, in effect, that the representing and computing cannot be separated, but must be treated together. This is a fertile ground for coding errors, and it is very helpful to have a simple necessary condition for correctness that is easily implemented and checked.

We propose a tool, the 'medal plot', that addresses both of these needs. It is almost obvious that the updated variance of any measured linear combination of the state vector has to be no larger than the smaller of its initial variance and the observation error variance. We prove this result for a collection of observations (Theorem 1). There is additional information available in the source of the bound (initial variance or observation error variance), and in the relation of the updated variance to its bound. This leads naturally to a visualisation tool in updates of random fields, for which the linear combinations are often localised in the domain. As no reference is made to the value of the observations, this diagnostic can be used before the observations are made available, for example in experimental design (e.g., Krause *et al.*, 2008).

Section 2 presents some theoretical results concerning Gaussian or, more generally, second-order updates. Section 3 describes our 'medal plot' for visualisation, whose structure and interpretation follows directly from our

theoretical results. Section 4 illustrates the utility of the medal plot in a large and complicated inference for mass trends in the Antarctic Ice Sheet.

## 2 Theoretical results

Let $\boldsymbol{X}$ be the collection of Gaussian random quantities, and $\boldsymbol{Y} := A\boldsymbol{X}$ be the known linear combinations which are measured, where $A$ is sometimes termed the 'incidence matrix'. Let $\boldsymbol{Z} := \boldsymbol{Y} + \boldsymbol{E}$ be the observations, including observation error $\boldsymbol{E}$. Denote the variance matrix of $\boldsymbol{Y}$ as $\Sigma$, and take the observation error to be independent of $\boldsymbol{X}$, with variance matrix $T$. If $V$ and $W$ are two variance matrices, write $V \leq W$ exactly when $W - V$ is non-negative definite. Then we have the following result, which applies not just in the Gaussian case, but also for more general second-order updating, such as the Bayes linear approach described in Goldstein and Wooff (2007).

**Theorem 1.** *Let* $\Sigma^* := \mathrm{Var}(\boldsymbol{Y} \mid \boldsymbol{Z})$ *be the updated variance matrix of* $\boldsymbol{Y}$*. If* $\Sigma + T$ *is non-singular, then* $\Sigma^* \leq \Sigma$ *and* $\Sigma^* \leq T$*.*

*Proof.* As $\mathrm{Cov}(\boldsymbol{Y}, \boldsymbol{Z}) = \Sigma$ and $\mathrm{Var}(\boldsymbol{Z}) = \Sigma + T$, the updated variance of $\boldsymbol{Y}$ satisfies

$$\Sigma^* = \Sigma - \Sigma(\Sigma + T)^{-1}\Sigma \tag{1}$$

(see, e.g., Mardia *et al.*, 1979, chapter 3). Hence $\Sigma^* \leq \Sigma$ because the second term on the righthand side of (1) is non-negative definite. If we can show that

$$\Sigma - \Sigma(\Sigma + T)^{-1}\Sigma = T - T(\Sigma + T)^{-1}T, \tag{2}$$

then (1) and the same reasoning implies that $\Sigma^* \leq T$, completing the proof.

Start with the identities

$$\mathbf{0} = \begin{cases} \Sigma - \Sigma(\Sigma + T)^{-1}(\Sigma + T), \\ T - T(\Sigma + T)^{-1}(\Sigma + T). \end{cases} \tag{3}$$

Equating the two terms on the righthand side gives

$$\Sigma - \Sigma(\Sigma + T)^{-1}\Sigma - \Sigma(\Sigma + T)^{-1}T = T - T(\Sigma + T)^{-1}T - T(\Sigma + T)^{-1}\Sigma.$$

The final terms on each side of this expression are equal, because they are symmetric, and (2) is proved. □

It is important that this result holds for singular $\Sigma$, provided that $\Sigma + T$ is non-singular. This is because there may be replications in the observations; e.g., the same component of $\boldsymbol{X}$ observed several times. This would be represented as duplicate rows in $A$. Alongside replications, there may be more observations than components of the state vector; e.g., for multiple instruments with overlapping footprints. This would be represented by an $A$ with more rows than columns. In both of these cases

$$\Sigma = A \operatorname{Var}(\boldsymbol{X}) A^T$$

would be singular (non-negative definite but not positive definite). No matter what the form of $A$, a non-singular $T$ is sufficient for $\Sigma + T$ to be non-singular (positive definite). Thus Theorem 1 always holds if there is measurement error.

A further useful result concerns the relationship between the *joint* update

$\mathrm{Var}(Y_i \mid \boldsymbol{Z})$ and the *local* update $\mathrm{Var}(Y_i \mid Z_i)$:

$$\mathrm{Var}(Y_i \mid \boldsymbol{Z}) \leq \mathrm{Var}(Y_i \mid Z_i) \leq T_{ii} \qquad \text{for each } i. \tag{4}$$

This ordering of global, local, and observation error variances is used in our visualisation tool, presented in Section 3. The first inequality is a standard result for second-order updates (see, e.g., Goldstein and Wooff, 2007, section 5.2), while the second inequality follows from Theorem 1. The second inequality in (4) can be verified by direct calculation:

$$\mathrm{Var}(Y_i \mid Z_i) = \Sigma_{ii} - \frac{\Sigma_{ii} \cdot \Sigma_{ii}}{\Sigma_{ii} + T_{ii}} = \frac{\Sigma_{ii} \cdot T_{ii}}{\Sigma_{ii} + T_{ii}} \leq T_{ii}. \tag{5}$$

This expression shows that there is a limit to how much relative effect a local update can have. Taking $T_{ii} \leq \Sigma_{ii}$, for concreteness, and writing $\kappa := T_{ii}/\Sigma_{ii}$,

$$\inf_{T_{ii} \leq \Sigma_{ii}} \frac{\mathrm{Var}(Y_i \mid Z_i)}{T_{ii}} = \inf_{\kappa \leq 1} \frac{1}{1+\kappa} = \frac{1}{1+1} = \frac{1}{2}. \tag{6}$$

In other words, information from $Z_i$ alone can push the updated variance of $Y_i$ down to half of its upper bound, and this occurs when $\Sigma_{ii} = T_{ii}$. Eq. (5) also shows that if $T_{ii} \ll \Sigma_{ii}$ then $\mathrm{Var}(Y_i \mid Z_i) \approx T_{ii}$. In other words, the variance of the local update tends to the observation error variance as the observation error variance becomes small relative to the initial variance.

The case where one of $\Sigma$ or $T$ is much larger than the other occurs frequently in practice, and it is interesting to consider the limiting case where, for concreteness, $T$ becomes vanishingly small relative to $\Sigma$. However, there is a difficulty with this case: if $\Sigma$ is singular, then a 'vanishingly small' $T$ will

ultimately conflict with the requirement that $\Sigma + T$ be non-singular. But, as explained above, it is common for $\Sigma$ to be singular. Therefore the following result has an additional condition relative to Theorem 1, but it is powerful when this condition holds.

**Theorem 2.** *Let $\Sigma$ and $T$ both be non-singular, and define $\kappa := \|T\Sigma^{-1}\|$, where $\|\cdot\|$ is any induced p-norm. If $\kappa < 1$ then*

$$\|\Sigma^* - T\| \leq \frac{\|\Sigma^{-1}\| \ \|T\|^2}{1 - \kappa}. \tag{7}$$

*Proof.* Start from (1) and the top branch of (3) to show that

$$\Sigma^* = \Sigma(\Sigma + T)^{-1}T.$$

Now under the conditions of the Theorem both $\Sigma$ and $T$ are non-singular, and this expression can be rearranged to show that

$$\Sigma^* = (\Sigma^{-1} + T^{-1})^{-1}$$

(see also Rue and Held, 2005, section 2.3.3). Then (7) follows from a standard result about inverses and perturbations (see, e.g., Golub and Van Loan, 1996, Theorem 2.3.4). □

In other words, if both $\Sigma$ and $T$ are non-singular then as $T$ becomes small relative to $\Sigma$, so the updated variance converges to $T$. However, it is important to appreciate that $T$ non-singular is not, on its own, sufficient for this convergence. This is seen in the following counter-example with a

singular $\Sigma$:

$$\mathrm{Var}(\boldsymbol{X}) = 10^6 \begin{pmatrix} 1.0 & 0.4 \\ 0.4 & 1.0 \end{pmatrix}, \quad A = \begin{pmatrix} 1.0 & 1.0 \\ 1.0 & 0.0 \\ 0.0 & 1.0 \end{pmatrix}, \quad T = \begin{pmatrix} 1.0 & 0.0 & 0.0 \\ 0.0 & 0.1 & 0.0 \\ 0.0 & 0.0 & 0.1 \end{pmatrix},$$

for which, informally, $T \ll \Sigma = A\,\mathrm{Var}(\boldsymbol{X})A^T$. But

$$\Sigma^* = \begin{pmatrix} 0.17 & 0.08 & 0.08 \\ 0.08 & 0.09 & -0.01 \\ 0.08 & -0.01 & 0.09 \end{pmatrix}.$$

It can be checked that $\Sigma^* \leq T$, as required by Theorem 1, but clearly $\Sigma^* \not\approx T$. This combination of a singular $\Sigma$ with both 'large footprint' imprecise observations and 'small footprint' precise observations occurs in our illustration in Section 4.

Finally, Theorems 1 and 2 provide a multivariate second-order generalisation of L.J. Savage's principle of stable inference (see, e.g., Savage *et al.*, 1962; Edwards *et al.*, 1963). We state it here for completeness, although it does not feature in what follows. The updated variance in this statement is a general second-order update, but applies in particular when $\boldsymbol{Y}$ and $\boldsymbol{E}$ are both Gaussian.

**Theorem 3** (Principle of stable inference, multivariate)**.** *Let* $\boldsymbol{Y}$ *be a vector of random quantities and* $\boldsymbol{Z}$ *be a vector of noisy observations on the components of* $\boldsymbol{Y}$*. Define the measurement error as* $\boldsymbol{E} := \boldsymbol{Z} - \boldsymbol{Y}$*. Let* $\boldsymbol{E} \perp\!\!\!\perp \boldsymbol{Y}$ *and* $\mathrm{Var}(\boldsymbol{E})$ *be non-singular. Then* $\mathrm{Var}(\boldsymbol{Y}|\boldsymbol{Z}) \leq \mathrm{Var}(\boldsymbol{E})$*. Furthermore, if* $\mathrm{Var}(\boldsymbol{Y})$ *is also non-singular and* $\mathrm{Var}(\boldsymbol{E}) \ll \mathrm{Var}(\boldsymbol{Y})$ *then* $\mathrm{Var}(\boldsymbol{Y} \mid \boldsymbol{Z}) \approx \mathrm{Var}(\boldsymbol{E})$*.*

This principle underlies the common 'plug-in' approximation

$$\boldsymbol{Y} \mid \boldsymbol{z}^{\mathrm{obs}} \overset{\mathrm{D}}{\approx} \boldsymbol{z}^{\mathrm{obs}} + \boldsymbol{E},$$

where '$\overset{\mathrm{D}}{\approx}$' denotes 'is approximately distributed as'. If the observation errors are uncorrelated, then the updated variance matrix is approximately diagonal. Our results indicate that the critical modelling judgement under which this approximation provides a conservative or approximate assessment of uncertainty about the true values $\boldsymbol{Y}$ is that the (additive) measurement error $\boldsymbol{E}$ is uncorrelated with $\boldsymbol{Y}$. The other conditions seem much less demanding in practice.

## 3 Visualisation: the 'medal plot'

We would like to visualise various features of the variance update, particularly for those components of $\boldsymbol{Y}$ which correspond to locations. These features include, for a specified $Y_i$: what the upper bound is, and where it comes from; what the updated variance is; and what contribution is made by observations other than $Z_i$.

### 3.1 Construction of the 'medal plot'

We assume that each $Y_i$ corresponds to a location in a domain which can be visualised: typically this would be the time domain (see Section 3.2) or the 2D spatial domain (see Section 4). Each $Y_i$ is represented by a medal of three concentric disks of decreasing radius:

1. A red/blue disk representing the upper bound on the updated variance of $Y_i$, either *red* where the prior provides the upper bound, or *blue* where the observation error provides the upper bound.

2. A *white* disk representing the updated variance using $Z_i$ alone (local update).

3. A *gold* disk representing the updated variance using all observations (joint update).

In all cases, the radius of the disk is proportional to the standard deviation.

The medals can be scaled so that when displayed they do not overlap by more than is necessary to preserve the systematic patterns. When there is an overlap, it is more effective to plot all of the red/blue disks first, and then to overplot with the white disks, and then with the gold disks. In some applications, including our illustration below, it is more effective to use semi-transparent colours, so that underlying map features are preserved.

For a given medal at location $i$, we might be particularly interested in the thickness of the white annulus. This thickness shows us how much of the update of $Y_i$ is coming from observations other than $Z_i$, with a thick annulus showing that other observations are making a large contribution (i.e., driving

the updated variance well below what is achieved by $Z_i$ alone). When we compare the medals across the domain of the observations we can see at a glance how the localisation of the update varies, by comparing the widths of the white annuli. This is illustrated in the next subsection.

### 3.2 Toy example

Here is a 1D example in which it is possible to visualise the individual uncertainties along with the medals. Figure 1 shows a stationary process which is updated by irregularly-spaced observations with varying error standard deviations. The prior process has a correlation length of 30, defined to be the distance at which the correlation drops to 0.05.

[Figure 1 about here]

Take the observation in the centre first (at location 55). The rim of its medal is blue, because the upper bound on the updated variance comes from the observation error. This location is not close enough to the other observations for the local update to be different from the global update. Therefore the white annulus between the inner and outer rings has no discernable thickness.

Now consider the observations on the lefthand side, and ignore, initially, the single observation with a small variance. As with the centre observation, the upper bound on the updated variances comes from the observation errors, and so the rims of the medals are blue. All of the rims are the same thickness, because the ratio of the prior variance to the observation error variance is constant. But the white annulus thickens towards the centre of the group, where there are more observations within the correlation length. These extra

observations drive down the variance relative to what can be achieved using only the observation itself.

The single observation with the small variance (at location 26) demonstrates another aspect of the variance update. This observation is in the centre of the group, and, by the logic above, it ought to have a thick white annulus. But its observation error variance is much smaller than its prior variance, and then Theorem 2 indicates that the updated variance (gold circle) will be close to the upper bound (thin blue rim), as shown. Effectively, the difference between a local and a global update has been squeezed out.

On the righthand side, the upper bound on the updated variances comes from the prior for the process, and so the rims are red. The thickness of the annulus follows the same pattern as for the lefthand side. The difference is that the prior variance and the observation error variance are almost the same size. In this case the rim attains nearly maximum relative thickness, as shown in (6). The maximum rim width is $1 - 1/\sqrt{2} \approx 0.3$ of the total radius.

Qualitatively, all of these properties can be read off Figure 1, once it is also known that the correlation length is 30. In other words, the medals in the 1D case provide a quick visual summary of other information which can also be found in the plot. The value of the medals becomes apparent in 2D spatial applications, where it is not possible to display, on one figure, the prior variances, the observation error variances, the updated variances, and the correlation lengths (which might vary spatially). But useful summaries of these features can be inferred, at least approximately, from the full set of medal plots. We illustrate this in Section 4, with a 2D spatial application for Antarctica.

## 4 Illustration

Our illustration is part of a mass-balance estimate for the Antarctic Ice Sheet (AIS), which is the world's largest freshwater reservoir. Here we provide a brief outline of our inference, which we describe in detail elsewhere; see Zammit-Mangion *et al.* (2014, 2015b) for the statistical modelling, and Zammit-Mangion *et al.* (2015a) for a non-technical summary. The illustration shows the iteration between visualisation, statistical modelling, and observations which is a necessary part of building confidence in the resulting inference.

In order to determine the AIS contribution to sea-level change, the change in height of the AIS over a fixed time period must be decomposed into the sum of four main processes: change in the height of the underlying rock, effect of ice dynamics, firn compaction (densification of past years' snow), and the net effect of surface processes (precipitation, run-off, melt, and refreeze). Quantifying the contribution to sea-level change requires summing the changes in height of ice, firn, and surface processes inside the grounding line (see the caption to Figure 2) over the AIS, and then mapping those changes to mass changes using specified densities.

We have observations from three types of instrument. First, a small number of GPS receivers on rocky outcrops, which give accurate observations for change in height of the underlying rock (at those outcrops). Second, satellite altimetry, which gives observations of height change (i.e., summing the four processes) along specified transects. Third, satellite gravimetry (Gravity Recovery and Climate Experiment, or GRACE), which provides measures

of mass change, and therefore sees a linear combination of change in the height of the underlying rock, the ice dynamics and surface processes (firn compaction changes height but not mass). These three instruments have very different spatial footprints, with GPS being a point observation, altimetry having a footprint of about $1\,\mathrm{km}^2$ (treated as a point observation), and gravimetry having a footprint of about $1600\,\mathrm{km}^2$.

This is an inherently statistical problem because: (i) we have three instruments for four fields; (ii) there are substantial observation errors; (iii) the footprints of the instruments are of such different sizes; (iv) the observations do not cover the whole of the AIS; and (v) uncertainty assessment is a crucial output for impact studies related to sea-level rise. The problem becomes soluble once we incorporate prior information about the processes, notably their variabilities and their characteristic length scales, both of which can vary spatially. As well as the four fields, our unknowns include statistical parameters for the processes and in the observation equation.

For this illustration we used finite element basis functions to model each of the four processes (see, e.g., Lindgren *et al.*, 2011), with variable resolution to account for greater heterogeneity near the coastline. We used a blocked Gibbs sampler to update the processes conditional on the statistical parameters, and to update the statistical parameters conditional on the processes. Then we plugged in the maximum *a posteriori* estimate of the statistical parameters (which were well-constrained), and redid the update of the fields, to draw the medal plots. We illustrate with a medal plot for the gravimetry observation footprints for 2006, shown in Figure 2. Recollect that the medals show the update from all observations—the gravimetry linear combinations

are updated not just by the gravimetry observations, but also by GPS and altimetry.

[Figure 2 about here]

We provide our rationalisation of the features of the 'final' medal plot, shown in Figure 2. Our rationalisation is expressed in terms of our understanding of the processes, the observations, and our modelling choices.

First, almost all of the medals have blue rims, showing that the upper bound on the updated variance comes from the observation error variance, not the initial variance. There is one exception, which is at about $(+800\,\mathrm{km}, +250\,\mathrm{km})$, which we checked. At this location our model for mass trends implies a small initial expectation and variance, because the ice velocity and expected accumulation is so small. The rims are all very thin, indicating (in the case of the blue rims) that the initial variance is much larger than the observation error variance. (In fact, in our plotting we expand the rims slightly, where they would otherwise be hardly visible.) This is consistent with our choice of a relatively large prior variance for each of the four fields.

Second, there are clear spatial patterns in the observation error uncertainties. These uncertainties are provided along with the GRACE observations, although we chose to aggregate to larger spatial footprints in order to decorrelate the measurement uncertainties. These patterns in the uncertainties are related to physical features that induce variations in the GRACE observations for successive overpasses of the satellites. For example, uncertainty outside the grounding line tends to be small, because the (floating) ice is in hydrostatic equilibrium. Around the grounding line the uncertainty tends to

be relatively large because of variations in precipitation and ice velocity. The large medal at the South Pole reflects a GRACE observation with a larger spatial footprint.

Third, the globally updated uncertainties are much smaller than the locally updated uncertainties, as indicated by a thick light-blue annulus (which would be white in the colour-scheme of Section 3). Therefore much of the reduction in uncertainty at each location is coming from other observations. Some of this will be from other GRACE observations, because GRACE detects height changes in the underlying rock, which has a very long correlation length (i.e., is spatially very stiff). But some of it might also come from the altimetry observations. The correlation length of surface processes varies spatially, but is it relatively large around the South Pole. This is interesting because the altimetry satellite cannot over-fly the South Pole, and so there is no local altimetry contribution to the South Pole GRACE medal. Nevertheless, even here the contribution from other observations is substantial.

Figure 2 and the description above represents the end-point of a long modelling process, during which time we produced several alarming-looking medal plots, some of which indicated implementation issues, and some of which indicated modelling issues. Our progress towards a plausible visualisation was important in building the team's confidence in the results: glaciologists and statisticians together. Medal plots can be computed at little cost, and we believe they will be helpful across the environmental sciences, where it is now common to assimilate observations within a hierarchical spatial-temporal framework (e.g., Cressie and Wikle, 2011).

## Acknowledgements

We would like to thank Nana Schön for collating and preparing the data for the application and reviewing an early draft of this article, and two reviewers for very helpful comments which improved the clarity of the paper. This work was supported by the Natural Environment Research Council (NERC) funded RATES project, grant NE/I027401/1. Andrew Zammit-Mangion is also an Honorary Research Fellow at the University of Bristol.

## References

A. Apte, C.K.R.T. Jones, A.M. Stuart, and J. Voss, 2008. Data assimilation: Mathematical and statistical perspectives. *International Journal of Numerical Methods in Fluids*, **56**, 1033–1046.

N. Cressie and C.K. Wikle, 2011. *Statistics for Spatio-Temporal Data*. John Wiley & Sons, Inc., Hoboken NJ, USA.

W. Edwards, H. Lindman, and L.J. Savage, 1963. Bayesian statistical inference for psychological research. *Psychological Review*, **70**(3), 193–242.

M. Goldstein and D.A. Wooff, 2007. *Bayes Linear Statistics: Theory & Methods*. John Wiley & Sons, Chichester, UK.

G.H. Golub and C.F. Van Loan, 1996. *Matrix Computations*. Johns Hopkins University Press, Baltimore MD, USA, 3rd revised edition.

E. Kalnay, 2002. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, Cambridge, UK.

A. Krause, A. Singh, and C. Guestrin, 2008. Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, **9**, 235–284.

F. Lindgren, H. Rue, and J. Lindström, 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society, Series B*, **73**(4), 423–498.

K.V. Mardia, J.T. Kent, and J.M. Bibby, 1979. *Multivariate Analysis*. Harcourt Brace & Co., London, UK.

H. Rue and L. Held, 2005. *Gaussian Markov Random Fields: Theory and Applications*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, Boca Raton FL, USA.

L.J. Savage *et al.*, 1962. *The Foundations of Statistical Inference*. Methuen, London, UK.

J.Q. Smith, 2010. *Bayesian Decision Analysis: Principle and Practice*. Cambridge University Press, Cambridge, UK.

A. Zammit-Mangion, J.L. Bamber, N. Schoen, and J.C. Rougier, 2015a. A data-driven approach for assessing ice-sheet mass balance in space and time. *Annals of Glaciology*, **56**, 175–183.

A. Zammit-Mangion, J.C. Rougier, J. Bamber, and N. Schön, 2014. Resolving the Antarctic contribution to sea-level rise: A hierarchical modelling framework. *Environmetrics*, **25**, 245–264.

A. Zammit-Mangion, J.C. Rougier, N. Schön, F. Lindgren, and J. Bamber, 2015b. Multivariate spatio-temporal modelling for assessing Antarctica's present-day contribution to sea-level rise. *Environmetrics*, **26**, 159–177.

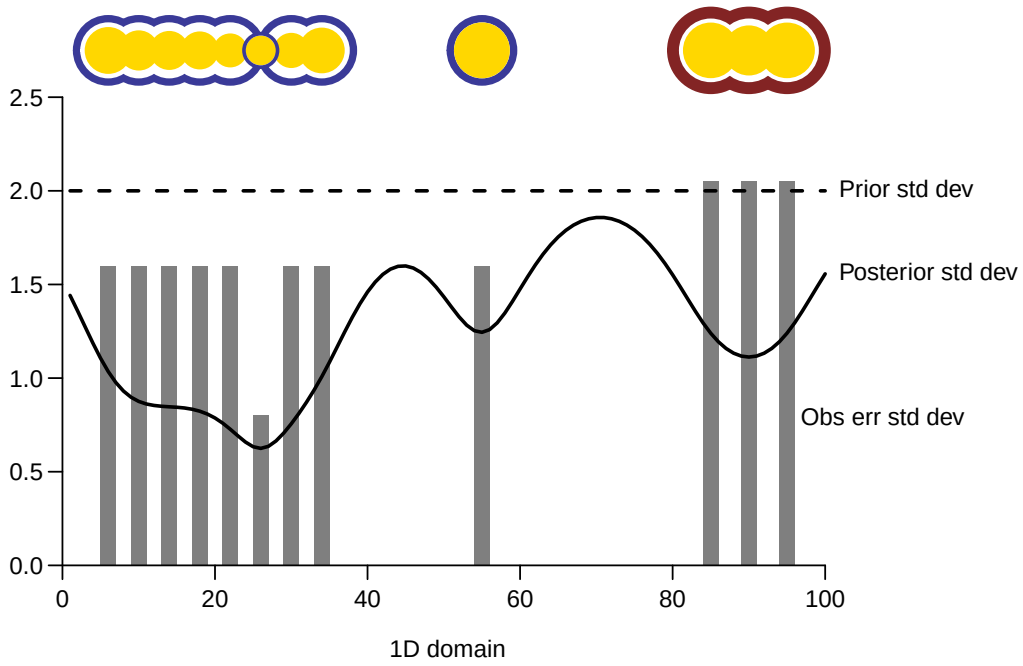


Figure 1: Toy problem showing medals for various configurations of the prior variance and observation error variance, and the proximity of other observations. The prior process is stationary with standard deviation 2 and correlation length 30. See Section 3.2 for details.

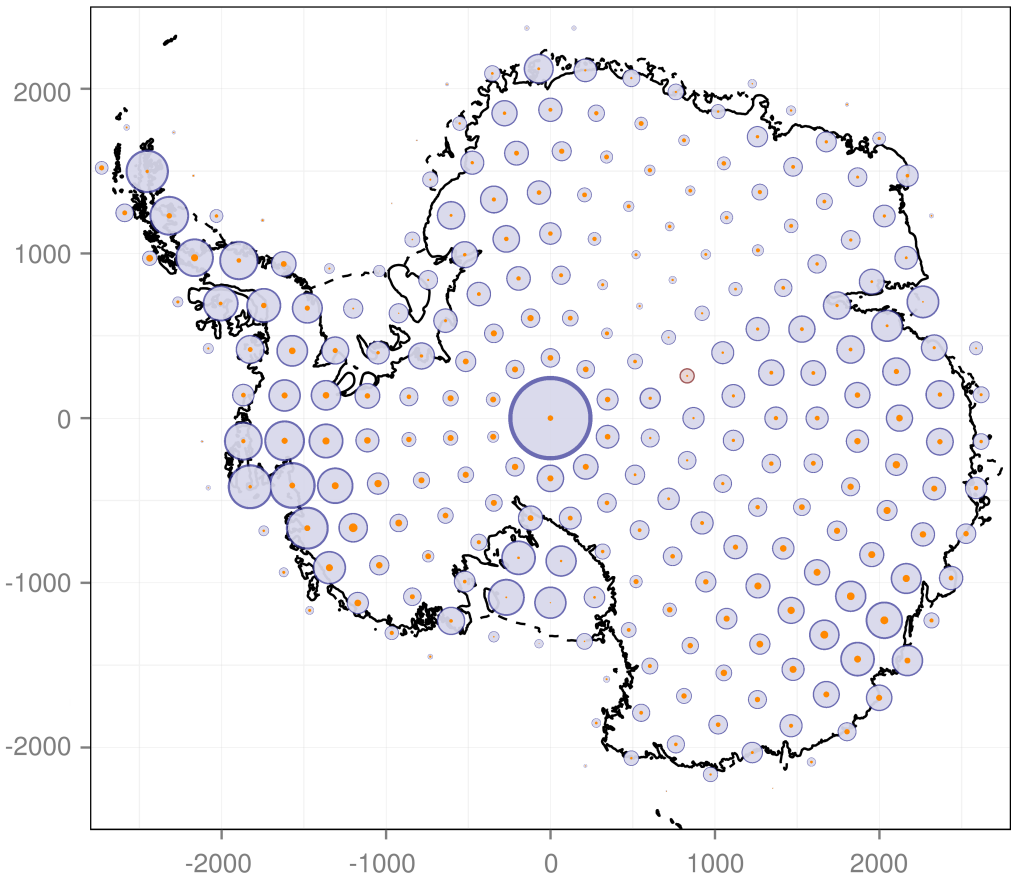


Figure 2: Medal plot for GRACE footprints over Antarctica, with distances in kilometres. The solid line is the grounding line (where the ice begins to float), and the dashed line is the coastline (which includes the floating ice). See Section 4 for details of the application and observations. We have used a semi-transparent blue instead of white for the annulus.

Name and address of corresponding author:

Jonathan Rougier
School of Mathematics
University of Bristol
University Walk
Bristol BS8 1TW
UK

Email j.c.rougier@bristol.ac.uk.