SELF-CONSITENT APPROACH TO ESTIMATING LONG-TERM AND INTERANNUAL VARIABILITY WITH COMBINED SPATIAL SCALES IN HISTORICAL DATA SETS

Alexey Kaplan *(alexeyk@ldeo.columbia.edu)*

Climate Modeling Group, Division of Ocean and Climate Physics

Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY 10964, USA

1. MOTIVATION

All interpolation procedures require the assumption of various a priori estimates: mean and covariances of the target fields as well as observational error covariances. All these parameters are usually being estimated from available observations. Posterior consistency of the anaysis results with a priori assumptions is not guaranteed and often not tested. Schneider (2001) proposed an Expectation-Maximization based iterative procedure that allows to aciheve consistency for the target field covariance estimate. Here we generalize this procedure for estimating means (climatology) too, while also relaxing his assumption of zero (or ridge-regression selected) observational error.

The practical implementation of the proposed procedure also requires the appropriate treatment of small scales, if working with spaceXspace covariance matrix is impractical. Dealing with small scales poses a bunch of new problems, like proper specification of their covariance and also correct evaluation of their total variance, that is useful for accurate modeling of observational error.

2. NEW PROCEDURE

Consistent estimation of sliding window climatology and covariances

The OI problem, i.e. estimating the field $\mathcal{T}$ from a first-guess (background) solution taken to be the sliding window climatology $\mathcal{T}^{\rm clim}$ and an incomplete set of observations $\mathcal{T}^{\rm obs}$ is given by:

$$\mathcal{T} = \mathcal{T}^{\rm clim} + \varepsilon^{\rm anom}, \quad \langle \varepsilon^{\rm anom} \rangle = 0, \quad \langle \varepsilon^{\rm anom} \varepsilon^{\rm anom\,T} \rangle = C,$$
$$\mathcal{T}^{\rm obs} = H\mathcal{T} + \varepsilon^{\rm obs}, \quad \langle \varepsilon^{\rm obs} \rangle = 0, \quad \langle \varepsilon^{\rm obs} \varepsilon^{\rm obs\,T} \rangle = R.$$

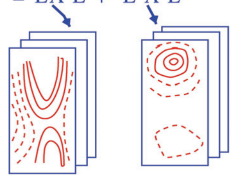
0. Initialization: $i = 0$. Use availble ad-hoc approaches to compute $\mathcal{T}_0^{\rm clim}$ and C_o from $\mathcal{T}^{\rm obs}$. Presumably only data for recent few decades can be used in this step.

Then apply to the entire period of historical data the iterations of the following cycle, for $i = 0, 1, 2, \cdots$, until convergence (similar to Expectation-Maximization approach, e.g. Schneider 2001):

1. Given the estimates of $\mathcal{T}_i^{\rm clim}$ and C_i, compute the OI solution, i.e. interpolated fields $\hat{\mathcal{T}}_i^{\rm OI}$ and error covariance estimate P_i:
$$\mathcal{T}_i^{\rm anom} = \mathcal{T}^{\rm obs} - \mathcal{T}_i^{\rm clim},$$
$$\hat{\mathcal{T}}_i^{\rm OI} = C_i H^T (R + HC_iH^T)^{-1} \mathcal{T}_i^{\rm anom},$$
$$P_i = C_i - C_iH^T(R + HC_iH^T)^{-1}HC_i$$
2. Adjust climatology (hereafter angle brackets denote sliding time window averaging):
$$\mathcal{T}_{i+1}^{\rm clim} = \mathcal{T}_i^{\rm clim} + \langle \hat{\mathcal{T}}_i^{\rm OI} \rangle$$
3. Compute new covariance estimate:
$$C_{i+1} = \left\langle \left(\hat{\mathcal{T}}_i^{\rm OI} - \langle \hat{\mathcal{T}}_i^{\rm OI} \rangle\right)\left(\hat{\mathcal{T}}_i^{\rm OI} - \langle \hat{\mathcal{T}}_i^{\rm OI} \rangle\right)^T \right\rangle + P_i.$$
4. Increase i:
$$i := i + 1$$
and go to step 1.

APPROXIMATING COVARIANCE

$$C = E\Lambda E^T + E'\Lambda' E'^T$$



Scale separation in a field estimate

OI problem: estimating a field $\mathcal{T}$ from a first-guess (background) solution $\mathcal{T}^b$ and an incomplete set of observations $\mathcal{T}^o$ is given by:

$$\mathcal{T}^b = \mathcal{T} + \varepsilon^b, \quad \langle \varepsilon^b \rangle = 0, \quad \langle \varepsilon^b \varepsilon^{b\,T} \rangle = C,$$
$$\mathcal{T}^o = H\mathcal{T} + \varepsilon^o, \quad \langle \varepsilon^o \rangle = 0, \quad \langle \varepsilon^o \varepsilon^{o\,T} \rangle = R.$$

The solution to this OI problem is a minimizer $\hat{\mathcal{T}}$ of the cost function

$$\mathbf{S}[\mathcal{T}] = (H\mathcal{T} - \mathcal{T}^o)^T R^{-1} (H\mathcal{T} - \mathcal{T}^o) + (\mathcal{T} - \mathcal{T}^b)^T C^{-1} (\mathcal{T} - \mathcal{T}^b).$$

$$\hat{\mathcal{T}} = CH^T(R + HCH^T)^{-1}\mathcal{T}^o.$$

$$C = E\Lambda E + E'\Lambda' E' = E\Lambda E + C'$$

$$\hat{\mathcal{T}} = E\hat{\alpha} + \Delta\hat{\mathcal{T}}.$$

$$\mathcal{T}^o = HE\alpha + \tilde{\varepsilon}^o, \qquad \langle \alpha\alpha^T \rangle = C', \qquad \langle \tilde{\varepsilon}^o \tilde{\varepsilon}^{o\,T} \rangle = HC'H^T + R.$$

$$\hat{\alpha} = \Lambda E^T H^T (HE\Lambda E^T H^T + HC'H^T + R)^{-1} \mathcal{T}^o.$$

Observational residual: $\Delta\mathcal{T}^o = \mathcal{T}^o - HE\hat{\alpha}$

$$\Delta\mathcal{T}^o = H\Delta\mathcal{T} + \varepsilon^o, \qquad \langle \Delta\mathcal{T}\Delta\mathcal{T}^T \rangle = C', \qquad \langle \varepsilon^o \varepsilon^{o\,T} \rangle = R.$$

$$\Delta\mathcal{T} = C'H(HC'H^T + R)^{-1}\Delta\mathcal{T}^o$$

3.EXAMPLE: RS SOLUTION

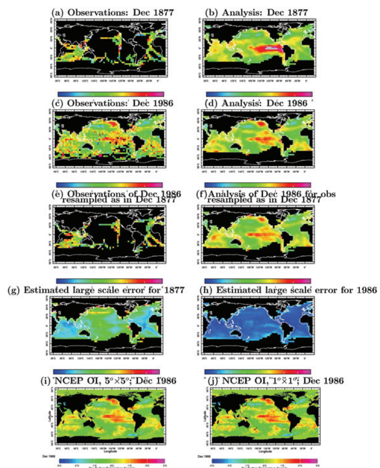
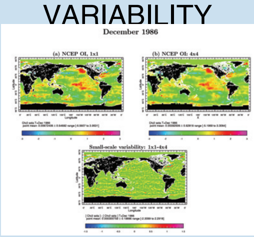


Figure 2: Available SST observations and their RS OS analysis for December 1877 (panels (a) and (b)) with verification through the experiment with 1986 data: simulated OS analysis for December 1986 using the data distribution of 1877 (panels (e) and (f)) versus the standard OS analysis for December 1986 with all available data (panels (c) and (d)). Also shown are large scale errors in the two reconstructions (panels (e) and (f)) and the NCEP OI December 1986 field presented in (i) $5^o \times 5^o$ and (j) $1^o \times 1^o$ resolution. Units are ^{o}C.

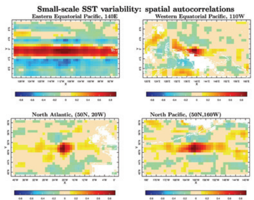
There is significant room for improvement , as far as reconstruction of small-scale variability is concerned.
We need to model it adequately

4. SMALL-SCALE VARIABILITY



Spatially non stationary covariance structure:

5.AUTOCORRELL. STRUCTURES

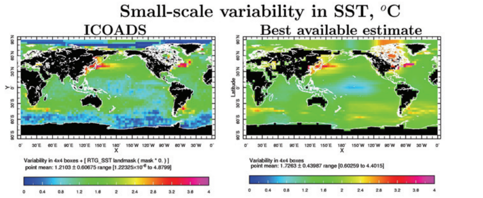


6.SMALL-SCALE VARIANCE

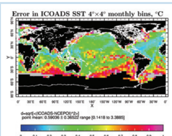
Small-Scale Variability Estimates

Traditional estimates [*Trenberth et al.* 1992, *Kaplan et al.*, 1997] of the small-scale variability (SSV) for the SST from ship data in the ICOADS data base [*Woodruff et al.* 1987] gives a noisy pattern with an obvious impact from sampling error in areas with poor coverage. Our recent attempts to compile a better estimate [*Kaplan et al.*, in prep.] use a few different sources for estimating the SSV: they combine estimates for the variability on the scales longer than 1^o and 1 week from the analysis by *Reynolds and Smith* [1994], which blends AVHRR and in situ data; variability on scales longer than 0.5^o and 1 day is taken from the RTG analysis by *Thiebaux et al.*, [2001], which uses various types of satellite data and is only available since 2001; yet shorter scale variability is inferred from *Kent et al.* [1999]. The resultant field is smoother and in general larger (particularly in high latitudes) than that from ICOADS. Unfortunately, it still underestimates the SSV.

Small-scale variability in SST, ^{o}C



Estimates of standard deviation of SSV in SST, ^{o}C, for $4^o \times 4^o$ monthly bins. Shown are: (left) estimate from in situ data (ICOADS) only; (right) best available estimate.



Computed as r = RMS[ICOADS − NCEPOI].

We took advantage of the 20 years of high quality *Reynolds and Smith* [1994] data, which blends AVHRR and in situ data, in order to create a benchmark 4^o resolution monthly data set of SST anomalies (1982-2001). This data set was compared with bin averages from raw ICOADS data. RMS difference between anomalies is shown in the figure.

7.SSV CONSISTENCY CHECK

Using an approximate formula for the expected error in the mean of $N_{\rm obs}$ observations contained in a space×time bin ("superobservations")

$$r = SSV/\sqrt{N_{\rm obs}}$$

[*Trenberth et al.* 1992, *Kaplan et al.*, 1997,2003a], we take

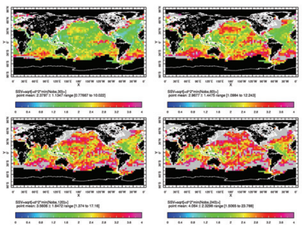
$$r = \text{RMS[ICOADS} - \text{NCEPOI]}$$

to back out the SSV values needed to explain the observed ICOADS–NCEPOI difference

$$SSV = \sqrt{\langle r^2 N_{\rm obs} \rangle}.$$

These values of SSV would include in situ measurement error, as a zero-scale contribution. While formulas above might be in error to some extent because of the possibility of a possibility of remaining satellite bias error and correlated in situ error, Figures suggest that the SSV must be much larger than our best estimates of SSV. Underestimation of SSV can have many explanations: smoothing of SSV by the [*Reynolds and Smith*, 1994] and [*Thiebaux et al.* 2001] analyses procedures, undersampling of the SSV by the ship data used by [*Kent et al.* 1999], etc. While it is possible to study in detail each of these explanations, the availability of densely sampled satellit SST data presents a unique opportunity to obtain accurate estimates of SSV in the SST fields.

SSV estimates required for consistency, ^{o}C



Estimates of standard deviation in SSV, ^{o}C, needed for consistency with ICOADS–NCEPOI difference. Computed as $SSV = \sqrt{\langle r^2 \min(N_{\rm obs}, N_{\rm cap}) \rangle}$ with values of $N_{\rm cap} = 30, 60, 120, 240$.

8.REFERENCES

Kaplan, A., Y. Kushnir, M. Cane, and M. Blumenthal, 1997: Reduced space optimal analysis for historical datasets: 136 years of Atlantic sea surface temperatures, *J. Geophys. Res.*, **102**, 27835–27860.

Kaplan, A., M. Cane, Y. Kushnir, A. Clement, M. Blumenthal, and B. Rajagopalan, 1998: Analyses of global sea surface temperature 1856-1991, *J. Geophys. Res.*, **103**, 18567-18589.

Kaplan A., M.A. Cane, and Y. Kushnir, 2003: Reduced space approach to the optimal analysis interpolation of historical marine observations: Accomplishments, difficulties, and prospects, in *Advances in the Applications of Marine Climatology: The Dynamic Part of the WMO Guide to the Applications of Marine Climatology*, WMO/TD-1081, World Meteorological Organization, Geneva, Switzerland, pp. 199-216

Kent, E.C., P. Challenor, and P. Taylor, 1999: A statistical determination of the random observational errors present in voluntary observing ships meteorological reports. *Journal of Atmospheric and Oceanic Technology*, **16**, 905-914.

Leith, C.E., 1973: The standard error of time-average estimates of climatic means. *Journal of Applied Meteorology*, **12**, 1066–1069.

Rayner, N.A., D.E. Parker, E.B. Horton, C.K. Folland, L.V. Alexander, D.P. Rowell, E.C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century *J. Geophys. Res.*, **108**(14), 10.1029/2002JD002670.

Reynolds, R. W. and T. M. Smith, 1994: Improved global sea surface temperature analyses using optimum interpolation, *J. Climate*, **7**, 929-948.

Schneider, T., 2001: Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values. *J. Climate*, **14**, 853-871.

Thiebaux, H. J., B. Katz, and W. Wang, 2001: New sea surface temperature analys is implemented at NCEP. *WCRP/SCOR Workshop on Intercomparison and Validation of Ocean-Atmosphere Flux Fields* (Bolger Center, Potomac, MD, USA, 21-24 May, 2001), WCRP-115, WMO/TD-No. 1083., pp. 90-94. [Available at `www.soc.soton.ac.uk/JRD/MET/WGASF/workshop/PDF/1p5Thiebaux.doc.pdf`].

Woodruff, S.D., R.J. Slutz, R.L. Jenne, and P.M. Steurer, 1987: A comprehensive ocean–atmosphere data set. *Bull. Amer. Meteor. Soc.*, **68**, 521-527.