Remote Sensing of Environment 164 (2015) 334–343

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Evaluation of different covariance models for the operational interpolation of high resolution satellite Sea Surface Temperature data over the Mediterranean Sea

B. Buongiorno Nardelli [a,b,*], A. Pisano [a], C. Tronconi [a], R. Santoleri [a]

[a] CNR—Istituto di Scienze dell'Atmosfera e del Clima-Gruppo Oceanografia da Satellite-Roma, Roma, Italy
[b] CNR—Istituto per l'Ambiente Marino Costiero, Napoli, Italy

ARTICLE INFO

Article history:
Received 7 November 2014
Received in revised form 10 April 2015
Accepted 21 April 2015
Available online 8 May 2015

Keywords:
Sea surface temperature
Satellite observations
Optimal interpolation
Covariance model
Mediterranean Sea

ABSTRACT

Daily gap-free Sea Surface Temperature (SST) fields at high resolution are required by several operational users working on monitoring and forecasting the status of the marine environment. Existing instruments cannot provide such fields, and careful interpolation of available data is required to provide gap-free SST estimates. In the framework of the MyOcean projects (now turning into the European Copernicus Marine Service), several satellite (and/or satellite and in situ) interpolated products are distributed in real time, and continuous research and development activities are carried out to improve their quality. In this paper, we describe the work done to improve the High Resolution (1/16°) interpolated product covering the Mediterranean Sea, that is obtained by combining all available satellite infrared images. Three Optimal Interpolation schemes based on different space–time covariance models and background fields are derived, compared and validated versus in situ drifting buoy measurements (leading to a minimum standard deviation error (STDE) of 0.46 °K and mean bias error (MBE) of −0.15 °K), as well as by adopting a holdout validation approach with artificial clouds (with STDE ranging between 0.22 °K and 0.55 °K, and MBE always between 0.03 °K and 0.05 °K, depending on the cloud configuration considered). Almost negligible differences are found between the three products, revealing only a slight improvement when using spatially varying covariance parameters and a daily climatology as background, which is less risky in case of prolonged cloudy conditions, though attaining the same performance with small/rapid clouds. This scheme has thus been implemented in the Mediterranean SST operational chain since April 2014.

1. Background

Sea Surface Temperature (SST) is a fundamental variable for many scientific and operational applications (see http://www.ghrsst.org). It is primarily required by the meteorological and marine operational forecasting systems to constrain their numerical prediction models (e.g. Chelton & Wentz, 2005; Dobricic et al., 2007), but also by public institutions and private companies working, for example, on marine environment and security managing, fisheries, tourism, marine transportation, offshore exploration and extraction. Most of these users need a high resolution (HR) gap-free estimate of the SST, known as level 4 (L4) data (as examples of environmental analyses requiring gap-free SST see Volpe, Buongiorno Nardelli, Cipollini, Santoleri, & Robinson, 2012; Liu et al., 2013; Bonanno et al., 2015). In situ instruments, however, only provide sparse point observations, and even satellite sensors are affected by both coverage and physical constraints. More specifically, space-borne infrared sensors cannot 'see' through clouds, while microwave measurements are contaminated by several factors (such as rain, land, ice etc.), so that satellite images are also always characterized by data voids (Robinson, 2004). Interpolation of available data is thus a crucial step to provide an accurate L4 SST field.

Within the GMES (Global Monitoring for Environment and Security)/Copernicus MyOcean projects, funded by the European Commission, a number of global and regional SST near-real-time (NRT) daily L4 products based on the combination of available satellite (and/or satellite and in situ) measurements have been developed and are currently distributed through a single user interface (http://www.myocean.eu/web/24-catalogue.php). All these products are obtained by applying a statistical technique known in literature as Optimal Interpolation (OI).

First introduced by Gandin (1965) and by Bretherton, Davis, and Fandry (1976) to the oceanographic community, classical optimal interpolation (OI) method can be applied in several different ways, depending on a number of assumptions, from the definition of the state vector to be reconstructed, which can require very different covariance models and calibration techniques, to the numerical algorithms used. Through OI, SST L4 data are obtained as a linear combination of the observations (namely, of the SST anomalies with respect to a background field, also

* Corresponding author at: CNR-IAMC Calata Porta di Massa, 80133, Napoli, Italy.

http://dx.doi.org/10.1016/j.rse.2015.04.025
0034-4257/

called *first guess* field), weighted directly with their correlation to the interpolation point and inversely with their cross-correlation and error. OI thus requires that the background error covariance and observation error covariance are known, and combines these error covariances and observations to provide the minimum variance solution (i.e. the solution that minimizes the mean square error of the analysis).

Though, theoretically, the background covariance matrix should be estimated directly from all available observations, the large amount of satellite SST data makes this approach impractical for NRT applications, and most of the SST L4 operational systems are thus based on predetermined parametric (analytical) covariance functions estimated from historical observations or from numerical model output (e.g. Donlon et al., 2012; Martin et al., 2012; Reynolds & Chelton, 2010; Roberts-Jones, Fiedler, & Martin, 2012, and references therein). Covariance functions depending only on spatial distance are also generally preferred. These OI schemes take the previous day analysis as the background field and daily increments with respect to this background are used as input data, eventually relaxing to daily/weekly climatologies in case of prolonged lack of observations.

Parametric approaches have the advantage of reducing the amount of data to be taken in input, as covariance functions clearly decay at increasing distance and observations that are found far from the interpolation point can be safely excluded from the analysis, which makes all these algorithms theoretically sub-optimal, but computationally efficient. However, approximated OI approaches can also be easily extended to higher-dimensional Euclidean spaces, including also temporal decorrelations (e.g. Le Traon, Nadal, & Ducet, 1998; Marullo et al., 2014), or even to higher dimensional spaces (e.g. Buongiorno Nardelli, 2012). More complex space–time models, even if requiring additional computational time, directly take into account temporal correlations instead of limiting to a spatial approach or hybrid spatial/climatological relaxation approach, which implies the discretional definition of a temporal e-folding scale. In fact, in 'truly' optimal space–time interpolation, one would consider a temporal sequence of the SST field as the state vector, and several realizations of these space–time data should be used to estimate the covariance. In the approximated space–time covariance models, this turns up into the need to define a mixed space–time covariance function and to estimate corresponding parameters. Simplified space–time models, however, allow simpler estimations by assuming the covariance function can be approximated as the product of space and time covariance functions, separately.

The Mediterranean Sea and Black Sea L4 SST products developed by the Consiglio Nazionale delle Ricerche-Istituto di Scienze dell'Atmosfera e del Clima-Gruppo di Oceanografia da Satellite (CNR-ISAC-GOS) within MyOcean projects, as well as in previous projects/research activities, have been obtained directly through a space–time OI approach (Buongiorno Nardelli, Colella, Santoleri, Guarracino, & Kholod, 2010; Buongiorno Nardelli, Tronconi, Pisano, & Santoleri, 2013; Buongiorno Nardelli et al., 2003; Marullo, Buongiorno Nardelli, Guarracino, & Santoleri, 2007; Santoleri, Marullo, & Böhm, 1991).

The most recent algorithm used for the Mediterranean (MED) HR L4 product (at 1/16° resolution) has been fully described and validated in Buongiorno Nardelli et al. (2013), and was operational within the MyOcean system until April 2014 (corresponding to the release of an upgraded MyOcean service, hereafter referred to as MyOcean V4). This algorithm was basically the same used by Marullo et al. (2007), briefly recalled hereafter. The background correlation function was assumed to depend separately from time and space lag:

$$C = C(\Delta r, \Delta t) = C(\Delta r) \cdot C(\Delta t) \quad (1)$$

and was defined as the product of two negative exponential functions with fixed and isotropic decorrelation space and time scales. These scales were estimated by a least-square fit of observed lagged-correlations in space and time, but were taken uniform all over the interpolation domain.

Since the MyOcean V4 release, a new covariance model has been adopted, relaxing this last assumption, i.e. using non-uniform space and time decorrelation scales, and also adopting a different functional dependence, based on the tests described hereafter. In particular, different background fields (and corresponding covariances) have also been tested, trying to improve the accuracy of the analysed field. This paper thus describes the new models tested and provides the CAL/VAL of the current (V4) operational MyOcean HR L4 SST data over the Mediterranean Sea.

2. Data

2.1. Satellite observations

All satellite data used for the SST L4 processing at CNR are provided by the Group for High-Resolution Sea Surface Temperature (GHRSST) Global and Regional Data Assembly Centres (GDAC, RDAC; see www.ghrsst.org for more details). The sensors (and platforms) ingested are: the Advanced Along Track Scanning Radiometer (AATSR, installed on the European ENVIronment SATellite, ENVISAT), MODerate resolution Imaging Spectroradiometer (MODIS, on both Aqua and Terra satellites), Advanced Very High Resolution Radiometer (AVHRR, on METeorological Operational satellite, METOP, and National Oceanic and Atmospheric Administration (NOAA) satellites), and Spinning Enhanced Visible and Infrared Imager (SEVIRI, installed on Meteosat Second Generation, MSG). The data considered consist of one year (2011) of the GHRSST Level 2P (L2P), that contain single sensor SST observations on the native grid/swath, geo-location data, error estimates (SSES, Single Sensor Error Statistics, i.e. bias error and standard deviation error), land and ice flags, as well as additional auxiliary fields for each pixel, referred to as dynamic flags. The information contained in the L2P files is used for a preselection of the data that are ingested in the processing chain, mapping super-collated (namely merged, multi-sensor) observations on the interpolation grid (these data will be referred to as L3S in the following). In particular, to avoid both diurnal warming effects and residual cloud contamination, we keep only the night-time observations that are flagged with quality level 5 (best observations) in the original data file. L3S data are then obtained by removing from each L2P image a large scale sensor bias estimated through an iterative procedure that adjusts all images to a reference sensor (AATSR or METOP) and keeps only one value per pixel, as detailed in Buongiorno Nardelli et al. (2013).

2.2. In situ measurements and matchup data

GHRSST group on satellite SST validation (STVAL) indicates drifting buoy measurements as the baseline for the calibration and validation of all SST products (https://www.ghrsst.org/ghrsst/tags-and-wgs/stval-wg/sses-common-principles/). The L4 validation activities described in the following sections are thus based on the compilation of a match-up database between satellite SST L4 and surface drifting buoys measurements distributed by the MyOcean In Situ Thematic Assembly Centre (In Situ-TAC), covering the year 2011. Standard quality flags are used in the CNR processing chain to exclude all suspect measurements, keeping only the highest quality data (for more details see also the MyOcean In Situ TAC Product User Manual, http://catalogue.myocean.eu.org/static/resources/myocean/pum/MYO2-INS-PUM-013-V1.2.pdf). Furthermore, specific quality control (QC) procedures have been set up and applied to drifter measurements before building our matchup dataset, as described in Buongiorno Nardelli et al. (2013). The matchup is restricted to night-time data (between 9 p.m. and 6 a.m. local time), keeping only one value per buoy for each date, in order to get fully independent measurements (autocorrelated information would be given by sequential measurements from the same drifter within the time period considered, see also Bayley and Hammersley, 1946). In fact, the single SST value per buoy/day used here is the closest (in time) to the position time (the drifting buoys collect SST data at

regular time intervals, while position measurements are recorded at longer time intervals, related to the satellite positioning system).

3. Updated optimal interpolation scheme

With respect to Buongiorno Nardelli et al. (2013) MED HR scheme, the OI schemes tested here include: 1) the definition of a new functional dependence for the spatial covariance (while the temporal covariance function remains unchanged); 2) the definition of spatially varying covariance parameters and 3) the use of two different background fields, namely the standard daily (pentad) climatological field described in Buongiorno Nardelli et al. (2013) and the previous day analysis. In this section, the new schemes are described in details. For all aspects related to the other steps involved in the MyOcean MED SST L4 operational pre-processing (namely the input data selection, quality control, cloud flagging and data merging prior to the interpolation), the readers can directly refer to Buongiorno Nardelli et al. (2013).

3.1. New background correlation function

A new functional model has been chosen for the spatial dependence, optimizing the fit to the correlations estimated from the observations.

The correlations were computed as a function of spatial and temporal lags, separately, considering sub-domains of the same size used within the optimal interpolation (namely restricting to a pre-defined influential radius, 300 Km in our case). The local spatial correlation was estimated as the correlation between all couples of L3S SST time series found within each spatial sub-domain, and successively averaged at fixed space separation bins. Similarly, the L3S time series autocorrelation at each location provided the temporal correlation as a function of a variable time lag. In both calculations, to remove the influence of eventual trends in SST at scales larger than the sub-domain, a moving average of the SST within the space–time sub-domain is removed before the computation. To reduce the effect of uneven L3S sampling within the influential radius on the filtering of this red-noise, the moving averages are estimated here from the corresponding L4 analyses obtained with the standard OI algorithm described in Buongiorno Nardelli et al. (2013).

It was found that a negative exponential correlation function provides the best fit for the temporal dependence:

$$C(\Delta t) = e^{-\frac{\Delta t}{\tau}} \tag{2}$$

with Δt representing temporal lag and τ a temporal decorrelation scale. Conversely, the spatial correlation is better approximated by a rational quadratic function than by a negative exponential (see Fig. 1), reducing corresponding goodness-of-fit statistic from >1 to less than 0.1 (with a corresponding reduction of the standard error by a factor >4). The function takes the form:

$$C(\Delta r) = \left(1 + \frac{(\Delta r)^2}{2\alpha L^2}\right)^{-\alpha} \tag{3}$$

where L and α are non-negative characteristic parameters and Δr represents spatial separation. The new covariance function can be interpreted as an infinite sum (integral) of squared exponential correlation functions, each with a different characteristic length-scale (Rasmussen & Williams, 2006).

3.2. New background covariance assumptions

Though the choice of the covariance functional dependence was done looking at basin/sub-basin averages, the new covariance models tested are assumed to be *locally* stationary, i.e. different characteristic space and time scales are considered at each location. However, similarly to most of the global operational SST L4 OI schemes, the correlation function used at each interpolation point is not changed within the pre-defined (space–time) influential radius during the optimal interpolation. This a priori assumption relaxes the model adopted in Buongiorno Nardelli et al. (2013), that considered a unique (basin-wide) space and time decorrelation scale. Local stationarity still represents an approximation of the true background covariance matrix, which would be unfeasible to estimate in NRT directly from the whole set of observations available, due to computational limitations.

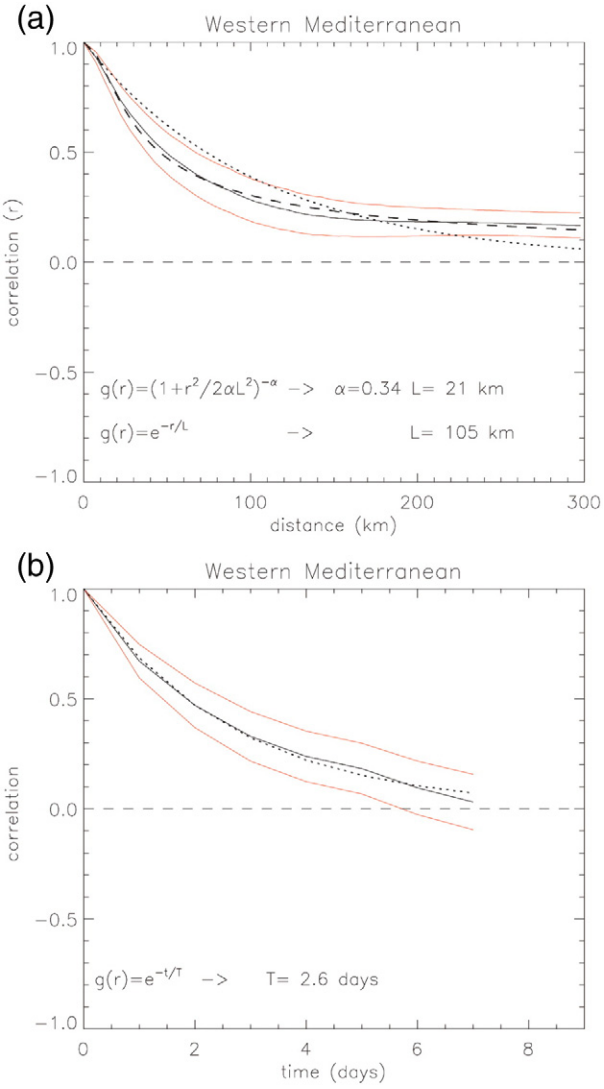

Fig. 1. Example of space (a) and time (b) Western Mediterranean sub-basin averaged lagged-correlations estimated from the observations (black continuous line). Red lines show the uncertainty ($\pm\sigma$) associated with the sub-basin averages. In (a) the dotted line shows the analytical correlation obtained as a best fit to a negative exponential, and the dashed line shows the best fit to a rational quadratic function. Lagged-correlations shown here have been computed from climatological anomalies (namely, using a pentad climatology as background).

A direct comparison of the background covariance function parameters estimated using the daily climatological (Fig. 2) and the previous day (Fig. 3) background fields is not straightforward. In fact, both L and α appreciably modify the correlation function shape, so that distinct pairs of these parameters can lead to very different characteristic spatial scales. However, it is the combined analysis of the spatial parameters and temporal decorrelation scales that gives some insight on the processes driving the covariance in the two cases. In fact, we should keep in mind that covariance will basically reflect the scale of the dominant processes observed in the anomalies with respect to the background field. In the case of the climatological background, L presents a more 'noisy' pattern, not directly relatable to average structures typical of

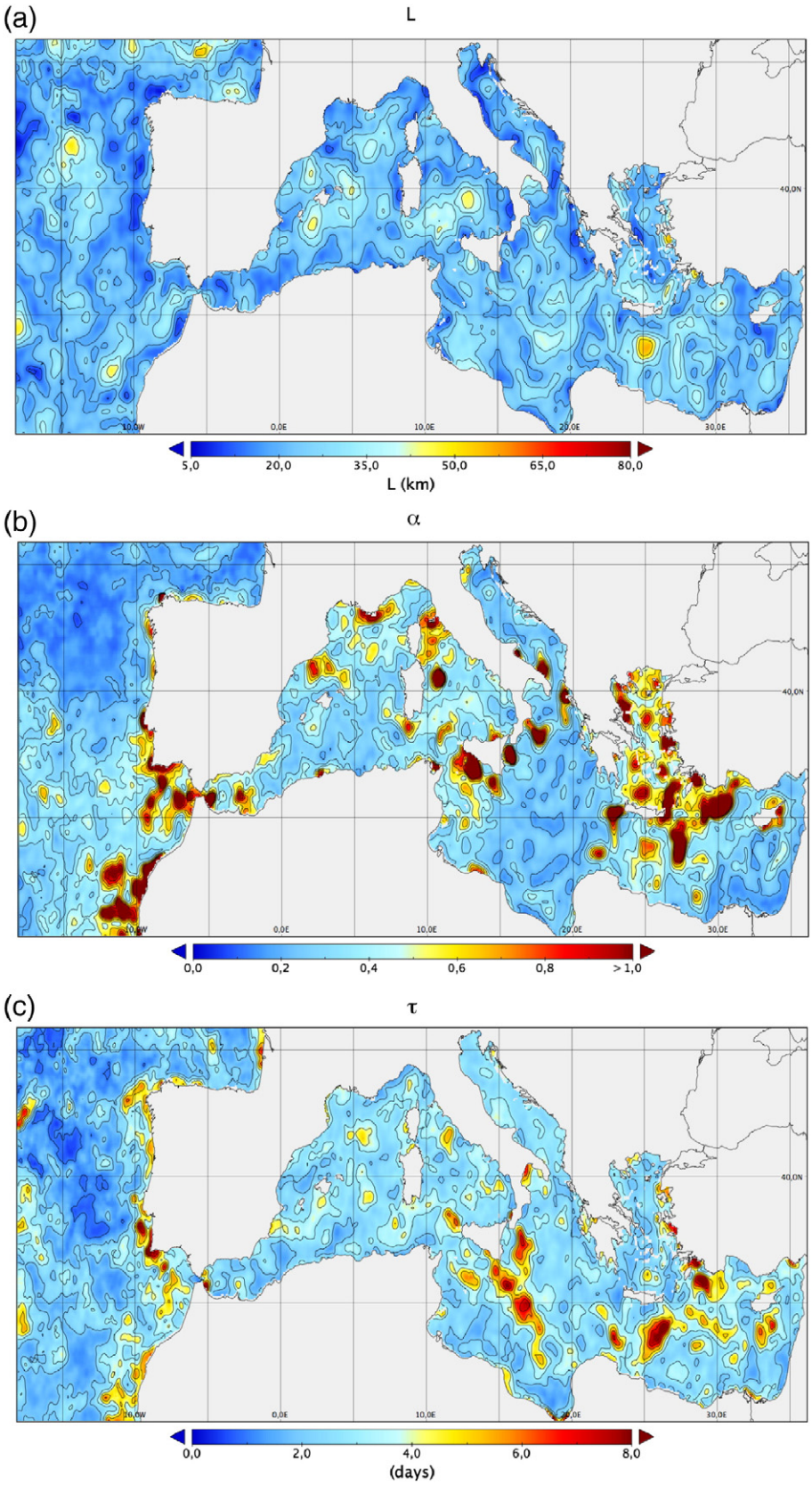

Fig. 2. Background covariance parameters defining the rational quadratic obtained using a pentad-climatology as first guess: (a) L, (b) α, and (c) τ.

Mediterranean basin, while temporal decorrelation occurs over few days to weekly scale (going up to 8 days, though attaining around 3 days on average). Conversely, temporal correlation decreases very rapidly when considering anomalies with respect to previous day (maximum <1.4 days, on average around 1 day), and average spatial correlation decays more slowly with increasing distance. However, in that case, lower *L* values are clearly associated with the areas where most of the energy is expected to be at the small scales, namely along

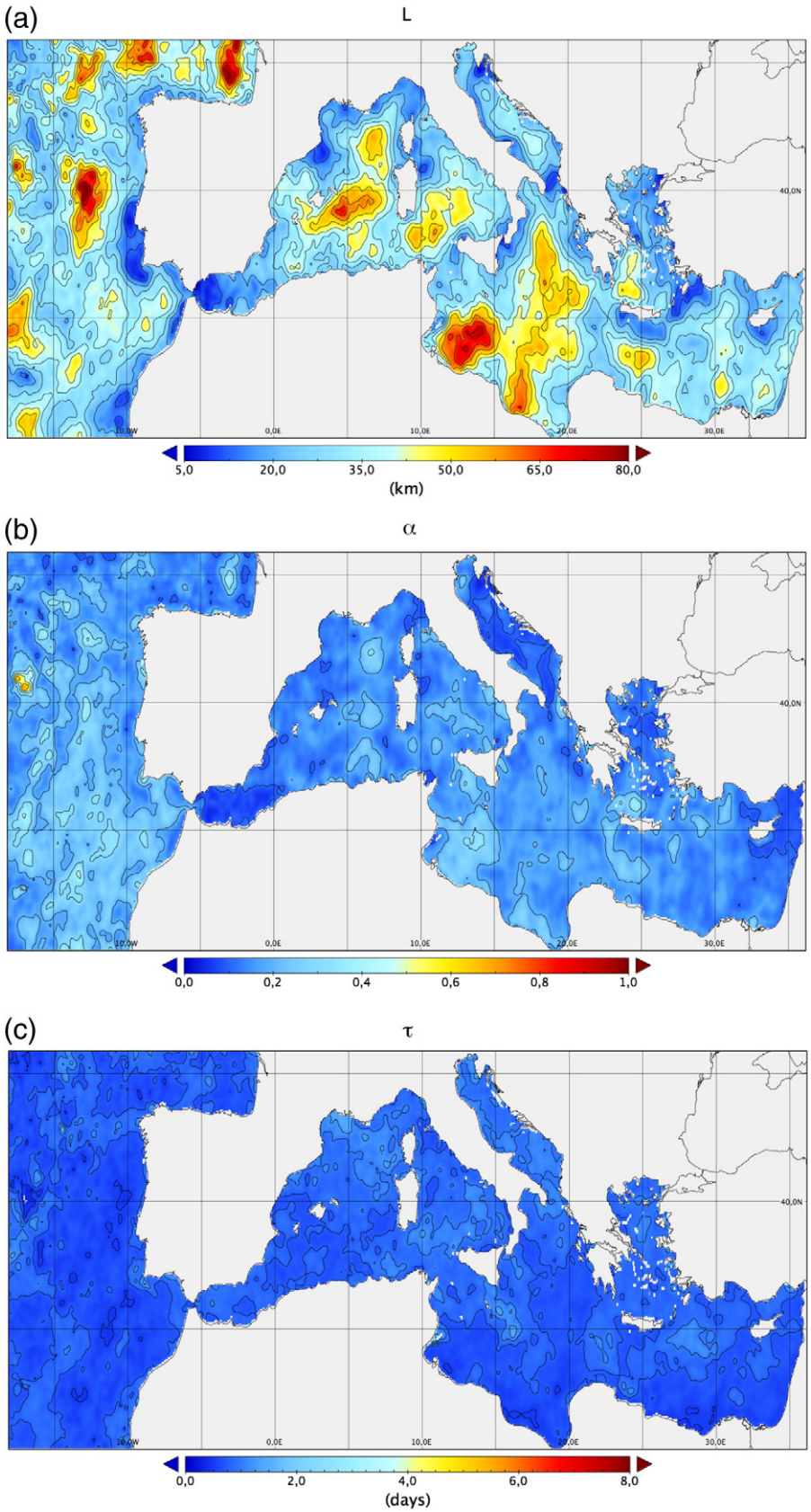

Fig. 3. Background covariance parameters defining the rational quadratic obtained using the previous day analysis as first guess: (a) L, (b) α, and (c)τ.

the coasts or in sub-basin structures characterized by intense (sub)mesoscale instabilities (e.g. around the Levantine gyres, in the Alboran Sea, in the Gulf of Lyons, etc.).

These patterns can be explained considering that most of the variance in the climatological case is given by the presence of ubiquitous mesoscale features in the anomaly fields, while only the slow evolution

of these features is revealed by the previous day analysis, which is conversely more strongly affected by the rapid large scale heating–cooling of the sea surface related to air–sea interactions (thus acting on larger sub-basin scales).

3.3. New signal variance and noise-to-signal ratio

Consistently with the definition of new background correlation functions, new spatially varying noise-to-signal ratios have been estimated. This computation was based on the three-way error analysis proposed by O'Carroll, Eyre, and Saunders (2008) and on the matchup dataset between satellite (L3S and climatological data) and in situ (drifter) observations set up by Buongiorno Nardelli et al. (2013). More specifically, assuming the errors from three data sources to be uncorrelated, it is possible to estimate the L3S observation variance (σ^2_{L3S}) from the variances of the difference between the three observation types (see also O'Carroll et al., 2008; Xu & Ignatov, 2010):

$$\sigma^2_{L3S} = \frac{1}{2}\left(\sigma^2_{L3S-drifter} + \sigma^2_{L3S-clim.} - \sigma^2_{drifter-clim.}\right). \quad (4)$$

In our case, basing on 3536 matchup data, it was possible to estimate $\sigma_{L3S\ -\ drifter} = 0.48°K$, $\sigma_{clim\ -\ drifter} = 0.89°K$, and $\sigma_{L3S\ -\ clim.} = 0.90°K$, which give $\sigma_{L3S} = 0.33\ °K$. Similarly, we estimated the climatology accuracy, $\sigma_{clim} = 0.83°K$, and the drifter accuracy, $\sigma_{drifter} = 0.32°K$, which is in agreement with previous estimates found in literature (Emery, Baldwin, Schlussel, & Reynolds, 2001; Gentemann, 2014; O'Carroll et al., 2008, Xu & Ignatov, 2010).

The (spatially varying) true signal variance as the difference between the variance of the difference between L3S and background data ($\sigma^2_{L3S\ -\ background}$), which is estimated from all available observation and background fields) and the noise variance (σ^2_{L3S}):

$$\sigma^2_{background} = \sigma^2_{background-L3S} - \sigma^2_{L3S}\left(+2r_{background,L3S}\sigma_{L3S}\sigma_{background}\right), \quad (5)$$

provided that the correlation between L3S and background errors ($2r_{background,L3S}$) can be assumed zero. This condition is essentially met when using a climatology as background field for two main reasons: 1) the climatology considered here covers a different time-period (1985–2005, see Marullo et al., 2007); 2) unless systematic errors are present in the observations (which is anyway an assumption of OI), the error of a climatological field is mostly given by the natural SST variability over the time considered, which is not necessarily related to the specific anomalies observed at a specific date, and generally much larger than typical L2P errors.

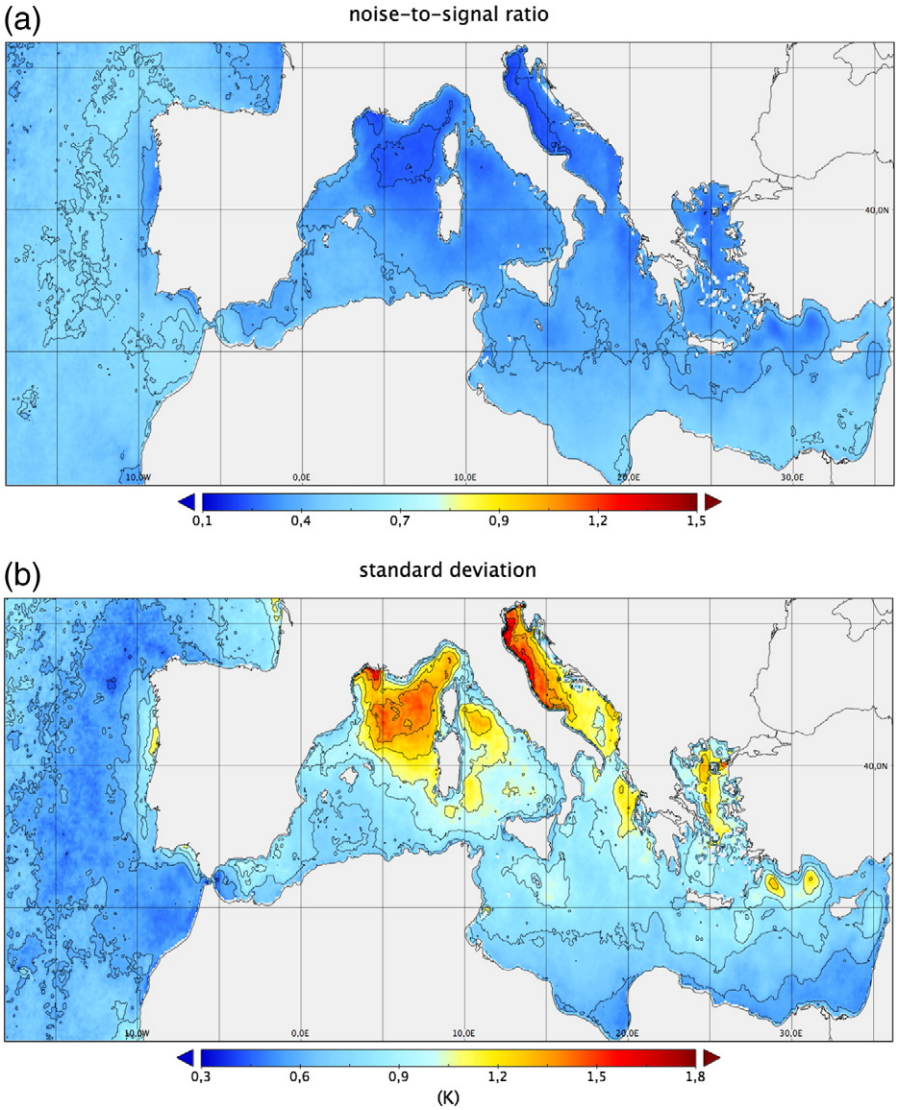

Fig. 4. Noise-to-signal (a) and signal standard deviation (b) estimated for the pentad-climatology first guess case.

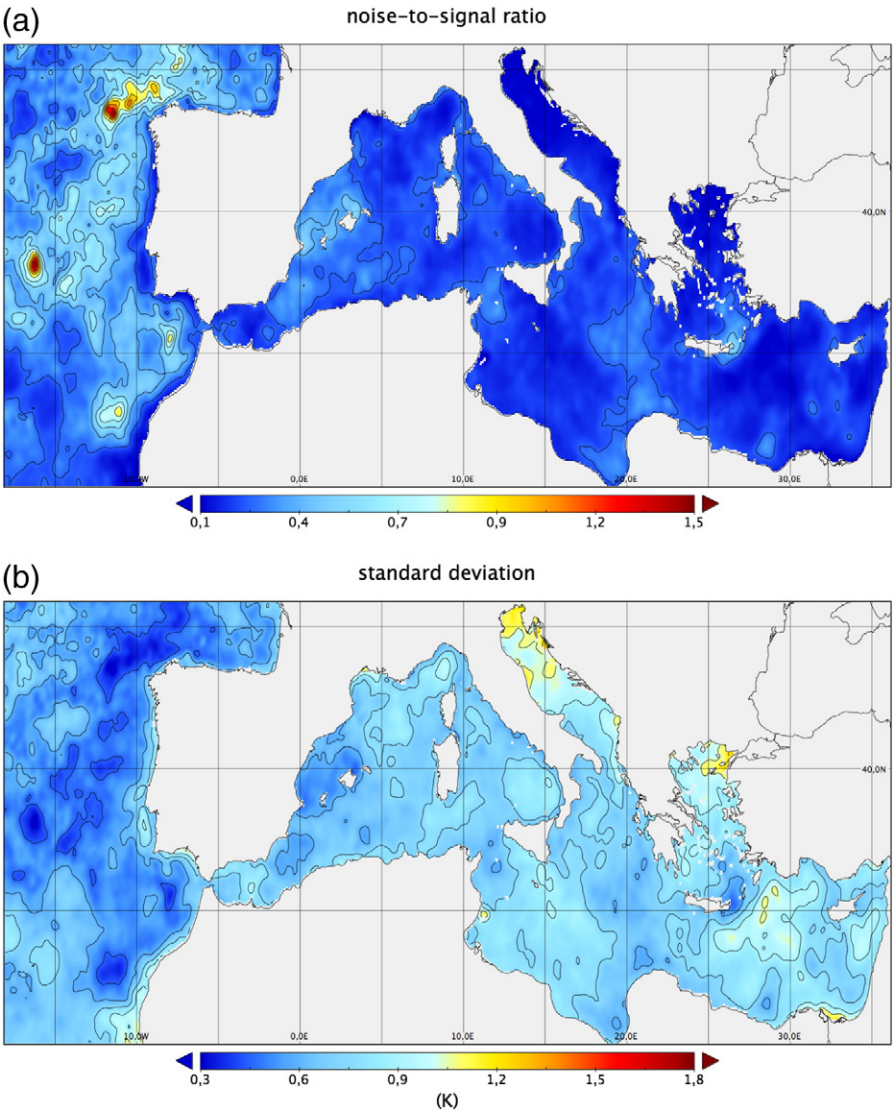

Fig. 5. Noise-to-signal (a) and signal standard deviation (b) estimated for the previous day analysis background case.

Computing the true signal variance in the case of the previous day background field gets more complicated for two reasons: firstly, it is likely that L3S and previous day analysis errors are correlated. In fact, L3S errors are often related to specific atmospheric conditions, possibly persisting for more than one day in specific areas. In that case, the background error would likely be underestimated, being correlated to the observation error. Secondly, as noise-to-signal ratio is clearly required by the OI, a first estimate of the background field is needed to compute it. Here, we have used the Version 3 MyOcean L4 product as an 'interim' previous day background field to be able to estimate the noise-to-signal levels needed by the new algorithm.

All the above might lead to less accurate noise-to-signal estimates, affecting the correct weighting of the new observations that enter the interpolation with respect to the background field, but we found no alternative way to estimate noise-to-signal until now.

Noise-to-signal ratio and signal variance (here showed as standard deviation) significantly differ depending on the first guess used (Figs. 4 and 5). Coherently with the larger variability observed in climatological anomalies than in daily increments, noise-to-signal ratio generally displays lower values in the first case (Fig. 4a), while it more often attains around (or even above) one in the latter (Fig. 5a). Consistently, signal variance is much smaller when considering daily increments (Figs. 4b and 5b).

4. Covariance models performance

The validation of the different covariance models was carried out following two different approaches, as detailed in the following sections. L4 data were first compared to co-located in situ drifter measurements taken from the matchup database described in Section 2.2, computing

Table 1
Definition of the OI configurations used for the tests.

OI schemes	Spatial covariance	Temporal covariance	Covariance parameters	Background field
V3	Negative exponential	Negative exponential	Constant	Pentad climatology
$V4_{clim}$	Rational quadratic	Negative exponential	Spatially variable	Pentad climatology
$V4_{prev}$	Rational quadratic	Negative exponential	Spatially variable	Previous day analysis

Table 2
Statistics of the differences between L4 SST estimates and drifter measurements (see text for details).

L4 vs. in situ data—MED HR 2011			
OI schemes	N_mup	MBE	STDE
V3	5698	−0.15 ± 0.01	0.46 ± 0.01
$V4_{clim}$	5698	−0.15 ± 0.01	0.47 ± 0.01
$V4_{prev}$	5698	−0.16 ± 0.01	0.48 ± 0.01

standard statistics of the differences. Successively, to get a more complete analysis of the impact of the different models in terms of space–time interpolation ability, a holdout validation approach has been followed, adding different types of synthetic clouds to the input data and comparing estimated L4 with the original L3S images under the simulated clouds. Confidence intervals for all the statistics considered have been estimated through a bootstrapping procedure (Efron and Tibshirani, 1993)

4.1. Comparison with in situ measurements

The three schemes considered are the standard OI described in Buongiorno Nardelli et al. (2013), using a spatially uniform covariance function and climatological first guess (hereafter referred to V3), and two OI using spatially varying covariance functions and climatological or previous day background fields, respectively (hereafter referred to as $V4_{clim}$ and $V4_{prev}$, see also Table 1). More than five thousands match-up data were used to estimate the statistics of the difference between satellite L4 and drifter SST. Somehow surprisingly, all schemes considered resulted almost equivalent (see Table 2), giving the same (at 95% confidence level) mean bias error (MBE = −.15 °K) and standard deviation of the difference (STDE = 0.46 °K). This result cannot, however, be considered conclusive, as direct comparison with in situ measurements does not provide information on the ability of the different schemes to correctly reconstruct surface structures that cannot be seen by the sparse sampling of drifters.

Table 3
Statistics of the differences between L4 SST estimates below artificial clouds and corresponding L3S data (see text for details).

Artificial Cloud Test #1: 12 clouds (253 × 10 pixels), daily cloud shift = 100 pixels (MED HR 2011)			
OI schemes	N_mup	MBE	
V3	3,849,112	0.0413 ± 0.0003	0.2299 ± 0.0004
$V4_{clim}$	3,849,112	0.0405 ± 0.0003	0.2201 ± 0.0004
$V4_{prev}$	3,849,112	0.0340 ± 0.0003	0.2242 ± 0.0004
Artificial Cloud Test #2: 6 clouds (253 × 20 pixels), daily cloud shift = 100 pixels (MED HR 2011)			
OI schemes	N_mup	MBE	STDE
V3	3,684,308	0.0448 ± 0.0003	0.2957 ± 0.0004
$V4_{clim}$	3,684,308	0.0436 ± 0.0003	0.2933 ± 0.0004
$V4_{prev}$	3,684,308	0.0404 ± 0.0003	0.2816 ± 0.0004
Artificial Cloud Test #3: 3 clouds (253 × 40 pixels), daily cloud shift = 100 pixels (MED HR 2011)			
OI schemes	N_mup	MBE	STDE
V3	3,461,275	0.0386 ± 0.0003	0.3638 ± 0.0004
$V4_{clim}$	3,461,275	0.0378 ± 0.0003	0.3636 ± 0.0004
$V4_{prev}$	3,461,275	0.0347 ± 0.0003	0.3598 ± 0.0004
Artificial Cloud Test #4 1 cloud (253 × 120 pixels), daily cloud shift = 100 pixels (MED HR 2011)			
OI schemes	N_mup	MBE	STDE
V3	3,488,326	0.0358 ± 0.0003	0.3912 ± 0.0004
$V4_{clim}$	3,488,326	0.0348 ± 0.0003	0.3923 ± 0.0004
$V4_{prev}$	3,488,326	0.0325 ± 0.0003	0.3920 ± 0.0004

4.2. Validation with synthetic clouds

In order to verify if any improvement could be found with one of the new schemes, a holdout validation strategy has been followed next. In practice, artificial clouds were added to the L3S observations before running the OI. L3S observations falling under the artificial clouds were then considered as true values and used to validate the different schemes. Two different types of synthetic clouds have been used. Firstly, similarly to Marullo et al. (2014), we used regularly spaced parallel longitudinal bands of various width that move from the western to the eastern boundary of the Mediterranean Sea. The dimension and persistence of the artificial clouds were chosen accordingly to the average cyclone characteristics of the Mediterranean Sea (Trigo & Trevor, 1999). Secondly, we considered random distributions of square clouds of increasing size, imposing different cloud persistence periods (1, 3, 6 days). Cloud distribution was thus randomly changed at the end of each persistence period in order to avoid uneven sampling of low/high variability regions, and added to real cloud coverage. The results of these holdout validations are summarized in Tables 3 and 4. They show that very similar errors are observed whatever the cloud

Table 4
Statistics of the differences between L4 SST estimates below artificial random clouds and corresponding L3S data (see text for details).

Artificial Random Cloud Test #1: 49 clouds (25 × 25 pixels)—MED HR 2011			
1 day persistency	N_mup	MBE	STDE
V3	3,424,186	0.0466 ± 0.0003	0.2851 ± 0.0004
$V4_{clim}$	3,424,186	0.0457 ± 0.0003	0.2816 ± 0.0004
$V4_{prev}$	3,424,186	0.0401 ± 0.0003	0.2739 ± 0.0004
3 day persistency	N_mup	MBE	STDE
V3	3,423,371	0.0341 ± 0.0003	0.2820 ± 0.0004
$V4_{clim}$	3,423,371	0.0337 ± 0.0003	0.2815 ± 0.0004
$V4_{prev}$	3,423,371	0.0360 ± 0.0003	0.2990 ± 0.0004
6 day persistency	N_mup	MBE	STDE
V3	3,452,859	0.0282 ± 0.0003	0.2880 ± 0.0004
$V4_{clim}$	3,452,859	0.0292 ± 0.0003	0.2878 ± 0.0004
$V4_{prev}$	3,452,859	0.0299 ± 0.0003	0.3217 ± 0.0004
Artificial Random Cloud Test #2: 12 clouds (50 × 50 pixels)—MED HR 2011			
1 day persistency	N_mup	MBE	STDE
V3	3,470,794	0.0398 ± 0.0003	0.3513 ± 0.0004
$V4_{clim}$	3,470,794	0.0388 ± 0.0003	0.3512 ± 0.0004
$V4_{prev}$	3,470,794	0.0333 ± 0.0003	0.3495 ± 0.0004
3 day persistency	N_mup	MBE	STDE
V3	3,508,708	0.0494 ± 0.0003	0.3607 ± 0.0004
$V4_{clim}$	3,508,708	0.0491 ± 0.0003	0.3627 ± 0.0004
$V4_{prev}$	3,508,708	0.0518 ± 0.0003	0.3822 ± 0.0004
6 day persistency	N_mup	MBE	STDE
V3	3,426,693	0.0359 ± 0.0003	0.3720 ± 0.0004
$V4_{clim}$	3,426,693	0.0381 ± 0.0003	0.3769 ± 0.0004
$V4_{prev}$	3,426,693	0.0492 ± 0.0003	0.4312 ± 0.0004
Artificial Random Cloud Test #3: 3 clouds (100 × 100 pixels)—MED HR 2011			
1 day persistency	N_mup	MBE	STDE
V3	3,813,800	0.0318 ± 0.0003	0.3862 ± 0.0004
$V4_{clim}$	3,813,800	0.0299 ± 0.0003	0.3865 ± 0.0004
$V4_{prev}$	3,813,800	0.0279 ± 0.0003	0.3808 ± 0.0004
3 day persistency	N_mup	MBE	STDE
V3	3,880,500	0.0310 ± 0.0003	0.4197 ± 0.0004
$V4_{clim}$	3,880,500	0.0315 ± 0.0003	0.4203 ± 0.0004
$V4_{prev}$	3,880,500	0.0326 ± 0.0003	0.4488 ± 0.0004
6 day persistency	N_mup	MBE	STDE
V3	3,958,965	0.0325 ± 0.0003	0.4708 ± 0.0004
$V4_{clim}$	3,958,965	0.0337 ± 0.0003	0.4742 ± 0.0004
$V4_{prev}$	3,958,965	0.0371 ± 0.0003	0.5506 ± 0.0004

configuration and scheme considered. In the first tests (Table 3), the STDE increased when adding larger clouds (from 0.22 °K to 0.39 °K), coherently with the necessity to extrapolate from more distant (less correlated) observations under the artificial clouds. MBE gave less clear indications, but the extremely small differences found (though statistically significant at 95% confidence level) have practically no impact from a physical point of view. Similarly, in the case of random clouds of increasing size and persistence (Table 4), the STDE ranged between 0.27 °K and 0.55 °K, with very minor differences between the three schemes in the case of small clouds (Random Clouds Test #1), even at 3–6 day persistence. Somehow surprisingly, in that test (actually only in that test), the MBE decreased with prolonged cloudiness, though differences still do not exceed 0.01 °K.

However, from Random Clouds Test #2 and #3, there is a more clear indication on the performance of the three schemes in case of data gaps persisting for more than 3 days. In fact, though no scheme outperformed the others, the $V4_{clim}$ only slightly improving with respect to V3, it is evident that the $V4_{prev}$ scheme is less accurate, with an MBE/STDE increase ranging between 5% and 10%.

5. Conclusions

Defining a proper covariance model for the optimal interpolation of satellite data is generally considered a crucial step to obtain accurate SST estimates for operational applications. Most of the operational schemes are based on simplifying approximations involving parametric covariance calibration and/or local low-dimensional covariance models (e.g. Martin et al., 2012, and references therein). In the framework of MyOcean research activities, we defined and tested two new multi-dimensional (space–time) covariance models for the operational interpolation of satellite infrared SST measurements at high resolution over the Mediterranean Sea (daily, 1/16° × 1/16°). The new schemes were compared with the standard approximated OI model described in Buongiorno Nardelli et al. (2013), which uses a uniform space–time separable covariance function for the whole Mediterranean Sea, and included the definition of a new functional dependence for the spatial covariance, the definition of spatially varying covariance parameters and the alternate use of two different background fields, namely a daily (pentad) climatological field or the previous day analysis. Corresponding noise-to-signal ratio have been computed by applying the three-way approach proposed by O'Carroll et al. (2008) to co-located L3S, drifter and climatological data. The three-way error analysis also allowed to evaluate the absolute accuracy of drifter ($\sigma_{drifter} = 0.32°K$), L3S ($\sigma_{L3S} = 0.33°K$) and climatological ($\sigma_{clim} = 0.83°K$) SST estimates, the former lying within the range of previous similar estimations (Emery et al., 2001; Gentemann, 2014; O'Carroll et al., 2008; Xu & Ignatov, 2010). Of course, it should be kept in mind that there are several potential sources of differences for these values. In particular, while we expect minor discrepancies related to individual in situ instruments accuracy, a significant difference is related to the in situ and satellite-derived data representativeness, which is clearly affected by the different product resolution and pre-processing steps.

The fields interpolated with the three OI schemes display no significant changes in terms of differences vs in situ drifter data. This can be explained by the relatively small differences in the covariance values that enter the OI weighting, so that, whenever a relatively high number of observations is available in space and time (so that the choice of the first guess becomes less important), equivalent reconstructions are obtained. However, as shown by performing a holdout validation with artificial clouds, using previous day analysis as the background and considering spatially varying covariance parameters leads to a slightly degraded surface feature reconstruction when dealing with data gaps persisting for at least 3 days or more (with a MBE/STDE increase of more than 5%). In fact, this result is not particularly surprising as, contrarily to other common interpolation schemes that only use daily innovations as input data to a purely spatial analysis, the interpolation here is carried out both in space and time. In fact, space–time OI schemes always include in the analysis all the observations found within a temporal influential radius (10 days in our case), whatever the background field considered. As a consequence, only minor differences are found in the weights used to combine the closest observations (which almost completely *hide* the background), while low-correlated observations (e.g. those far in both space and time) only provide minor corrections to the background, and this is a possible explanation of why previous day field is less accurate than climatology in case of prolonged data voids. The new scheme using spatially variable covariance parameters and the daily climatology as background has thus been implemented in the MyOcean Mediterranean SST operational processing chain since April 2014, and will be maintained in the framework of European Copernicus Marine Service.

Holdout validation obtained through artificial clouds also proved as a powerful tool to investigate algorithms' performance, and might in the future be used for extended validation exercises, including an analysis of SST gradients under the synthetic clouds, and possibly complementing more classical approaches for the evaluation of effective product resolution (e.g. through spatial wavenumber spectra).

Acknowledgements

This work has been carried out in the framework of the MyOcean-2 project ("Prototype Operational Continuity for the GMES Ocean Monitoring and Forecasting Service", funded within the EU call FP7-SPACE-2011-1SPA.2011.1.5-01: Prototype operational continuity of GMES services in the Marine Area (Grant agreement: 283367)) and MyOcean Follow-on project (funded within a European Commission Horizon 2020 Framework Programme Grant Agreement) call EU H2020-Adhoc-2014-20 (Grant agreement: 633085). All satellite data used were taken from the Group for High-Resolution Sea Surface Temperature (GHRSST) Global and Regional Data Assembly Centres (GDAC, RDAC; see www.ghrsst.org for more details).

References

Bayley, G. V., & Hammersley, J. M. (1946). The "effective" number of independent observations in an autocorrelated time series. *Supplement to the Journal of the Royal Statistical Society*, *8*(2), 184–197.

Bonanno, A., Zgozi, S., Basilone, G., Hamza, M., Barra, M., Genovese, S., et al. (2015). Acoustically detected pelagic fish community in relation to environmental conditions observed in the Central Mediterranean sea: A comparison of Libyan and Sicilian – Maltese coastal areas. *Hydrobiologia*. http://dx.doi.org/10.1007/s10750-015-2234-0.

Bretherton, F. P., Davis, R. E., & Fandry, C. B. (1976). A technique for objective analysis and design of oceanographic experiments applied to MODE-73. *Deep Sea Research*, *23*, 559–582.

Buongiorno Nardelli, B. (2012). A novel approach for the high-resolution interpolation of in situ sea surface salinity. *Journal of Atmospheric and Oceanic Technology*, *29*, 867–879. http://dx.doi.org/10.1175/JTECH-D-11-00099.1 (http://journals.ametsoc.org/doi/abs/10.1175/JTECH-D-11-00099.1).

Buongiorno Nardelli, B., Colella, S., Santoleri, R., Guarracino, M., & Kholod, A. (2010). A re-analysis of Black Sea surface temperature. *Journal of Marine Systems*, *79*, 1–2. http://dx.doi.org/10.1016/j.jmarsys.2009.07.001 (50–64).

Buongiorno Nardelli, B., Larnicol, G., D'Acunzo, E., Santoleri, R., Marullo, S., & Le Traon, P. Y. (2003). Near real time SLA and SST products during 2-years of MFS pilot project: Processing, analysis of the variability and of the coupled patterns. *Annales Geophysicae*, *20*, 1–19 (2).

Buongiorno Nardelli, B., Tronconi, C., Pisano, A., & Santoleri, R. (2013). High and ultra-high resolution processing of satellite sea surface temperature data over southern European seas in the framework of MyOcean project. *Remote Sensing of Environment*, *129*, 1–16. http://dx.doi.org/10.1016/j.rse.2012.10.012.

Chelton, D. B., & Wentz, F. J. (2005). Global microwave satellite observations of sea-surface temperature for numerical weather prediction and climate research. *Bulletin of the American Meteorological Society*, *86*, 1097–1115.

Dobricic, S., Pinardi, N., Adani, M., Tonani, M., Fratianni, C., Bonazzi, A., et al. (2007). Daily oceanographic analyses by Mediterranean Forecasting System at the basin scale. *Ocean Science*, *3*, 149–157.

Donlon, C. J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E., & Wimmer, W. (2012). The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system. *Remote Sensing of Environment*, *116*, 140–158. http://dx.doi.org/10.1016/j.rse.2010.10.017.

Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall/CRC 456 pp.

Emery, W. J., Baldwin, D. J., Schlussel, P., & Reynolds, R. W. (2001). Accuracy of in situ sea surface temperatures used to calibrate infrared satellite measurements. *Journal of Geophysical Research*, *106*(C2), 2387–2405.

Gandin, L. S. (1965). *Objective analysis of meteorological fields.* Jerusalem: Israel Program for Scientific Translation, 242.

Gentemann, C. L. (2014). Three way validation of MODIS and AMSR-E sea surface temperatures. *Journal of Geophysical Research*, *119*, 2583–2598. http://dx.doi.org/10.1002/2013JC009716.

Le Traon, P. Y., Nadal, F., & Ducet, N. (1998). An improved mapping method of multisatellite altimeter data. *Journal of Atmospheric and Oceanic Technology*, *15*, 522–533.

Liu, G., Rauenzahn, J. L., Heron, S. F., Eakin, C. M., Skirving, W., Christensen, T. R. L., et al. (2013). NOAA coral reef watch 50 km satellite sea surface temperature-based decision support system for coral bleaching management. *NOAA technical report NESDIS. 143*, (35 pp.).

Martin, M., Dash, P., Ignatov, A., Banzon, V., Beggs, H., Brasnett, B., et al. (2012). Group for High Resolution Sea Surface temperature (GHRSST) analysis fields intercomparisons. Part 1: A GHRSST multi-product ensemble (GMPE). *Deep Sea Research Part II: Topical Studies in Oceanography*, 77–80. http://dx.doi.org/10.1016/j.dsr2.2012.04.013 (21–30).

Marullo, S., Buongiorno Nardelli, B., Guarracino, M., & Santoleri, R. (2007). Observing the Mediterranean sea from space: 21 years of pathfinder-AVHRR sea surface temperatures (1985 to 2005). Re-analysis and validation. *Ocean Science*, *3*, 299–310.

Marullo, S., Santoleri, R., Ciani, D., Le Borgne, P., Péré, S., Pinardi, N., et al. (2014). Combining model and geostationary satellite data to reconstruct hourly SST field over the Mediterranean Sea. *Remote Sensing of Environment*, *146*, 11–23. http://dx.doi.org/10.1016/j.rse.2013.11.001.

O'Carroll, A. G., Eyre, J. R., & Saunders, R. W. (2008). Three-way error analysis between AATSR, AMSR-E, and in situ sea surface temperature observations. *Journal of Atmospheric and Oceanic Technology*, *25*, 1197–1207. http://dx.doi.org/10.1175/2007JTECHO542.1.

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning, chapter 4: Covariance functions.* The MIT Press (ISBN 026218253X).

Reynolds, R. W., & Chelton, D. B. (2010). Comparisons of daily sea surface temperature analyses for 2007–08. *Journal of Climate*, *23*, 3545–3562. http://dx.doi.org/10.1175/2010JCLI3294.1.

Roberts-Jones, J., Fiedler, E. K., & Martin, M. J. (2012). Daily, global, high-resolution SST and sea ice reanalysis for 1985–2007 using the OSTIA system. *Journal of Climate*, *25*, 6215–6232. http://dx.doi.org/10.1175/JCLI-D-11-00648.1.

Robinson, I. S. (2004). *Measuring the oceans from space: The principles and methods of satellite oceanography.* Berlin, Germany: Springer/Praxis Publishing (669 pp.).

Santoleri, R., Marullo, S., & Böhm, E. (1991). An objective analysis scheme for AVHRR imagery. *International Journal of Remote Sensing*, *12*(4), 681–693.

Trigo, I. F., & Trevor, D. D. (1999). Objective climatology of cyclones in the Mediterranean region. *Journal of Climate*, *12*, 1685–1696.

Volpe, G., Buongiorno Nardelli, B., Cipollini, P., Santoleri, R., & Robinson, I. S. (2012). Seasonal to interannual phytoplankton response to physical processes in the Mediterranean Sea from satellite observations. *Remote Sensing of Environment*, *117*, 223–235.

Xu, F., & Ignatov, A. (2010). Evaluation of in situ sea surface temperatures for use in the calibration and validation of satellite retrievals. *Journal of Geophysical Research Oceans*, *115*, 1–18. http://dx.doi.org/10.1029/2010JC006129.