A Neural Network–Based Observation Operator for Coupled Ocean–Acoustic Variational Data Assimilation

ANDREA STORTO,[a] GIOVANNI DE MAGISTRIS,[a] SILVIA FALCHETTI,[a] AND PAOLO ODDO[a]

[a] *NATO/STO Centre for Maritime Research and Experimentation, La Spezia, Italy*

(Manuscript received 26 September 2020, in final form 31 March 2021)

ABSTRACT: Variational data assimilation requires implementing the tangent-linear and adjoint (TA/AD) version of any operator. This intrinsically hampers the use of complicated observations. Here, we assess a new data-driven approach to assimilate acoustic underwater propagation measurements [transmission loss (TL)] into a regional ocean forecasting system. TL measurements depend on the underlying sound speed fields, mostly temperature, and their inversion would require heavy coding of the TA/AD of an acoustic underwater propagation model. In this study, the nonlinear version of the acoustic model is applied to an ensemble of perturbed oceanic conditions. TL outputs are used to formulate both a statistical linear operator based on canonical correlation analysis (CCA), and a neural network–based (NN) operator. For the latter, two linearization strategies are compared, the best-performing one relying on reverse-mode automatic differentiation. The new observation operator is applied in data assimilation experiments over the Ligurian Sea (Mediterranean Sea), using the observing system simulation experiments (OSSE) methodology to assess the impact of TL observations onto oceanic fields. TL observations are extracted from a nature run with perturbed surface boundary conditions and stochastic ocean physics. Sensitivity analyses indicate that the NN reconstruction of TL is significantly better than CCA. Both CCA and NN are able to improve the upper-ocean skill scores in forecast experiments, with NN outperforming CCA on the average. The use of the NN observation operator is computationally affordable, and its general formulation appears promising for the adjoint-free assimilation of any remote sensing observing network.

SIGNIFICANCE STATEMENT: Deep learning algorithms are now widely spread in a diverse range of fields to help with solving automatic classification and regression problems. Here, we present and assess a strategy aimed at introducing an observation operator based on neural networks in data assimilation. Linearization of such an operator, required by variational schemes, is also discussed and implemented. The methodology is applied to the coupled oceanic–acoustic data assimilation problem, and provides promising results. Our approach may be extended in the future to assimilate any remotely sensed type of observations.

KEYWORDS: Acoustic measurements/effects; Neural networks; Variational analysis

1. Introduction

Assimilating remote sensing data into numerical weather and ocean prediction models is known to be significantly beneficial to the accuracy of the following forecasts (e.g., Eyre et al. 2020; Storto et al. 2013). However, the optimal assimilation of such observations requires advanced techniques to filter out observations with large inaccuracies (quality control) and correct possible systematic errors of the observations (bias correction). Both procedures generally rely on identifying predictors for the gross error occurrence and the bias, respectively. Further to these preprocessing procedures, the assimilation of remote sensing data relies on several simplifications of the observation operator, which is the function projecting the model state onto observation space, required by all assimilation methods. For instance, satellite radiances are usually assimilated into numerical weather prediction models after calculating regression coefficients from full physics line-by-line radiative transfer models (e.g., Saunders et al. 2018); satellite altimetry data are commonly assimilated in oceanographic models assuming local hydrostatic balance (e.g., Storto et al. 2011). Additionally, variational assimilation techniques require the tangent-linear and adjoint versions of the observation operator to ensure quadraticity of the cost function and speed up the minimization, respectively. Thus, all these three components of the assimilation of remote sensing data—quality control, bias correction, and observation operators—contain, to different extent, some empirically formulated features.

With the recent popularity of data-driven procedures for regression and classification problems in complex dynamical systems (machine and deep learning), there is a growing number of available algorithms that can be adapted to specific tasks of the data assimilation problem. Preliminary examples consist in the use of neural networks for bias correcting dust observations in a real analysis system (Jin et al. 2019) or the use of relevance vector machine in the bias correction of sea surface temperature data (Storto and Oddo 2019). The data selection (subsampling) and quality control of Earth observations were also investigated through machine learning algorithms (Lary et al. 2016, 2018). Finally, the coding of complex observation operators may be substituted by data-driven algorithms (e.g., Xue and Forman 2017; Fang and Li 2019; Kwon et al. 2019) in specific applications. All these ideas may help to optimize the assimilation of high-resolution observing networks and greatly

Storto's current affiliation: Institute of Marine Sciences, National Research Council, Rome, Italy.

Corresponding author: Andrea Storto, andrea.storto@cnr.it

DOI: 10.1175/MWR-D-20-0320.1

facilitate the assimilation of new observation types into regional assimilation systems. In this work, we aim at demonstrating the potential advantage of using artificial neural networks in the construction of observation operators and their tangent-linear and adjoint models. Specifically, we test this idea in the framework of coupled oceanic-acoustic data assimilation, where the remotely sensed observations are, in particular, the underwater acoustic propagation measurements.

Oceanic and acoustic measurements may provide complementary information to be exploited in the context of operational analyses and forecasts. The link between oceanic and acoustic variables is provided by the sound propagation in the ocean, which strongly depends on space–time sound speed fields, and thus seawater temperature and, to a lesser extent, salinity. The underwater acoustic propagation is sensitive to oceanic scales that are much smaller than those resolved by most regional models (Castor et al. 2004). Therefore, acoustic observations appear promising in complementing oceanic in situ observations for sampling the ocean meso- and submesoscales. In particular, for high-resolution regional and coastal applications, the accuracy of the predictions is significantly hampered by the scarceness of in situ profiles sampling the subsurface oceans. Underwater acoustic measurements may thus be inverted to provide small-scale corrections (see, e.g., Howe et al. 2019 for an overall discussion).

The assimilation of acoustic data into ocean models can introduce corrections into the temperature, salinity and current fields in order to preserve the acoustic characteristic of the waveguide. Although the estimation of the oceanic environment and the acoustic fields is considered as a coupled data assimilation problem (Robinson et al. 2002; Culver and Camin 2008), there are only few works that attempt to build a truly coupled ocean–acoustic analysis and forecast system.

Ocean acoustic tomography, which was introduced as an approach to ocean observations in the 1980s (Munk and Wunsch 1982; Munk 1986; Spiesberger and Metzger 1992; Munk et al. 1995), is the field in which most of the experience in assimilating acoustic observables to infer ocean parameters (temperature or currents) has occurred. The essential information provided by tomography is the average temperature, either a profile average over range or an average over both range and depth (Dushaw et al. 1993; Cornuelle and Worcester 1996; Dushaw 1999; Sagen et al. 2016; Dushaw 2019). The acoustic travel times, being an integral over space, are indeed considered a robust measure of spatially averaged oceanographic variables (Lewis et al. 2005). The determination of the averaged oceanographic variables is derived from its inverse estimate, which can be computed in different ways, from simple two-dimensional least squares fit (Dushaw and Sagen 2016), to the use of numerical models (e.g., Gaillard 1992; Lebedev et al. 2003; Lewis et al. 2005).

Little work exists that exploit the assimilation of acoustic pressure or transmission loss (TLs), besides the acoustic travel times employed in acoustic tomography. Li et al. (2014) used a variational method to invert the sound speed via acoustic pressure data assimilation. An internal wave model was used for the specific shallow water environment, and the effect of sediment on the acoustic propagation was included in the inversion algorithm. An adjoint of a parabolic propagation model (Hursky et al. 2014) was used to invert acoustic pressure data for sound speed perturbation in shallow water environment and for short ranges (of the order of 2–3 km). Following these works, Ngodock et al. (2017) describe the theoretical framework for the variational assimilation of the acoustic pressure measurements using an adjoint version of the acoustic propagation model RAM (Collins 1989a,b, 1994). The assimilation of TLs was also explored in a twin experiment by Lermusiaux and Chiu (2002) and Lermusiaux (2006) using the error subspace statistical estimation (ESSE).

The present work aims at further exploring the assimilation of TLs to correct the ocean fields (temperature and salinity) using a state of the art data assimilation scheme and an observation operator based on canonical correlations or neural networks. The assimilation of the TLs is here explored using multivariate data assimilation. Observing system simulation experiments (OSSEs; e.g., Halliwell et al. 2017; Dushaw et al. 2016) are used to demonstrate the feasibility of assimilating simulated low-frequency signal (i.e., 75 Hz) TLs in the Ligurian Sea (Mediterranean Sea), extending the previous assessment of the impact of oceanic analysis schemes on the accuracy of underwater sound propagation prediction (Storto et al. 2020). The chosen frequency (75 Hz) is representative of ship noise, and as such the synthetic measurements are conceived as coming from ships of opportunity.

The paper is structured as follows: after this introduction, data and methods are presented in section 2, including the analysis and forecast system, while section 3 presents methods for assimilating transmission loss observations. Selected results from the assimilation of underwater acoustic data are shown in section 4, while section 5 discusses and concludes.

2. Data and methods

In this section, the integrated ocean–acoustic modeling system, including the oceanographic data assimilation scheme and the acoustic scenario, are introduced.

a. The oceanic and acoustic forecasting system

The analysis and forecast system used in this study is the Ligurian Sea modeling suite developed at the Centre for Maritime Research and Experimentation (Storto et al. 2019; Storto and Oddo 2019). It includes the NEMO (v3.6) ocean model (Madec et al. 2017) forced at the lateral boundaries by the Copernicus Marine Environment Monitoring Service (CMEMS) Mediterranean Sea forecasts (Clementi et al. 2017), and at the sea surface by the European Centre for Medium-Range Weather Forecasts (ECMWF) meteorological analyses through the bulk formulas of Large and Yeager (2004). The model has a horizontal resolution of about 1.8 km and 91 vertical depth levels and partial steps. Figure 1 shows the extension of the NEMO model domain, together with the bathymetry interpolated from the General Bathymetric Chart of the Oceans (GEBCO) database (Weatherall et al. 2015). The NEMO configuration includes sea surface relaxation to the SST analyses provided by CMEMS (Buongiorno Nardelli et al. 2013). Analysis increments from the data assimilation system, described in the next

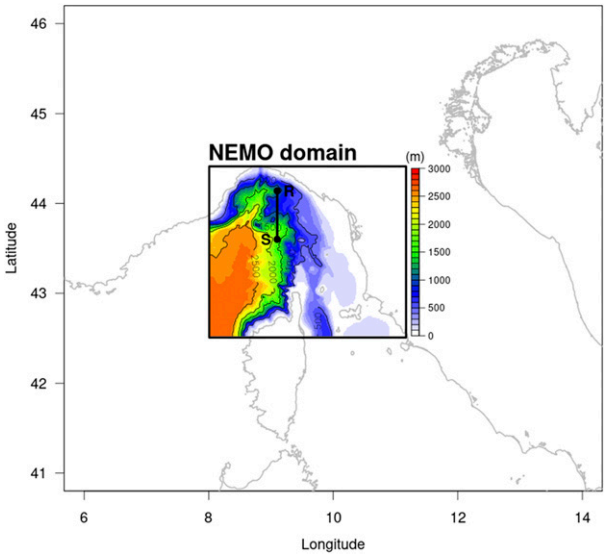

FIG. 1. Acoustic propagation scenario used in this study: the Ligurian Sea NEMO computational domain of the oceanographic model is indicated by the thick box on top of the gray coastlines, together with the associated bathymetry (black contours every 500 m, from GEBCO) and locations of the acoustic source (*S*) and receivers (*R*) in the Ligurian Sea.

section, are applied through the incremental analysis update scheme (Bloom et al. 1996).

Acoustic modeling is performed through the Range-dependent Acoustic Model (RAM) parabolic equation model (Collins 1989a,b), which uses the split-step Padé approximation for the solution of the parabolic equation (Collins 1993). Geoacoustic characteristics at the sea bottom have been setup according to the NOAA Deck41 database (Bershad and Weiss 1976) for the Ligurian Sea, and corresponding to clayey silt parameters, taken from Lurton (2010, p. 116). RAM takes as input the sound speed fields interpolated onto the acoustic propagation transect from the NEMO model outputs. More details about the RAM configuration are available in Storto et al. (2020).

b. *The variational data assimilation scheme*

Data assimilation is based on a three-dimensional variational scheme with first guess at appropriate time (3DVAR/FGAT) scheme. Details of the analysis and forecast system are provided by Storto and Oddo (2019), and are briefly recalled here. In its incremental formulation (Courtier 1997), the 3DVAR cost function reads as follows:

$$J(\mathbf{v}) = \frac{1}{2}\mathbf{v}^{\mathrm{T}}\mathbf{v} + \frac{1}{2}(\mathbf{HVv} - \mathbf{d})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{HVv} - \mathbf{d}), \qquad (1)$$

where $\mathbf{v}$ is the minimization control variable, such that $\delta\mathbf{x} = \mathbf{Vv}$ is the model state increment [equal to the analysis increment $\delta\mathbf{x}^a$ at the minimum of $J(\mathbf{v})$], with $\mathbf{x}$ and $\mathbf{x}^a$ as the state vector (the three-dimensional grid of temperature, salinity and sea surface height) and the analysis, respectively. In Eq. (1), $\mathbf{B} = \mathbf{VV}^{\mathrm{T}}$ is the background-error covariance matrix, and $\mathbf{d}$ is the vector of misfits, which is equal to $\mathbf{d} = \mathbf{y} - H(\mathbf{x}^b)$, with $\mathbf{y}$ the vector of observations and $\mathbf{x}^b$ the background, taken at the same time of the observations; $H(\,)$ is the observation operator that maps the state of the ocean into observation space. Finally, $\mathbf{H}$ is the tangent-linear approximation of the observation operator, linearized around $\mathbf{x}^b$.

The left square root of the background-error covariance matrix $\mathbf{V}$ includes several operators to account for vertical covariances, horizontal correlations, and cross covariances (Storto et al. 2018, 2020). Background-error covariances are formulated as stationary vertical empirical orthogonal functions (EOFs) and calculated from a dataset of anomalies with respect to the seasonal mean. Horizontal correlations are modeled through a first-order recursive filter. Finally, a dynamic height operator is embedded in $\mathbf{V}$ to estimate the balanced component of the sea level increment.

c. *The underwater acoustic propagation scenario*

The acoustic scenario used in this study is an individual transect located in the Ligurian Sea, with sound source and receivers at the extremes of the transect (see Fig. 1). The source is located around the middle of the model domain (43.6°N, 9.1°E), at 10 m of depth. The receivers are displaced 60 km far from the source along the south–north direction. There are 18 receivers located at 44.1°N, 9.1°E with a regular vertical spacing of 10 m of depth, up to 180 m of depth. The configuration aims to mimic a mooring equipped with hydrophones along its chain. The acoustic propagation frequency chosen for this study is 75 Hz. The acoustic propagation at 75-Hz frequency was found to be more sensitive than higher frequencies to the oceanographic data assimilation schemes in previous studies (Storto et al. 2020); namely, we are in the range of the acoustic spectrum where we can expect the largest impact of the acoustic observations in the ocean forecasting system. A 10-m source at 75 Hz represents a typical noise from ships, i.e., conceptually representing ship-of-opportunity underwater acoustic observations.

3. Assimilation of transmission loss data

Noting that modifications of underwater acoustic propagation characteristics have no impact on the physical ocean state, the problem of assimilating acoustic observations in a physical oceanographic systems reduces to the definition of the observation operator $H^{\mathrm{AC}}(\,)$ and its tangent-linear version $\mathbf{H}^{\mathrm{AC}}$, which are able to map the temperature onto transmission loss values, with its adjoint $(\mathbf{H}^{\mathrm{AC}})^{\mathrm{T}}$ operating the backward transformation, which is required within the variational data assimilation system. In general,

$$\mathbf{y}^{\mathrm{TL}} = H^{\mathrm{AC}}(\mathbf{x}) + \boldsymbol{\varepsilon}^{\mathrm{TL}}, \qquad (2)$$

where $\mathbf{y}^{\mathrm{TL}}$ is the vector of acoustic observations and $\boldsymbol{\varepsilon}^{\mathrm{TL}}$ is the associated error, and

$$H^{\mathrm{AC}}(\mathbf{x}) - H^{\mathrm{AC}}(\mathbf{x}^b) \cong \mathbf{H}^{\mathrm{AC}}(\mathbf{x} - \mathbf{x}^b) = \mathbf{H}^{\mathrm{AC}}\delta\mathbf{x}, \qquad (3)$$

where the $\mathbf{H}^{\mathrm{AC}}$ is linearized around the background state $\mathbf{x}^b$. In (2) we have explicitly referred to $\mathbf{y}^{\mathrm{TL}}$ to indicate transmission loss observations. In the remainder of the work, we will

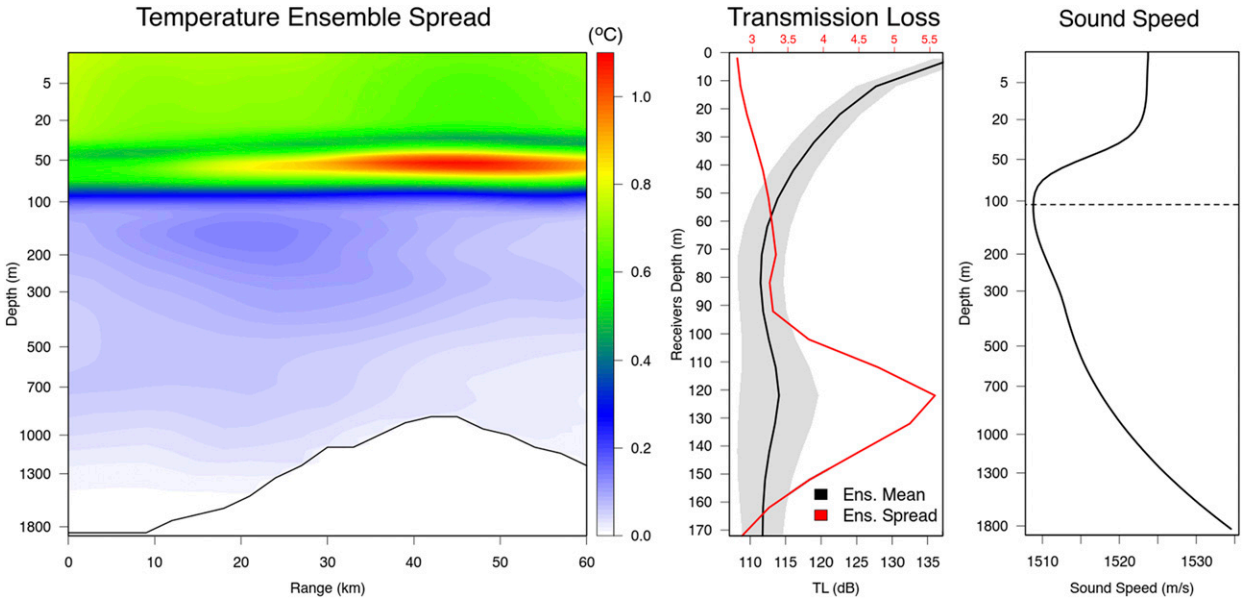

FIG. 2. (left) Ensemble spread for temperature across the transect of Fig. 1 (with the black line indicating the sea bottom), and (center) corresponding mean and ensemble spread for the one-way coupled transmission loss data at the receiver depths (from 2 to 172 m of depth, every 10 m), located 60 km north of the sound source, as computed by the RAM acoustic propagation model. Note that the depth is logarithmic on the *y*-axis of the left panel. (right) Averaged sound speed profile along the cross section of Fig. 1 during the period from 14 Oct to 11 Nov 2017, from the ''truth'' experiment (see section 3d). The dashed line corresponds to the depth with minimum sound speed.

consider that $\mathbf{x}$ is the temperature field; namely, neglecting the weak dependence of the sound speed on the salinity. This also halves the size of the input matrix of the observation operator. This is equivalent to assume that the derivative of H^{AC} with respect to salinity and sea surface height [the other ocean state variables included in $\mathbf{x}$, see Eq. (1)] is equal to zero.

The fully nonlinear observation operator $H(\,)$ is in practice the underwater acoustic propagation model RAM that, given a certain geometry, sound speed fields, frequency, and boundary conditions (i.e., geo-acoustic characterization), provides an estimate of the transmission loss recorded at the receiver.

Regarding the formulation of the tangent-linear operator, several approaches are in general available. One possible choice would be to code the tangent-linear and adjoint observation operator by manually differentiating the RAM model, as it was previously proposed by e.g., Ngodock et al. (2017). Such choice is quite demanding in terms of developing, debugging and maintenance of the code, thus requiring preliminary evidence of the positive impact of acoustic observations in operational oceanographic forecasts. Furthermore, the validity of the tangent-linear approximation of a parabolic equation acoustic model has been shown to be limited to short ranges (e.g., less than 5 km), due to the high nonlinearity of the acoustic propagation and its interactions with surface and bottom boundaries (Hursky et al. 2014). More accurate tangent-linear versions of the parabolic equations might be formulated (Lin 2013) but have never been tested in real-world applications. For these reasons, the analytical formulation of the tangent-linear operator may have several limitations, and formulating a data-driven tangent-linear operator for our acoustic problem is particularly attractive.

There exist already adjoint-free strategies where the tangent-linear and adjoint are approximated by regression statistics or explicit numerical differentiation of the full nonlinear model (e.g., Bishop et al. 2017; Mattern and Edwards 2019). However, these methods are best suited to quasi-linear problems and may not satisfy the along-range nonlinear evolution of the underwater acoustic propagation. Here, we will first formulate a classical linear data-driven observation operator, based on canonical correlation analysis, and then introduce the artificial neural network operator. We will refer to these observation operators as $\mathbf{H}^{\mathrm{AC}}_{\mathrm{cca}}$ and $\mathbf{H}^{\mathrm{AC}}_{\mathrm{nn}}$, respectively, whose formulation is detailed in the next two sections.

a. Observation operator through canonical correlation analysis

Acoustic variables at the receiver location are in general integrated quantities, which depend on all physical conditions from the source along the propagation path, i.e., the two-dimensional (transect) sound speed field. Among the many possible techniques, we have considered as a simple linear formulation the canonical correlation analysis (CCA), whose aim is to find the modes of (co)variability that maximize the cross correlation between different sets of variables. CCA is suitable for problems of medium and high dimensions like ours (e.g., Haddad et al. 2015) and was successfully applied by Jansen et al. (2019) to the problem of skin SST data assimilation. Canonical correlation analysis was first introduced by

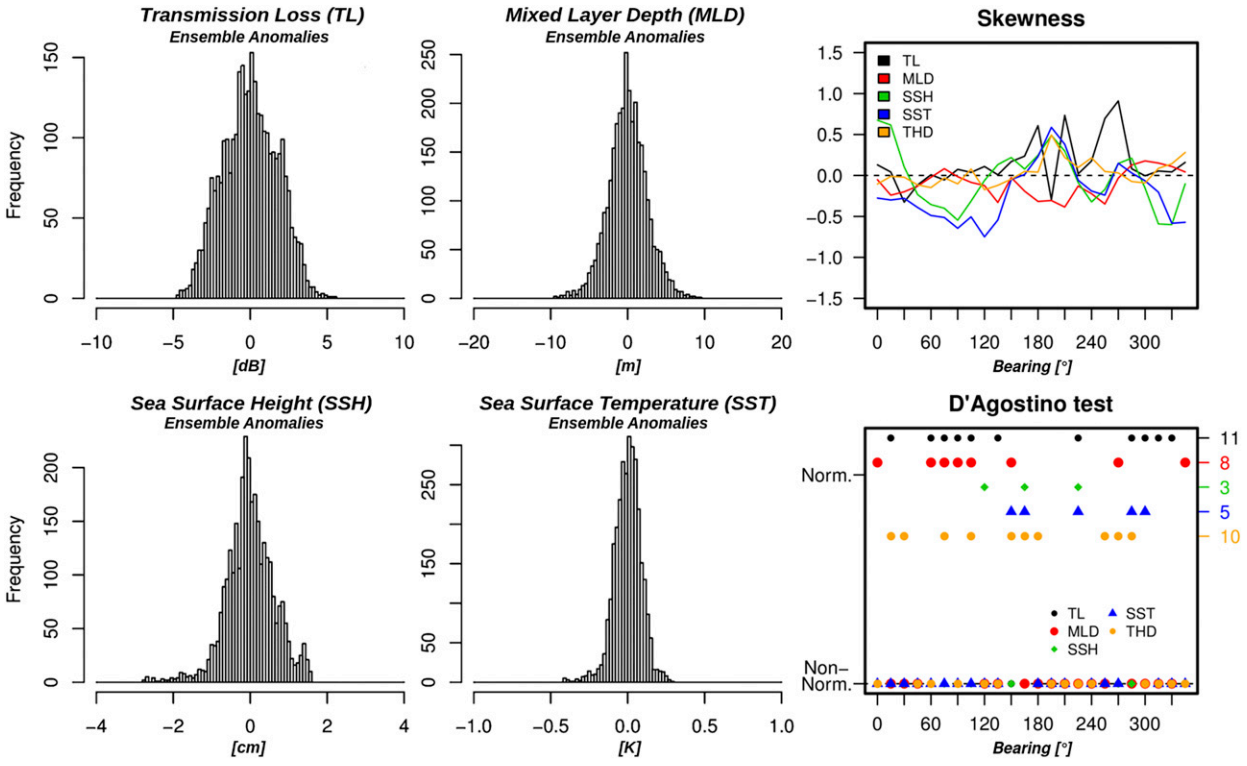

FIG. 3. (left),(center) Histograms of ensemble anomalies for selected acoustic and oceanic variables. (top right) Values of skewness as a function of bearing angle (transect departing from the source in Fig. 1) for the selected acoustic and oceanic variables. (bottom right) Results of the D'Agostino normality tests for the selected acoustic and oceanic variables, as a function of bearing angle. The numbers on the right axis of the bottom right panel show how many cases passed the normality test for the selected variables over the total 24 transects [transmission loss (TL), mixed layer depth (MLD), sea surface height (SSH), sea surface temperature (SST), and thermocline depth (THD)].

Hotelling (1936), and considers two matrices $\mathbf{X}$ and $\mathbf{Y}$ whose columns represent the variables, after normalization by their standard deviation, and the rows contain different realizations of such variables.

The goal of CCA is to find some matrices $\mathbf{A}$ and $\mathbf{B}$ through which $\mathbf{X}$ and $\mathbf{Y}$ are transformed into maximally correlated canonical variables, such that

$$\mathbf{F} = \mathbf{X}'\mathbf{A}; \quad \mathbf{G} = \mathbf{Y}'\mathbf{B}, \tag{4}$$

where $\mathbf{X}'$ and $\mathbf{Y}'$ contain the anomalies of $\mathbf{X}$ and $\mathbf{Y}$; namely, the mean over each column is subtracted to the raw data, and $\mathbf{F}$ and $\mathbf{G}$ are the canonical variables. The matrices $\mathbf{X}'$ and $\mathbf{Y}'$ have sizes of $m \times p$ and $n \times p$, respectively, with m the number of temperature grid points within the acoustic propagation section, n the number of TL observations, and p the number of cases to calculate CCA. Several pairs of canonical variables can be obtained through CCA, any following pair maximizing the remaining correlation between $\mathbf{X}$ and $\mathbf{Y}$, with the maximum number of canonical variable pairs being equal to the minimum dimension of the $\mathbf{X}$ and $\mathbf{Y}$ matrices. The transformation matrices were calculated through the procedure described by Björck and Golub (1973) by means of QR and singular value (SVD) matrix decompositions, which, applied to the dimensions of our problem, was computationally inexpensive.

Once the transformation matrices $\mathbf{A}$ and $\mathbf{B}$ are known, it is possible to calculate $\mathbf{Y}'$ from $\mathbf{X}'$ by means of transformations through the canonical space. In particular, defining the matrix $\mathbf{S} = \mathbf{A}\mathbf{B}^{-1}$, we have that $\mathbf{Y}' \approx \mathbf{X}'\mathbf{S}$. The tangent-linear observation operator for acoustic TL is thus redefined as the matrix $\mathbf{H}_{\text{cca}}^{\text{AC}} \overset{\text{def}}{=} \mathbf{S}$, and the adjoint of such operator as $\mathbf{S}^{\text{T}}$. Doing this, the problem of identifying an oceanic–acoustic observation operator is equivalent to the problem of the canonical correlation analysis given a dataset of oceanic and acoustic training data.

b. *Observation operator through neural networks*

Unlike CCA, which formulates a linear relationship between the input and output data, the NN model nonlinearly relates the temperature fields and the transmission loss data. The NNs are excellent nonlinear function approximations, already adopted and validated in several domains, such as image processing (Krizhevsky et al. 2017), speech recognition (Graves et al. 2013), robotics (Inoue et al. 2017), etc. In atmospheric and oceanic sciences, successful applications of NN have been shown, among others, for the postprocessing of ensemble forecasts (Scheuerer et al. 2020), for substituting

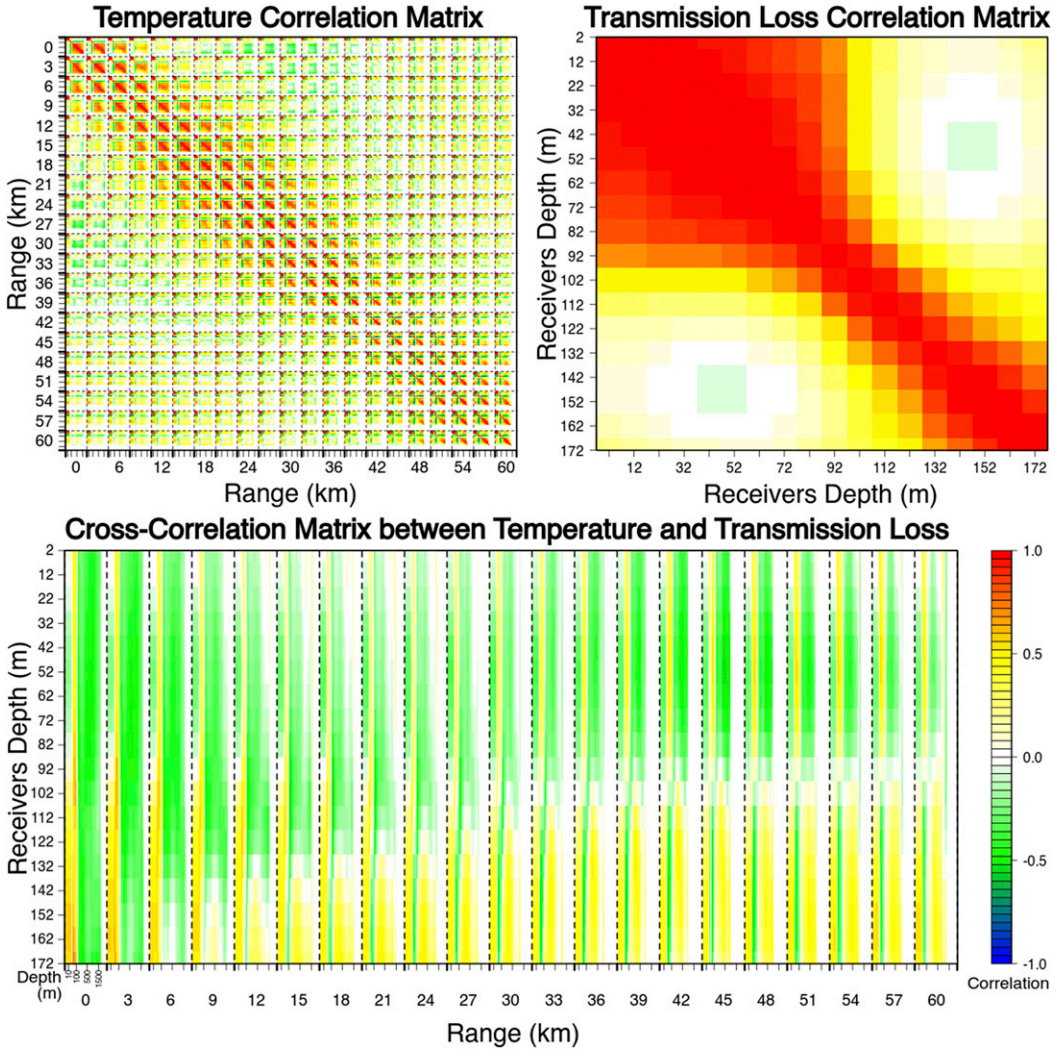

FIG. 4. (top) Correlation matrices calculated from the (left) temperature and (right) transmission loss training data. (bottom) The temperature–transmission loss cross-correlation matrix. For temperature, the 2D cross section is unrolled with depth as fast-varying dimension and range as slow-varying dimension, as explicated in the x axis of the bottom panel.

complex and computationally expensive parameterization schemes in numerical models (Krasnopolsky et al. 2005), or for reconstructing ocean subsurface fields from surface data only (Sammartino et al. 2018; Bao et al. 2019).

The input of the NN is the vector of the temperature fields followed by two fully connected layers of several neurons with a Rectified Linear Unit (ReLU) activation function. This nonlinear activation function helps the network to learn complex patterns in the data. The activation function decides whether a neuron should be activated or not by calculating the summed weighted input from the neuron, and further adding bias to it. The output layer contains n neurons according to the dimension of loss data followed by the linear activation function. The NN is trained using a root mean squared error as loss function; to speed up the learning of the optimal parameters, the input data are normalized in the range of -1 to 1. A number of hyperparameters need to be setup, such as the number of layers and neurons, the batch size and the number of epochs, which together control the complexity of the NN, the dimension and the number of simulation iterations, respectively. The sensitivity of the TL reconstructed by NN to these hyperparameters is discussed in section 4a.

As NN is a nonlinear model, it is not directly usable in variational data assimilation, whose cost function requires linear observation and model functions to preserve its quadraticity and, thus, the uniqueness of the analysis solution. Therefore, a tangent-linear approximation has to be formulated for the NN. A simple way to derive the tangent-linear of the neural network observation operator is through numerical differentiation, provided that the size of our problem (input and output variables) is much smaller than the data assimilation problem itself, and the prediction step of the neural network algorithm

is relatively fast from a computational point of view. To this end, we have tuned the Richardson's extrapolation algorithm (Richardson 1911) implemented for numerical differentiation (Dubeau 2019). It is an iterative method where the step size, i.e., the temperature increment in the evaluation of the transmission loss predicted by the neural network, is decreased sequentially and the derivative of the neural network is extrapolated for the increment tending toward zero. Another possibility is offered by the Tensorflow package (Abadi et al. 2015), which embeds automatic differentiation capability based on reverse-mode automatic differentiation. The package supports automatic differentiation for use e.g., in neural networks backpropagation. Reverse-mode automatic differentiation (or reverse accumulation mode) propagates derivatives starting from a given output back to the input (Bartholomew-Biggs et al. 2000; Baydin et al. 2017). In our application, automatic differentiation is switched on during the neural network model compilation and prediction step, and the tangent-linear version of the observation operator can be found through running the "gradients" function in Tensorflow,[1] linearized around the flow-dependent temperature background fields. Comparison between these two strategies will be shown in section 4. Formally, the tangent-linear observation operator is then given by

$$\mathbf{H}_{\mathrm{nn}}^{\mathrm{AC}} \stackrel{\mathrm{def}}{=} \left.\frac{\partial \mathbf{TL}_{\mathrm{nn}}}{\partial \mathbf{x}}\right|_{\mathbf{x}^b}. \tag{5}$$

Namely, it is the derivative of the transmission loss data given by the neural network model ($\mathbf{TL}_{\mathrm{nn}}$) with respect to the input field ($\mathbf{x}$), in practice the temperature fields, and evaluated for the background temperature ($\mathbf{x}^b$).

c. *The training dataset*

The methodologies described in sections 3a and 3b formulate a relationship between the input physical parameters and the output acoustic TL values. However, in order for the machine learning algorithms to provide an observation operator to be used within the variational data assimilation scheme, a relatively large amount of training data is required. To this end, we have exploited a 1-month 24-member ensemble system of NEMO model simulations, each member of which has been coupled to the RAM model through provision of input sound speed fields. During the period from 14 October to 14 November 2017 (32 days) and extracting data from 6-hourly outputs, a total of 3072 ensemble pairs of input physical and output acoustic variables was used. The evolution of the ensemble spread in an ensemble system with perturbed physics and boundary conditions mimics to some extent the error evolution of the deterministic counterpart system (Storto and Randriamampianina 2010), provided that the perturbations span the real-world uncertainties. Thus, using ensemble anomalies appears an appealing strategy to form the training dataset. For generating the ensemble, a stochastic physics package was developed and implemented in the NEMO ocean general circulation model (Storto and Andriopoulos 2021), similar to the stochastically perturbed parameterization tendencies (SPPT) of Palmer et al. (2009), leading to perturbation fields that are correlated in time and space (Storto et al. 2020).

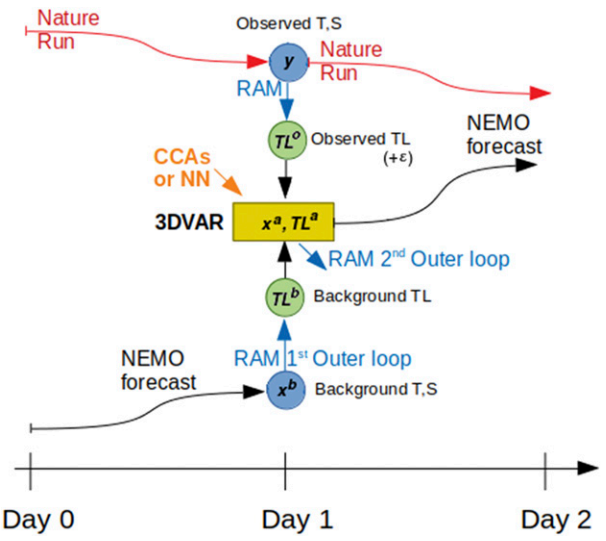

FIG. 5. Sketch of the assimilation experiment, as explained in the text. TL is the transmission loss.

Figure 2 shows the temperature (left panel) and TL ensemble mean and spread at the receivers' location (center panel). The plot highlights the uncertainty of temperature fields in proximity of the thermocline, especially in the northernmost part of the transect. TL data show large uncertainty around 100–120 m of depth, due to the vertical variability of the sound speed minimum (see right panel of Fig. 2), although the TL is highest toward the sea surface.

As a preliminary step to assess the feasibility of coupled oceanic-acoustic data assimilation, we have investigated the statistical characteristics of the physical and acoustic variables. Ensemble-derived probability density functions of acoustic and oceanic parameters revealed that using Gaussian assimilation methods is feasible to a large extent; namely, misfits of TL observations may be approximated to a Gaussian probability density function (pdf), subject to outlier rejection. Figure 3 provides a summary of these diagnostics. In particular, we compared the pdf, skewness and Gaussianity through the D'Agostino test (D'Agostino 1970) of several physical variables and the acoustic transmission loss coming from the ensemble dataset previously introduced, along several transects departing from the sound source and identified by the bearing angle (every 15°). The left panels of Fig. 3 show the distribution of the ensemble anomalies, in order to visualize which distribution shape each parameter has. The skewness of the distributions is reported in the top-right panel of Fig. 3, for each section (bearing angle), with zero corresponding to the perfectly symmetric pdf case. Significant skewness of transmission loss emerges only for 4 transects, while that of e.g., SST occurs in more cases (7 transects). Based on the D'Agostino normality test (bottom-right panel of Fig. 3), transmission loss data are found Gaussian in 11 transects out

[1] https://www.tensorflow.org/api_docs/python/tf/gradients.

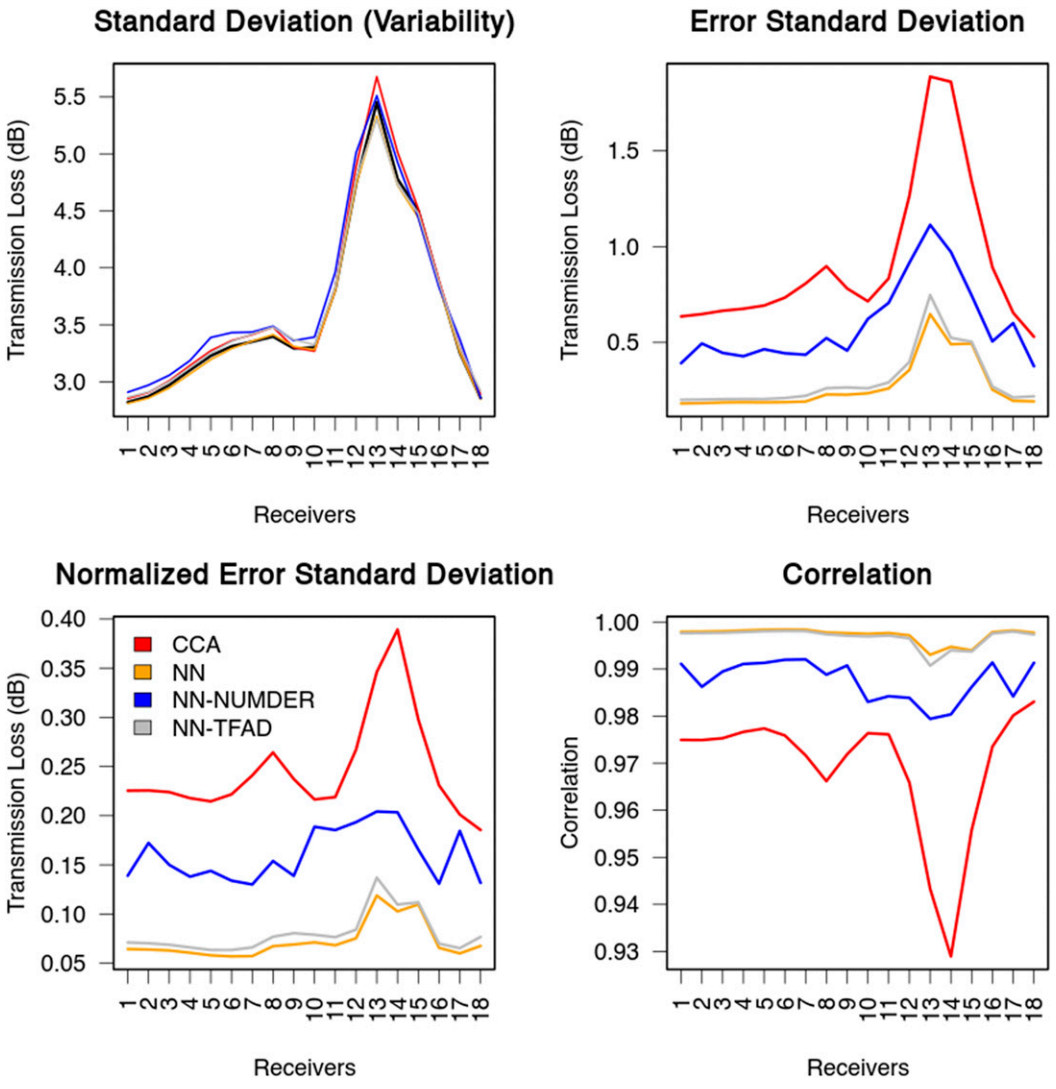

FIG. 6. Variability (standard deviation), root-mean-square error (RMSE), normalized RMSE, and correlation of the reconstructed transmission loss data vs transmission loss validating data, as a function of the 18 hydrophones (output variables). The normalized RMSE is defined as the RMSE divided by the standard deviation (i.e., natural variability) of the validating data. The methods used are canonical correlation analysis (CCA), neural network (NN), neural network linearized through numerical differentiation using Richardson's extrapolation (NN-NUMDER), and neural network linearized through the Tensorflow reverse-mode automatic differentiation.

of 24, against sea surface height (only 3 transects out of 24) and sea surface temperature (5 out of 24). These results indicate that the occasional non-Gaussianity of TL data is not a limiting factor, at least considering other physical variables for which the Gaussian assumption is usually assumed to hold within variational data assimilation systems.

Figure 4 shows the correlation and cross-correlation matrices of and between the physical and acoustic parameters. For temperature, the axes report for each range the model levels, as indicated in the bottom panel. The temperature correlation matrix shows in general large positive correlations between pairs of temperature either above or below the mixed layer. Below the mixed layer, along-range correlations hold within about 10 km of horizontal distance (in forward or backward directions). Beyond these distances, the correlations generally drop. Transmission loss data show on the contrary very high correlation among data within about 100 m of receiver depth, while, below this depth, the data appear generally uncorrelated between pairs located at more than 10 m of depth. The cross-correlation matrix shows significant correlations especially within the first 10 km of range, with transmission loss data from deep receivers significantly positively correlated with upper-ocean temperature, and negatively with temperature below the mixed layer. For higher values of range, the correlation structure becomes more complex. In general, there exist significant cross

TABLE 1. Depth-averaged normalized RMSE for different CCA and NN experiments. For the CCA experiments, the number identifies how many canonical correlations were used. For the NN, different setups of neurons, layers and hyper-parameters were used, and we report here selected configurations: the baseline is NN1 (3 layers, 64 neurons, 5000 epochs, and batch size equal to 128), then NN2 (as NN1, but with 128 neurons), NN3 (as NN1, but with 32 neurons), NN4 (as NN1, but with 4 layers), NN5 (as NN1, but with 1028 layers), NN6 (as NN1, but with 2000 epochs), NN7 (as NN1, but with batch size equal to 256), NN8 (as NN1, but with 256 neurons), NN9 (as NN1, but with 128 neurons and 4 layers), and NN10 (as NN1, but with 256 layers and 4 layers).

	2 CCAs	6 CCAs	10 CCAs	14 CCAs	18 CCAs
RMSE	0.888	0.309	0.246	0.240	0.240
	NN1	NN2	NN3	NN4	NN5
RMSE	0.072	0.067	0.088	0.070	0.070
	NN6	NN7	NN8	NN9	NN10
RMSE	0.075	0.075	0.068	0.069	0.074

correlations, whose shape depends on the position along the propagation path. Note that these cross-correlation features are involved directly in the formulation of the CCA transformation matrix but not in the NN model.

d. Experimental configuration

The assimilation experiments assessed in the next section cover the 29-day period from 14 October to 11 November 2017. The assimilation time-window is daily, and each day a 1-day ocean forecast is performed to bring forward in time the ocean state. The sketch of the assimilation scheme is reported in Fig. 5. The methodology used to test the assimilation of acoustic data is based on observing system simulation experiments (OSSE, see e.g., Atlas 1997). A NEMO nature run—with perturbed boundary conditions and a stochastic physics scheme on top of the nominal model configuration—was previously run. From the nature run's temperature and salinity fields, the RAM model is run to provide the "true transmission loss" data from which the synthetic observations are extracted. A Gaussian random error is added to such observations, with vertically uniform standard deviation equal to 1.0 dB. Within the experiments, the forecasted (background) temperature and salinity fields ($\mathbf{x}^b$) provide the sound speed conditions for RAM to obtain the background TL ($\mathbf{TL}^b$). Observed minus background TL (misfits) are used within the 3DVAR scheme together with the machine learning algorithms and all other ancillary information (observation and background error covariances, grid geometry) to evaluate the analyzed temperature and salinity (and analyzed TL, $\mathbf{TL}^a$). From such corrected sound speed state, the second RAM simulation is run (second outer loop) for the sake of validation against the nature run, while from the analyzed temperature and salinity state, the NEMO forecast is run to reach the following day and provide the background state for the subsequent analysis step, and so on for the entire experimental period.

On top of the variational minimization, a simple background quality check is performed, where TL misfits are rejected if their square value exceeds 9 times the sum of the background and error variances. This is done in order to exclude observations suspected of gross errors and avoid too large misfits that will translate in unrealistically large increments (see e.g., Storto 2016), provided that the tangent-linear observation operators are designed to capture small covariations of physical and acoustic fields. Note that using 9 as variance threshold is equivalent to exclude non-Gaussian observations at 99% confidence level. Further to the two assimilation experiment (CCA and NN hereafter, with the observation operator based on canonical correlations and neural networks, respectively), we run also a control experiment (Ctrl) with no data assimilation. An experiment where the NN is linearized around the Ctrl fields (for sake of assessing the impact of the linearization) will also be introduced in section 4d, and named NN-C. In all data assimilation experiments, the TL observational errors were rescaled (50%) from the TL ensemble spread (Fig. 2, middle panel), assuming that the variability of the TL simulations provides a realistic representation of the acoustic observation uncertainty and characterizes the different uncertainties between the depths where the hydrophones are located.

4. Results

In this section, we first assess the offline transmission loss reconstruction accuracy performed through the CCA and NN methods, and then we evaluate the impact of the assimilation of acoustic observations in physical ocean data assimilation experiments, on both physical and acoustic variables.

a. Reconstruction of transmission loss data

The ability of data-driven algorithms to reconstruct the transmission loss fields from input temperature has been first assessed with the ensemble anomaly data (training data, section 3c). We calculated a set of canonical correlations to form the transformation matrix (CCA) or train the neural network model (NN) from a fraction of this dataset. The correlation values associated with the first 18 canonical variables (CCA method) are all significant and range between 0.99 and 0.75 (not shown). The use of 18 canonical variables is due to the minimum dimension between the input and output variables, here being 18; namely, the number of receivers that the acoustic mooring is equipped with, which is in turn the maximum number of possible canonical correlations. On the contrary, the adoption of NN is more complex, as it requires sensitivity tests to identify the optimal choice of several hyperparameters (number of layers/neurons, number of epochs, batch size, activation function, etc.), which drive the ability of NN to reconstruct TL data. Such tests are presented later at the end of this section.

For the purpose of validating the data-driven algorithms, we have followed the approach of randomly selecting 80% of the 3072 ensemble anomalies for calculating CCAs or NN

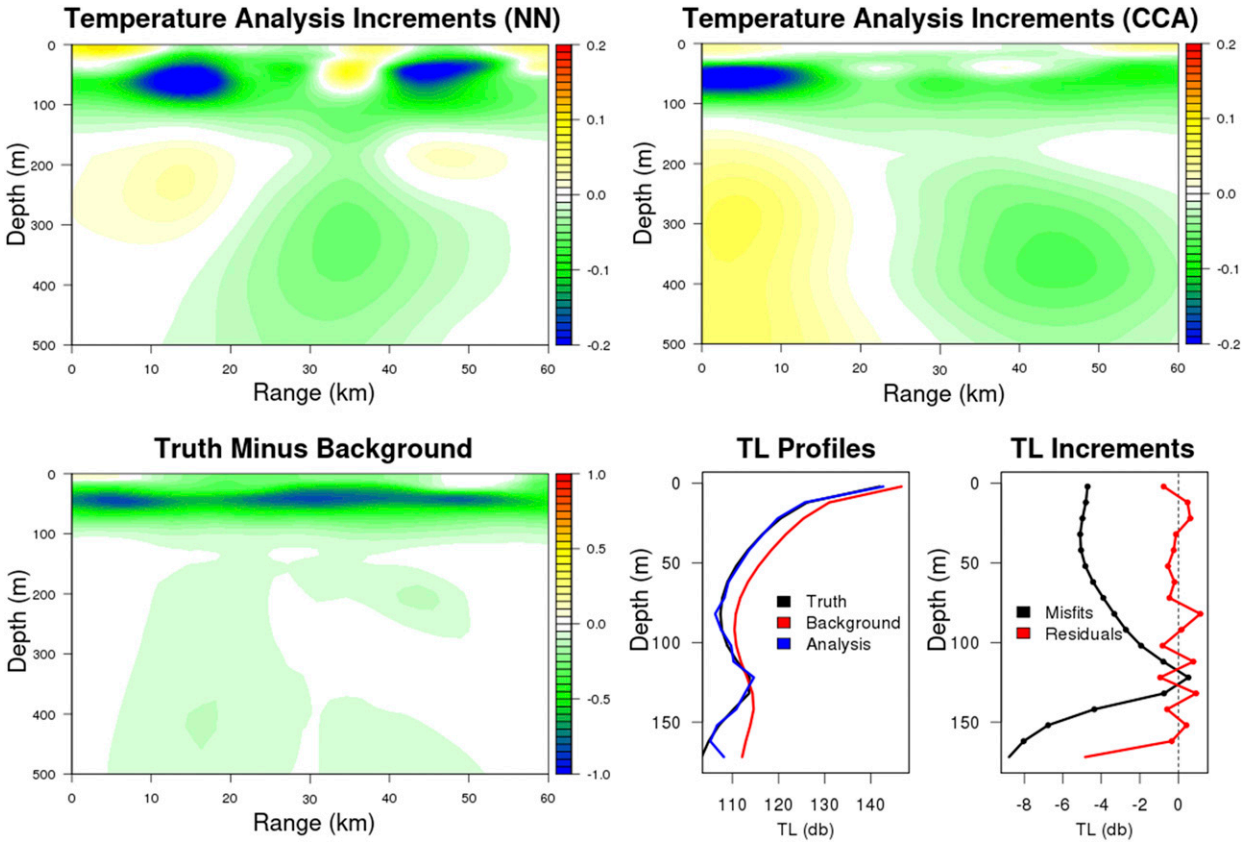

FIG. 7. Example of NN-based inversion of transmission loss data through single-profile data assimilation experiment. (top) Analysis increments of temperature along the cross section corresponding to the propagation path in the top 500 m of depth for the (left) NN and (right) CCA observation operators. (bottom left) Difference between the nature run (truth) and the background. (bottom right) Transmission loss profiles and observation misfits (observation minus background) and residuals (observation minus analysis). Black and red dots correspond to the transmission loss observations actually assimilated; namely, the observation at around 175 m has been rejected by the assimilation system based on the background quality check. The single-profile assimilation experiment refers to 28 Oct 2017.

(training data) and kept 20% of the ensemble anomalies (test data) to validate the method. Results are summarized in Fig. 6 in terms of output data variability (standard deviation), root-mean-square error (RMSE), normalized RMSE (i.e., RMSE divided by the verifying data standard deviation) and correlation coefficients, all as function of the acoustic receivers. All the methods show variability of the output acoustic data close to that of the verifying test data (black line). The accuracy of the CCA reconstruction is below 1 dB for most receivers except between 120 and 160 m of depth (receivers 11 to 16), which correspond to values of normalized RMSE occasionally exceeding 30%. The correlation is significant and high (greater than 0.9) for all depths and greater than 0.96 for most depths except those between 120 and 160 m. Note that, by construction, CCA-based reconstructions maximize the correlation of the reconstructed data rather than their absolute mean or variability value.

The NN-based reconstruction accuracy (orange lines) is remarkably higher than that of CCA at all receivers, with values of normalized RMSE below 10% and correlation values all greater than 0.99. Such skill scores suggest the superiority of the NN method with respect to the CCA. For our application, is also crucial to validate the linearization strategy that is used within the variational data assimilation scheme. We compare in particular the two methods for formulating the tangent-linear approximation introduced in section 3; namely, the numerical derivation based on the Richardson's extrapolation (NN-NUMDER) and that making use of the Tensorflow reverse mode automatic differentiation (NN-TFAD). It turns out that the automatic differentiation strategy leads to transmission loss accuracy approximately as high as the fully nonlinear NN, with slight degradation on normalized RMSE not exceeding 1% and values of the correlation coefficient unchanged. Conversely, numerical differentiation degrades the skill scores, leading to normalized RMSE and correlation coefficients for NN-NUMDER generally at intermediate values between NN and CCA. Numerical differentiation is then found to degrade the accuracy of the reconstruction. For this reason, in the remainder of the paper, we only consider the implementation of NN with tangent-linear model provided by the reverse mode automatic differentiation.

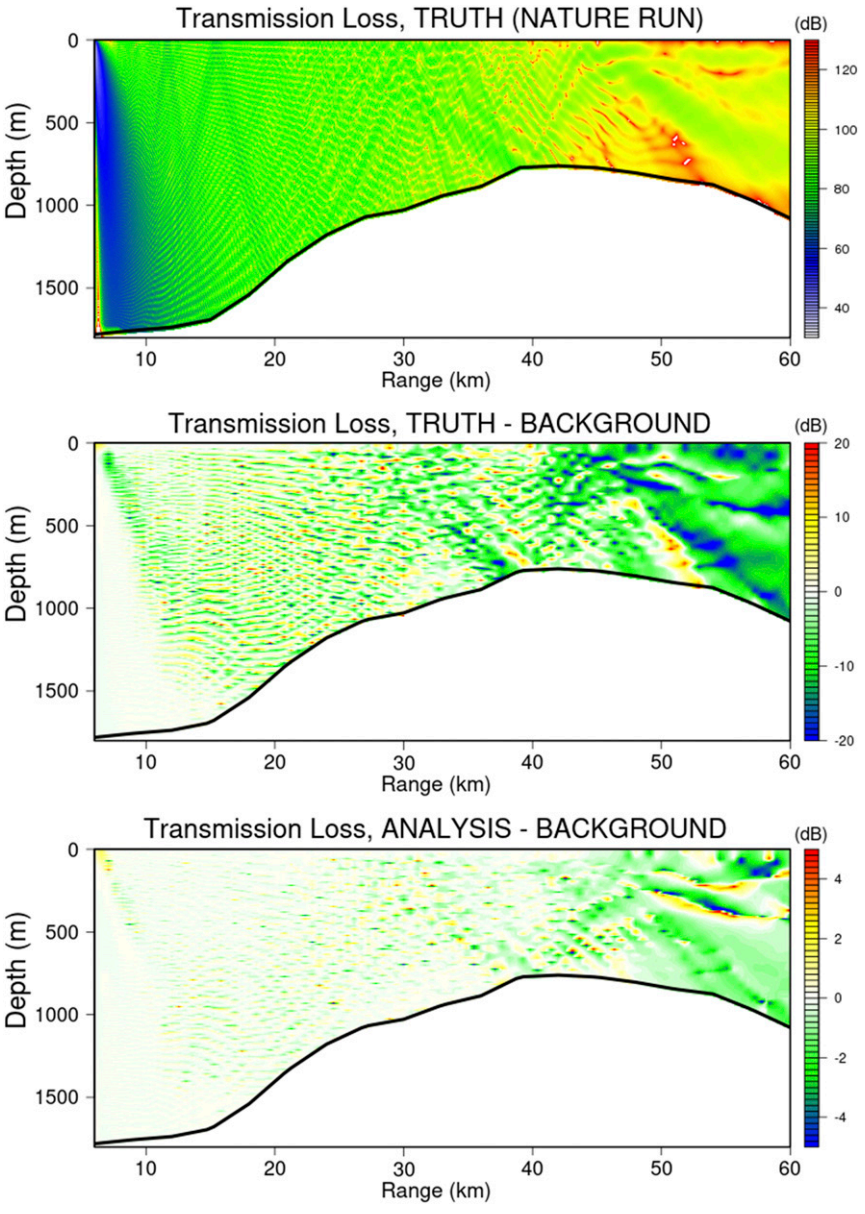

FIG. 8. Transmission loss (dB) along the transect, for the case shown in Fig. 7. (top) From the nature run (truth). (middle) Difference between the truth and the background. (bottom) Difference between the analysis (NN case) and the background. Note that the color bars are different.

While the setup presented in Fig. 6 shows the best configuration for each method (CCA and NN), we have also performed sensitivity experiments, summarized in Table 1. Shortly, using the largest number of CCA provides the best accuracy among the CCA-based experiments, although no significant difference is found for a number of CCA greater than 14. Accuracy of NN within the tested configurations ranges between 6.7% and 8.8%, depending on the number of layers, neurons and hyper-parameters of the neural network. The best-scoring reconstruction considers 3 layers, 128 neurons, 5000 epochs, and batch size equal to 128, and therefore is the one used in the data assimilation experiments. We found in particular that reducing the number of neurons (e.g., to 32) or the number of epochs (e.g., to 2000), deteriorated the skill scores, suggesting in turn that this medium-sized problem requires for the NN a relatively large number of degrees of freedom and iterations to converge toward the optimal model configuration.

The different accuracy of the TL reconstructions using either CCA or NN suggests that their use in real assimilation experiments may require different specification of the representation errors of the TL observations. Indeed, the

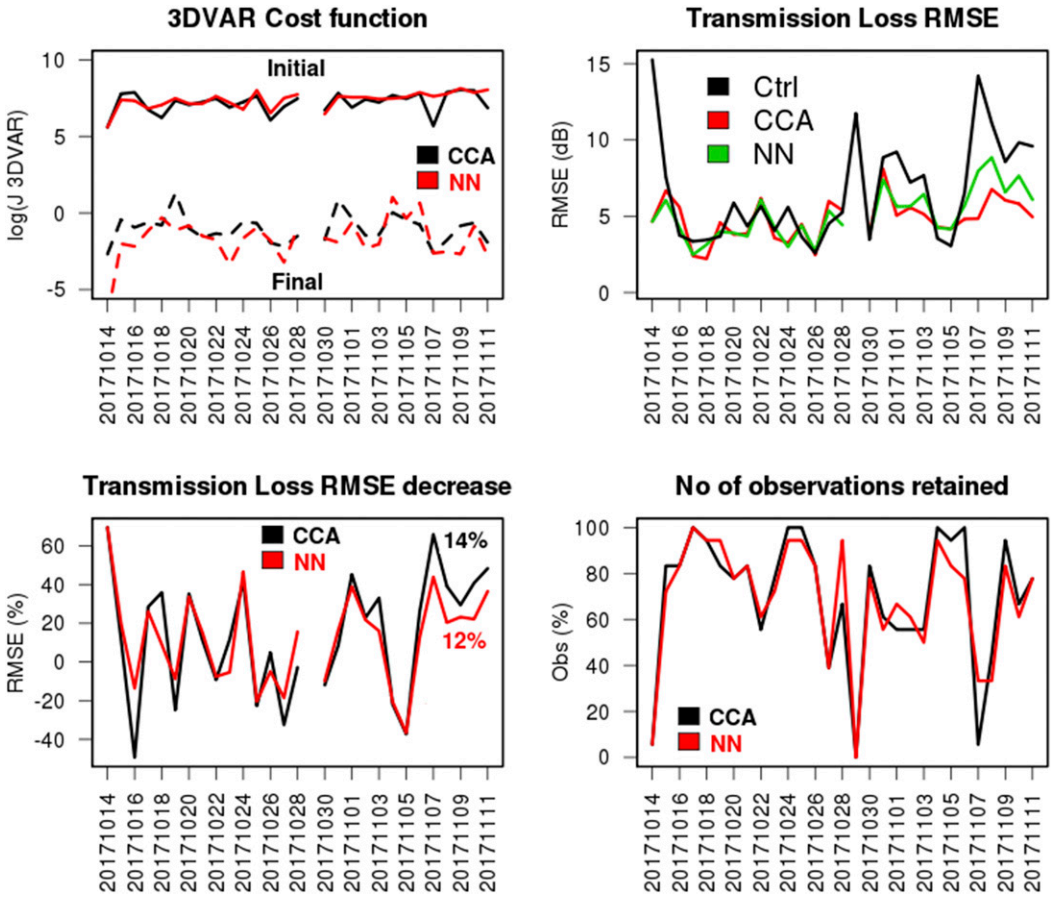

FIG. 9. Assimilation diagnostics in observation space. (top left) Initial and final values of the 3DVAR cost function as a function of time, in logarithmic scale. (top right) RMSE of observations minus background for the three experiments indicated in the legend. (bottom left) Transmission loss RMSE decrease with respect to the Ctrl run misfits; the percentage values represent the time-average RMSE decrease with respect to the Ctrl (in black for CCA and in red for NN). (bottom right) Percentage of retained transmission loss observations as a function of time.

representation error is known to include any possible inaccuracy or approximation implicit in the observation operator (see, e.g., Liu and Rabier 2002; Janjić et al. 2018). Preliminary sensitivity studies with or without the inclusion of the additional representation error (not reported here) have shown, however, that this has negligible impact or no impact on the CCA-based and NN-based assimilation experiments, respectively. For sake of simplicity, the assimilation experiments assessed in the next sections are performed without the representation error; namely, with the error definition as given in section 3d.

b. *Inversion of transmission loss data*

Inversion of temperature data from synthetic TL observations is exemplified in this section through a single-profile assimilation experiment. To visualize the output of the physical data assimilation, we report in Fig. 7 a cross section of temperature analysis increments (top panels) coming from the assimilation of a profile of TL data through the neural network observation operator (top-left panel) or the CCA observation operator (top-right panel). In particular, among the 18 TL observations at different depths, 17 were retained by the assimilation system, while 1 (relative to the receiver at around 175 m of depth) was rejected based on the background quality check. The bottom-left panel reports the difference between the nature run (truth) and the background, while the bottom-right panels show the profiles of TL (background, analysis and truth) and the misfit and residual profiles, in the case of the NN observation operator experiment (the CCA experiment providing very similar profiles, not shown). It turns out that the variational minimization with the observational errors discussed earlier leads to very small values for the residuals (observation minus analysis), compared to the values of the misfits that range between −8 to 0 dB. In particular, the profile of misfits exhibits negative values, i.e., the truth showing smaller TL than the background, corresponding to the consistently colder section in the truth, particularly around 50 m of depth, that slows down the underwater acoustic propagation. Nonnegligible colder state for the nature run occurs also below 200 m of depth, between approximately

5 and 25 km of range. Note that for the nonassimilated receiver, vertical correlations implied by the assimilation system lead to nonnegligible increments, although smaller than those at the assimilated receivers.

The resulting temperature increments are mostly located in the top 100 m with generally negative patterns, which, however, differ between the CCA and the NN case. Both exhibit negative increments around 50 m of depth, but with peaks that are located at different ranges. Nonnegligible increments are also present below 100 m of depth, with NN reproducing to some extent the truth minus background patterns, and CCA providing some positive increments at the beginning of the cross section, and displacing the negative ones toward the end of the transect. In the corresponding TL transect (Fig. 8), the NN analysis increments resemble the difference between the truth and the background, although with smaller amplitude, within around the last 15 km of the section, indicating that the assimilation provides qualitative agreement with the innovations also in the space of the transmission loss. The way the transmission loss misfits translate onto temperature increments is shaped by both the adjoint of the observation operator, and the temperature background-error covariances that constrain the spatial spread of the increments and have crucial impact on the inversion results.

c. *Assimilation diagnostics*

Assessing the impact of the TL synthetic observations in the coupled forecasts can be done through several diagnostics. We start the evaluation exercise looking at assimilation output diagnostics, which are summarized in Fig. 9. The cost function and its decrease though the minimization appears rather stable along time. On the average, the final 3DVAR cost function is about 3000 (4500) times smaller than the initial cost function for CCA (NN). The two experiments provide close diagnostics, NN slightly providing a more pronounced cost function decreases. The root-mean-square error in the transmission loss space (dB) is shown in the top-right panel as time-dependent receiver-averaged values. The panel reports the RMSE calculated for the Ctrl experiment (black, as RMSE after the first outer loop of the data assimilation step) and after rerunning the RAM model using the corrected temperature and salinity (background plus analysis increments) as basis for the sound speed fields (red and green for CCA and NN, respectively). This latter diagnostics quantifies the impact of correcting physical variables with acoustic observations on the acoustic predictions themselves. There is indeed a positive impact, which is comparable between the two experiments (6.7 dB of RMSE in the Ctrl against 4.8 and 5.0 in CCA and NN, respectively). In particular, the bottom-left panel shows that on the average the physical ocean analysis reduces by 14% (CCA) and 12% (NN) the transmission loss error with respect to the Ctrl experiment, after rerunning the RAM underwater acoustic propagation model. Indeed, we obtain a positive impact that increases with time. It is likely that for longer experimental period, the system can better adjust to the transmission loss observations and further gain in accuracy. Finally, the bottom-right panel shows

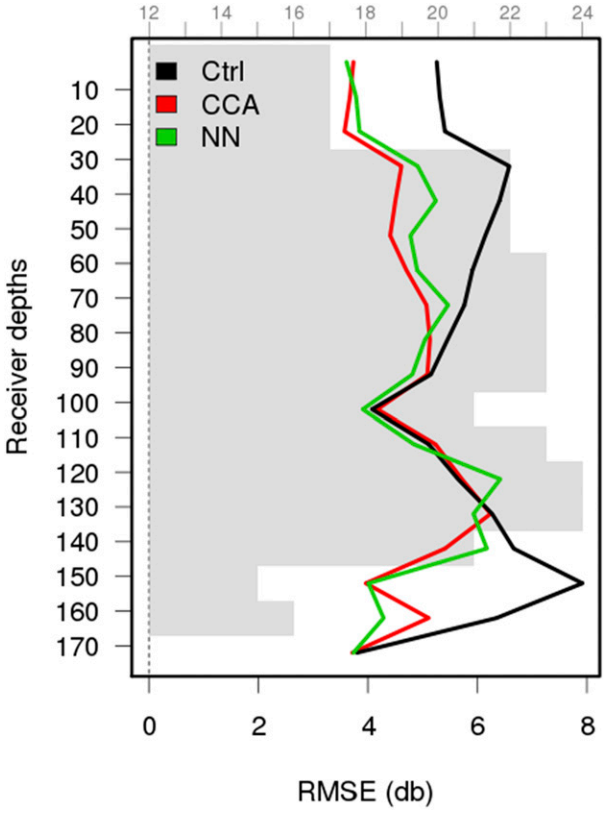

FIG. 10. Transmission loss RMSE as a function of receiver depth for the period 1–11 Nov 2017, for the three experiments presented in the text. Gray shades correspond to the number of observations assimilated for each depth (top axis).

the percentage of retained TL observations. After an initial shock (large misfits leading to numerous rejections, also visible in large RMSE values at the beginning of the experimental period, section 4d), the system adjusts to values that are generally larger than 50% (except two individual events). On average, the percentage of retention of TL observations is equal to about 70% in both experiments.

To further see the benefits of TL data assimilation on the acoustic propagation predictions, Fig. 10 reproduces the time-averaged RMSE skill scores as a function of the receiver depth. We found a large improvement of the corrected transmission loss compared to the uncorrected data within the top 100 m of depth and between 140 and 170 m of depth. Between 100 and 130 m of depth, there is no impact, and this is likely linked with the depth of the minimum sound speed located at that depth (see the right panel of Fig. 2), meaning that the middle depth of the sound channel is not shifted during the assimilation procedure. The decrease of RMSE between the Ctrl and either assimilation experiment indicates that the forecasting system is able to retain the information of the acoustic observations till the next assimilation cycle (1-day forecasts), implying that the temperature corrections coming from acoustic observations improve the acoustic propagation predictability in short-range forecasts.

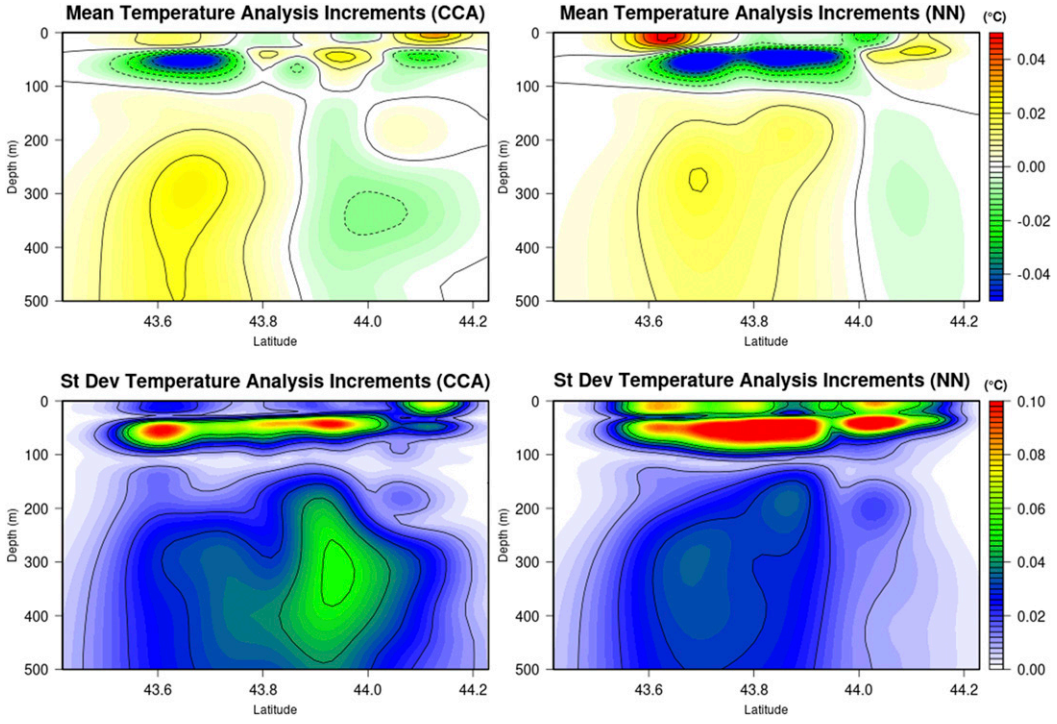

FIG. 11. Mean and standard deviation of temperature analysis increments during the experimental period from 14 Oct to 11 Nov 2017, along the source–receiver propagation path of Fig. 1, for the two experiments presented in the text.

Furthermore, while the control experiment exhibits a RMSE profile increasing with depth and peaking at around 150 m of depth, the RMSE from the assimilation experiments shows a more homogeneous behavior along the vertical; namely, the assimilation leads to a rather uniform vertical accuracy, which is a desirable feature for real-world deployments.

Mean and standard deviation of temperature analysis increments during the entire experimental period are reported in Fig. 11, for both experiments CCA and NN. Such diagnostics sketch the expected impact of the assimilation of TL observations on the mean state and variability of the temperature fields, respectively. Averaged increments are in general low (smaller than 0.1°C in absolute values). While CCA induces a notable average cooling concentrated in the first 20 km and at depths between 50 and 100 m, NN provides extended cooling up to about 40 km of range. At the sea surface there is a qualitatively similar impact, with warming at the beginning and end of the transect, although location and amplitude importantly differ between the two experiments. Below 100 m of depth, the mean corrections are also in agreement (warming and then cooling along the cross section), with NN having the transition shifted toward north. More importantly, the analysis increment standard deviations indicate that corrections are concentrated in the layer 50–100 m, and are larger in NN than CCA, with CCA exhibiting smaller impact than NN also in the top 50 m. Below 100 m of depth, corrections lead to reduced variability, peaking to 0.06°C only in CCA, while NN shows very limited impact (less than 0.04°C).

d. *Impact of acoustic observation on temperature and mixed layer depth*

The way analysis increments impact the ocean state is shown in Fig. 12. Difference of temperature for the two experiments are shown with respect to the Ctrl experiment, which provides a better visualization of the impact of the two schemes compared to the differences with respect to the nature run. In both experiments, the cooling between 50 and 100 m of depth in the analyses is dynamically shifted northward. However, NN exhibits larger cooling (up to −0.5°C) than CCA. Consequently, the mixed layer depth in that region (solid red line) rises more pronouncedly in NN than CCA with respect to the mixed layer depth in the Ctrl experiment (dashed red line), and becomes closer to the mixed layer depth from the nature run (dashed black line). Indeed, the use of NN is able to better thinning the mixed layer compared to CCA.

The main validation exercise is performed for the physical variables, validated against the nature run from which TL synthetic observations were extracted, in order to assess the effectiveness of the coupled assimilation scheme. Temperature RMSE profiles (and their improvement with respect to Ctrl) are shown in Fig. 13. Further to the CCA and NN experiments, also the RMSE profile of the experiment NN-C, having the neural network linearized around the Ctrl experiment fields, is

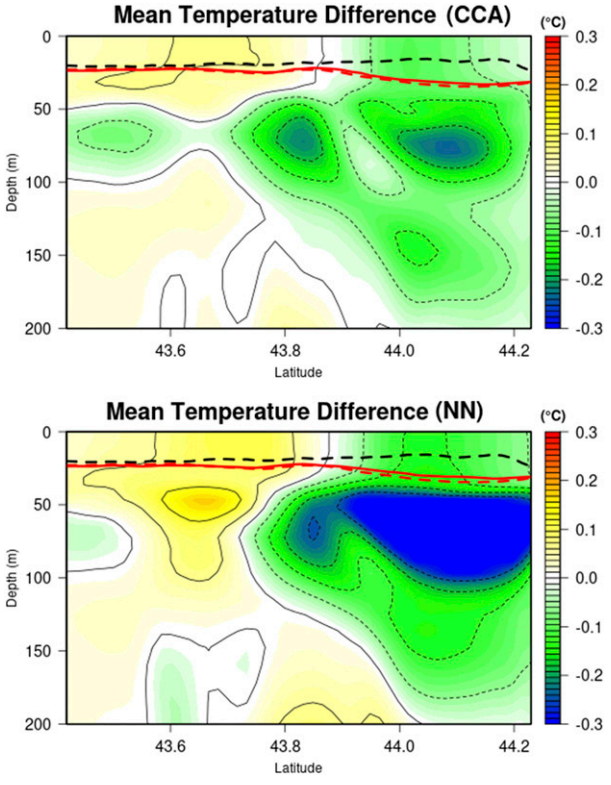

FIG. 12. Mean temperature difference with respect to the Ctrl experiment for the two experiments CCA and NN during the experimental period from 14 Oct to 11 Nov 2017, along the source-receiver propagation path of Fig. 1. Solid red lines correspond to the time-averaged mixed layer depth of the two experiments, while the red dashed line corresponds to the MLD of the Ctrl experiment and the black dashed line corresponds to the MLD of the nature run.

shown for comparison. All the experiments show improvements between approximately 30 and 200 m of depth. However, NN improvements are the largest, peaking up to 0.3°C RMSE decrease (12% improvement) in the thermocline against around 0.15°C for CCA (6% improvement). NN-C still provides a positive impact but smaller than the other two experiments, thus proving the importance of linearizing the observation operator around the actual flow-dependent fields. The RMSE decrease is statistically significant for all experiments between around 50 and 150 m of depth. Salinity RMSE profiles exhibit a negligible impact, and are not shown, while sound speed RMSE profiles closely resemble those of temperature (not shown). The main mechanism for the RMSE reduction is the attenuation of the warm bias (right panel of Fig. 13) around the thermocline, performed at most by NN and consistent with the large cold analysis increments therein (Fig. 12).

Figure 14 shows the time series of RMSE for temperature at 50 m of depth and mixed layer depth. For the former, the impact of TL assimilation is visible after about one week from the experiment initialization, due to slow adjustment of the system to the new observations, and amplifies with time afterward. MLD skill scores are even slower than temperature ones to start exhibiting impact (after about two weeks). On the average, the RMSE of MLD is reduced from about 16.8 (Ctrl) to 16.0 m (CCA, 5% reduction) or 15.5 m (NN, 8% reduction), while NN-C shows the smallest reduction of RMSE (16.1 m, 4% reduction). Note that the error growth with time, particularly visible in the MLD skill scores, is driven by the thickening of the mixed layer, which is largely overestimated by the ocean model especially during autumn, causing an increase of the RMSE as shown also by Storto and Oddo (2019). Without the assimilation of other physical observing network, this feature emerges clearly during the experimental period, suggesting in turn the importance of the synergistic assimilation of physical and acoustic observations.

Although on the average NN outperforms CCA in both the temperature and MLD skill scores, occasionally CCA provides comparable or better results than NN, for instance during the last days of the simulations. Indeed, CCA proves also a robust adjoint-free method to perform the assimilation of acoustic observations into regional oceanic numerical models.

5. Conclusions

In this work, we have investigated the feasibility of assimilating underwater acoustic observations into regional oceanographic prediction systems, with focus on the use of data-driven observation operators. We have considered the case of assimilating TL data from a "source of opportunity" such as a 75-Hz noise-producing ship. The scenario we have set up includes a source and a vertical array of receivers that a mooring may be equipped with. The problem of assimilating acoustic parameters for correcting temperature fields is in general equivalent to the optimal construction of an observation operator (and its adjoint, in the context of variational assimilation schemes) that maps physical fields onto acoustic quantities. While there exist several approaches to do that, we have investigated the feasibility of using neural networks (NN) to build the observation operator, and compared to simpler and more classical linear approaches such as canonical correlation analysis (CCA). We have then assessed the ability of such scheme to correct temperature fields along the source-receiver transmission path by assimilating synthetic observations extracted from a nature ("truth") run, following the approach commonly known as observing system simulation experiments (OSSE).

Training data for estimating both the CCA and the NN models were extracted from an ensemble of oceanic simulations with stochastic physics, to each member of which corresponds an underwater acoustic propagation simulation. This specific application requires input variables ("features") covering two-dimensional transects and, thus, training data are conveniently defined in the model space; however, other applications may rely on observed measurements also for the training dataset.

The validation of the two methods on test data indicated the superiority of NN compared to CCA within our specific application. In the assimilation experiments, results suggest that the assimilation of TL data with either observation operator helps improve the top-200-m temperature variability and, to

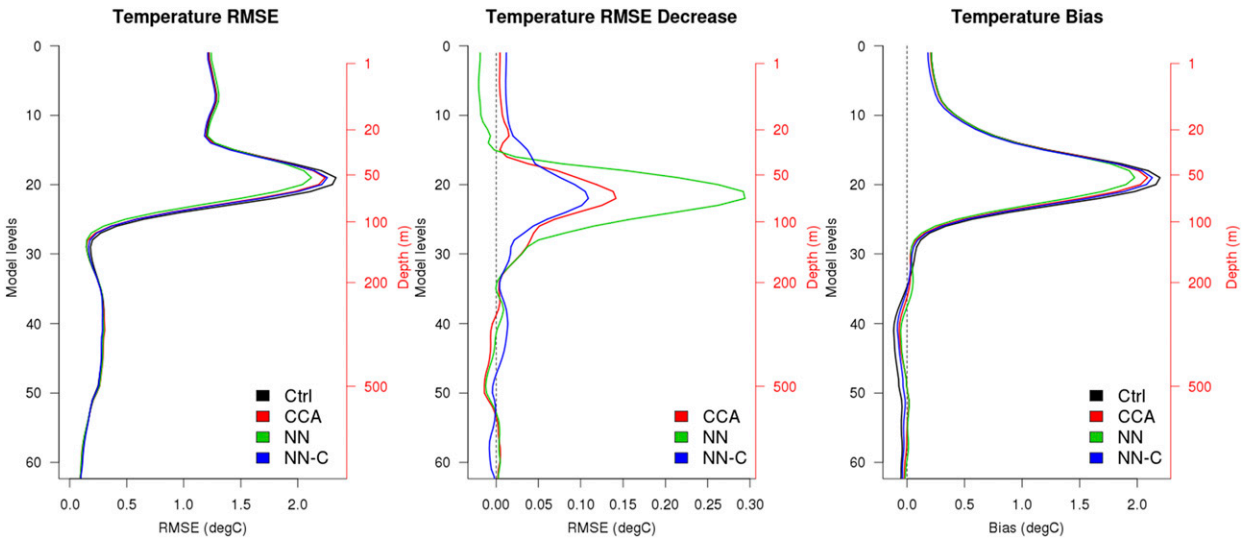

FIG. 13. (left) Temperature RMSE profiles vs the nature run along the cross section of Fig. 1, (center) temperature RMSE difference profiles with respect to the Ctrl experiment, and (right) temperature bias profiles. Bias is computed as model minus observations. The experiments are Ctrl (black), CCA (red), NN (green), and NN-C (blue). All the statistics are calculated over the entire experimental period from 14 Oct to 11 Nov 2017.

some extent, the mean state and variability of the mixed layer depth, thus contributing to an enhanced representation of the upper-ocean vertical structure. Temperature around the thermocline is impacted at most, showing a significant reduction of forecast errors.

Acoustic data may therefore complement other observing networks such as remotely sensed sea surface temperature observations, which are able, in general, to correct only the upper ocean; namely, the mixed layer. Note also that for the geometry chosen here, the sea bottom is quite deep and acoustic rays may not be absorbed by it. Therefore, our configuration represents an acoustic scenario where we can expect maximum impact, and the specific impacts of the acoustic data assimilation cannot be readily generalized as they depend on the chosen geometry. Additionally, the assimilation improves the acoustic propagation prediction performed with the assimilation-corrected temperature fields, implying that the assimilation is self-consistent and provides benefits to the acoustic prediction itself.

From the methodological perspective, we have shown for the first time that neural networks can be embedded in variational data assimilation scheme to form observation operators, and that the performances of such observation operator is on the average superior than a purely linear CCA-based operator. This proof-of-concept paves the way to the substitution of several operators used in data assimilation schemes with machine learning ones, and we argue that for many operators that are built empirically or that contain several approximations, the nonlinearity implied by neural network models can provide significant improvements. Another conclusion is about the importance of the linearization strategy. First, using numerical derivation instead of automatic differentiation limits the accuracy of the neural network model, with automatic differentiation providing almost the same accuracy as the original (fully nonlinear) neural network model. Second, linearizing the observation operator around a field different from the actual background—in our case the control experiment, but also climatological fields or other options may be used—jeopardizes to a large extent the benefits of the neural network observation operator. Practically, this finding calls for a closer interconnection between the data assimilation and the machine learning software, in order to interoperate sequentially in real-time.

Another implicit advantage of our approach is that neural networks capture the temporal changes of the covariations between input (temperature) and output (transmission loss) data. A method like CCA could be in principle reformulated to be flow-dependent, but this would imply forming at each assimilation cycle a large training dataset; namely, a very large ensemble that is impractical in real-world applications. It is worth noting that also the use of CCA as a linear data-driven observation operator is able to capture most of the benefits of the TL data assimilation; for some periods and diagnostics (for instance the RMSE in TL space, Fig. 10), CCA provides slightly better results than NN, although on the average the verification of physical variables indicates that NN outperforms CCA.

The use of data-driven formulations for constructing observation operators appears promising for a large variety of observations—remote sensing data, integrated physical quantities, etc.—namely for all observations for which operators are anyway empirically described or approximated or require too complicated tangent-linear and adjoint counterparts. Our approach can also be adapted to other data assimilation components, for instance balance (or cross-covariance) operators in cases of poor performance of analytical formulations, as for instance the assimilation of altimetry data in shallow or coastal waters.

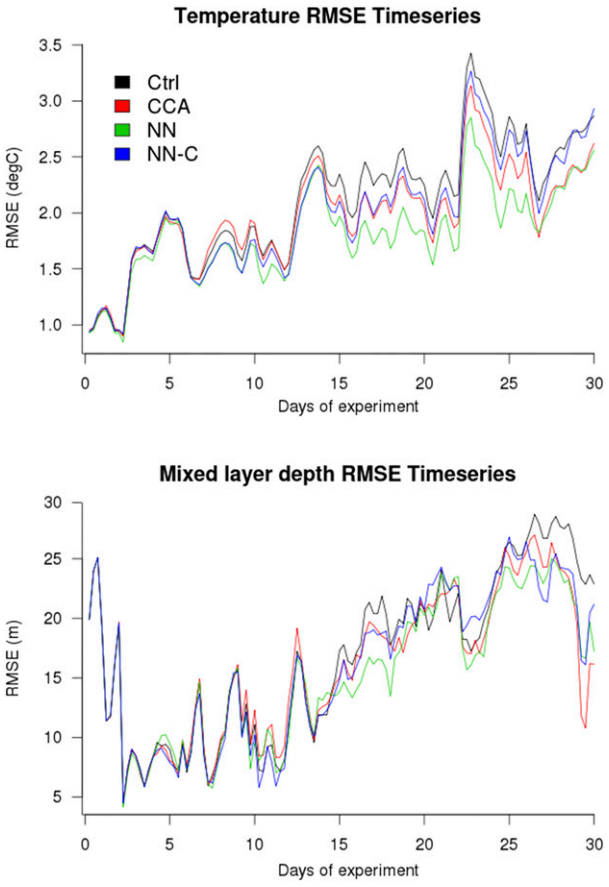

FIG. 14. RMSE time series of thermocline temperature (at 50 m of depth) and mixed layer depth vs the nature run.

Acknowledgments. This work is supported by the project "Sensing and predicting underwater noise using robotic platforms and forecast models for Maritime ISR (EKOE2/MISR2)" funded by the NATO Allied Command Transformation (ACT). We are grateful to Alberto Alvarez (CMRE) for fruitful preliminary comments to this work. Acknowledgement is made for the use of the European Centre for Medium-Range Weather Forecasts (ECMWF) computing and archive facilities in this research. We are grateful to three anonymous reviewers and the editor for their valuable suggestions and comments.

Data availability statement. Training data for the CCA and NN model construction and output model data from the assessment experiments are archived in the ECMWF File Storage system (ECFS), and are available to any interested researcher upon request to the authors.

REFERENCES

Abadi, M., and Coauthors, 2015: TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Accessed 13 May 2021, tensorflow.org.

Atlas, R., 1997: Atmospheric observations and experiments to assess their usefulness in data assimilation. *J. Meteor. Soc. Japan*, **75**, 111–130, https://doi.org/10.2151/jmsj1965.75.1B_111.

Bao, S., R. Zhang, H. Wang, H. Yan, Y. Yu, and J. Chen, 2019: Salinity profile estimation in the Pacific Ocean from satellite surface salinity observations. *J. Atmos. Oceanic Technol.*, **36**, 53–68, https://doi.org/10.1175/JTECH-D-17-0226.1.

Bartholomew-Biggs, M., S. Brown, B. Christianson, and L. Dixon, 2000: Automatic differentiation of algorithms. *J. Comput. Appl. Math.*, **124**, 171–190, https://doi.org/10.1016/S0377-0427(00)00422-2.

Baydin, A. G., B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, 2017: Automatic differentiation in machine learning: A survey. *J. Mach. Learn. Res.*, **18**, 5595–5637.

Bershad, S., and M. Weiss, 1976: Deck41 Surficial Seafloor Sediment Description Database. National Geophysical Data Center, NOAA, accessed 13 May 2021, https://doi.org/10.7289/V5VD6WCZ.

Bishop, C. H., S. Frolov, D. R. Allen, D. D. Kuhl, and K. Hoppel, 2017: The local ensemble tangent linear model: An enabler for coupled model 4D-Var. *Quart. J. Roy. Meteor. Soc.*, **143**, 1009–1020, https://doi.org/10.1002/qj.2986.

Björck, Å., and G. H. Golub, 1973: Numerical methods for computing angles between linear subspaces. *Math. Comput.*, **27**, 579–594, https://doi.org/10.2307/2005662.

Bloom, S. C., L. L. Takacs, A. M. da Silva, and D. Ledvina, 1996: Data assimilation using incremental analysis updates. *Mon. Wea. Rev.*, **124**, 1256–1271, https://doi.org/10.1175/1520-0493(1996)124<1256:DAUIAU>2.0.CO;2.

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

Castor, K., P. Gerstoft, P. Roux, W. A. Kuperman, and B. E. McDonald, 2004: Long-range propagation of finite-amplitude acoustic waves in an ocean waveguide. *J. Acoust. Soc. Amer.*, **116**, 2004–2010, https://doi.org/10.1121/1.1756613.

Clementi, E., P. Oddo, M. Drudi, N. Pinardi, G. Korres, and A. Grandi, 2017: Coupling hydrodynamic and wave models: First step and sensitivity experiments in the Mediterranean Sea. *Ocean Dyn.*, **67**, 1293–1312, https://doi.org/10.1007/s10236-017-1087-7.

Collins, M. D., 1989a: Applications and time-domain solution of higher-order parabolic equations in underwater acoustics. *J. Acoust. Soc. Amer.*, **86**, 1097–1102, https://doi.org/10.1121/1.398101.

——, 1989b: A higher-order parabolic equation for wave propagation in an ocean overlying an elastic bottom. *J. Acoust. Soc. Amer.*, **86**, 1459–1464, https://doi.org/10.1121/1.398706.

——, 1993: A split-step Padé solution for the parabolic equation method. *J. Acoust. Soc. Amer.*, **93**, 1736–1742, https://doi.org/10.1121/1.406739.

——, 1994: Generalization of the split-step Pade solution. *J. Acoust. Soc. Amer.*, **96**, 382–385, https://doi.org/10.1121/1.410488.

Cornuelle, B. D., and P. F. Worcester, 1996: Ocean acoustic tomography: Integral data and ocean models. *Modern Approaches to Data Assimilation in Ocean Modeling*, P. Malanotte-Rizzoli, Ed., Elsevier Science, 97–115, https://doi.org/10.1016/S0422-9894(96)80007-9.

Courtier, P., 1997: Variational methods. *J. Meteor. Soc. Japan*, **75**, 211–218, https://doi.org/10.2151/jmsj1965.75.1B_211.

Culver, R. L., and H. J. Camin, 2008: Sonar signal processing using probabilistic signal and ocean environmental models. *J. Acoust. Soc. Amer.*, **124**, 3619–3631, https://doi.org/10.1121/1.3006379.

D'Agostino, R. B., 1970: Transformation to normality of the null distribution of g1. *Biometrika*, **57**, 679–681, https://doi.org/10.1093/biomet/57.3.679.

Dubeau, F., 2019: A remark on Richardson's extrapolation process and numerical differentiation formulae. *J. Comput. Phys.: X*, **2**, 100017, https://doi.org/10.1016/j.jcpx.2019.100017.

Dushaw, B. D., 1999: Inversion of multi mega meter range acoustic data for ocean temperature. *IEEE J. Oceanic Eng.*, **24**, 215–223, https://doi.org/10.1109/48.757272.

——, 2019: Ocean acoustic tomography in the North Atlantic. *J. Atmos. Oceanic Technol.*, **36**, 183–202, https://doi.org/10.1175/JTECH-D-18-0082.1.

——, and H. Sagen, 2016: A comparative study of the properties of moored/point and acoustic tomography/integral observations of Fram Strait using objective mapping techniques. *J. Atmos. Oceanic Technol.*, **33**, 2079–2093, https://doi.org/10.1175/JTECH-D-15-0251.1.

——, P. F. Worcester, B. D. Cornuelle, and B. M. Howe, 1993: Variability of heat content in the central North Pacific in summer 1987 determined from long-range acoustic transmissions. *J. Phys. Oceanogr.*, **23**, 2650–2666, https://doi.org/10.1175/1520-0485(1993)023<2650:VOHCIT>2.0.CO;2.

——, H. Sagen, and A. Beszczynska-Möller, 2016: On the effects of small-scale variability on acoustic propagation in Fram Strait: The tomography forward problem. *J. Acoust. Soc. Amer.*, **140**, 1286–1299, https://doi.org/10.1121/1.4961207.

Eyre, J. R., S. J. English, and M. Forsythe, 2020: Assimilation of satellite data in numerical weather prediction. Part I: The early years. *Quart. J. Roy. Meteor. Soc.*, **146**, 49–68, https://doi.org/10.1002/qj.3654.

Fang, M., and X. Li, 2019: An artificial neural networks-based tree ring width proxy system model for paleoclimate data assimilation. *J. Adv. Model. Earth Syst.*, **11**, 892–904, https://doi.org/10.1029/2018MS001525.

Gaillard, F., 1992: Evaluating the information content of tomographic data: Application to mesoscale observations. *J. Geophys. Res.*, **97**, 15 489–15 505, https://doi.org/10.1029/92JC01295.

Graves, A., A. Mohamed, and G. Hinton, 2013: Speech recognition with deep recurrent neural networks. *2013 IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Vancouver, British Columbia, Canada, IEEE, 6645–6649, https://doi.org/10.1109/ICASSP.2013.6638947.

Haddad, Z. S., J. L. Steward, H. C. Tseng, T. Vukicevic, S.-H. Chen, and S. Hristova-Veleva, 2015: A data assimilation technique to account for the nonlinear dependence of scattering microwave observations of precipitation. *J. Geophys. Res. Atmos.*, **120**, 5548–5563, https://doi.org/10.1002/2015JD023107.

Halliwell, G. R., M. Mehari, L. K. Shay, V. H. Kourafalou, H. Kang, H.-S. Kim, J. Dong, and R. Atlas, 2017: OSSE quantitative assessment of rapid-response prestorm ocean surveys to improve coupled tropical cyclone prediction. *J. Geophys. Res. Oceans*, **122**, 5729–5748, https://doi.org/10.1002/2017JC012760.

Hotelling, H., 1936: Relations between two sets of variates. *Biometrika*, **28**, 321–377, https://doi.org/10.1093/biomet/28.3-4.321.

Howe, B. M., J. Miksis-Olds, E. Rehm, H. Sagen, P. F. Worcester, and G. Haralabus, 2019: Observing the oceans acoustically. *Front. Mar. Sci.*, **6**, 426, https://doi.org/10.3389/fmars.2019.00426.

Hursky, P., M. B. Porter, W. S. Hodgkiss, and W. A. Kuperman, 2014: Adjoint modeling for acoustic inversion. *J. Acoust. Soc. Amer.*, **115**, 607–619, https://doi.org/10.1121/1.1636760.

Inoue, T., G. De Magistris, A. Munawar, T. Yokoya, and R. Tachibana, 2017: Deep reinforcement learning for high precision assembly tasks. *2017 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, Vancouver, British Columbia, Canada, IEEE, 819–825, https://doi.org/10.1109/IROS.2017.8202244.

Janjić, T., and Coauthors, 2018: On the representation error in data assimilation. *Quart. J. Roy. Meteor. Soc.*, **144**, 1257–1278, https://doi.org/10.1002/qj.3130.

Jansen, E., S. Pimentel, W. H. Tse, D. Denaxa, G. Korres, I. Mirouze, and A. Storto, 2019: Using canonical correlation analysis to produce dynamically based and highly efficient statistical observation operators. *Ocean Sci.*, **15**, 1023–1032, https://doi.org/10.5194/os-15-1023-2019.

Jin, J., H. X. Lin, A. Segers, Y. Xie, and A. Heemink, 2019: Machine learning for observation bias correction with application to dust storm data assimilation. *Atmos. Chem. Phys.*, **19**, 10 009–10 026, https://doi.org/10.5194/acp-19-10009-2019.

Krasnopolsky, V. M., M. S. Fox-Rabinovitz, and D. V. Chalikov, 2005: New approach to calculation of atmospheric model physics: Accurate and fast neural network emulation of longwave radiation in a climate model. *Mon. Wea. Rev.*, **133**, 1370–1383, https://doi.org/10.1175/MWR2923.1.

Krizhevsky, A., I. Sutskever, and G. E. Hinton, 2017: ImageNet classification with deep convolutional neural networks. *Commun. ACM*, **60**, 84–90, https://doi.org/10.1145/3065386.

Kwon, Y., B. A. Forman, J. A. Ahmad, S. V. Kumar, and Y. Yoon, 2019: Exploring the utility of machine learning-based passive microwave brightness temperature data assimilation over terrestrial snow in high mountain Asia. *Remote Sens.*, **11**, 2265, https://doi.org/10.3390/rs11192265.

Large, W. G., and S. G. Yeager, 2004: Diurnal to decadal global forcing for ocean and sea-ice models: The data sets and flux climatologies. NCAR Tech. Note TN-460+STR, NCAR, https://doi.org/10.5065/D6KK98Q6.

Lary, D. J., A. H. Alavi, A. H. Gandomi, and A. L. Walker, 2016: Machine learning in geosciences and remote sensing. *Geosci. Frontiers*, **7**, 3–10, https://doi.org/10.1016/j.gsf.2015.07.003.

——, and Coauthors, 2018: Machine learning applications for Earth observation. *Earth Observation Open Science and Innovation*, P. P. Mathieu and C. Aubrecht, Eds., ISSI Scientific Report Series, Vol. 15, Springer, 165–218, https://doi.org/10.1007/978-3-319-65633-5_8.

Lebedev, K. V., M. Yaremchuk, H. Mitsudera, I. Nakano, and G. Yuan, 2003: Monitoring the Kuroshio Extension through dynamically constrained synthesis of the acoustic tomography, satellite altimeter and in situ data. *J. Oceanogr.*, **59**, 751–763, https://doi.org/10.1023/B:JOCE.0000009568.06949.c5.

Lermusiaux, P. F. J., 2006: Uncertainty estimation and prediction for interdisciplinary ocean dynamics. *J. Comput. Phys.*, **217**, 176–199, https://doi.org/10.1016/j.jcp.2006.02.010.

——, and C. S. Chiu, 2002: Four-dimensional data assimilation for coupled physical—Acoustical fields. *Impact of Littoral Environmental Variability of Acoustic Predictions and Sonar Performance*, N. G. Pace and F. B. Jensen, Eds., Springer, 417–424.

Lewis, J. K., J. Rudzinsky, S. Rajan, P. J. Stein, and A. Vandiver, 2005: Model-oriented ocean tomography using higher frequency, bottom-mounted hydrophones. *J. Acoust. Soc. Amer.*, **117**, 3539–3554, https://doi.org/10.1121/1.1893355.

Li, J. L., L. L. Jin, and W. Xu, 2014: Inversion of internal wave-perturbed sound-speed field via acoustic data assimilation. *IEEE J. Oceanic Eng.*, **39**, 407–418, https://doi.org/10.1109/JOE.2013.2255975.

Lin, Y.-T., 2013: A higher-order tangent linear parabolic-equation solution of three-dimensional sound propagation. *J. Acoust. Soc. Amer.*, **134**, EL251–EL257, https://doi.org/10.1121/1.4813852.

Liu, Z.-Q., and F. Rabier, 2002: The interaction between model resolution, observation resolution and observation density in data assimilation: A one-dimensional study. *Quart. J. Roy. Meteor. Soc.*, **128**, 1367–1386, https://doi.org/10.1256/003590002320373337.

Lurton, X., 2010: *An Introduction to Underwater Acoustics: Principles and Applications*. 2nd ed. Springer-Verlagerg, 680 pp.

Madec, G., and Coauthors, 2017: NEMO ocean engine. Zenodo, accessed 13 May 2021, https://doi.org/10.5281/ZENODO.1472492.

Mattern, P. J., and C. A. Edwards, 2019: A simple finite difference-based approximation for biogeochemical tangent linear and adjoint models. *J. Geophys. Res. Oceans*, **124**, 4–26, https://doi.org/10.1029/2018JC014283.

Munk, W. H., 1986: Acoustic monitoring of ocean gyres. *J. Fluid Mech.*, **173**, 43–53, https://doi.org/10.1017/S0022112086001064.

——, and C. Wunsch, 1982: Observing the ocean in the 1990s. *Philos. Trans. Roy. Soc. London*, **307A**, 439–464.

——, P. F. Worcester, and C. Wunsch, 1995: *Ocean Acoustic Tomography*. Cambridge University Press, 433 pp.

Ngodock, H., M. Carrier, J. Fabre, R. Zingarelli, and I. Souopgui, 2017: A variational data assimilation system for the range dependent acoustic model using the representer method: Theoretical derivations. *J. Acoust. Soc. Amer.*, **142**, 186–194, https://doi.org/10.1121/1.4989541.

Palmer, T. N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. Shutts, M. Steinheimer, and A. Weisheimer, 2009: Stochastic parametrization and model uncertainty. ECMWF Tech. Memo. 598, 42 pp., https//doi.org/10.21957/ps8gbwbdv.

Richardson, L. F., 1911: The approximate arithmetical solution by finite differences of physical problems including differential equations, with an application to the stresses in a masonry dam. *Philos. Trans. Roy. Soc. London*, **210A**, 307–357, https://doi.org/10.1098/rsta.1911.0009.

Robinson, A. R., P. Abbot, P. F. J. Lermusiaux, and L. Dillman, 2002: Transfer of uncertainties through physical-acoustical-SONAR end-to-end systems: A conceptual basis. *Impact of Littoral Environmental Variability of Acoustic Predictions and SONAR Performance*, N. G. Pace and F. B. Jensen, Eds., Springer, 603–610.

Sagen, H., B. D. Dushaw, E. K. Skarsoulis, D. Dumont, M. A. Dzieciuch, and A. Beszczynska-Möller, 2016: Time series of temperature in Fram Strait determined from the 2008–2009 DAMOCLES acoustic tomography measurements and an ocean model. *J. Geophys. Res. Oceans*, **121**, 4601–4617, https://doi.org/10.1002/2015JC011591.

Sammartino, M., S. Marullo, R. Santoleri, and M. Scardi, 2018: Modelling the vertical distribution of Phytoplankton biomass in the Mediterranean Sea from satellite data: A neural network approach. *Remote Sens.*, **10**, 1666, https://doi.org/10.3390/rs10101666.

Saunders, R., and Coauthors, 2018: An update on the RTTOV fast radiative transfer model (currently at version 12). *Geosci. Model Dev.*, **11**, 2717–2737, https://doi.org/10.5194/gmd-11-2717-2018.

Scheuerer, M., M. B. Switanek, R. P. Worsnop, and T. M. Hamill, 2020: Using artificial neural networks for generating probabilistic subseasonal precipitation forecasts over California. *Mon. Wea. Rev.*, **148**, 3489–3506, https://doi.org/10.1175/MWR-D-20-0096.1.

Spiesberger, J. L., and K. Metzger, 1992: Basin-scale ocean monitoring with acoustic thermometers. *Oceanography*, **5**, 92–98, https://doi.org/10.5670/oceanog.1992.15.

Storto, A., 2016: Variational quality control of hydrographic profile data with non-Gaussian errors for global ocean variational data assimilation systems. *Ocean Modell.*, **104**, 226–241, https://doi.org/10.1016/j.ocemod.2016.06.011.

——, and R. Randriamampianina, 2010: Ensemble variational assimilation for the representation of background error covariances in a high-latitude regional model. *J. Geophys. Res.*, **115**, D17204, https://doi.org/10.1029/2009JD013111.

——, and P. Oddo, 2019: Optimal assimilation of daytime SST retrievals from SEVIRI in a regional ocean prediction system. *Remote Sens.*, **11**, 2776, https://doi.org/10.3390/rs11232776.

——, and P. Andriopoulos, 2021: A new stochastic ocean physics package and its application to hybrid-covariance data assimilation. *Quart. J. Roy. Meteor. Soc.*, **147**, 1691–1725, https://doi.org/10.1002/qj.3990.

——, S. Dobricic, S. Masina, and P. Di Pietro, 2011: Assimilating along-track altimetric observations through local hydrostatic adjustment in a global ocean variational assimilation system. *Mon. Wea. Rev.*, **139**, 738–754, https://doi.org/10.1175/2010MWR3350.1.

——, S. Masina, and S. Dobricic, 2013: Ensemble spread-based assessment of observation impact: Application to a global ocean analysis system. *Quart. J. Roy. Meteor. Soc.*, **139**, 1842–1862, https://doi.org/10.1002/qj.2071.

——, P. Oddo, A. Cipollone, I. Mirouze, and B. Lemieux, 2018: Extending an oceanographic variational scheme to allow for affordable hybrid and four-dimensional data assimilation. *Ocean Modell.*, **128**, 67–86, https://doi.org/10.1016/j.ocemod.2018.06.005.

——, ——, E. Cozzani, and E. Ferreira Coelho, 2019: Introducing along-track error correlations for altimetry data in a regional ocean prediction system. *J. Atmos. Oceanic Technol.*, **36**, 1657–1674, https://doi.org/10.1175/JTECH-D-18-0213.1.

——, S. Falchetti, P. Oddo, Y.-M. Jiang, and A. Tesei, 2020: Assessing the impact of different ocean analysis schemes on oceanic and underwater acoustic predictions. *J. Geophys. Res. Oceans*, **125**, e2019JC015636, https://doi.org/10.1029/2019JC015636.

Weatherall, P., and Coauthors, 2015: A new digital bathymetric model of the world's oceans. *Earth Space Sci.*, **2**, 331–345, https://doi.org/10.1002/2015EA000107.

Xue, Y., and B. A. Forman, 2017: Integration of satellite-based passive microwave brightness temperature observations and an ensemble-based land data assimilation framework to improve snow estimation in forested regions. *2017 IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, Fort Worth, TX, IEEE, 311–314, https://doi.org/10.1109/IGARSS.2017.8126958.