



# A method to generate fully multi-scale optimal interpolation by combining efficient single process analyses, illustrated by a DINEOF analysis spiced with a local optimal interpolation

J.-M. Beckers<sup>1,3</sup>, A. Barth<sup>1,2</sup>, I. Tomazic<sup>1</sup>, and A. Alvera-Azcárate<sup>1</sup>

<sup>1</sup>GeoHydrodynamics and Environment Research, MARE, University of Liège, Sart-Tilman B5, 4000 Liège, Belgium

<sup>2</sup>Research Associate, F.R.S.-FNRS, Liège, Belgium

<sup>3</sup>Honorary Research Associate, F.R.S.-FNRS, Liège, Belgium

Correspondence to: J.-M. Beckers (jm.beckers@ulg.ac.be)

Received: 7 March 2014 – Published in Ocean Sci. Discuss.: 18 March 2014

Revised: 19 September 2014 – Accepted: 22 September 2014 – Published: 30 October 2014

**Abstract.** We present a method in which the optimal interpolation of multi-scale processes can be expanded into a succession of simpler interpolations. First, we prove how the optimal analysis of a superposition of two processes can be obtained by different mathematical formulations involving iterations and analysis focusing on a single process. From the different mathematical equivalent formulations, we then select the most efficient ones by analyzing the behavior of the different possibilities in a simple and well-controlled test case. The clear guidelines deduced from this experiment are then applied to a real situation in which we combine large-scale analysis of hourly Spinning Enhanced Visible and Infrared Imager (SEVIRI) satellite images using data interpolating empirical orthogonal functions (DINEOF) with a local optimal interpolation using a Gaussian covariance. It is shown that the optimal combination indeed provides the best reconstruction and can therefore be exploited to extract the maximum amount of useful information from the original data.

## 1 Introduction

Optimal interpolation (in the following noted OI) is well established (e.g., Gandin, 1965; Delhomme, 1978; Bretherton et al., 1976) and a reference tool when analyzing satellite images. The method has therefore been applied in a large number of scientific studies (e.g., Kawai et al., 2006) and operational setups (e.g., Stark et al., 2007; Donlon et al.,

2012; Nardelli et al., 2013). To be optimal, the method requires the correct specification of covariance matrices, most of the time given by parametric functions (e.g., Reynolds and Smith, 1994). Another approach for analyzing a set of satellite images (data interpolating empirical orthogonal functions – DINEOF) uses the data to create a truncated empirical orthogonal function (EOF) representation of the data set to fill in missing data (e.g., Beckers and Rixen, 2003; Alvera-Azcárate et al., 2005, 2007). The latter method has been favorably compared to OI and has been exploited in a series of applications (e.g., Sheng et al., 2009; Ganzedo et al., 2011; Nikolaidis et al., 2014; Wang and Liu, 2014), including operational setups (e.g., Volpe et al., 2012). In some situations, however, the truncation of the EOFs series can reject some interesting small-scale features that only give a small contribution to the total variance, and that can often be split into several modes (Sirjacobs et al., 2008). This is due, on one hand, to the fact EOF truncation is related to the percentage of variance that would be associated with noise and, on the other hand, to the limits of EOF decomposition itself in identifying evolving mesoscale features in a single mode (actually, any feature propagating across the spatial domain is split into several modes) and under clouds. Hence, there might remain small-scale information not fully exploited. One natural approach would be to first analyze the large scales with DINEOF and then add an analysis of the residuals using a local optimal interpolation. We will show that this simple approach provides good results but is still suboptimal. The purpose of the present paper is thus to recover additional pieces of

information in the reconstruction process by optimally combining the DINEOF approach with a local small-scale optimal interpolation approach. The method we will propose is however more general in the sense that we will show how to deal with a situation in which a combination of processes at different scales needs to be analyzed, possibly using different software optimized to analyze a specific process. We will then illustrate the approach in the particular situation with DINEOF analyses for larger scales and OI for smaller scales.

Section 2 will present the theory of optimally combining the analyses of two processes. Then a synthetic example in Sect. 3 will serve as a test bed to provide guidelines in selecting an analysis order among a series of possible options identified. The algorithm called DINEOFI is then spelled out for the combination of DINEOF and OI in Sect. 4. Finally, the application of the method to real data with a validation exercise will be presented in Sect. 5.

## 2 Multi-scale optimal interpolation

We will present the method in an optimal interpolation framework and use the following notations and assumptions. Without loss of generality, data are considered anomalies with respect to a reference field and are stored in vector  $\mathbf{d}$ . The observations are considered imperfect or as including representativity errors so that the observational error is characterized by covariance matrix  $\mathbf{R}$ . The analysis will generally be performed on regular grids, hence in locations not necessarily coinciding with the observation points. The so-called observing operator  $\mathbf{H}$  (e.g., Kalnay, 2002) allows the definition of how to retrieve the analysis values at the observational points from the analysis grid. For filling pixels in satellite images, the application of this matrix to the analysis containing all pixels would simply retrieve the pixels where data were originally present.

The field we try to recover is considered to be created by different independent processes with each process characterized by a specific covariance. On the analysis grid, each process therefore defines a different background covariance matrix. We note these matrices  $\mathbf{B}_s$  for a specific process  $s$ . Since we are in the presence of different independent processes, each one leading to a different covariance matrix  $\mathbf{B}_s$ , the covariance of the total field is  $\mathbf{B} = \sum_s \mathbf{B}_s$  and includes the contribution of processes at all scales (e.g., Wackernagel, 2003). Here we start with the premise that the covariances have already been chosen as there is a wide range of literature on how to model and calibrate covariances (using parametric functions, reduced rank representation, ensemble approaches, smoothness or diffusion operators, see for example Gasparini and Cohn, 1999 and Weaver and Mirouze, 2013). Hence, the purpose of the present paper is not to defend or propose one particular version.

In this case, with covariances defined, the best analysis  $\phi$  of data  $\mathbf{d}$  including all scales in the sense of optimal

interpolation (e.g., Cushman-Roisin and Beckers, 2011) would be given by  $\phi = \bar{\mathbf{K}}\mathbf{d}$  with gain matrix

$$\begin{aligned} \bar{\mathbf{K}} &= \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \\ &= \sum_s \mathbf{B}_s\mathbf{H}^T \left( \sum_i \mathbf{H}\mathbf{B}_i\mathbf{H}^T + \mathbf{R} \right)^{-1} = \sum_s \bar{\mathbf{K}}_s \end{aligned} \quad (1)$$

and

$$\bar{\mathbf{K}}_s = \mathbf{B}_s\mathbf{H}^T \left( \sum_i \mathbf{H}\mathbf{B}_i\mathbf{H}^T + \mathbf{R} \right)^{-1}. \quad (2)$$

At this stage, we observe that the optimal interpolation demands the inversion of  $(\sum_i \mathbf{H}\mathbf{B}_i\mathbf{H}^T + \mathbf{R})$ . If this can be achieved for the chosen representation of the different  $\mathbf{B}_s$ , then the problem of the multi-scale analysis is solved since the optimal interpolation is provided by Eq. (1).

However, sometimes some of the individual matrices  $\mathbf{B}_s$  are not explicitly calculated (for example, in spline methods, e.g., Brasseur et al., 1996, or 3DVar-NMC implementations, e.g., Parrish and Derber, 1992; Fisher, 2003) and cannot be added to other matrices before inversion. In other cases, using a single  $\mathbf{B}_s$  can lead to very efficient inversions (for example, when reduced rank approaches are possible, e.g., Kaplan et al., 2000; Beckers et al., 2006, or localization is used, e.g., Reynolds and Smith, 1994) while when trying the inversion with the sum of  $\mathbf{B}_s$ , one cannot exploit the particular structure of the individual  $\mathbf{B}_s$  anymore. In such situations, we can therefore only suppose that we have efficient tools to calculate  $\mathbf{K}_s\mathbf{x}$ , i.e., we are able to apply an analysis tool with a single specific background covariance  $\mathbf{B}_s$  to any data array  $\mathbf{x}$ . Formally, the Kalman gain matrix for the sole process  $s$  is<sup>1</sup>

$$\mathbf{K}_s = \mathbf{B}_s\mathbf{H}^T(\mathbf{H}\mathbf{B}_s\mathbf{H}^T + \mathbf{R})^{-1}. \quad (3)$$

Hence, the problem is to find a way to calculate  $\bar{\mathbf{K}}_s\mathbf{x}$  when we can only calculate  $\mathbf{K}_s\mathbf{x}$ .

Here we present the solution for two processes, but by recursion more scales can be taken into consideration. If we have only two processes, we can provide the optimal analysis for each one of them as  $\bar{\mathbf{K}}_1\mathbf{d}$  and  $\bar{\mathbf{K}}_2\mathbf{d}$  and also the overall optimal interpolation  $\bar{\mathbf{K}}_1\mathbf{d} + \bar{\mathbf{K}}_2\mathbf{d}$  exploiting the following matrix identities (proof in Appendix A):

$$\bar{\mathbf{K}}_1 = \mathbf{K}_1 - \mathbf{K}_1\mathbf{H}(\mathbf{I} - \mathbf{K}_2\mathbf{H}\mathbf{K}_1\mathbf{H})^{-1}\mathbf{K}_2(\mathbf{I} - \mathbf{H}\mathbf{K}_1), \quad (4)$$

$$= \mathbf{K}_1 - \mathbf{K}_1\mathbf{H}\mathbf{K}_2(\mathbf{I} - \mathbf{H}\mathbf{K}_1\mathbf{H}\mathbf{K}_2)^{-1}(\mathbf{I} - \mathbf{H}\mathbf{K}_1) \quad (5)$$

Method P1a,

$$= (\mathbf{I} - \mathbf{K}_1\mathbf{H}\mathbf{K}_2\mathbf{H})^{-1}\mathbf{K}_1(\mathbf{I} - \mathbf{H}\mathbf{K}_2), \quad (6)$$

$$= \mathbf{K}_1(\mathbf{I} - \mathbf{H}\mathbf{K}_2\mathbf{H}\mathbf{K}_1)^{-1}(\mathbf{I} - \mathbf{H}\mathbf{K}_2) \quad \text{Method P1b.} \quad (7)$$

Similar relationships hold for  $\bar{\mathbf{K}}_2$  (as we can just interchange indices 1 and 2). We therefore now have a way to calculate

<sup>1</sup>Carefully note the absence of  $-$  to distinguish it from Eq. (2).

the optimal interpolation exploiting the individual analysis tools  $\mathbf{K}_1$  and  $\mathbf{K}_2$ . All formulations are mathematically equivalent but they have the problem that  $(\mathbf{I} - \mathbf{K}_2\mathbf{H}\mathbf{K}_1\mathbf{H})^{-1}$  or similar inversions are needed. These are not accessible or lead to an expensive inversion for the same reasons we did not want to invert a matrix involving the sum of  $\mathbf{B}_s$ . So it seems we have only displaced the problem. However, the important point to observe is that gains  $\mathbf{K}_s\mathbf{H}$  or  $\mathbf{H}\mathbf{K}_s$  are “smaller than 1”<sup>2</sup> and the matrix inversion can therefore be implemented by a series expansion: for  $\mathbf{\Xi}$  smaller than 1, we have the convergent series (e.g., Young, 1981):

$$(\mathbf{I} - \mathbf{\Xi})^{-1} = (\mathbf{I} + \mathbf{\Xi} (\mathbf{I} + \mathbf{\Xi} (\mathbf{I} + \mathbf{\Xi} (\dots)))) \quad (8)$$

When applied to a vector  $\mathbf{x}$ , this immediately provides the algorithm to calculate  $\mathbf{y} = (\mathbf{I} - \mathbf{\Xi})^{-1}\mathbf{x}$  as follows:

---

**Algorithm 1.** Approximate matrix inversion.

---

```

y ← x
Loop
y ← x + Ξ y
End loop
    
```

(9)

---

This will converge to the desired vector  $\mathbf{y}$ . In our case, the calculation of  $\mathbf{\Xi}\mathbf{y}$  only involves successive analyses using  $\mathbf{K}_1$  and  $\mathbf{K}_2$ , which by hypothesis can be done efficiently. Then of course, if we limit the number of iterations, the four formulations (4)–(7) might lead to different results. In fact, it is easy to show that  $n$  iterations used in Eqs. (4) and (5) will yield the same results since

$$\begin{aligned} & (\mathbf{I} + \mathbf{\Xi}_1 + \mathbf{\Xi}_1^2 + \dots + \mathbf{\Xi}_1^n) \mathbf{K}_2 \\ &= \mathbf{K}_2 (\mathbf{I} + \mathbf{\Xi}_2 + \mathbf{\Xi}_2^2 + \dots + \mathbf{\Xi}_2^n), \end{aligned} \quad (10)$$

with  $\mathbf{\Xi}_1 = \mathbf{K}_2\mathbf{H}\mathbf{K}_1\mathbf{H}$  and  $\mathbf{\Xi}_2 = \mathbf{H}\mathbf{K}_1\mathbf{H}\mathbf{K}_2$ . Similarly, iterative versions of Eqs. (6) and (7) will also lead to the same analysis. The only difference is possible in terms of computational load. Since generally the number of observations is much smaller than the number of points on the output grid of the analysis, using Eqs. (5) and (7) during iterations will work on smaller files/arrays and also could benefit from quicker analysis if the analysis tool is faster when asked to output values only at data locations. Hence, Eqs. (5) and (7) are two formulations which we will retain and which will differ if iterations on the inversion of  $(\mathbf{I} - \mathbf{H}\mathbf{K}_1\mathbf{H}\mathbf{K}_2)$  or  $(\mathbf{I} - \mathbf{H}\mathbf{K}_2\mathbf{H}\mathbf{K}_1)$  are not pursued until convergence. For  $\bar{\mathbf{K}}_2$ ,

---

<sup>2</sup>A formal definition of “smaller than 1” for the matrices demands the module of their eigenvalues being all smaller than 1; see page 127 in Daley (1993) for a proof that  $\mathbf{H}\mathbf{K}$ , which is also sometimes called the hat matrix or influence matrix, has eigenvalues between 0 and 1.

we now write out the two equivalent relevant versions:

$$\bar{\mathbf{K}}_2 = \mathbf{K}_2 - \mathbf{K}_2\mathbf{H}\mathbf{K}_1(\mathbf{I} - \mathbf{H}\mathbf{K}_2\mathbf{H}\mathbf{K}_1)^{-1}(\mathbf{I} - \mathbf{H}\mathbf{K}_2) \quad (11)$$

Method P2a,

$$= \mathbf{K}_2(\mathbf{I} - \mathbf{H}\mathbf{K}_1\mathbf{H}\mathbf{K}_2)^{-1}(\mathbf{I} - \mathbf{H}\mathbf{K}_1) \quad (12)$$

Method P2b,

The cost for calculating an analysis with  $\bar{\mathbf{K}}_1$  or  $\bar{\mathbf{K}}_2$  using  $n$  iterations is  $3 + 2n$  analyses using  $\mathbf{K}_1$  and  $\mathbf{K}_2$  in Eqs. (5) or (11), whereas the use of Eqs. (7) or (12) demands  $2 + 2n$  standard analyses. In the following, P1a and P2a will refer respectively to the application of Eqs. (5) and (11) with iterations, and P1b and P2b refer respectively to the use of Eqs. (7) or (12).

We also note that even when using iterated versions, using Eqs. (5) and (12) leads to

$$\bar{\mathbf{K}}_1 = \mathbf{K}_1 (\mathbf{I} - \mathbf{H}\bar{\mathbf{K}}_2), \quad (13)$$

which will be exploited in our final algorithm.

$\bar{\mathbf{K}}_1$  and  $\bar{\mathbf{K}}_2$  allow us to calculate the best analysis for each of the processes. A natural but suboptimal estimate for each process would of course be obtained when using  $\mathbf{K}_1$  for process 1 and  $\mathbf{K}_2$  and for process 2. A slightly more elaborated idea would rather be to use  $\mathbf{K}_1(\mathbf{I} - \mathbf{H}\bar{\mathbf{K}}_2)$  and  $\mathbf{K}_2(\mathbf{I} - \mathbf{H}\bar{\mathbf{K}}_1)$  so as to get the analysis of one process only using data from which we tried to subtract the information from the other process. Therefore, we can also design several simple ways to sub-optimally estimate the overall analysis using the following gain matrices:

$$\mathbf{S}_{00} = \mathbf{K}_1 + \mathbf{K}_2, \quad (14)$$

$$\mathbf{S}_{01} = \mathbf{K}_1 + \mathbf{K}_2(\mathbf{I} - \mathbf{H}\mathbf{K}_1), \quad (15)$$

$$\mathbf{S}_{10} = \mathbf{K}_1(\mathbf{I} - \mathbf{H}\mathbf{K}_2) + \mathbf{K}_2, \quad (16)$$

$$\mathbf{S}_{11} = \mathbf{K}_1(\mathbf{I} - \mathbf{H}\mathbf{K}_2) + \mathbf{K}_2(\mathbf{I} - \mathbf{H}\mathbf{K}_1). \quad (17)$$

These simple versions ( $\mathbf{S}_{**}$ ) can also be compared later to the iterated versions.

### Error covariance fields

The estimation of uncertainties is very valuable information that is to be added to the analysis itself. The formal derivation of it is well established for the general case: the analysis error covariance matrix  $\mathbf{P}^a$  when using any gain matrix  $\mathbf{K}$  is given by (e.g., Cushman-Roisin and Beckers, 2011)

$$\mathbf{P}^a = \mathbf{B} - \mathbf{B}\mathbf{H}^T\mathbf{Z}^{-1}\mathbf{H}\mathbf{B} \quad (18)$$

$$+ (\mathbf{B}\mathbf{H}^T - \mathbf{K}\mathbf{Z})\mathbf{Z}^{-1}(\mathbf{H}\mathbf{B} - \mathbf{Z}\mathbf{K}^T),$$

$$\mathbf{Z} = (\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}). \quad (19)$$

When the gain is optimal, the last term vanishes and we get the minimal error of the optimal interpolation approach. Hence, we can also assess the errors of the different analyses we propose.

For example, if we are interested in the analysis of process 1 and use the exact form of  $\bar{\mathbf{K}}_1$ , the expected error covariance on the analysis of process 1 reads

$$\mathbf{P}_1^a = \mathbf{B}_1 - \bar{\mathbf{K}}_1 \mathbf{H} \mathbf{B}_1. \quad (20)$$

Similarly, the error of the analysis of process 2 alone when using the exact form of  $\bar{\mathbf{K}}_2$  reads

$$\mathbf{P}_2^a = \mathbf{B}_2 - \bar{\mathbf{K}}_2 \mathbf{H} \mathbf{B}_2. \quad (21)$$

If we are interested in the overall analysis including both processes, the error on the sum of the two analyses is

$$\mathbf{P}^a = \mathbf{B}_1 + \mathbf{B}_2 - (\bar{\mathbf{K}}_1 + \bar{\mathbf{K}}_2) (\mathbf{H} \mathbf{B}_1 + \mathbf{H} \mathbf{B}_2) \quad (22)$$

since we used  $\mathbf{K} = \bar{\mathbf{K}}_1 + \bar{\mathbf{K}}_2$ ; we note that the error on the full analysis is not the sum of the errors on each process but will be lower.

In all cases, since we have a way to replace  $\bar{\mathbf{K}}_1$  and  $\bar{\mathbf{K}}_2$  as functions of  $\mathbf{K}_1$  and  $\mathbf{K}_2$ , the error covariance fields can also be assessed in any desired location by applying our new method to a pseudo-data set containing, in fact, covariances.

### 3 Synthetic example

The objective of the synthetic test case is see which formulations are the most interesting, and we do so by using a Monte Carlo approach. In order to allow for a large number of realizations, a simplified framework is used in which only spatial interpolation will be performed instead of a full time–space analysis. This approach is not limiting since the time interpolation amounts to add a dimension to the system. As long as the same types of correlation functions are also used over time, the timescale will play exactly the same role as correlation length, and the conclusions of the spatial interpolation will also hold for the higher dimensional case (see also Nardelli, 2012). Also, the conclusion we will derive from the synthetic case will be shown a posteriori to perform well in the real application including the time dimension.

Focusing on spatial interpolation, we generate two random fields of different covariances (“true states”) on a domain 100 grid points wide, sample the fields, sum them up, add noise and then analyze these pseudo-data with the different formulations. Then we can calculate the squared norm of the difference between the analysis and the true fields (each field individually and the overall field). This exercise is repeated a large number of times (100 000) to have a statistically significant estimate and also error distribution. To simplify the interpretation, the error estimates are scaled by the expected analysis errors Eqs. (20)–(22) so that if the analysis is indeed optimal, the average scaled error should be 1.

We can then look at the quality of the analysis for each individual process and the overall analysis for a series of different values of the parameters describing the covariance structures of each process. These parameters will characterize the

correlation length of the processes and the signal-to-noise ratio (e.g., Troupin et al., 2010): the signal-to-noise ratio for process 1 is defined as  $\lambda_1 = \text{trace}(\mathbf{B}_1) / \text{trace}(\mathbf{R})$  and similar definitions hold for the signal-to-noise ratio for process 2 and the total signal-to-noise ratio.

We made a series of tests with different length scales and signal-to-noise ratios for both processes the results of which are provided in Appendix C for completeness. Interestingly, very clear and robust guidelines emerged.

### Resulting guidelines

From what we saw, the best we can do is to label the process with the highest signal-to-noise ratio as 1 and the other process as 2 and apply P1a + P2b if we want to achieve the best analysis.

If both processes have a similar signal-to-noise ratio, we should label the larger-scale process as process 1 so that the same formulas are still the best.

For simpler analyses Eq. (15) defining  $\mathbf{S}_{01}$  remains a good option with this choice of numbering.

For individual processes, P1a should be used for the iterated versions, whereas  $\mathbf{K}_1$  is the best choice for a simple analysis for process 1; for process 2 version P2b and  $\mathbf{K}_2(\mathbf{I} - \mathbf{H}\mathbf{K}_1)$  are indicated respectively for the iterated approach or a simple approach.

For iterative methods, if scales are well separated or at least one of the processes has a small signal-to-noise ratio, only very few iterations are needed. In this case, the simple formulation can also be competitive.

On the other hand, in difficult situations with overlapping scales and strong signals, the iterated versions can always be made convergent by using more iterations and the optimal interpolation will result. Alternatively if scales are very similar, a single analysis with average correlation length could be used.

### 4 DINEOF + OI

We now come to the application of our ideas to the better reconstruction of satellite images. DINEOF has proven to be efficient, particularly when looking at larger scales since the interpolation exploits dominant large-scale EOF structures to reconstruct missing data under clouds. Those EOF structures generally capture a very high percentage of the data variance and in this sense we can assume that the signal-to-noise ratio is high.

If we try to add a local covariance structure with an optimal interpolation approach, the signal-to-noise ratio for this additional process we want to extract is probably much lower. Hence, we are in the situation covered in synthetic case (cases 3 and 4 of Appendix C) and the iterative version of P1a + P2b is the natural choice, where  $\mathbf{K}_1$  stands for the DINEOF-based analysis and  $\mathbf{K}_2$  for the local OI application.

The analyzed field  $\phi$  can be easily obtained from P1a + P2b, called hereafter DINEOFI, programmed as

---

**Algorithm 2.** Combination of two analysis tools for an optimal interpolation.

---

```

 $\phi \leftarrow \mathbf{K}_1 d$ 
 $w_1 \leftarrow \mathbf{H}\phi$  (Use only analysis at data points without clouds)
 $w_1 \leftarrow (d - w_1)$  (Residuals)
    Calibrate or check OI parameters at this point using  $w_1$ 
 $w_2 \leftarrow w_1$  (Start of iterative matrix inversion)
Loop  $n$  times
     $w_2 \leftarrow \mathbf{H}\mathbf{K}_2 w_2$  (Apply tool 2 and retrieve solution at data points)
     $w_2 \leftarrow \mathbf{H}\mathbf{K}_1 w_2$  (Apply tool 1 and retrieve solution at data points)
     $w_2 \leftarrow w_1 + w_2$ 
End loop
 $\omega_1 \leftarrow \mathbf{K}_2 w_2$ 
     $\omega_1$  now contains scale 2 field  $\overline{\mathbf{K}_2 d}$ ; can be saved for other uses
 $\phi \leftarrow \phi + \omega_1$  (Add it to  $\mathbf{K}_1 d$ )
 $w_1 \leftarrow \mathbf{H}\omega_1$ 
 $\omega_1 \leftarrow \mathbf{K}_1 w_1$  (Calculates  $\mathbf{K}_1 \mathbf{H}\overline{\mathbf{K}_2 d}$ )
 $\phi \leftarrow \phi - \omega_1$  (Final analysis)
    If scale 1 solution is requested, subtract saved scale 2 solution from
     $\phi$  and save it
    
```

---

This shows that the simple application of existing analysis tools to data arrays and a few working arrays/files ( $w_1, w_2, \omega_1$ ) will solve our problem and provide the analysis  $\phi$ . This demands  $2 + n$  applications of  $\mathbf{K}_1$  and  $1 + n$  applications of  $\mathbf{K}_2$ .

For the calibration of OI parameters it is probably safe to work on residuals after step 3 (available in intermediate results  $w_1$ ) for fitting or verification of the correlation length and signal-to-noise ratio since we showed in the synthetic case that if scale separation is good,  $\mathbf{K}_2(\mathbf{I} - \mathbf{H}\mathbf{K}_1)d$  is a good proxy of the analysis for process 2.

There remains to formulate how  $\mathbf{K}_1$  and  $\mathbf{K}_2$  are defined. If during step 1 of the algorithm  $\phi$  could be interpreted as the DINEOF result (and hence  $\mathbf{K}_1$  the DINEOF tool applied to the data), in the subsequent part of the algorithm we cannot just replace  $\mathbf{K}_1$  by “apply DINEOF” since the EOFs calculated are data dependent. Hence, when working with residuals in the later steps, we would not use the same covariances as we would generate new covariances from the residuals instead of using the covariances from the large-scale processes. So instead of choosing “apply DINEOF”, we must find a way to exploit the large-scale covariances recovered by the original DINEOF application. For this purpose we can use the EOFs calculated by DINEOF on the raw data to specify covariance matrices as in Beckers et al. (2006). For such an optimal interpolation using EOFs from DINEOF, we recall that DINEOF provides the singular value decomposition (SVD) of the data<sup>3</sup> stored in a  $m \times n$  matrix  $\mathbf{X}$  with  $m$  pixels for each cloud-free image and  $n$  images as

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T. \quad (23)$$

---

<sup>3</sup>We stress again that averages have been subtracted to work with anomalies.

The  $N$  spatial EOFs (stored in the  $m \times N$  matrix  $\mathbf{U}$ ) and temporal EOFs (stored in the  $n \times N$  matrix  $\mathbf{V}$ ) can be used together with the singular values (stored in the diagonal matrix  $\mathbf{\Sigma}$ ) to estimate the spatial covariances

$$\mathbf{B}^{(s)} = \frac{1}{n} \mathbf{U}\mathbf{\Sigma}^2 \mathbf{U}^T. \quad (24)$$

These spatial covariances then allow the spatial interpolating/analysis of each  $n$  image: for a spatial analysis of image  $j$ , the gain matrix reads indeed

$$\mathbf{K}^{(s)j} = \mathbf{B}^{(s)} \mathbf{H}^T \left( \mathbf{H}\mathbf{B}^{(s)} \mathbf{H}^T + \mathbf{R} \right)^{-1}. \quad (25)$$

Note that for each image  $j$ , another observation operator  $\mathbf{H}$  is used as the cloud coverage changes. In order not to overload notations we do not add an index but remember that  $\mathbf{H}$  is image dependent.

The computational efficiency of this method stems from the fact that the application of the Woodbury matrix identity (e.g., Golub and Van Loan, 2012) translates this into a least square fitting of  $N$  modes (e.g., Beckers et al., 2006). Assuming independent and identical observational errors of variance  $\epsilon^2$ , this leads to

$$\mathbf{K}^{(s)j} = \mathbf{L} \left( \epsilon^2 \mathbf{I} + \mathbf{L}_p^T \mathbf{L}_p \right)^{-1} \mathbf{L}_p^T, \quad (26)$$

with  $\mathbf{L} = \mathbf{U}\mathbf{\Sigma}/\sqrt{n}$  and the corresponding version for the data present in each image  $\mathbf{L}_p = \mathbf{H}\mathbf{L}$  (note that here  $\mathbf{H}$  still changes for each image). The computational efficiency is now understood as for each image the matrix to invert is of size  $N \times N$  when using Eq. (26) instead of  $p \times p$  when using Eq. (25), with the number  $N$  of retained EOFs much smaller than the number  $p$  of pixels in an image. So from the EOFs found by DINEOF, we can perform spatial interpolations at low cost yet retaining the original spatial covariances.

By symmetry we can also choose to perform a temporal interpolation of each of the  $m$  pixels by using for pixel  $i$  the time interpolation from covariance

$$\mathbf{B}^{(t)} = \frac{1}{m} \mathbf{V}\mathbf{\Sigma}^2 \mathbf{V}^T \quad (27)$$

and use the computationally efficiently calculated gain matrix

$$\mathbf{K}^{(t)i} = \mathbf{T} \left( \epsilon^2 \mathbf{I} + \mathbf{T}_p^T \mathbf{T}_p \right)^{-1} \mathbf{T}_p^T, \quad (28)$$

with  $\mathbf{T} = \mathbf{V}\mathbf{\Sigma}/\sqrt{m}$  and  $\mathbf{T}_p = \mathbf{H}\mathbf{T}$  (here  $\mathbf{H}$  changes for each of the pixels and selects the images in which the pixel is not clouded).

Then we can further improve the covariance model exploiting both temporal and spatial correlations as follows: instead of choosing one or the other method (spatial or temporal interpolation alone) we can again apply the basic idea of the present paper allowing the combination of two analysis

tools. We can indeed use as a covariance the combination of the spatial and temporal covariances and for a pixel  $i$  of image  $j$ , the covariance with a pixel  $i'$  of image  $j'$  is modeled as  $\mathbf{B}_{ii'}^{(s)}\delta_{jj'} + \mathbf{B}_{jj'}^{(t)}\delta_{ii'}$  with the standard Kronecker<sup>4</sup> symbol  $\delta_{ii'}$ . In other words, as we are in the presence of the combination of two covariances, without actually creating the sum and full gain matrix, we can use again the iterated versions P1a + P2b where process 1 is now provided by the spatial interpolation tool and process 2 by the temporal interpolation tool. This “inner” iterative combination provides then our EOF-based analysis tool  $\mathbf{K}_1$  for the “outer” combination method with a local OI.

So if it is clear that in the later steps of Algorithm 2 we use  $\mathbf{K}_1$  as just described (the combination of spatial and temporal EOF-based covariances), the first step could still be taken as the DINEOF analysis as it could be regarded as some kind of best first guess. This option will be looked at later.

For  $\mathbf{K}_2$ , a standard optimal interpolation tool is used. In practice, the OI application for satellite images repeated at the same locations can be strongly optimized since it is easy to collect the data points which are within reach of the covariance function at any given point in which the analysis is demanded. As the correlation length will be small, the number of data points involved will also be small and the matrix inversion very quick for each analysis point. The typical parametric correlation function  $c$  used in optimal interpolation is a Gaussian

$$c = e^{-\left(\frac{\delta x}{L_x}\right)^2} e^{-\left(\frac{\delta y}{L_y}\right)^2} e^{-\left(\frac{\delta t}{T}\right)^2}, \quad (29)$$

where  $\delta x$ ,  $\delta y$  and  $\delta t$  are the difference in space and time coordinates between the two points for which we want to calculate the covariance.  $L_x$  and  $L_y$  are then the correlation length scale in  $x$  and  $y$  direction and  $T$  the correlation timescale. The correlation function has to be multiplied by variance  $\sigma_2^2$  to provide  $\mathbf{B}_2$ . For the implementation of this optimal interpolation we used the open-source code `optiminterp` available at <http://octave.sourceforge.net/optiminterp/overview.html><sup>5</sup>. Its use was computationally optimized by exploiting the fact that for each point in which the analysis is demanded the gridded nature of the data points allows a direct selection of possibly involved data points and a rejection of all others. So it means one can feed the analysis only with data in a box of size  $4L_x \times 4L_y \times 4T$  centered around the analysis point, which greatly reduces sorting and selecting time and also allows for a parallel execution.

## 5 Test case

For the real case application, we use an interesting sea surface temperature (SST) hourly data set derived from

<sup>4</sup> $\delta_{ii'}$  is 1 when  $i = i'$ , 0 otherwise.

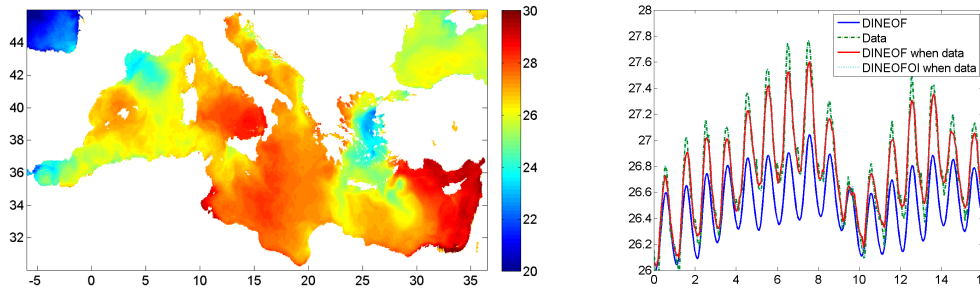
<sup>5</sup>Or [http://modb.oce.ulg.ac.be/mediawiki/index.php/Optimal\\_interpolation\\_Fortran\\_module\\_with\\_Octave\\_interface](http://modb.oce.ulg.ac.be/mediawiki/index.php/Optimal_interpolation_Fortran_module_with_Octave_interface)

Spinning Enhanced Visible and Infrared Imager (SEVIRI) onboard the Meteosat Second Generation (MSG) geostationary satellite and produced operationally by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) Ocean and Sea Ice Satellite Application Facility (OSI-SAF) at Meteo-France/Centre de Meteorologie Spatial (CMS) in Lannion (EUMETSAT, 2011; Le Borgne et al., 2011). Because of its geostationary orbit, high-frequency sampling is available and analyses resolving the daily cycle appeared (e.g., Marullo et al., 2010, 2014). Some high-frequency content is expected to be filtered by standard DINEOF applications and therefore the data should provide a good test bed for our new method. Furthermore, capturing the diurnal cycle in using satellite-derived sea surface temperature is considered an important challenge (Stuart-Menteth et al., 2003) and several studies have been performed recently using hourly data (Karagali et al., 2012; Le Borgne et al., 2012; Eastwood et al., 2011).

Excellent results when using SEVIRI data have been obtained (Marullo et al., 2014): by combining a dynamical model resolving the daily cycle with the SEVIRI data, their reconstruction of hourly data covering the whole Mediterranean Sea during the 3-month summer period of 2011 showed rms (root mean square) errors compared to in situ data of 0.64 K with offshore moorings and 0.47 K with drifter data, with a well-resolved daily cycle. A synthetic cloud cross-validation exercise also provided an rms difference of only 0.16 K between the artificially clouded SEVIRI points and their reconstruction. In their case, the artificial clouds were constructed by moving a 200 km wide meridional band of “clouds” westward across the domain over around 24 h so that no persistent artificial clouds were included.

For our study focusing on the methodology rather the reconstruction itself, we selected a smaller data set, covering the period of 1–16 August 2013. This data set has some clear images with daily cycles on which artificial clouds can be added for validation purposes. The spatial resolution of the data is  $0.05^\circ$ , available at an hourly frequency, leading to  $n = 16 \times 24 = 384$  images. In all cases, we used all data with quality flags 4 or 5 (acceptable and best quality). In order to assess the robustness of the method and the conclusions of this experiment, we look at four different cases.

Two different spatial coverages were considered, one looking only at the western Mediterranean Sea, the other to the whole Mediterranean Sea. This approach permits checking the robustness of the statistical results and the effect of different EOF structures, and also shows the effect of larger data variability. For the western Mediterranean Sea, we retrieved images of  $327 \times 217 = 70\,959$  pixels from which 29 380 fall on land, so that the maximum number of pixels in each image useful for analysis is  $m = 41\,579$ . Without additional artificial clouds, this data set presents an average cloud coverage of 29.70 %. For the whole Mediterranean Sea, images have  $850 \times 320$  pixels, from which 120 343 fall onto the sea so that  $m = 151\,566$  (see Fig. 1). The average cloud coverage is



**Figure 1.** Mean spatial structure of standard DINEOF analysis (left panel) and spatially averaged time evolution (right panel) of the analyzed field (solid blue line), the original data (dashed green line) and the DINEOF analysis at pixels with data present (red line). The lower temperatures in the DINEOF reconstruction are simply due to the fact that mostly cold regions were covered by clouds and the data average only used available (warmer) pixels. More importantly, we see however that the DINEOF reconstruction has a lower-amplitude cycle than the data. With the new methods DINEOFOI explained later, we recover the amplitudes found in the data (dotted line coinciding with the dashed line).

23.53 % in this case. We also note that for the setup over the whole Mediterranean Sea, data from the Black Sea and the Atlantic Ocean are included, making the test more challenging as the EOF structures will need to cope with different, mostly unconnected, variabilities.

For cross-validation purposes, additional clouds will be added to the original data. The so-called cross-validation data are never used in the calculations (neither in the DINEOF decomposition nor the different optimal interpolations) but allow for the computation of metrics characterizing the quality of the analysis under clouds. The rms error  $\delta_D$  of the DINEOF reconstruction under these additional clouds can be used as a reference error which we want to improve. We can then define a skill-score  $S$  as

$$S = 1 - \frac{\delta^2}{\delta_D^2}, \quad (30)$$

which measures the relative increase in precision (a 0 skill-score value indicates no better results than DINEOF whereas a value of 1 means a perfect reconstruction). A second standard metric for the validation is the correlation coefficient between the analysis and the cross-validation points noted  $r$ .

Two different kinds of artificial cloud sets allows one to check the effect of more or less persistent cloud coverage on the reconstruction quality and the improvements brought by our new method. In one situation, clouds from the last 2 weeks of July 2013 are superimposed onto the already present clouds and the pixels masked in this way taken aside as cross-validation data. In the other situation, we generate artificial clouds as in Marullo et al. (2014): the slab of clouds moves from east to west across the domain in about 1 day and then starts again from the east. The fraction of overlap between successive images was taken similar to the one in Marullo et al. (2014).

From these two different spatial coverages and two different cloud coverages we create our four test cases:

### Setup 1: western Mediterranean Sea, added July clouds

For cross-validation purposes, clouds from the last 2 weeks of July were superimposed, resulting in total average cloud coverage which increases from 29.70 to 42.06 %. The data set is then analyzed with DINEOF, providing  $N = 38$  EOFs, the singular values and the DINEOF reconstruction explaining 91 % of the variance. As expected, DINEOF captures a large part of the variability. A first diagnostic can be performed by calculating the root mean square difference between this reconstruction and the artificially clouded pixels (the cross-validation points, here we have 1 973 438 of them), and we obtain an associated standard deviation  $\delta_D = 0.496$  K close to the noise level in the data, as the SEVIRI SST standard deviation is around 0.5 K (Brisson et al., 2002). For DINEOF we obtain  $r = 0.9371$ , which is therefore the baseline to improve upon in this case.

For this case and the following cases, the metrics of the standard DINEOF application just mentioned are summarized in Table 1 and for the following cases we only describe the cross-validation approach.

### Setup 2: western Mediterranean Sea, added banded clouds

In the first setup, analyses under persistent large-scale clouds can of course not be improved by the local optimal interpolation. To check if we can reconstruct data under fast-moving clouds, we take a band of westward-moving “clouds”.

### Setup 3: whole Mediterranean Sea, added July clouds

To check how the reconstructions behave when a larger range of observations are present, we look at the whole Mediterranean Sea. The full Mediterranean Sea case exhibits generally less clouds (23.53 %); with the July clouds added to the original data, the cloud coverage increases to 35.29 %.

**Table 1.** Skill scores and correlation coefficients for the four SEVIRI setups and the different methods. DINEOF stands for the direct use of standard DINEOF results. DINEOF +  $\mathbf{K}_2$  uses the simple approach of adding a local optimal interpolation of residuals. DINEOF + 10 it stands for the method in which we start from DINEOF results and then apply the iterative version using  $\mathbf{K}_1$  based on the EOFs. Version  $\mathbf{K}_1$  uses only the EOF-based covariances,  $\mathbf{S}_{01}$  stands for the simple approach when we add the local optimal interpolation of residuals to the previous solution and finally the last two lines correspond with the full method with no iterations or 10 iterations (for inner iterations within  $\mathbf{K}_1$  combining space- and time-EOF-based interpolations, iterations are maintained in all cases).

	Setup 1		Setup 2		Setup 3		Setup 4	
	$S$	$r$	$S$	$r$	$S$	$r$	$S$	$r$
$\delta_D$	0.496 K		0.308 K		0.492 K		0.285 K	
Cloud fraction	42.06 %		34.09 %		35.29 %		27.02 %	
$N$ EOF modes	38		37		38		40	
DINEOF	0.00	0.9371	0.00	0.9699	0.00	0.9707	0.00	0.9824
DINEOF + $\mathbf{K}_2$	0.21	0.9507	0.45	0.9835	0.21	0.9769	0.40	0.9895
DINEOF + 10 it	-1.25	0.8476	-1.79	0.9149	-0.51	0.9559	-0.97	0.9650
$\mathbf{K}_1$	0.24	0.9516	0.22	0.9764	0.18	0.9760	0.18	0.9855
$\mathbf{S}_{01}$	0.33	0.9576	0.47	0.9839	0.28	0.9791	0.42	0.9899
DINEOFOI 0 it	0.34	0.9579	0.47	0.9840	0.29	0.9792	0.43	0.9899
DINEOFOI 10 it	0.37	0.9602	0.49	0.9845	0.31	0.9798	0.44	0.9901

#### Setup 4: whole Mediterranean Sea, added banded clouds

Finally, we complete the comparisons by using the whole Mediterranean Sea, but with a fast-moving westward band of “clouds”.

To proceed, we now need to specify the observational error variance and the variance  $\sigma_2^2$  for the local optimal interpolation in each case. As we will show later, the conclusions remain robust in all cases and we only detailed results of the most challenging Setup 3 in Appendix D. The parameter optimization of the other cases was performed in a similar way with similar results.

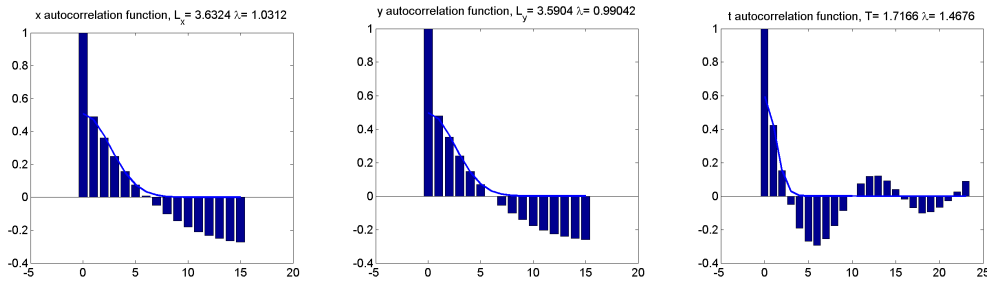
From the calibration we note in Fig. 2 that the time autocorrelation spectrum exhibits an anticorrelation at 6 h and maximum autocorrelation at 12 h. This reflects, in fact, the asymmetry of the diurnal cycle. When analyzing the raw data instead of residuals, the spectrum and autocorrelation (not reproduced here) indeed shows an anticorrelation at 12 h and correlation at 24 h. Thus, the residuals will allow for following the higher-frequency content.

The observational noise variance was fixed as  $0.25 \text{ K}^2$ . A sensitivity analysis was performed, changing this value into the range of 0.1 to  $10 \text{ K}^2$ . In all cases, the conclusion in terms of ranking of the different methods remained the same; only the relative part of the solution picked up by the EOF-based covariances or local optimal interpolation was affected. As the value of  $0.25 \text{ K}^2$  is the typical value associated with SEVIRI data (Brisson et al., 2002), we kept this value even if better skill scores were obtained for other values used in sensitivity analyses.

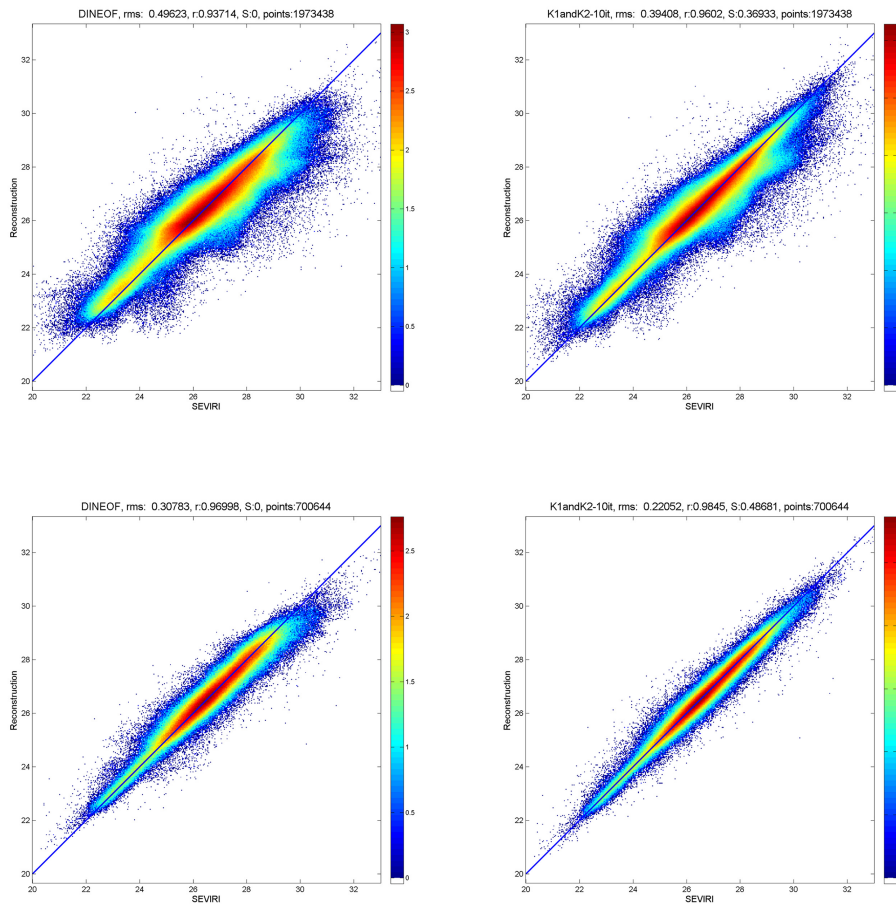
With all parameters defined, we proceed to the optimal interpolation and compare the different techniques to the standard DINEOF solution for the four different setups. As seen from Table 1, robust conclusions can be drawn:

- Results for cases working on a larger domain had better correlations than the cases working on the western Mediterranean Sea alone because of the larger data range, but ranking between the methods remain the same, independently of the domain covered. Conclusions are therefore robust with respect to the domain size.
- Results with fast-moving slabs of cloud are always better than with more realistic clouds, but again the ranking of the methods is not changed between the two situations and conclusions robust with respect to the type of cross-validation clouds added.
- In all cases, our new iterated method provides the best skill scores and correlation coefficient. Using 0 iterations instead of 10 iterations only slightly degrades results and keeps the non-iterated method ranked second in terms of skill scores and correlation coefficient.
- The simple method of using DINEOF followed by a local optimal interpolation of residuals (DINEOF +  $\mathbf{K}_2$ ) also provides good skill scores. In cases with fast-moving artificial clouds, the skill score is close to the one of our new version, but adding real clouds using the new method clearly increases the skill further.
- Starting from DINEOF and then applying iterations with  $\mathbf{K}_1$  based on the DINEOF EOFs (called DINEOF + it) is not a good idea since in all cases negative skills and lower correlations are observed. Hence, in this case the inconsistency of the first step in our algorithm with the rest of the steps deteriorates results.
- Finally, the simple method  $\mathbf{S}_{01}$  also yields better results than DINEOF +  $\mathbf{K}_2$ .





**Figure 2.** Spectrum analysis leading to autocorrelation functions, for longitude, latitude and time (from left to right). Autocorrelation is given as a function of distance expressed in number of grid points or as a function of time-delay expressed in number of time steps (hours). The continuous line corresponds with the fitted Gaussian correlation function.

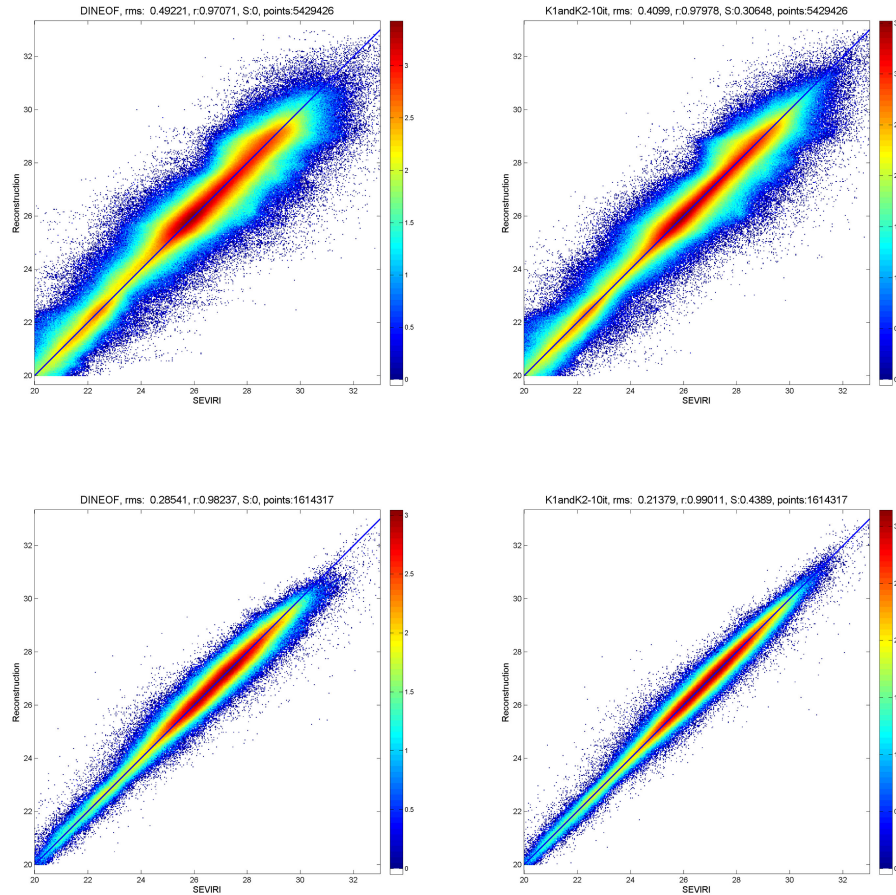


**Figure 3.** Correlation at cross-validation points using DINEOF alone and the new method DINEOFI in Setup 1 (top) and Setup 2 (below).

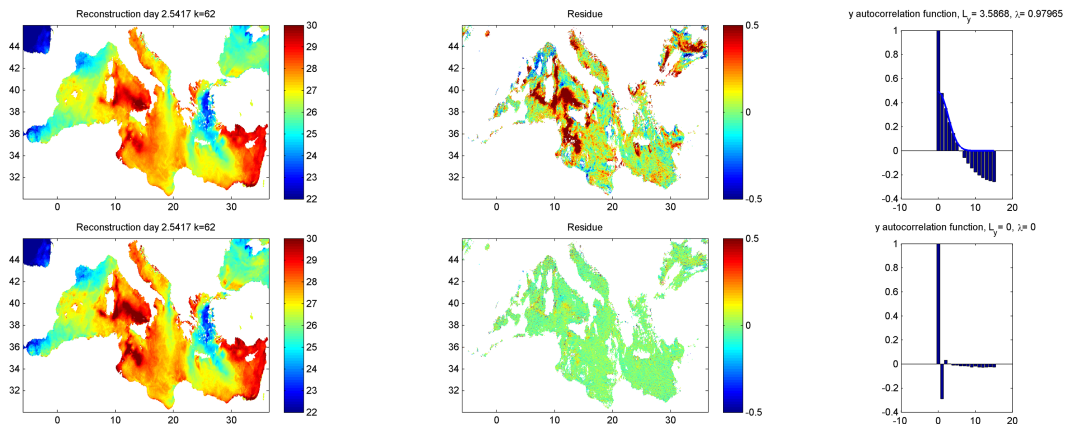
The higher correlation found with the banded fast-moving clouds is also seen in the correlation plots of Figs. 3 and 4. When July clouds are used, some persistent large-scale clouds do not permit improving DINEOF reconstructions in these regions. This leads to larger dispersion. When fast-moving clouds are used, the local optimal interpolation can infer missing information from the surrounding pixels and correlations are much higher. We also notice that even in the July cloud case, the new method filters the extreme

values less than the original DINEOF reconstruction (the S-shaped tendency of DINEOF reconstructions dampening the extreme variations is reduced with the new method).

Since the computational load of the new method is a fraction of the computational cost of the DINEOF decomposition (typically 10 % for the iterated version) it is the natural choice to improve DINEOF results when good skill scores and correlations are sought. If one does not want to program



**Figure 4.** Correlation at cross-validation points using DINEOF alone and the new method DINEOFOI in Setup 3 and Setup 4.



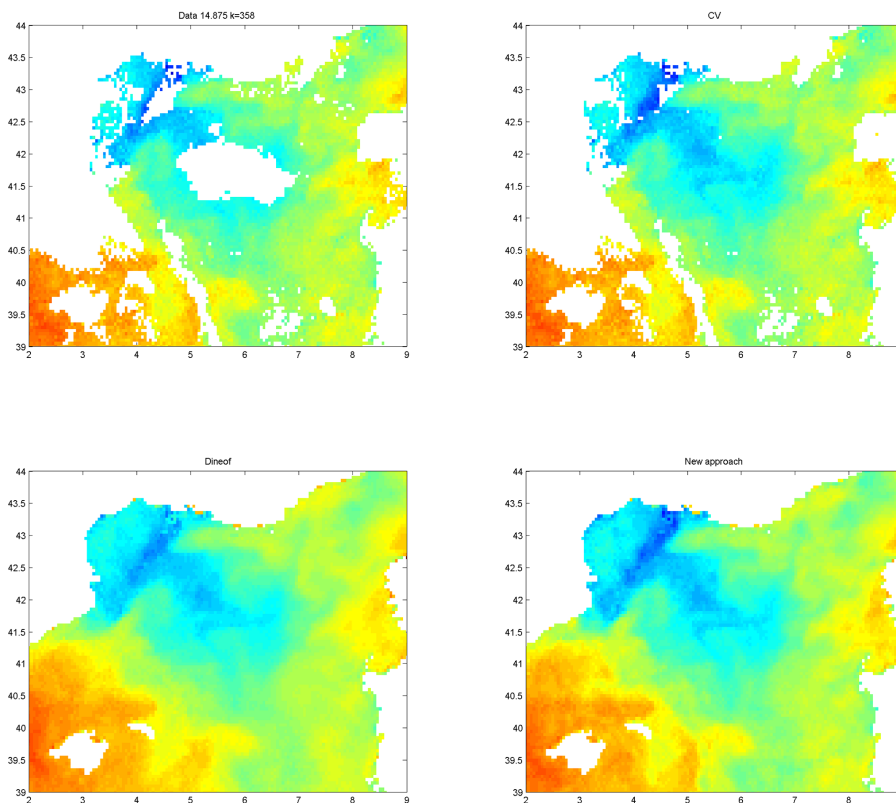
**Figure 5.** DINEOF reconstruction, residual and autocorrelation of residuals. DINEOF (top) and new method DINEOFOI (lower images).

$\mathbf{K}_1$  and iterations<sup>6</sup>, the simple approach of using DINEOF followed by the local optimal interpolation of residuals is a simple alternative.

<sup>6</sup>To avoid reprogramming  $\mathbf{K}_1$  and iterations, the code will be made available at <http://modb.oce.ulg.ac.be/mediawiki/index.php/DINEOF>

Skill scores and correlation coefficients are not the only criteria to select a method, and we can further check if the new method also leads to better feature reconstruction.

Figure 5 shows the analysis, residuals and the  $y$  autocorrelation for the standard DINEOF approach and our new iterated solution DINEOFOI. While DINEOF residuals exhibit



**Figure 6.** Data used (upper left panel), data including cross-validation data (upper right), DINEOF reconstruction (lower left) and new approach DINEOFOI with 10 iterations (lower right).

patterns identified in the autocorrelation function, the residuals of the new method are almost pure spatially uncorrelated noise. For the other methods, not shown here,  $\mathbf{K}_1$  and DINEOF + it both contain a similarly autocorrelated signal as DINEOF. All other methods are able to retrieve this structure and leave mostly spatially uncorrelated residuals. We also see that the amplitude of the daily cycle is now closely following the data (Fig. 1).

Looking into the northwestern Mediterranean Sea (Fig. 6), the typical lower temperature structures are also used in model validations (e.g., Millot, 1987; Beckers et al., 2002). Here the comparison of the reconstructions with the cross-validation data not used also reveals that some features are now better visible in the reconstruction with the new method, in particular the form of the cold patch in the center of the domain.

Instead of looking at the full analysis, we can also focus on individual processes 1 and 2. For example, the time evolution at latitude  $34.375^\circ$  N in the eastern Mediterranean Sea (Fig. 7) shows clearly how scale 1 solution is dominated by the daily cycle (we even see the earlier heating in the eastern part during the day). Scale 2, on the other hand, has a 12 h period component whose structure could lead to interesting interpretations.

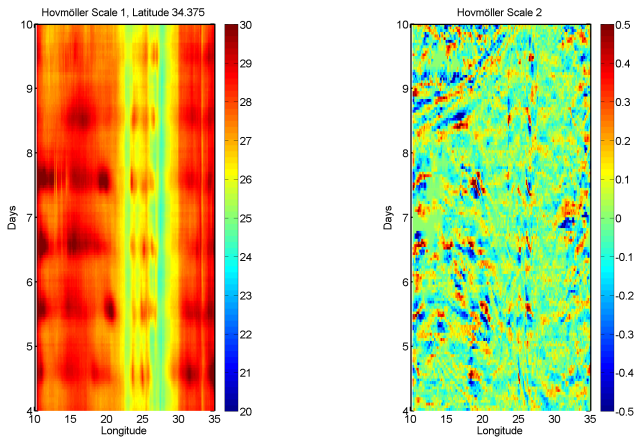
## 6 Conclusions

We presented a general framework for multi-scale analysis using optimal combinations of analysis tools focusing on different scales and processes. It provides this multi-scale analysis in an efficient way by combining the analyses obtained with different simpler tools based on the assumption that only a single process is present. The method also allows one to focus on individual scales by isolating the analysis of a single process in the presence of another one.

With a synthetic test case, we were able to provide guidelines for which formulations to choose among a series of possible algorithms. In particular, we showed that by naming process 1 as the process which has the highest signal-to-noise ratio (or largest correlation length when the other process has a similar signal-to-noise ratio), process 1 can then be efficiently analyzed by a well-identified iterative method<sup>7</sup>. The other process then is optimally reconstructed with another iterative formula<sup>8</sup> and the combination of the two provides an analysis which is close to the theoretical optimal interpolation even with a reduced number of iterations.

<sup>7</sup>P1a in the text.

<sup>8</sup>P2b in the text.



**Figure 7.** Hovmöller diagram in the eastern Mediterranean at latitude  $34.375^{\circ}$  N. Temperature as a function of longitude (in degrees) and time (in days) for process 1 (left panel) and process 2 (right panel).

Then we showed how we can exploit large-scale covariances together with small-scale covariances in a feasible way without the need to explicitly invert huge matrices that would arise from a direct application of the OI formula. This was possible through the particular choice of the covariance model (EOF-based large-scale covariance and localized parametric covariance) and the new iterative method we proposed.

This was then implemented in a real application where it was shown that with the new approach, we can indeed

retrieve fine-scale features previously filtered out. The iterative version with a combination of DINEOF-based EOF covariances and local optimal interpolation called DINEOFOI behaved best, but an alternative cheaper version is using the DINEOF analysis itself augmented by the analysis of residuals with a local optimal interpolation. But we have shown that the latter, quite natural approach was suboptimal in all situations. Other simple methods have also been identified: a time–space interpolation using EOFs from DINEOF alone<sup>9</sup> possibly augmented by an additional local optimal analysis of residuals<sup>10</sup>. Both yield good results, but they also remain suboptimal. Hence, DINEOFOI provided the optimal solution at a still reasonable cost.

Further fine-tuning of the real application could be performed but is out of the scope of the present paper, focusing on how to combine two or more analysis tools, each one exploiting different covariance specifications. To provide the best possible SST products, the following improvements could be included: we can adopt more complicated expressions of the observational error covariances, exploiting the preliminary outliers detection offered by DINEOF and the quality-flag value of the original data. In this case, suspect pixels (near clouds or with statistically too-large residuals) can be flagged and see their observational error increased. This should lead to further improvement in noise filtering. The temporal covariance functions used in the local optimal interpolation could be modulated by a cosine function to take into account the 12 h cycle identified in the residuals and provide better estimates for clouds present over several hours.

<sup>9</sup> $\mathbf{K}_1$  in the text.

<sup>10</sup> $\mathbf{S}_{01}$  in the text.

**Appendix A: Proof of scale-selective analysis formula**

With two processes, if we want to isolate the best analysis for process 1, it can be obtained with the gain matrix Eq. (2) in the presence of two scales:

$$\bar{\mathbf{K}}_1 = \mathbf{B}_1 \mathbf{H}^T (\mathbf{H} \mathbf{B}_1 \mathbf{H}^T + \mathbf{H} \mathbf{B}_2 \mathbf{H}^T + \mathbf{R})^{-1}. \quad (\text{A1})$$

We will now prove Eq. (6) rewritten as

$$\begin{aligned} \mathbf{B}_1 \mathbf{H}^T (\mathbf{H} \mathbf{B}_1 \mathbf{H}^T + \mathbf{H} \mathbf{B}_2 \mathbf{H}^T + \mathbf{R})^{-1} \\ = \mathbf{G}^{-1} \mathbf{K}_1 (\mathbf{I} - \mathbf{H} \mathbf{K}_2), \end{aligned} \quad (\text{A2})$$

with  $\mathbf{G} = \mathbf{I} - \mathbf{K}_1 \mathbf{H} \mathbf{K}_2 \mathbf{H}$  supposed to be invertible and  $\mathbf{K}_1$  and  $\mathbf{K}_2$  defined by Eq. (3).

Multiplying each side on the left by  $\mathbf{G}$  and on the right by  $(\mathbf{H} \mathbf{B}_1 \mathbf{H}^T + \mathbf{H} \mathbf{B}_2 \mathbf{H}^T + \mathbf{R})$ , we need to prove that

$$\mathbf{G} \mathbf{B}_1 \mathbf{H}^T = \mathbf{K}_1 (\mathbf{I} - \mathbf{H} \mathbf{K}_2) (\mathbf{H} \mathbf{B}_1 \mathbf{H}^T + \mathbf{H} \mathbf{B}_2 \mathbf{H}^T + \mathbf{R}), \quad (\text{A3})$$

which, by using the definitions of  $\mathbf{G}$  and  $\mathbf{K}_1$  is true if

$$\begin{aligned} \mathbf{I} - (\mathbf{H} \mathbf{B}_1 \mathbf{H}^T + \mathbf{R}) \mathbf{H} \mathbf{K}_2 \mathbf{H} \mathbf{B}_1 \mathbf{H}^T = (\mathbf{H} \mathbf{B}_1 \mathbf{H}^T + \mathbf{R})^{-1} \\ (\mathbf{I} - \mathbf{H} \mathbf{K}_2) (\mathbf{H} \mathbf{B}_1 \mathbf{H}^T + \mathbf{H} \mathbf{B}_2 \mathbf{H}^T + \mathbf{R}), \end{aligned} \quad (\text{A4})$$

which, multiplying each side by  $(\mathbf{H} \mathbf{B}_1 \mathbf{H}^T + \mathbf{R})$  is true if

$$\begin{aligned} \mathbf{H} \mathbf{B}_1 \mathbf{H}^T + \mathbf{R} - \mathbf{H} \mathbf{K}_2 \mathbf{H} \mathbf{B}_1 \mathbf{H}^T \\ = (\mathbf{I} - \mathbf{H} \mathbf{K}_2) (\mathbf{H} \mathbf{B}_1 \mathbf{H}^T + \mathbf{H} \mathbf{B}_2 \mathbf{H}^T + \mathbf{R}), \end{aligned} \quad (\text{A5})$$

which is true if

$$\begin{aligned} -\mathbf{H} \mathbf{K}_2 \mathbf{H} \mathbf{B}_1 \mathbf{H}^T = \mathbf{H} \mathbf{B}_2 \mathbf{H}^T \\ - \mathbf{H} \mathbf{K}_2 (\mathbf{H} \mathbf{B}_1 \mathbf{H}^T + \mathbf{H} \mathbf{B}_2 \mathbf{H}^T + \mathbf{R}), \end{aligned} \quad (\text{A6})$$

which in turn is true if

$$0 = \mathbf{H} \mathbf{B}_2 \mathbf{H}^T - \mathbf{H} \mathbf{K}_2 (\mathbf{H} \mathbf{B}_2 \mathbf{H}^T + \mathbf{R}). \quad (\text{A7})$$

Since the latter is true by virtue of the definition of  $\mathbf{K}_2$ , we have proven Eq. (6). Then, using Eq. (B1) we also prove Eq. (7). From there, by interchanging indices 1 and 2 we also prove Eq. (12).

We can also prove that Eqs. (4) and (6) are equivalent. We need to show that with  $\mathbf{F} = \mathbf{I} - \mathbf{K}_2 \mathbf{H} \mathbf{K}_1 \mathbf{H}$ ,

$$\mathbf{K}_1 - \mathbf{K}_1 \mathbf{H} \mathbf{F}^{-1} \mathbf{K}_2 (\mathbf{I} - \mathbf{H} \mathbf{K}_1) = \mathbf{G}^{-1} \mathbf{K}_1 (\mathbf{I} - \mathbf{H} \mathbf{K}_2), \quad (\text{A8})$$

which is true if

$$\mathbf{G} \mathbf{K}_1 - \mathbf{G} \mathbf{K}_1 \mathbf{H} \mathbf{F}^{-1} \mathbf{K}_2 (\mathbf{I} - \mathbf{H} \mathbf{K}_1) = \mathbf{K}_1 (\mathbf{I} - \mathbf{H} \mathbf{K}_2), \quad (\text{A9})$$

itself true if

$$\mathbf{K}_1 \mathbf{H} \mathbf{K}_2 (\mathbf{I} - \mathbf{H} \mathbf{K}_1) = \mathbf{G} \mathbf{K}_1 \mathbf{H} \mathbf{F}^{-1} \mathbf{K}_2 (\mathbf{I} - \mathbf{H} \mathbf{K}_1), \quad (\text{A10})$$

which is the case if

$$\mathbf{K}_1 \mathbf{H} = \mathbf{G} \mathbf{K}_1 \mathbf{H} \mathbf{F}^{-1} \quad (\text{A11})$$

or if

$$\mathbf{K}_1 \mathbf{H} \mathbf{F} = \mathbf{G} \mathbf{K}_1 \mathbf{H}, \quad (\text{A12})$$

which is true. Hence, we proved Eq. (4). Then application of Eqs. (B2) to (4) proves Eq. (5). From there, as before, interchanging indices 1 and 2 yields Eq. (11) and all our identities are verified.

**Appendix B: Useful matrix identity**

If the inverse matrices involved in the following expression exist, than the following identity holds for any (size compatible) matrices  $\mathbf{H}$ ,  $\mathbf{K}_1$  and  $\mathbf{K}_2$

$$(\mathbf{I} - \mathbf{K}_1 \mathbf{H} \mathbf{K}_2 \mathbf{H})^{-1} \mathbf{K}_1 = \mathbf{K}_1 (\mathbf{I} - \mathbf{H} \mathbf{K}_2 \mathbf{H} \mathbf{K}_1)^{-1}. \quad (\text{B1})$$

This can be easily shown by an application of the Woodbury formula or a direct proof obtained by multiplying the right side by  $(\mathbf{I} - \mathbf{H} \mathbf{K}_2 \mathbf{H} \mathbf{K}_1)$  and the left side by  $(\mathbf{I} - \mathbf{K}_1 \mathbf{H} \mathbf{K}_2 \mathbf{H})$ . A similar expression can be obtained by interchanging indices 1 and 2:

$$(\mathbf{I} - \mathbf{K}_2 \mathbf{H} \mathbf{K}_1 \mathbf{H})^{-1} \mathbf{K}_2 = \mathbf{K}_2 (\mathbf{I} - \mathbf{H} \mathbf{K}_1 \mathbf{H} \mathbf{K}_2)^{-1}. \quad (\text{B2})$$

**Appendix C: Detailed comparison for synthetic case**

*Case 1:* As a first case, we use two processes with unit signal-to-noise ratios but very different correlation length. For process 1 a length scale of 33 grid points was taken whereas for process 2 a length scale of 4.5 grid points was used. The data cover the original grid entirely and the analysis is therefore only checking the filtering properties of the different formulations. Hence, we only look at average behaviors, but distinguish errors in reconstructing process 1, process 2 and the overall field, as detailed in Tables C1–C4. For process 1 (Table C1), we note that it is not beneficial to try to subtract the small-scale information before the analysis by a simple residual approach (column 3) and that the iterative version P1a even without iterations yields the best results. When performing iterations, Table C1 shows convergence and the superiority of P1a. For small-scale process 2 (Table C2), the simple version of first subtracting the large-scale analysis from the data performs best and is equivalent to a non-iterated version of P2b. All iterative versions converge and iterative version P2b shows its superiority for small scales.

For the overall analysis, the simple approaches in Table C3 exhibit interesting facts. The best approach is to combine the analysis of one process with the analysis of the residual for the other process. Furthermore, even if we showed that process 1 is best captured by  $\mathbf{K}_1$  and process 2 by analyzing

**Table C1.** Average error for different analysis formulations of process 1. Column 5 of the upper part shows the same result as column 3 since it can be shown that without iterations, the formulations are equivalent. So column 5 of the upper part of the table essentially gives confidence that the implementation is correct. The average relative error variance is based on 100 000 independent realizations.

Case	$\mathbf{K}_1$	$\mathbf{K}_1(\mathbf{I} - \mathbf{HK}_2)$	0 it P1a	0 it P1b
1	1.3969	3.2775	1.2212	3.2775
2	1.8056	2.0853	1.6121	2.0853
3	1.1970	1.4024	1.0634	1.4024
4	1.0620	7.4532	1.0215	7.4532
5	1.0055	1.2405	1.0038	1.2405
6	1.9071	2.9159	1.6922	2.9159
7	1.9422	2.1873	1.8183	2.1873

Case	2 it P1a	2 it P1b	20 it P1a	20 it P1b
1	1.0916	2.2260	1.0007	1.0057
2	1.4012	1.8109	1.0281	1.0779
3	1.0205	1.1012	1.0024	1.0024
4	1.0044	2.6406	0.9982	1.0021
5	1.0021	1.1195	0.9997	1.0069
6	1.5009	2.5738	1.0894	1.4318
7	1.6686	2.0194	1.2510	1.3978

**Table C2.** Average error for different analysis formulations of process 2. As in Table C1, column 5 of the upper part gives the same result as column 3, as it should. The average relative error variance is based on 100 000 independent realizations.

Case	$\mathbf{K}_2$	$\mathbf{K}_2(\mathbf{I} - \mathbf{HK}_1)$	0 it P2a	0 it P2b
1	2.5580	1.1923	2.1361	1.1923
2	2.0783	1.6684	1.9256	1.6684
3	1.1882	1.0355	1.0890	1.0355
4	6.0804	1.0225	3.7195	1.0225
5	6.1443	1.0814	4.0191	1.0814
6	1.5339	1.2145	1.4921	1.2145
7	2.0680	1.7672	1.9825	1.7672

Case	2 it P2a	2 it P2b	20 it P2a	20 it P2b
1	1.6118	1.0669	1.0016	0.9991
2	1.6931	1.4126	1.0672	1.0278
3	1.0321	1.0115	1.0024	1.0024
4	2.1335	1.0066	1.0037	1.0004
5	2.5873	1.0415	1.0946	1.0008
6	1.4230	1.1522	1.1580	1.0348
7	1.8451	1.6062	1.3362	1.2211

residuals via  $\mathbf{K}_2(\mathbf{I} - \mathbf{HK}_1)$ , the overall analysis performs as well when doing the inverse and adding the results up. For the iterated version (Table C4), we see a similar pattern; adding up formulations of the same type (P1a + P2a or P1b + P2b) does not perform as well as adding up formulations of different types (P1a + P2b or P1b + P2a), but convergence is observed in all cases. The natural choice for this case is

**Table C3.** Average error for different simplified analysis formulations for the overall analysis. The average relative error variance is based on 100 000 independent realizations.

Case	$\mathbf{S}_{00}$	$\mathbf{S}_{10}$	$\mathbf{S}_{01}$	$\mathbf{S}_{11}$
1	6.2802	1.0088	1.0085	5.6440
2	21.6620	1.0165	1.0200	19.8658
3	1.4088	1.0371	1.0357	1.3120
4	4.7596	2.2262	1.0231	6.5037
5	1.5216	1.1202	1.0015	1.3122
6	2.5735	1.2828	1.1630	2.7218
7	19.8920	1.0089	1.0373	19.0295

**Table C4.** Average error for iterative versions for the overall analysis: 0, 2 and 20 iterations. The average relative error variance is based on 100 000 independent realizations.

Case	0 it P1a + P2a	0 it P1a + P2b	0 it P1b + P2a	0 it P1b + P2b
1	4.6928	1.0220	1.0385	5.6440
2	18.0134	1.0784	1.0119	19.8658
3	1.1679	1.0071	1.0559	1.3120
4	2.9738	1.0055	2.8756	6.5037
5	1.3133	1.0016	1.0656	1.3122
6	2.3563	1.1118	1.2926	2.7218
7	18.0300	1.0737	1.0139	19.0295

Case	2 it P1a + P2a	2 it P1a + P2b	2 it P1b + P2a	2 it P1b + P2b
1	2.9137	1.0017	1.0203	3.3676
2	13.1811	1.0130	1.0077	14.2558
3	1.0560	1.0014	1.0073	1.0823
4	1.8108	1.0015	1.2031	2.4070
5	1.1680	1.0005	1.0195	1.1583
6	2.0934	1.0696	1.2255	2.3798
7	15.3637	1.0211	1.0113	16.0365

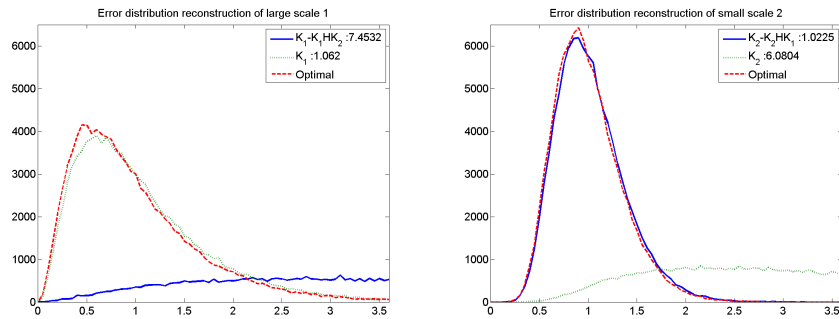
Case	20 it P1a + P2a	20 it P1a + P2b	20 it P1b + P2a	20 it P1b + P2b
1	1.0086	0.9998	0.9999	1.0102
2	2.0564	0.9991	0.9996	2.1293
3	1.0006	1.0006	1.0006	1.0006
4	1.0034	1.0009	1.0010	1.0043
5	1.0101	1.0000	1.0006	1.0094
6	1.3308	1.0056	1.0290	1.3799
7	6.6037	1.0004	1.0024	6.7399

therefore clearly P1a for the larger scales, equation P2b for the shorter scales and their sum for the overall analysis.

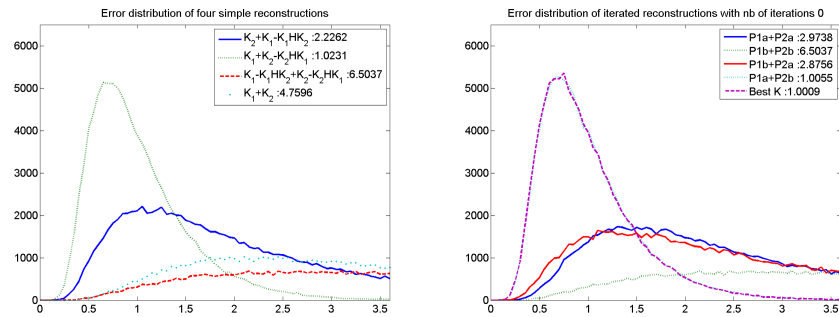
For *case 2*, we use the same parameters but examine the behavior when the scales are less well-separated and use a correlation length of 16 grid points instead of 4.5 for process 2. Compared to case 1, the conclusions remain mostly unchanged, but we observe that more iterations are needed to reach the same quality in the iterated version.

For *case 3* and the following cases, we only use observations in parts of the domain. In one region, sampling is done at grid resolution and in other regions with larger parts void of data (up to 20 grid points). This mimics the situation we encounter with satellite images. Otherwise, the case is





**Figure C1.** Histogram of scaled errors for reconstruction of process 1 (left) and process 2 (right). We see that for process 1, it is better to use the full data set, whereas for process 2 it is better to work with residuals, i.e., data from which we subtract the analysis of the process 1.



**Figure C2.** Histogram of scaled errors for simple reconstructions of the total field (left) and with 0 iterations in the new formulations (right). For reference, the distribution of the optimal solution is also shown, proving that not only is the average error of the new method combining P1a and P2b optimal, but also that the error distribution falls along the optimal one.

identical to case 1<sup>11</sup>. Conclusions for case 3 are similar to case 1, and we see that even with few iterations we capture the maximum information from the data available.

In *case 4* we take the same parameters as in case 3 (partial observation, very different scales), but assume that process 1 has a much higher signal-to-noise ratio: we use a value of 20 instead of 1. In this case, P1a again provides the best analysis for process 1 and P2b, for process 2 alone. For the simple approaches for the total field (Table C3), now the selection of the field used to create residuals is important, and as one might have guessed, one has first to analyze the large-scale process with the high signal-to-noise ratio and then add the small-scale analysis of the residuals (meaning  $S_{01}$  is used for the global analysis). For the iterated version, it is now clearly the combination P1a and P2b which outperforms the others.

Since this case is also similar to our original question, we can now also have a look at the error distribution of the 100 000 realizations (Figs. C1 and C2). They confirm that when the average error of the formulations is close to the error of the truly optimal solution, the histogram of errors also

almost coincide, providing evidence that we are indeed very close to the optimal interpolation.

To complete the analysis, some additional cases have been examined: in *case 5* we use the same parameters as in case 4 but use the short correlation length for process 1 and the larger one for process 2. Despite this inversion, the same conclusions as in case 4 remain. This means that it is the process with the highest signal-to-noise ratio which should be labeled as process 1 so that P1a and P2b remain optimal.

For *case 6* we use the same situation as in case 3, but now both processes have the same high signal-to-noise ratio of 20. Similar conclusions still hold, but the simpler approaches now degrade.

Finally, *case 7* uses the same situation as case 6 (two high signal-to-noise ratios), but in addition the correlation scales are now closer to each other (30 grid points for process 1 and 16 grid points for process 2). This is the most difficult situation and requires more iterations to converge. In this case, one could anyway question whether a scale separation approach is meaningful or if a single correlation length scale should be used.

<sup>11</sup>Note that when comparing the results of case 1 and case 3, the error has been scaled by the best analysis possible with the data available. Hence, a better score in case 3 does not mean the analysis itself is better than the analysis in case 1 (which uses more data) but that we are closer to the best we can get from the data.

#### Appendix D: Calibration of covariance functions and signal-to-noise ratio

To specify the covariance function to be used with the local optimal interpolation, we can take advantage of two particularities of the data we will analyze: we expect them to contain mainly small scales (as we work on residuals) and we have a rather large amount of data available on a regular grid to fit the parametric covariance function. The following approach exploits these two particularities.

For the longitude correlation length, we can randomly select a point in the data set and verify if a prescribed number of consecutive data is present in the longitudinal direction. Since we are looking at small scales, typically 20–50 points of consecutive data would be enough to capture signals, and the probability to find such a chunk of complete data is high enough. Once such a chunk is found, we can apply a fast Fourier transform to it and calculate the squared Fourier amplitude. This is repeated a large number of times (10 000 in our case), and an average squared amplitude spectrum can be assessed. The autocovariance function can then be calculated

as the inverse Fourier transform of this square of the amplitude spectrum of the signal (e.g., Bracewell, 1986). This autocovariance from the data can then finally be used to fit the analytical function and estimate the signal-to-noise ratio. In practice, the fitting is done by trying to fit the Gaussian function  $\alpha \exp(-\beta \delta_x^2)$  to the points of the data-based autocorrelation function up to the zero-crossing but eliminating the zero-lag point. The signal-to-noise ratio can finally be estimated by (see Fig. 2)

$$\lambda_2 = \frac{\alpha}{1 - \alpha}, \quad (\text{D1})$$

where  $\alpha$  is the fitted value representing the signal and  $1 - \alpha$  is the part not explained by the correlation function, hence the noise.

The same approach is then also applied to latitude and time, and since the fitting is not perfect, the estimated overall signal-to-noise ratio is conservatively taken as the lowest one found from the three directions.



*Acknowledgements.* DINEOF was developed by projects funded by the Belgian Science Policy (BELSPO) in the context of RECOLOUR (SR/00/111), BELCOLOUR-2 (SR/00/104), HiSea (SR/12/140), BESST (SR/12/158) and GEOCOLOR (SR/00/139). All projects were part of the Research Program For Earth Observation “STEREO II”. This research was also supported by the SANGOMA Project (European FP7-SPACE-2011 project, grant no. 283580). The F.R.S.-FNRS is acknowledged for providing supercomputing access. This is a MARE publication.

Edited by: E. J. M. Delhez

## References

- Alvera-Azcárate, A., Barth, A., Rixen, M., and Beckers, J.-M.: Reconstruction of incomplete oceanographic data sets using Empirical Orthogonal Functions. Application to the Adriatic Sea Surface Temperature, *Ocean Model.*, 9, 325–346, doi:10.1016/j.ocemod.2004.08.001, 2005.
- Alvera-Azcárate, A., Barth, A., Beckers, J.-M., and Weisberg, R. H.: Multivariate reconstruction of missing data in sea surface temperature, chlorophyll and wind satellite field, *J. Geophys. Res.*, 112, C03008, doi:10.1029/2006JC003660, 2007.
- Beckers, J.-M. and Rixen, M.: EOF calculation and data filling from incomplete oceanographic datasets, *J. Atmos. Ocean. Tech.*, 20, 1839–1856, 2003.
- Beckers, J.-M., Rixen, M., Brasseur, P., Brankart, J.-M., El-moussaoui, A., Crépon, M., Herbaut, C., Martel, F., Van den Berghe, F., Mortier, L., Lascaratos, A., Drakopoulos, P., Korres, P., Pinardi, N., Masetti, E., Castellari, S., Carini, P., Tin-tore, J., Alvarez, A., Monserrat, S., Parrilla, D., Vautard, R., and Speich, S.: Model intercomparison in the Mediterranean. The MedMEx simulations of the seasonal cycle, *J. Marine Syst.*, 33–34, 215–251, 2002.
- Beckers, J.-M., Barth, A., and Alvera-Azcárate, A.: DINEOF reconstruction of clouded images including error maps – application to the Sea-Surface Temperature around Corsican Island, *Ocean Sci.*, 2, 183–199, doi:10.5194/os-2-183-2006, 2006.
- Bracewell, R.: *The Fourier Transform and its Applications*, 2nd Edn., rev., international student ed Edn., McGraw-Hill, New York, 1986.
- Brasseur, P., Beckers, J.-M., Brankart, J.-M., and Schoenauen, R.: Seasonal temperature and salinity fields in the Mediterranean Sea: climatological analyses of a historical data set, *Deep-Sea Res. Pt. I*, 43, 159–192, doi:10.1016/0967-0637(96)00012-X, 1996.
- Bretherton, F. P., Davis, R. E., and Fandry, C.: A technique for objective analysis and design of oceanographic instruments applied to MODE-73, *Deep-Sea Res.*, 23, 559–582, doi:10.1016/0011-7471(76)90001-2, 1976.
- Brisson, A., Le Borgne, P., and Marsouin, A.: Results of one year of preoperational production of sea surface temperatures from GOES-8, *J. Atmos. Ocean. Tech.*, 19, 1638–1652, 2002.
- Cushman-Roisin, B. and Beckers, J.-M.: *Introduction to Geophysical Fluid Dynamics, Physical and Numerical Aspects*, Academic Press, 2011.
- Daley, R.: *Atmospheric Data Analysis*, Vol. 2, Cambridge University Press, 1993.
- Delhomme, J.: Kriging in the hydrosociences, *Adv. Water Resour.*, 1, 251–266, doi:10.1016/0309-1708(78)90039-8, 1978.
- Donlon, C. J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E., and Wimmer, W.: The operational sea surface temperature and sea ice analysis (OSTIA) system, *Remote Sens. Environ.*, 116, 140–158, 2012.
- Eastwood, S., Le Borgne, P., Péré, S., and Poulter, D.: Diurnal variability in sea surface temperature in the Arctic, *Remote Sens. Environ.*, 115, 2594–2602, 2011.
- EUMETSAT: Geostationary sea surface temperature product user manual, available at: <http://www.osi-saf.org> (last access: 18 March 2014), 2011.
- Fisher, M.: Background error covariance modelling, in: *Seminar on Recent Development in Data Assimilation for Atmosphere and Ocean*, 45–63, 2003.
- Gandin, L. S.: *Objective Analysis of Meteorological Fields*, Tech. rep., Israel Program for Scientific Translations, Jerusalem, 1965.
- Ganzedo, U., Alvera-Azcárate, A., Esnaola, G., Ezcurra, A., and Saenz, J.: Reconstruction of sea surface temperature by means of DINEOF: a case study during the fishing season in the Bay of Biscay, *Int. J. Remote Sens.*, 32, 933–950, doi:10.1080/01431160903491420, 2011.
- Gasparini, G. and Cohn, S.: Construction of correlation functions in two and three dimensions, *Q. J. Roy. Meteor. Soc.*, 125, 723–757, 1999.
- Golub, G. H. and Van Loan, C. F.: *Matrix Computations*, Vol. 3, JHU Press, 2012.
- Kalnay, E.: *Atmospheric Modeling, Data Assimilation and Predictability*, 1st Edn., Cambridge University Press, 2002.
- Kaplan, A., Kushnir, Y., and Cane, M. A.: Reduced space optimal interpolation of historical marine sea level pressure: 1854–1992, *J. Climate*, 13, 2987–3002, doi:10.1175/1520-0442(2000)013<2987:RSOIOH>2.0.CO;2, 2000.
- Karagali, I., Høyer, J., and Hasager, C.: SST diurnal variability in the North Sea and the Baltic Sea, *Remote Sens. Environ.*, 121, 159–170, 2012.
- Kawai, Y., Kawamura, H., Takahashi, S., Hosoda, K., Murakami, H., Kachi, M., and Guan, L.: Satellite-based high-resolution global optimum interpolation sea surface temperature data, *J. Geophys. Res.-Oceans*, 111, C06016, doi:10.1029/2005JC003313, 2006.
- Le Borgne, P., Roquet, H., and Merchant, C.: Estimation of sea surface temperature from the Spinning Enhanced Visible and Infrared Imager, improved using numerical weather prediction, *Remote Sens. Environ.*, 115, 55–65, 2011.
- Le Borgne, P., Legendre, G., and Péré, S.: Comparison of MSG/SEVIRI and drifting buoy derived diurnal warming estimates, *Remote Sens. Environ.*, 124, 622–626, 2012.
- Marullo, S., Santoleri, R., Banzon, V., Evans, R., and Guarra-cino, M.: A diurnal-cycle resolving sea surface temperature product for the tropical Atlantic, *J. Geophys. Res.-Oceans*, 115, C05011, doi:10.1029/2009JC005466, 2010.
- Marullo, S., Santoleri, R., Ciani, D., Borgne, P. L., Péré, S., Pinardi, N., Tonani, M., and Nardone, G.: Combining model and geostationary satellite data to reconstruct hourly SST field over the Mediterranean Sea, *Remote Sens. Environ.*, 146, 11–23, doi:10.1016/j.rse.2013.11.001, 2014.
- Millot, C.: Circulation in the Western Mediterranean Sea, *Oceanol. Acta*, 10, 143–149, 1987.

- Nardelli, B.: A Novel Approach for the High-Resolution Interpolation of In Situ Sea Surface Salinity, *J. Atmos. Ocean. Tech.*, 29, 867–879, 2012.
- Nardelli, B. B., Tronconi, C., Pisano, A., and Santoleri, R.: High and ultra-high resolution processing of satellite sea surface temperature data over Southern European Seas in the framework of MyOcean project, *Remote Sens. Environ.*, 129, 1–16, doi:10.1016/j.rse.2012.10.012, 2013.
- Nikolaidis, A., Georgiou, G., Hadjimitsis, D., and Akylas, E.: Filling in missing sea-surface temperature satellite data over the Eastern Mediterranean sea using the DINEOF algorithm, *Central European Journal of Geosciences*, 6, 27–41, doi:10.2478/s13533-012-0148-1, 2014.
- Parrish, D. and Derber, J.: The National Meteorological Center's spectral statistical interpolation analysis system, *Mon. Weather Rev.*, 120, 1747–1763, 1992.
- Reynolds, R. W. and Smith, T. M.: Improved global sea surface temperature analyses, *J. Climate*, 7, 929–948, 1994.
- Sheng, Z., Shi, H.-Q., and Ding, Y.-Z.: Missing satellite-based sea surface temperature data reconstructed by DINEOF method, *Advances in Marine Science*, 2, 017, ISSN: 1671-6647, 2009.
- Sirjacobs, D., Alvera-Azcárate, A., Barth, A., Park, Y., Nechad, B., Ruddick, K., and Beckers, J.-M.: Cloud filling of total suspended matter, chlorophyll and sea surface temperature remote sensing products by the data interpolation with empirical orthogonal functions methodology, application to the BELCOLOUR-1 database, ESA Special Publication, Vol. SP666, available at: <http://orbi.ulg.ac.be/handle/2268/6552> (last access: 18 March 2014), 2008.
- Stark, J. D., Donlon, C. J., Martin, M. J., and McCulloch, M. E.: OSTIA: an operational, high resolution, real time, global sea surface temperature analysis system, in: *OCEANS 2007 – Europe*, IEEE, 1–4, 2007.
- Stuart-Menteth, A. C., Robinson, I. S., and Challenor, P. G.: A global study of diurnal warming using satellite-derived sea surface temperature, *J. Geophys. Res.-Oceans*, 108, 3155, doi:10.1029/2002JC001534, 2003.
- Troupin, C., Machin, F., Ouberdous, M., Sirjacobs, D., Barth, A., and Beckers, J.-M.: High-resolution Climatology of the North-East Atlantic using Data-Interpolating Variational Analysis (DIVA), *J. Geophys. Res.*, 115, C08005, doi:10.1029/2009JC005512, 2010.
- Volpe, G., Colella, S., Forneris, V., Tronconi, C., and Santoleri, R.: The Mediterranean Ocean Colour Observing System – system development and product validation, *Ocean Sci.*, 8, 869–883, doi:10.5194/os-8-869-2012, 2012.
- Wackernagel, H.: *Multivariate Geostatistics*, Springer, 2003.
- Wang, Y. and Liu, D.: Reconstruction of satellite chlorophyll-a data using a modified DINEOF method: a case study in the Bohai and Yellow seas, China, *Int. J. Remote Sens.*, 35, 204–217, doi:10.1080/01431161.2013.866290, 2014.
- Weaver, T. and Mirouze, I.: On the diffusion equation and its application to isotropic and anisotropic correlation modelling in variational assimilation, *Q. J. Roy. Meteor. Soc.*, 139, 242–260, 2013.
- Young, N.: The rate of convergence of a matrix power series, *Linear Algebra Appl.*, 35, 261–278, doi:10.1016/0024-3795(81)90278-0, 1981.