Skills of different hydrographic networks in capturing changes in the Mediterranean Sea at climate scales

J. Llasses1,*, G. Jordà1, D. Gomis1,2

1Mediterranean Institute for Advanced Studies (IMEDEA, UIB-CSIC), c/Miquel Marquès 21, 07190 Esporles, Spain
2University of the Balearic Islands, Physics Department, Cra.Valldemossa km 7.5, 07122 Palma, Spain

ABSTRACT: The skills of 5 observational networks are explored in the context of the monitoring of climate signals in the Mediterranean Sea. Namely we explore the capabilities of hydrographic surveys and ships of opportunity, of Argo buoys, of a (virtual) regularly distributed mooring network, of the present-day observational system (which makes use of the 3 kinds of observations) and of a targeted future system. The skills of each observational network are quantified as follows: first, the output of a realistic regional circulation model (considered here as the virtual truth) is sampled at the same time and location of the actual observations gathered by each observational network. An objective analysis scheme based on Optimal Statistical Interpolation is then applied to the pseudo-observations to obtain gridded products, which are compared to the model output in order to infer the capability of each sampling to capture the true fields. We do it for different periods (for 1962−2000 and for the whole 21st century) and for different parameters (temperature, salinity and the rate of deep water formation in the Western Mediterranean). Results indicate that the skills to reproduce large scale climatic signals depend on the depth and variable, ranging from >90% of explained monthly variance and <5% relative trend errors for the upper (0−100 m) and intermediate layer (100−400 m) temperature fields, to <60% of variance and 30% relative trend errors for the upper layer salinity field. When averaging temperature and salinity over the whole basin volume, both annual values and long term trends are properly captured by all the networks, though the deep water formation rate in the Western Mediterranean is largely overestimated. Conversely, regional features are missed by all the sampling networks, since none of them has an adequate spatial distribution to capture small scale processes.

KEY WORDS: Climate variability · Mediterranean · Monitoring · Observing systems · Climate change

1. INTRODUCTION

Climate monitoring is nowadays a crucial issue, particularly in 'hot spot' regions such as the Mediterranean Basin (see e.g. Giorgi 2006), where climate changes are expected to be large. Adaptation to future climate scenarios in a fast and effective way requires the combination of climate projections and proper monitoring of the climate signal. Concerning the temperature and salinity of the Mediterranean Sea, there are 2 key factors that hinder climate change detection: the fact that climate changes are still of the same order or smaller than inter-annual natural variability, and the paucity and non-homogeneous spatio-temporal distribution of observations (e.g. Vargas-Yáñez et al. 2012).

Regarding the available observations, different authors (see e.g. Tsimplis et al. 2011, Jordà & Gomis 2013) have analyzed historical databases such as MEDATLAS (MEDAR Group 2002) and concluded that the number and distribution of observations spanning the last decades of the 20th century are in-
sufficient to properly characterize the hydrographic climate variability and trends at different depths. More specifically, they have shown that in those databases the intra-annual variability is poorly represented, the interannual variability may not be representative of the actual variability and the inferred trends are underestimated.

In such a context, one must ask how far the present observational networks are able to properly monitor Mediterranean climate variability and, in particular, climate change signals. Several authors have analyzed the capabilities of different observational networks to capture features of ocean variability. For instance, Ruiz et al. (2007) evaluated the optimal number of Argo floats needed to recover the North-East Atlantic temperature fields; more recently, Juza et al. (2012) have estimated the distortion of the mixed layer property distributions induced by the Argo sampling, and L’Hévéder et al. (2013) have investigated the potential skills of a fleet of gliders to capture mesoscale and submesoscale variability. However, to our knowledge, similar studies focused on Mediterranean climate variability do not exist. The goal of this work is to fill that void by quantifying the skills of different monitoring networks in capturing climate variability, and make some suggestions regarding the monitoring of climate change in the Mediterranean Sea over the next decades.

The methodology of this study was as follows. (1) We extracted a set of temperature and salinity pseudo-observations from numerical simulations spanning the 1962–2100 period. This covered a period representative of the present climate (1962–2000) and a period affected by climate change (2000–2100). The numerical simulations were considered as the (virtual) ‘truth’, and the pseudo-observations had the same spatio-temporal distribution as the actual observations acquired by different sampling strategies: oceanographic cruises, Argo buoys, a (virtual) mooring network, and the present-day observational system (which makes use of the 3 kinds of observations above). A target future observational network was also considered. (2) We applied an objective analysis scheme (Optimal Statistical Interpolation) to the extracted pseudo-observations in order to obtain gridded products similar to those made available from different databases. The suitability of the 5 observational distributions to capture the climate signal (from interannual variability to long-term trends) was evaluated by comparing the analyzed fields with the numerical simulation. This methodology is described in detail in Section 2.

2. DATA AND METHODS

2.1. Numerical simulations

The numerical simulations used in this study have been carried out with NEMOMED8 (Sevault et al. 2009, Beuvier et al. 2010), a regional configuration of the 3D baroclinic ocean model NEMO (Madec 2008). NEMOMED8 covers the whole Mediterranean Sea plus a buffer zone in the near Atlantic Ocean, but does not explicitly include the Black Sea. The model is eddy-permitting and has a horizontal resolution of 1/8° in longitude × 1/8°cos(φ) in latitude, where φ represents the latitude (i.e. it uses a squared mesh with ~9 km resolution to the north of the domain and ~12 km to the south). Near the Gibraltar Strait the grid is tilted and stretched in order to properly reproduce the SW–NE axis of the real strait (the resolution is up to 6 km in that region). The vertical layover consists of 43 unevenly spaced levels (where Δz = 6 to 200 m from the upper layers to deepest layers, with 25 levels in the first 1000 m). The bathymetry is based on the ETOPO5’x5’ database (Smith & Sandwell 1997). NEMOMED8 has been shown to be particularly suitable to simulate the hydrodynamics of the Mediterranean Sea. For instance, Beuvier et al. (2010) have used it to model the interannual variability for the period 1961–2000, showing that the model is able to reproduce it in the upper and intermediate layers and to reasonably reproduce the Eastern Mediterranean Transient. Herrmann et al. (2010) have used the same model to analyze deep convection events in the Western Mediterranean, after demonstrating the good agreement between the simulation and deep convection observations.

Here NEMOMED8 is forced with the atmospheric fields of a dynamical downscaling of the CNRM-CM3 global model. The downscaling was carried out with the ARPEGE regional model, which has a spatial resolution of ~50 km over the Mediterranean. The atmospheric model follows the observed greenhouse gas and aerosol concentrations up to year 2000 (control period), and the SRES A2 scenario from 2001 to 2100 (scenario period). For a more detailed description of the forcings the reader is referred to Somot et al. (2006), Jordà et al. (2012) or Albouy et al. (2013). Fig. 1 shows the evolution of the 3D annual basin averaged temperature and salinity during the whole period; both variables show a clear increase (1.1°C and 0.36 psu, respectively) during the period 2000–2100. Complementarily, Fig. 2 shows the spatial distribution of the 5 m temperature trends computed for the same period. The values are similar
over the whole basin, with the exception of large values obtained in small areas (e.g. to the north of the Balearic Sea or in the Aegean Sea). At deeper layers (not shown) the interannual variability and trends are smaller than at upper layers for both temperature and salinity. Fig. S1 (in the Supplement at www.intres.com/articles/suppl/c063p001_supp.pdf) shows the mean temperature and salinity profiles averaged over the whole basin and over a 25 yr window. It can be seen that the temperature and salinity increases are specially marked in the last part of the period (2075–2100) at all depths. Because we are interested in long-term signals, the ‘true values’ will consist of the monthly mean values extracted from the simulation.

2.2. Virtual observational network

For each observational network (Table 1), pseudo-observations were extracted from the model monthly outputs at the given position and time. Then Gaussian noise was added to the pseudo-observations in order to account for both observational errors and short scales (in time and space) not resolved by the model monthly fields but that would be present in real observations. The added noise is scaled accordingly to the decrease of the variability with depth (Table 2). The description of each observational network is as follows.

(1) Oceanographic cruises (hereafter referred to as the ‘cruise’ sampling). This first distribution intends to evaluate the advantages and disadvantages of the observations gathered by dedicated surveys and ships of opportunity. These observations are normally not designed for climate studies and therefore

<table>
<thead>
<tr>
<th></th>
<th>Mean number of observations</th>
<th>Horizontal coverage</th>
<th>Vertical coverage</th>
<th>Temporal coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moorings</td>
<td>16</td>
<td>Regularly distributed</td>
<td>Only a few levels sampled. The rest are linearly interpolated</td>
<td>High frequency</td>
</tr>
<tr>
<td>Cruises</td>
<td>393</td>
<td>Inhomogeneously distributed. Much of the observations correspond to campaigns located in a certain region</td>
<td>Complete but with a high decrease in the number of observations &lt;400 m (see Table 3)</td>
<td>Very low frequency. Campaigns usually do not last long in certain regions</td>
</tr>
<tr>
<td>Argo</td>
<td>96</td>
<td>Drifters move throughout the basin except shallow regions or closed seas</td>
<td>Complete but with a high decrease in the number of observations &lt;700 m (see Table 3)</td>
<td>Worse than moorings but much better than cruises</td>
</tr>
</tbody>
</table>

*Mean monthly observations at 5 m depth*
their spatio-temporal coverage may not be adequate for that goal. Thus, a major objective here is to quantify how the spatio-temporal inhomogeneity that characterizes the cruise observation distribution can affect the capture of climate signals. The location and time of each observation were obtained from the cruises included in the MEDATLAS database (MEDAR Group 2002), which is considered to be the most complete data base of the Mediterranean Sea. More precisely, we considered the CTD casts corresponding to the period 1980–1990 as the most representative of the recent decades in terms of coverage (MEDATLAS was built up in 2000 and therefore some of the cruises carried out during the decade 1990–2000 were not included in the database). Each year of the analysis period 1962-2100 was sampled with the observations corresponding to a randomly selected year of the 1980-1990 period. The number of observations of the cruise distribution strongly depends on depth: a mean of 393 observations were gathered at 5 m depth within 1 mo, a large part of them (~20%) located over the continental shelf.

Table 3 shows an abrupt decrease in the number of observations between 400 and 600 m depth because many oceanographic cruises, especially those devoted to biological surveys, do not sample deeper layers.

(2) Argo buoys. Argo is an international project that deploys an array of free-drifting profiling floats over the global ocean. It constitutes a pilot project of the Global Ocean Observing System and a major contributor to the Global Climate Observing System (Howard et. al. 2010). In the Mediterranean Sea, Argo buoys are set to measure temperature and salinity from the surface to 2000 m depth, in a 10 d period: at Day 5 the buoy collects data from the surface to 700 m depth; At Day 10, data is collected down to 2000 m depth. To simulate a typical coverage of Argo observations, we use the Argo profiles corresponding to the period 2005–2011 (as extracted from the Coriolis database at: www.coriolis.eu.org/). The mean number of observations per month is 96 in the upper 700 m and less than half that at deeper levels (due to the number of buoys over regions with shallower depths, see Table 3). As for the cruises, the annual spatio-temporal distribution of observations for the chosen 6 yr period is repeated randomly to cover the whole simulated period.

(3) Moorings. The third distribution does not correspond to an actual sampling network, but to a virtual sampling strategy: it consists of a spatially regular network (with a mean separation distance of ~300 km) of 16 moorings collecting data continuously in time. With this distribution we aimed to evaluate the capabilities of a dedicated monitoring system collecting data at a few locations but with a homogeneous spatio-temporal distribution. Unlike for the cruises and Argo buoys, only a set of vertical levels are measured (5, 10, 20, 50, 100, 500, 1000 and 2000 m), simulating an actual mooring array. We applied a linear interpolation to obtain the in-between level values.

(4) Realistic network. The 4th sampling method is the actual observational network, and will therefore illustrate the capabilities of the present monitoring of the Mediterranean Sea. It is based on the addition of cruises, Argo buoys and 12 actual mooring stations whose locations have been obtained from the CIESM-Hydrochanges program (see www.ciesm.org/marine/programs/hydrochanges.htm): 9 of the moorings are located on the continental shelf, while the other 3 acquire data at deeper layers also. The sampling rate and the Gaussian noise added to the pseudo-observations are the same as for the cruises and Argo networks.

(5) Target network. The last network is an optimistic but plausible future network where the number of CTD casts and Argo buoys are doubled with respect to the realistic network. Furthermore, the moorings of the realistic network are extended with additional moorings that are expected to be operational in the next years and decades, according to the CIESM-Hydrochanges and Dyfamed programs.
The objective of testing this network is to quantify the improvements (always in terms of climate signal detection) with respect to the present network.

Additionally, we also test a sampling configuration with a high spatial and temporal density (horizontal separation distance of ~70 km and 25 levels from the surface to the bottom, collecting data every month). We will refer to this distribution as the ‘reference’ sampling, as it represents an ideal but not realistic system that will serve as a reference to compare the results of the networks described above. That is, the reference sampling is used here to quantify the best results that could be obtained with our analysis scheme. A typical example of the spatial coverage of the 6 networks and the gridded fields generated with each of them is presented in Fig. 3.

2.3. Mapping algorithm

The mapping technique is basically the same as the one used for the generation of MEDATLAS products (Rixen et al. 2005), EN3 products (Ingleby & Huddleston 2007) and ISHII products (Ishii & Kimoto 2009). The algorithm is based on the Optimal Statistical Interpolation (OI) scheme originally described by Gandin (1963). In OI, the value of the analysis at a given point \( r_j \), \( \psi_a(r_j) \), is obtained by linearly weighting the deviations of observations \( \psi_o(r_i) \) from the values of a first guess or background field at the same observation points \( r_i \), \( \psi_b(r_i) \) plus the value of the background field at the analysis point \( \psi_b(r_j) \):

\[
\psi_a(r_j) = \psi_b(r_j) + \sum_{i=1}^{N} w_i (\psi_o(r_i) - \psi_b(r_i)) = \psi_b(r_j) + \sum_{i=1}^{N} w_i d_i \quad (1a)
\]

where \( N \) is the number of observations and \( w_i \) is the weight assigned to the deviation or anomaly \( d_i \). Eq. (1a) can be written as the scalar product of 2 \( N \)-vectors, \( w \) and \( d \), containing the \( N \) weights and the \( N \) anomalies respectively:

\[
\psi_a(r_j) = \psi_b(r_j) + w^T d \quad (1b)
\]

where \( T \) denotes transposition. Furthermore, Eq. (1b) can be written so as to obtain the analysis at a set of

Fig. 3. Temperature maps corresponding to a single month (September 2099) of (a) the true field, and the difference between the true field and the analysis fields generated from the (b) reference network; (c) target network; (d) realistic network; (e) cruise network; (f) Argo buoys network; and (g) virtual mooring network. Black dots: observations gathered by each sampling during that month. Note: color bar in (b) also applies to panels (c−g).
M points (e.g. nodes of a grid) instead of at a single point. Considering the output values as the components of an M-vector $\psi_a$, it is easy to show that:

$$\psi_a = \psi_b + W^T d$$

(1c)

where $\psi_b$ is the M-vector of background values at the analysis points $r_i$ and $W$ is an $N \times M$ matrix whose column $j$ contains the vector $w$ defined in Eq. (1b).

The weights constituting matrix $W$ are determined under the constraint of minimizing (in a statistical sense) the differences between the analyzed values and the true field. It can be demonstrated that, under that constraint, $W^T$ is given by:

$$W^T = S^T D^{-1}$$

(2)

where the $N \times N$ matrix $D$ is the so called observation covariance matrix ($D = dd^T$) and $S$ is an $N \times M$ matrix containing the covariance between the observed anomalies at the $N$ observation points and the true anomalies at the $M$ analysis points. Assuming that observational errors are uncorrelated with the background field, $D$ can be written as the sum of 2 matrices:

$$D = (B + R)$$

(3)

with the elements of $B$ being the covariance between pairs of true field anomaly values at observation points, and the elements of $R$ being observational error covariances. If observational errors are assumed to be spatially uncorrelated (a common assumption for the networks analyzed in this work), then $R$ becomes a diagonal matrix. Furthermore, if the variance of the true field is assumed to be spatially homogeneous, the covariance elements of all matrices can be divided by the variance of the signal, so that they become correlations in the case of matrices $W$ and $B$. In the case of $R$, assuming that the observation error covariance is also spatially homogeneous, all the elements of the diagonal will be equal to the ratio between the noise variance and the field variance (the so-called noise-to-signal ratio).

Because in practice we lack a proper statistic to compute covariances or correlations, a final assumption is that correlation is homogeneous and isotropic, that is, it only depends on the distance between the correlated points. In that case, the correlation elements of matrices $B$ and $S$ will be determined from a function that is equal to 1 at zero distance and decreases to zero with increasing distance. Namely we will use a Gaussian function defined as:

$$\text{Corr}(\delta) = \exp\left(-\frac{\delta^2}{2C_L^2}\right)$$

(4)

where $\delta$ represents the distance between the 2 points and $C_L$ is the so called correlation length or correlation scale.

A crucial issue is how to determine the parameters involved in the mapping scheme. Here we use a large value for the correlation length (800 km). The reason is that large correlations are suitable to fill unsampled regions at the expenses of losing regional details (which could not be recovered anyway by the sampling networks). Moreover, the climatic signals we are interested in are mostly associated with large scale signals. Sensitivity experiments have been performed changing $C_L$ and the results suggest that 800 km is an optimal value. No temporal correlation has been implemented; instead, the analyses are performed monthly and considering all the observations available within each month as ‘simultaneous’. The noise-to-signal ratio has been scaled by depth and is higher for salinity than for temperature following the same criteria as in MEDATLAS (Brasseur et. al. 1996); the noise-to-signal values not only consider instrumental errors, but also the contribution of small scale structures that cannot be resolved by the sampling.

The mapping algorithm has been implemented by boxes, in order to avoid the influence of observations gathered in geographically near regions but that are separated from the analysis point by land (e.g. in order to avoid the mutual influence between the Tyrrenian Sea and the Adriatic Sea observations). The analysis grid has a $1/5^\circ \times 1/5^\circ$ horizontal resolution and 25 vertical levels (with $\Delta z = 5$ m at upper layers and progressively increasing up to $\Delta z = 500$ m at the deeper layers). The fields of the numerical simulation were interpolated onto the analysis grid to directly compare the results of the mapping with the true field; the comparison will be performed by means of a set of diagnostics described in the next section.

A central issue of the analysis scheme is the background field. The theory of OI states that a proper background field would be the mean of many independent realizations with identical statistical properties to the one we intend to interpolate. For this reason, a fixed climatology of the last decades can be considered a reasonable approach in a stable climate. In an evolving climate, however, the climatology must be updated in order to approach as much as possible the statistical mean of the interpolated fields. The time-evolving monthly climatology used here is calculated using OI from the observations gathered during the 30 yr preceding the time at which each analysis is performed (i.e. to compute the analysis in January we use data gathered in the last
30 January). Some sensitivity tests comparing the results using this approach to the results using a "perfect" background suggest that the impact of the background choice has a small impact provided it is evolving in time.

Fig. 3 shows an example of the mapping process for the temperature at 5 m depth for a single month. Namely this figure shows the true field for September 2099 and the differences between that field and the OI analyses of the pseudo-observations extracted for each of the 5 observational networks. The difference between the true field and the reference field is also plotted, showing values very close to zero over most of the basin, as expected. For the other cases, the results are clearly worse in regions not covered by observations, where the influence of the nearest observations is small, and therefore the values given by the interpolation scheme approach the background. In the northern Adriatic Sea, for instance, there are neither moorings (only 1 in the southern Adriatic) nor Argo observations during the analyzed month, and therefore the analysis essentially reproduces the background field and departs significantly from the true field (see Fig. 3f,g). Conversely, Fig. 3b,d show that the presence of a single coastal mooring in the Northern Adriatic Sea is enough to influence the whole region and approach the true field.

2.4. Evaluation of skills and process analysis

The skills of each observational network will be evaluated on the basis of 3 diagnostics:

1. The point-by-point root mean square difference between the true field and the analyzed field averaged over a certain period:

$$\text{RMSE}(r_j) = \sqrt{\text{var} [\psi_s(r_j) - \psi_t(r_j)]}$$  

(5)

where $\psi_s(r_j)$ denotes the analyzed field, $\psi_t(r_j)$ denotes the true (simulation) values and the overbar denotes a temporal mean. RMSE will be either presented in the form of maps or further averaged over the whole domain to obtain single values.

2. The percentage of explained variance, defined as:

$$\text{VEXP}(r_j) = 100 \left( 1 - \frac{\text{var} [\psi_s(r_j) - \psi_s(r_j)]}{\text{var} [\psi_t(r_j)]} \right)$$  

(6)

The explained variance quantifies the percentage of the true temporal variability accounted for by the analysis at each point. It will also be presented either in the form of maps or averaged over the whole domain.

3. The error in linear trend estimations (TE):

$$\text{TE}(r_j) = \text{abs} \left( \text{trend} [\psi_t(r_j)] - \text{trend} [\psi_s(r_j)] \right)$$  

(7a)

where ‘abs’ denotes the absolute value. The TE is also given for every point in the form of maps or averaged over the whole domain. It will be given either as an absolute error trend (as in Eq. 7a) or in the form of a relative trend error (RTE):

$$\text{RTE}(r_j) = 100 \cdot \{ \text{TE}(r_j)/\text{trend} [\psi_t(r_j)] \}$$  

(7b)

Finally we will also estimate the skills of the observational networks to capture the deep water formation rate in the Western Mediterranean. The reason for doing this is that the deep water formation rate is a crucial parameter for the evolution of the Mediterranean climate (Millot 1999). Also, the results will serve us to evaluate the skills of the analysis to capture a particular process, rather than basin averages. We will consider the Western Mediterranean as a recently formed deep water mass with density values larger than a density threshold, which will be defined as the 1500 m density averaged over the region delimited by the coordinates 42.2°N–43°N, 3.3°E–7.5°E, averaged over 25 yr periods and linearly interpolated between the averaged values (which are assigned to the central year of each period). Due to the salinization of the basin during the 21st century (see Fig. S2 in the Supplement at www.int-res.com/articles/suppl/c063p001_supp.pdf), the threshold density increases in time.

3. RESULTS

3.1. Temperature

Fig. 4 shows the different diagnostics (RMSE, explained variance and TE) applied to the 100 m temperature fields of the 2000−2100 period obtained for each of the 5 observational networks. The RMSE basin average ranges from 0.31°C (target network) to 0.40°C (Argo network). With respect to the spatial structure of RMSE, all networks behave similarly in the Tyrrhenian Sea, where they show the best results (errors of the order of 0.2°C). On the other hand, the 5 networks exhibit larger RMSE (in some cases >0.5°C) in regions poorly sampled or where the small scale not solved by the analysis dominates the variability (e.g. in the Algerian basin). Fig. 4 also shows the standard deviation of the true field, whose basin average (0.81°C) is a reference for the RMSE. The explained variance is >80% for all the observational networks when averaged over the whole basin. Val-
Fig. 4. Root mean square error (RMSE), explained variance (VEXP) and trend error for the 100 m temperature inferred from the cruises, moorings, Argo buoys, realistic and target networks. Fourth column maps: true SD and true trend maps. All maps correspond to the 2000–2100 period.
ues <80% are obtained in the same regions where the RMSE are larger.

With respect to the TEs all networks perform similarly. The errors are <0.1°C 100 yr⁻¹ over most of the basin, and averaged relative errors are <5% for all networks. The worst results are obtained in the regions with the highest true trends, where errors can exceed 0.3°C 100 yr⁻¹ and even 0.6°C 100 yr⁻¹ as for the Alboran Sea.

Fig. 5 complements the results of Fig. 4 in the sense that it shows basin averaged results at all depths. Namely, Fig. 5a shows the profiles of the basin averaged values of RMSE, explained variance and trend errors corresponding to the 2000–2100 period obtained for each observational network, as well as for the reference sampling. The standard deviation of the true field is also plotted in the RMSE panel so as to infer how significant the errors are (for scaling reasons the standard deviation is out of range from surface to 80 m; a maximum value of 4°C is reached at surface). The RMSE increases from 5 to 30 m for all the networks, reaching maximum values between 0.57° and 0.73°C (except for the reference sampling, which shows a maximum value of 0.53°C). At lower levels the RMSE profiles decrease following the standard deviation of the field. For the cruise sampling, however, the RMSE increases from 400 to 600 m (from 0.13° to 0.19°C). Below 600 m the RMSE remains almost constant, slightly increasing near the bottom (>2000 m).

The explained variance shows maximum values (almost 100%) at the surface, decreasing to ~75–85% at ~100 m. From 100 to 400 m, the explained variance increases again for the 5 distributions, reaching values of 80 to 90% at 400 m. From 400 to 600 m the explained variance of the cruise sampling decays abruptly to ~60%, while the other samplings slowly increase down to 2000 m (where they reach 90 to 95%). Near the bottom all profiles decrease again, especially for the Argo buoys, which only reach 2000 m.

TE exhibits a similar behavior for the 5 observational distributions, going from ~0.05° to ~0.1°C 100 yr⁻¹ from the surface down to 30–50 m, and then back to 0.05°C 100 yr⁻¹ at 400 m. Below this point, TEs remain fairly constant except for the cruise sampling, which shows a marked increase from 400 to 1000 m (where it reaches 0.2°C 100 yr⁻¹). Regarding RTEs, they are in all cases <10% except for the cruises sampling (25% at 1500 m) and below 2000 m (>10%, reaching 21% in the case of the Argo sampling).

Fig. 6a,c present the time-series of the temperature 3D basin averages obtained from the 6 networks for the period 1962–2100. The absolute value is shown in Fig. 6a while the differences with respect to the true value are presented in Fig. 6c. The first 3 rows of Table 4 summarize the skills of each observational network regarding the recovery of the 3D basin averaged temperature (except in the case of the trends, which are computed for the 2000–2100 period). The mooring network overestimates the true mean value by about 0.025°C, while the cruise network underestimates the mean value, especially for the last decades, when differences are >0.1°C. The Argo, realistic and target networks show very accurate mean values. Concerning the trends, all the networks are capable of properly capturing the true trend, with the exception of the cruise sampling, due to the progressive underestimation starting by 2020 and accentuating during the second half of the 21st century. The reasons for such behavior will be discussed in Section 4.

### 3.2. Salinity

Fig. 7 shows the RMSE, explained variance and TE maps obtained for the 100 m salinity corresponding to the 2000–2100 period. As in the case with temperature, the 4 sampling networks exhibit similar RMSE basin averages: from 0.13 psu for the target network to 0.17 psu for the Argo network. The basin average of explained variance ranges from 67.8% for the target sampling to 45.8% for the Argo network, and the RTE is >40% for all samplings.

Fig. 5b shows the basin averaged RMSE, explained variance and RTE profiles obtained for the 2000–2100 period; as in the temperature case, the standard deviation profile of the true field is also plotted in the RMSE panel. The main difference with respect to temperature is that for salinity the highest values of the RMSE (between 0.15 and 0.20 psu) are obtained in the upper layers. All samplings display a similar behavior from the surface to the bottom: a smooth decrease down to 150 m followed by a pronounced decrease down to 400 m, where they are all <0.05. From there to the bottom, the RMSE of all samplings remains almost constant except for the cruises sampling, which shows a small increase up to 0.05 psu. In terms of explained variance, the best results of the single-kind networks are shown by the moorings, with values in the order of 60% from surface to ~150 m; obviously the realistic and target networks show better performances (65 and 70% respectively). In contrast, the Argo and cruises networks show values between 40 and 50% from the surface to 150 m. It
is worth noting that the explained variance of the reference sampling is always <80% until 150 m. From there to 400 m there is a marked increase (all samplings reach values >90% at 400 m) followed by rather constant values from 400 m to the bottom except for the cruise sampling, which shows smaller values. The TE remains below 0.05 psu 100 yr⁻¹ from the surface to the bottom except for the cruise sampling, which reaches 0.07 psu 100 yr⁻¹ at 1200 m. The RTE presents strong variations in all cases, reaching values of almost 100% below the surface and decreasing to 20% at intermediate and deep layers.

The time-series of the salinity 3D basin averages obtained from the 6 networks for the period 1962–
Fig. 6. Time-series of the 3D basin averages obtained from the 6 networks (target, cruises, moorings, Argo buoys, realistic and reference) and the true 3D basin averages for temperature (left column) and salinity (right column) for the period 1962−2100. (a,b) Absolute value and (c,d) difference with respect to the true value.

Table 4. Skills of the 5 observational networks to reproduce the 3D basin averaged temperature (T3D) and salinity (S3D) as well as the volume of deep water formation (DWF) in the Western Mediterranean, compared to the numerical simulation (True) and reference sampling. They have been quantified in terms of temporal average and trend, and in terms of the RMSE between the true values and the analysis values. All results have been obtained for the 1962−2100 period, except the trends, which are computed for the last century. The uncertainty of the trend estimate is the SE of the linear regression.
Fig. 7. Root mean square error (RMSE), explained variance (VEXP) and trend error for the 100 m salinity inferred from cruises, moorings, Argo buoys, realistic and target networks for the 2000–2100 period.
2100 and the differences with respect to the true field are shown in Fig. 6b,d; rows 4 to 6 of Table 4 summarize the skills of each observational network (averages, RMSE and trends between the true annual values and the series of Fig. 6b; as with temperature, trend values correspond to the 2000–2100 period). The behavior is very similar to the one obtained for temperature, except that for salinity, the moorings underestimate the true values instead of overestimating them. The cruise network shows the largest departures, which increase with time and reach a 0.04 psu underestimation during the last decade of the 21st century.

When comparing the skills of the networks to reproduce the past and future variability (see Fig. S3 in the Supplement at www.int-res.com/articles/supp/c063p001_supp.pdf showing the profiles of VEXP, RMSE and TEs for 1962–2000) we find that, with the exception of the surface and subsurface temperature, the variance explained for the past decades is smaller than for future decades. The reason for this will be discussed in Section 4.

3.3. Deep water formation rates

Fig. 8 shows the time-series of the Western Mediterranean deep water formation rate for the 1962–2100 period as inferred from the true field, the background field, the 5 observational networks and the reference sampling. Rows 7 to 9 of Table 4 summarize the diagnostics (average, RMSE and SD). Here the mooring network exhibits the best results: bias of 0.08 Sv and RMSE of 0.15 Sv (not shown). The worst results are obtained with the cruises sampling (bias of 0.50 Sv and RMSE of 0.56 Sv; not shown). All networks overestimate the true formation rate. The background field shows an almost constant rate for all the period. In terms of reproducing the true SD, the moorings, realistic and target networks perform identical results (0.18 Sv instead of 0.17 Sv of the true field). Cruises and Argo networks show higher SD values (0.28 and 0.24 Sv, respectively).

4. DISCUSSION

Gridded products generated from 5 different observational networks have been used to infer the capabilities of each network for monitoring climate signals such as the evolution of the mean temperature and salinity or the deep water formation rate in the Mediterranean Sea. A first point to note is that the skills of the different networks in monitoring climate depend on their particular configuration (i.e. spatial and temporal coverage), but also on the characteristics of the observed fields. In particular, the characteristic spatial scale \( L \) of the dominant features of the field is the main factor that determines the ability of the different networks to capture a significant fraction of the field variance (climatic time scales are assumed to be much larger than the time sampling rate of all networks). To illustrate this feature we have decomposed the temperature and salinity anomaly fields in their large \( (L > 800 \text{ km}) \) and short \( (L < 800 \text{ km}) \) scale components using a Gauss-
ian filter. The time series of each component is plotted in Fig. 9 for a point located at 38.8° N, 6.5° E.

Near the surface, the time variability of the temperature field is clearly dominated by the long scale component. The reason is that surface heat fluxes, which are the dominant forcing of the upper layer temperature, are mostly determined by atmospheric patterns (which are large scale features compared to oceanic scales). This means that a few observations are enough to capture the large scale features. Additionally, the marked trend of surface temperature is considered here part of the variability, and the trends are well recovered by all the samplings. Altogether this allows the surface temperature field

![Fig. 9. Temperature (left column) and salinity (right column) monthly time series at a point located at 38.8° N, 6.5° E at 5, 50, 100 and 500 m depth. The signal has been split in its large scale (>800 km; blue line) and small scale (<800 km; red line) signals](image-url)
variance to be well captured by all the networks (almost 100% of VEXP, see Fig. 5a). In terms of interannual variability (illustrated by the RMSE profiles in Fig. 5a) the results are also better than for subsurface scales.

At subsurface (~30–50 m), the variability of the large and small scales are comparable (Fig. 9), differing only in that the first shows a trend. The reason for having a similar interannual variability is that the influence of surface heat fluxes decreases with depth, and the circulation is dominated by mesoscale features (e.g. eddies, filaments or current meandering). At that depth all observational networks have problems, which makes the explained variance decrease by ~10 to 15% and the RMSE increase by ~0.2°C compared to the surface values (see Fig. 5a). One reason is because of the poorer performance of the networks at subsurface compared to surface results is that data voids prevent the capturing of mesoscale features. A second reason is that the long correlation length used to obtain the gridded products prevent the features being reproduced (note that even the performance of the reference sampling is poor at that depth, showing an RMSE of 0.53°C at 30 m depth). Reducing the correlation length scale is not a solution; it would only improve the performance of the reference network, while it would strongly degrade the performance of all the other networks. From 50 m to the bottom, the variance associated with mesoscale features is weaker, and the signal of the basin-wide warming, which is a large scale process, dominates the series (Fig. 9). This explains why all the networks can capture again a large part of the variance without major problems.

For salinity the picture is similar except at the surface, where the relative importance of large and short scales is now similar. This makes the networks' performance from the surface to 100 m poorer than for temperature, ranging from a 65.8% of explained variance for the target network to values of 50% or below for cruises and Argo. Below 100 m, the variability of the salinity field associated with mesoscale structures decreases and the basin-wide salinization, which is a large scale process related to the increase of the freshwater deficit of the basin, increases its relative importance. Thus, the performance of all the networks is improved at those depths and the differences between them are small.

Besides the characteristics of the field, the other factor affecting the performance of each network is obviously its spatial coverage. Thus, the mooring network is characterized by a few, but regularly spaced profiles that are moreover continuous in time, and therefore it will be able to correctly capture any signal with a spatial scale larger than the mooring separation (~300 km) except in areas where there are no moorings (e.g. the Adriatic). Conversely, the Argo network has a large number of profiles, but many of them are clustered (those provided by the same drifter are usually close to each other, see Fig. 3); this can result in large spatial gaps during certain periods which obviously penalize the network performance. The same happens in principle with the cruise network, for which the profiles are often obtained in the framework of dedicated experiments and are therefore concentrated in relatively small areas. Additionally, the number of cruise observations strongly decreases at deep layers (see Table 3). The realistic network has a much larger number of profiles, so that the problem of data voids is partially overcome, and the overall result is improved with respect to the cruises and Argo samplings. Finally, the target network literally doubles the number of observations of the realistic network, which translate in an improvement of the results with respect to present capabilities.

The performance of the networks at deep levels (>500 m) deserves particular attention. In the case of the cruise sampling there is a sudden degradation of the RMSE/explained variance results from 400 to 600 m depth. The explanation of this behavior is directly related to the sudden decrease in the number of observations at those levels (see Table 3). The Argo network also shows a decrease in the number of observations with depth, but the effect is not as drastic as for the cruises network. The reason is that, despite the very similar number of deep observations from Argo buoys and from cruises, the spatio-temporal distribution is more homogeneous for the Argo network. Note that the deepest observations of the Argo floats are obtained at 2000 m depth; below that depth the analysis reflects exclusively the background field.

Concerning the trends, all databases slightly underestimate the true trends because of the presence of background values in data voids. This effect is greater for the cruises and at deeper layers, where data voids are larger and the background plays a major role. This problem is partially overcome using a time-evolving climatology as a background field, which therefore accounts for a large fraction of the trend. The fact that the background values that fill data voids are always ‘delayed’ in time with respect to the analyzed field is the reason why both temperature and salinity are underestimated, particularly for the cruise sampling (see Fig. 6). The temperature
and salinity trends (warming and salinization of the entire basin) show some acceleration with time, making the problem mentioned above more important. This explains why the underestimation of the cruise sampling increases with time.

The recovery of the 3D basin averaged temperature and salinity shown in Fig. 6 is also related to the vertical sampling characteristics. This is the reason why the mooring network presents an overestimation (underestimation) of the temperature (salinity) fields during all the analyzed period: since the deepest instrument of the virtual mooring is set to 2000 m, below that depth the gridded values come from the extrapolation of 2000 m values, which are hence warmer and fresher than true values.

The skills of the 5 samplings to reproduce the deep water formation rate are worse than for the 3D basin averages of temperature and salinity (see the comparison in Table 4). This result is somehow expected, as the deep water formation rate is a regional process that demands a better spatio-temporal coverage to be captured than does the 3D basin average of temperature or salinity. The best results are given by the mooring and target networks. The reason is that both networks have fixed observations near the deep water formation region, which helps to capture the events. In the case of the realistic, cruises and Argo networks, the yearly results strongly depend on the particular spatio-temporal configuration of the networks when the deep water formation events take place, since none of them are fixed nearby the events.

With regards to the differences between the monitoring of past and future climate, the main difference is due to the accuracy of the background field. If the evolving background field were perfect, it would accurately reproduce the warming and salinization of the basin. In that case the anomaly field to be captured by the observational networks would have the same statistical characteristics as it has in the present climate, and the performance of the different networks would not depend on the analyzed period. However, in practice the background can only be estimated using the last decades, so it cannot capture the whole warming and salinization of the basin. Therefore, the anomaly field also includes a part of that signal, which is a large scale signal. In consequence, the characteristic length of the anomaly field is larger for the future climate than for the present, and hence more easily captured by the different networks. In consequence the explained variance is larger for the future than for the present (compare the VEXP profiles of Figs. 5 & S3 in the Supplement). In terms of interannual variability, there are no differences between the past and the future periods (see the RMSE profiles of Figs. 5 & S3 in the Supplement). This is because the time-evolving background properly captures the long term evolution, and the anomalies both in present and future periods have a similar variability. Concerning the TE, the results for the past and future periods are similar for salinity. For temperature the errors are slightly larger when analyzing the period 1962–2000. This is, however, because the sampling period is shorter and the statistical significance of trends is reduced.

5. CONCLUSIONS

The results shown in this paper suggest that present day monitoring systems can capture, at least to a certain degree of accuracy, the evolution of the 3D basin averaged temperature and salinity. Thus, all the tested networks recover a 1962–2100 3D basin averaged temperature of 13.85°C, in agreement with the true field, and a trend of 1.01 ± 0.30°C 100 yr⁻¹ in the worst case, which falls within the statistical uncertainty of the true trend (1.11 ± 0.33°C 100 yr⁻¹).

A similar accuracy is obtained for salinity, with a recovered 1962–2100 average of 38.7 psu and a trend of 0.33 ± 0.10 psu 100 yr⁻¹ in the worst case, which is also in good agreement with the true trend (0.36 ± 0.11 psu 100 yr⁻¹).

The performances are worse and dependent on depth and variable when it comes to 2D fields analysis. Thus, the temperature variability and basin averaged trends are fairly well captured in the upper layer (almost 100% of the monthly field variance is recovered and the RTE is <5%). At intermediate layers the performances are poorer: 80 to 90% of explained variance and a RTE of 5%. At the bottom layer the performance improves again (85 to 95% of explained variance and RTE of 3%). For intermediate and bottom layers the explained variance of the salinity field is >90% while the RTE is <5%. The worst results are obtained for surface salinity, with an explained variance of <80% in all cases and <60% in some cases, and a RTE of ~18%.

The results for deep water formation rate in the Western Mediterranean are poorer than for the 3D averages of temperature and salinity, exhibiting an overestimation of the true values. Finally, regional features are completely lost by all samplings since none of them has an adequate spatial distribution to capture small scale processes.
The realistic network analysed here represents the actual monitoring system, while the target network represents an optimistic future monitoring system where the number of CTD cruises and Argo buoys are doubled with respect to the present, and the planned future mooring network is implemented. This means the best results are obtained with the target network, but the differences with respect to the present observational network are not very large: 0.04°C (in the first 30 m) and 0.02 psu (at 30 to 50 m) for temperature and salinity RMSE, and 11% (at 800 m) and ~8% (at 50 m) for temperature and salinity explained variance. With regards to capturing 3D temperature and salinity averages, the results are identical; and with regards to the deep water formation rate, the results are only slightly better for the target network. The virtual mooring network only improves the results of the present monitoring system in the case of the deep water formation rate, and provided that there is a mooring close to the formation region.

As an overall conclusion, a proper monitoring of Mediterranean climate evolution would imply maintaining and, if possible, improving the existing observation systems, as well as favouring the merging of all the observations gathered by the different networks. The results obtained for the 3 networks analyzed in this work (cruises, Argo buoys and moorings) can be used to design a better monitoring system. Thus, the virtual mooring network illustrates the value of a regular spatio-temporal sampling, particularly when regional processes account for more variance than large scale processes (e.g. salinity at upper layers or deep water formation). The combination of the present monitoring system with the capabilities of capturing regional processes of a mooring network is envisaged as an optimal solution; no doubt better than increasing the number of CTD cruises and Argo buoys.

Acknowledgements. This work has been carried out in the framework of the projects VANIMEDAT-2 (CTM2009-10163-C02-01, funded by the Spanish Marine Science and Technology Program and the E-Plan of the Spanish Government) and ESCENARIOS (funded by the Agencia Estatal de METeoroología, AEMET). G.J. acknowledges a JAE-Doc contract funded by the Spanish Research Council (CSIC) and the European Science Foundation and a post-doctoral grant from the Conselleria d’Educació, Cultura i Universitats of the Balearic Islands Government and the European Social Fund. We thank S. Somot and F. Sevault for kindly providing the outputs of the NEMOMED8 simulation. Finally, we are very grateful to the editor and the reviewers for their supporting and detailed revision of the paper, which has definitely helped to improve the quality of the published manuscript.

LITERATURE CITED


Madec G (2008) NEMO, the ocean engine. Note Pôle Model 27, Inst Pierre-Simon Laplace (IPSL), Paris


Editorial responsibility: Eduardo Zorita, Geesthacht, Germany

Submitted: January 8, 2014; Accepted: November 1, 2014
Proofs received from author(s): February 13, 2015