Estimating and presenting uncertainties in an historical sea-surface temperature analysis

John Kennedy, Nick Rayner, Michael Saunby, Rob Smith, David Parker.

Introduction

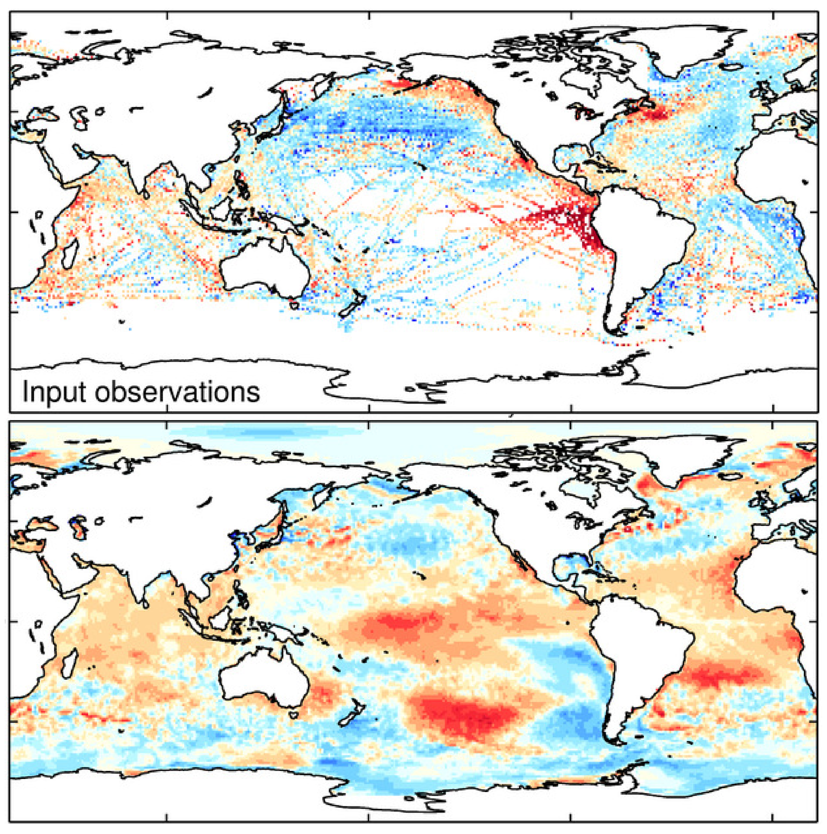
Historical sea-surface temperature (SST) analyses are used in a wide variety of applications. For many users, the ideal SST analysis would have very high temporal and spatial resolution and would look like the best fields that can currently be produced using high-resolution satellite data. Unfortunately, given the sparseness of historical *in situ* observations, this is not possible.

In order to encompass the 'true' value of the SST, analyses are often presented as a combination of a best-estimate together with an error range. This is a convenient way of expressing the magnitude of the uncertainty in the data, but it does not tell the user how the errors covary. The error covariance tells us a lot about the structure of the errors. In addition, the best estimate fields do not look like the high resolution fields derived from satellite data and have a tendency to be too 'smooth'.

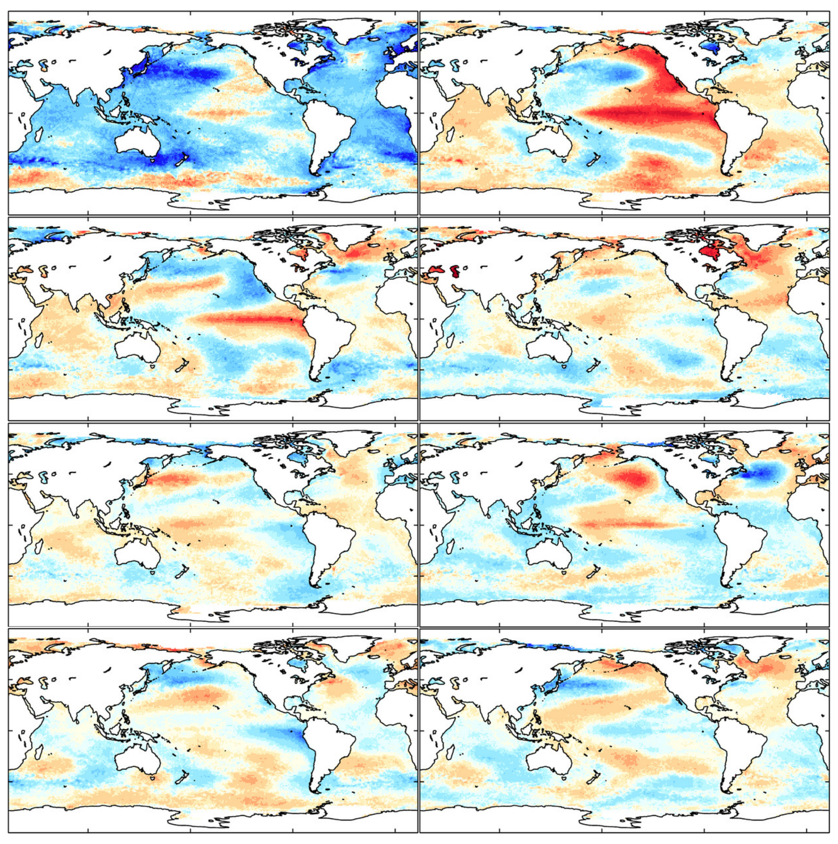
In developing the HadISST2 data set our aim has been to create an ensemble of equi-probable realisations of the data which are consistent with the available data and known covariance structure of SST, span the range of uncertainty, and which contain 'realistic' temporal and spatial variability at the target resolution.

The challenge

The aim of the method is to make sparse, *in situ* data such as those shown in the figure to the right look like fields produced from a high-resolution satellite data set such as OSTIA (shown to the right). This is done by drawing samples from the posterior distributions of the reconstructions and bias corrections in a multi-step process in which successively greater levels of detail are added to the reconstruction.

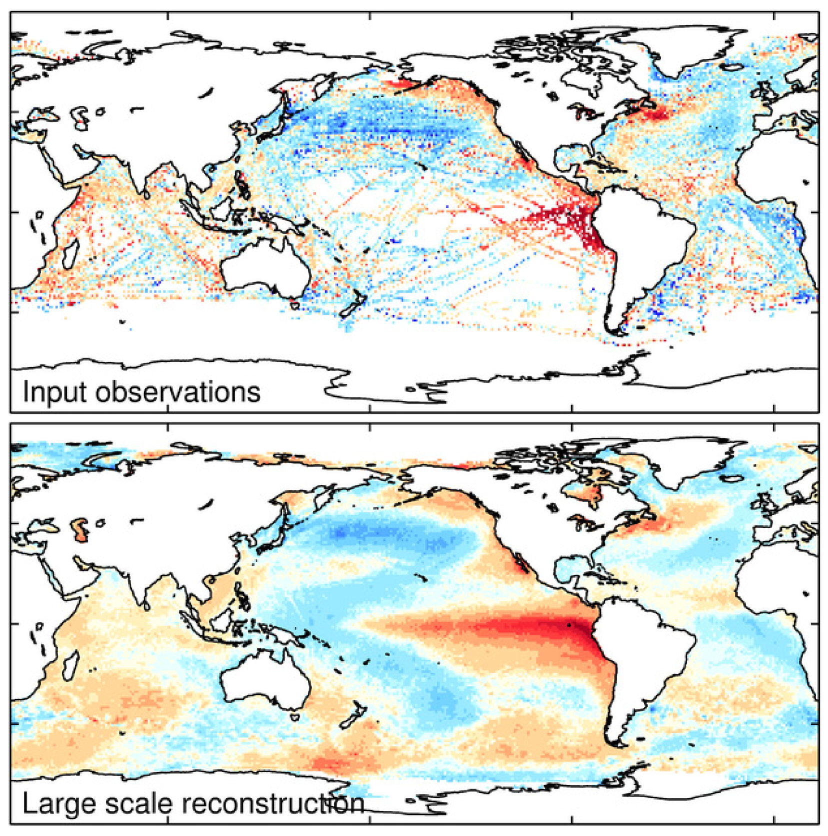


(upper panel) *in situ* SST anomalies for June 1983 gridded at 1° latitude by 1° longitude resolution. Large areas of the southern hemisphere and tropical Pacific contain few observations however, certain patterns are clear, such as the strong El Nino event. Even in the northern hemisphere the fields tend to be affected by measurement errors which add white-noise. (lower panel) OSTIA SST anomalies for January 2010. The fields are globally complete and are shown here reduced to a 1° latitude by 1° longitude resolution.

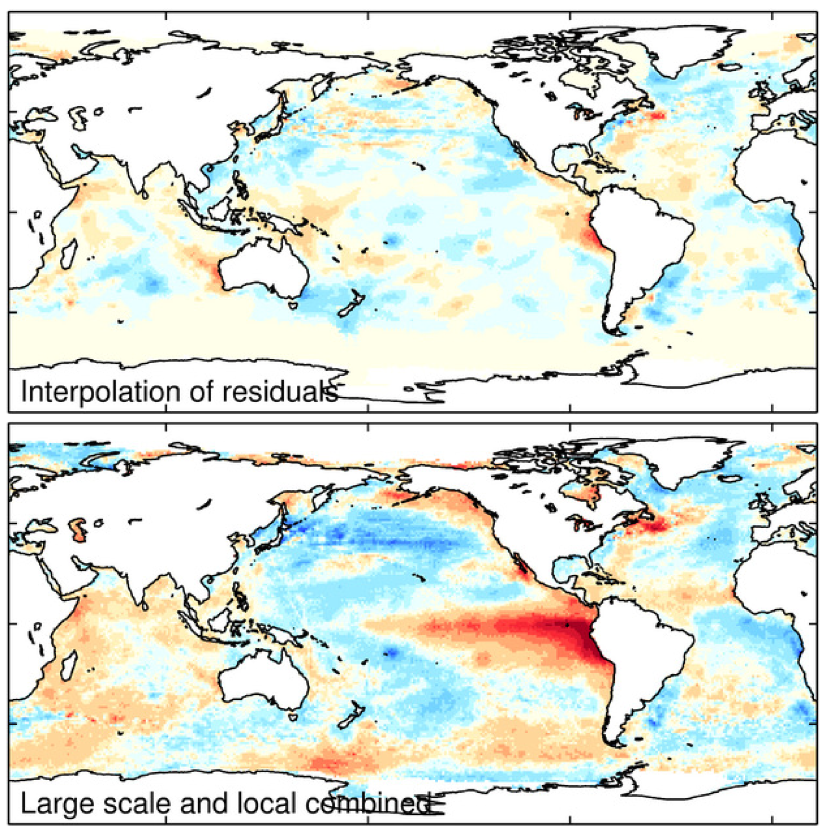


Step 1 Large scale: The observations are reconstructed using Variational Bayesian Principal Component Analysis (Ilin and Kaplan 2009*). VBPCA iteratively improves both the reconstruction and the estimated leading order covariance structures. 21 of these patterns were used to reconstruct the data and the first 8 are shown above. The number of patterns is kept relatively low in order to capture only variations at the largest spatial scales and to avoid over-fitting to the observational data.

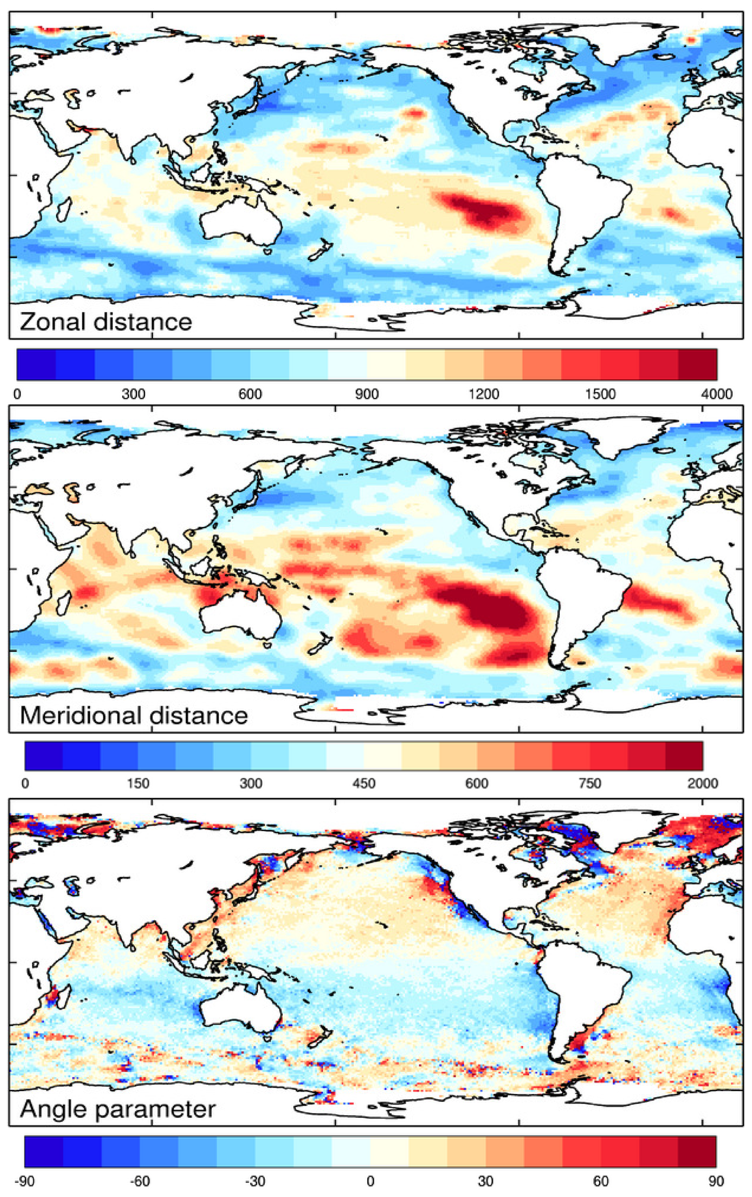
*A. Ilin and A. Kaplan. (2009) Bayesian PCA for Reconstruction of Historical Sea Surface Temperatures. In Proc. of the Int. Joint Conf. on Neural Networks (IJCNN 2009), pp. 1322-1327, Atlanta, USA, June 2009.

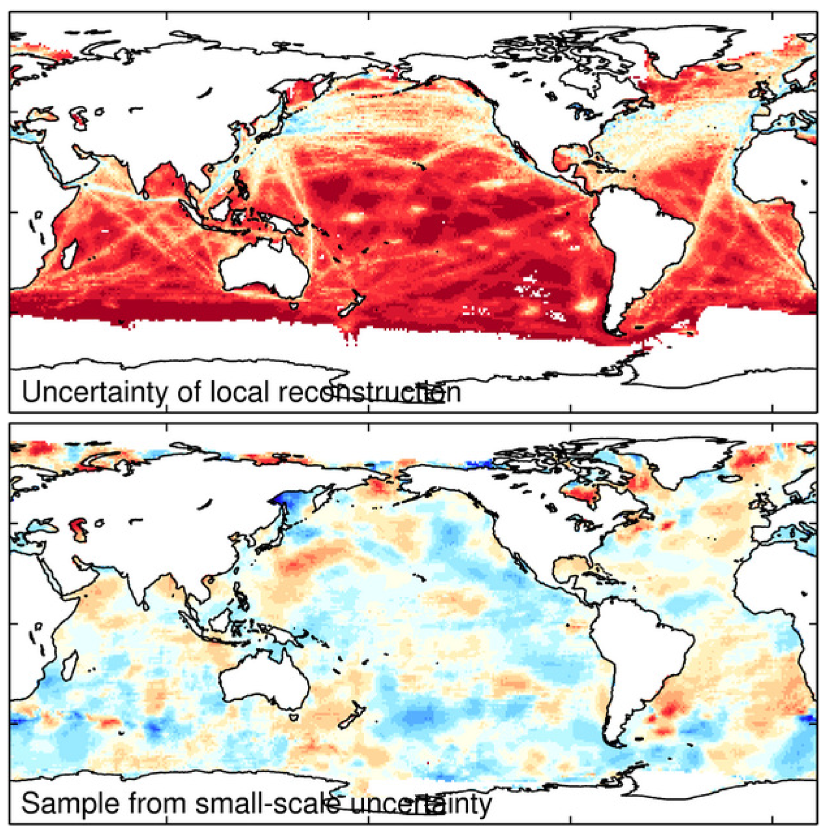


The upper panel shows the *in situ* observations for June 1983. The lower panel shows a single sample drawn from the posterior distribution of the large scale reconstruction. In 1983 there are sufficient data to constrain the large scale reconstruction, but in earlier years, particularly the 1860s, the reconstruction is less well constrained and samples from the large scale reconstruction can vary significantly whilst remaining consistent with the available observations (see panel at the bottom of the poster for more reconstructions).



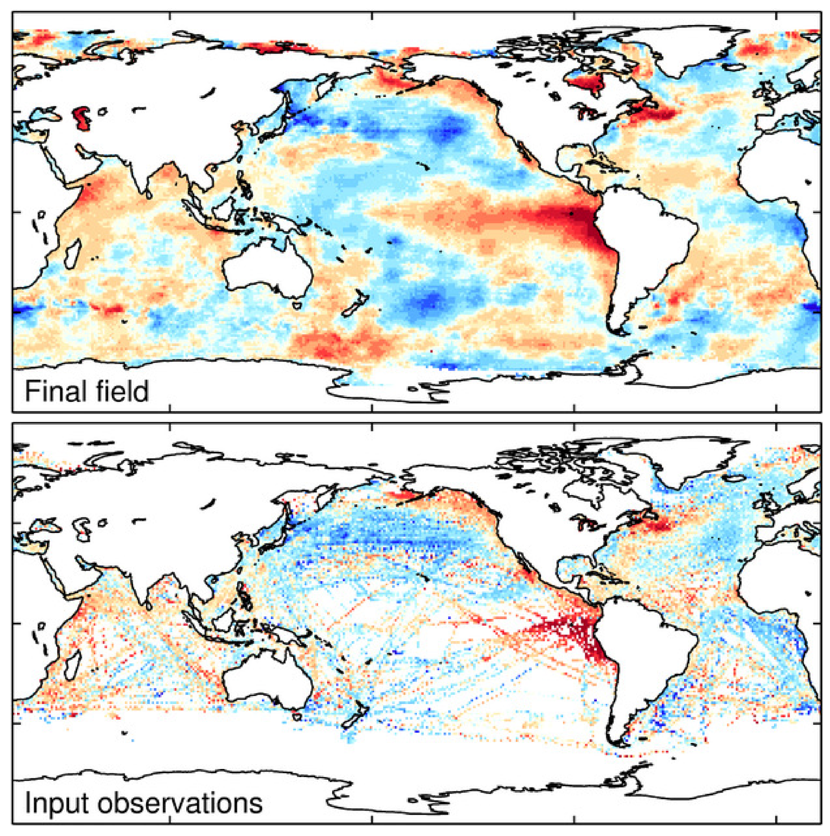
Step 2 Small scale : the large scale reconstruction is subtracted from the *in situ* observations and a local optimal interpolation (OI) method is used to create a small-scale analysis of the residuals (upper panel). The OI scheme uses three parameters to build a local covariance matrix (two orthogonal length scales and the angle of the principal axis from the line of longitude as in Karspeck et al. (submitted), shown to the right). The OI solution is then added to the large scale reconstruction (lower panel). Note that in areas with few observations the best-estimate OI gives very smooth fields.



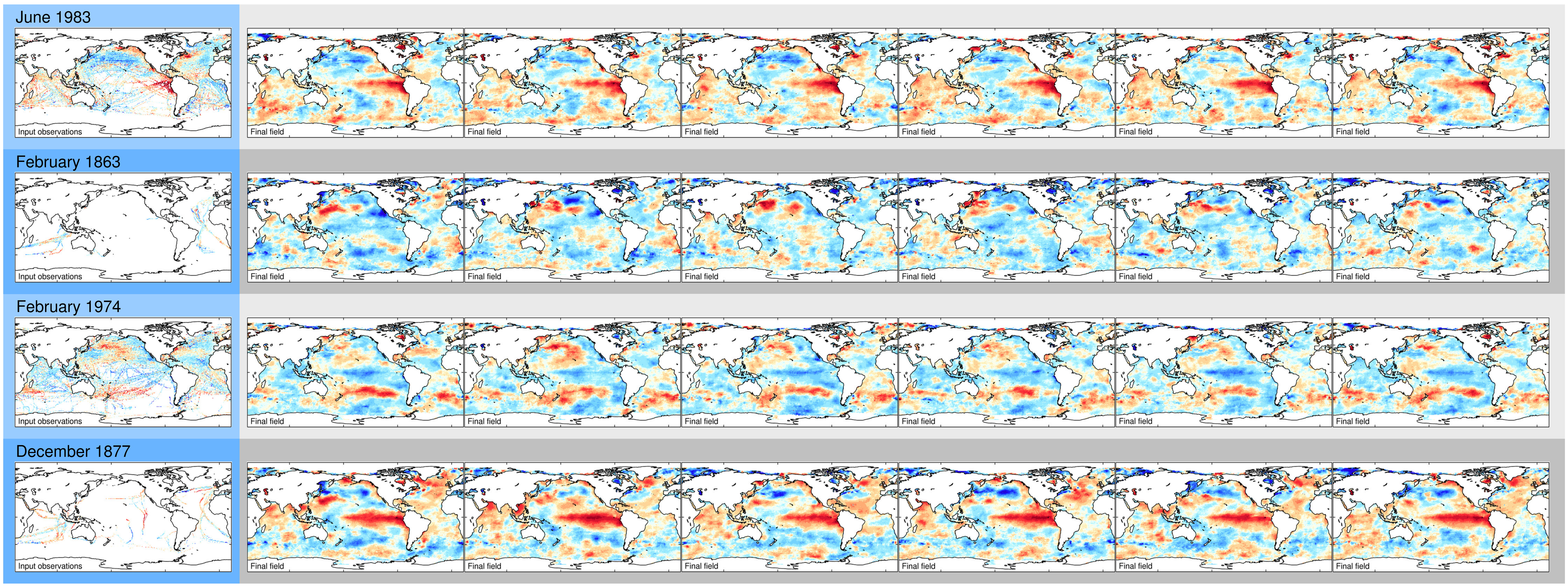


Step 3 Residual: the upper panel shows a measure of uncertainty in the local OI. Uncertainties are lowest where there are many observations. The uncertainties are used to develop a posterior covariance matrix from which a single sample is drawn (lower panel). The sample of the small scale variability adds 'detail' in areas with short spatial scales such as the western boundary currents, and the Agulhas retroflection region.

The sample is then added to the combined reconstruction and the resulting final field is shown in the next panel, to the right.



Step 4 Repeat: The process is repeated many times for each month allowing representative distributions to be built up. Some examples are shown at the bottom of the poster. In each case independent samples were drawn at each stage of the reconstruction. Where observations are plentiful, much of the between sample variance comes from the local reconstruction. At times when there are fewer observations, there can be large between-sample variance owing to differences in the large scale reconstruction. See for example the range of El Nino response consistent with the observations made in December 1877.



Met Office Hadley Centre, FitzRoy Road, Exeter, Devon, EX1 3PB United Kingdom
Tel: 01392 885105 Fax: 01392 885681
Email: john.kennedy@metoffice.gov.uk