Computers & Geosciences 72 (2014) 201–209

Contents lists available at ScienceDirect

Computers & Geosciences

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

Neural approach to inverting complex system: Application to ocean salinity profile estimation from surface parameters

Mbaye Babacar Gueye [a,b,*], Awa Niang [a], Sabine Arnault [b], Sylvie Thiria [b,c], Michel Crépon [b]

[a] *Laboratoire de Traitement de l'Information/ESP, Université Cheikh Anta Diop, BP 5085, Dakar Fann, Senegal*
[b] *Sorbonne Universités, UPMC Univ Paris 06, LOCEAN/IPSL, F-75005 Paris, France*
[c] *UVSQ, 78035 Versailles, France*

ARTICLE INFO

Article history:
Received 6 September 2013
Received in revised form
26 June 2014
Accepted 16 July 2014
Available online 28 July 2014

Keywords:
Salinity profile estimation
Neural network
Kohonen map
Inverse problem

ABSTRACT

A neural network model is proposed for reconstructing ocean salinity profiles from sea surface parameters only. The method is applied to the tropical Atlantic. Prior data mining on a complete dataset shows that latitude and sea surface salinity are the most relevant surface parameters in the prediction of salinity profiles. A classification using a self-organizing map learned on a large multivariate dataset is able to retrieve the most probable vertical salinity profiles from the surface parameters only. Both in situ and modelled oceanic data are used to evaluate the results. The reconstruction misses some salinity features in areas with high time-space variability in which the limited available dataset was unable to provide the complete variability ranges during the learning process. However, apart from these restricted areas, the salinity profiles are reproduced with correlations greater than 0.95 for most of the profiles of the test set.

1. Introduction

Geophysical systems (ocean, atmosphere, etc.) have many complex properties due to interactive processes governed by non-linear laws. An important aspect of physical sciences is the inference of physical parameters from observations. In general, the laws of physics permit the computation of data values from physical variables through a model (the direct model). This is called the forward problem. The inverse problem aims at estimating the physical variables and/or the parameters of the direct model from a set of measurements through the use of an inverse model (Snieder and Trampert, 1999). While the forward problem has (in deterministic physics) a unique solution, the inverse problem may not (Tarantola, 2005), if the model is non-linear. Inverse problems of geophysics are therefore generally ill-posed in the sense of Hadamard (Starostenko and Zavorot'ko, 1996; Tarantola, 2005), i.e their solution is no more a single-valued function. In an ill-posed inverse problem, a classical least-square minimum distance or maximum likelihood solution may not be uniquely defined. Moreover, the sensitivity of such solutions to slight perturbations in the data is often unacceptably large. Alternative methods based on knowledge related to the observations through a learning process have become widely developed and used. Our approach belongs to these latter methods.

Our study aims at retrieving ocean vertical salinity profiles from ocean surface measurements. The density of satellite observations has permitted a quasi-continuous observation of the global ocean surface. The two-dimensional images provided by satellite sensors contain information on physical or biogeophysical sea-surface variables but not directly on their associated vertical profiles. Inverting the sea-surface data remotely sensed by satellite to obtain the vertical profiles of biogeophysical variables requires inverting a numerical modelling of their relationships. The difficulty is that it is found experimentally that several different profiles might correspond to the same surface measurements. Moreover, as is described in Section 4, the problem is highly under-determined, the profile data being much larger in quantity than the associated surface data. Such inverse models are, however, often faced with problems of non-linearity, complexity and incomplete knowledge of the mechanisms that govern the formation of these profiles. Therefore the reconstruction of the vertical profiles from the surface information often leads to an ill posed problem, as explained here above. In this paper we propose a statistical approach that addresses this difficult inverse problem; i.e. the determination of the ocean 3-D structure from surface data, using advanced statistical methods.

Water exchanges in the global hydrological cycle are determined by complex mechanical and thermodynamic constraints

* Corresponding author. Present address: Laboratoire de Traitement de l'Information, Ecole Supérieure Polytechnique, Université Cheikh Anta Diop de Dakar, BP 5085 Dakar Fann, Senegal. Tel.: +221 775708811.
E-mail address: mbayebabacar@gmail.com (M.B. Gueye).

http://dx.doi.org/10.1016/j.cageo.2014.07.012
0098-3004/

that form the basis of the climate system dynamics. Ocean salinity (S) is one of the variables in this cycle which is the most difficult to observe (U.C. Office, 2007), especially in the Atlantic, where it presents a strong spatial variability. *S* is also an indicator of the intensity and location of water exchange (evaporation and precipitation) at the ocean–atmosphere interface (Dessier and Donguy, 1994). Like temperature (T), *S* has an impact on the water density, and therefore any integrated quantity depending on it, such as sea-surface height (SSH), which is now commonly provided by satellite altimetry. Therefore any algorithm and model linking easy-to-observe surface parameters, such as SSH, sea-surface temperature (SST), and sea-surface salinity, to salinity profiles would be a valuable asset in providing information to centres studying ocean and weather variability.

A lot of inverse methods, based on in situ measurements, have already been proposed in order to solve this problem; they differ by the information used, the statistical method applied and the area where their performances were tested. Among them, we may mention:

- Agarwal et al. (2007) reconstructed subsurface *S* profiles in the Indian Ocean from January 2000 to July 2004. They used a combination of Empirical Orthogonal Function (EOF) analysis and a genetic algorithm connecting an *S* profile to the SST and the SSS.
- Maes and Behringer (2000) estimated the vertical variability of the *S* field by using a weighted least-square procedure. They computed vertical modes of *T* and *S* provided by an EOF analysis of in situ measurements, SSS climatology, remotely sensed sea level anomaly (SLA) and SST. The method was tested on two Tropical Atmosphere-Ocean (TAO) moorings along 165°E in the western Pacific Ocean for the period 1993–1998, giving a correlation between observed and estimated time-series of about 0.7 in the equatorial band.
- Thacker (2008) proposed an empirical relationship between *S* and *T* with the aid of local regression. His strategy was to estimate the $T-S$ relationship, at each depth level in sixteen $20^{\circ} \times 5^{\circ}$ sub-regions of different latitude and longitude, by a linear regression. The method applied a non-linear fit to the data since it dealt with the square of *T* together with T. Local regression models were applied to the South Atlantic.
- Guinehut et al. (2012) described an observation-based approach that combined *T* profiles, *S* profiles, satellite altimeter sea-level anomalies and satellite SST using statistical methods. In a first step they derived synthetic three-dimensional temperature fields from altimeter and sea-surface temperature observations, and three-dimensional salinity fields from altimeter observations and temperature fields, through multiple/simple linear-regression methods. The second step of their method consists in combining the synthetic fields with in situ temperature and salinity profiles using an optimal interpolation method.

These examples show that each inversion uses part of the available surface parameters and/or introduces the geographical coordinates. All of them dealt with statistical approaches that are different from each other. Most of the time, they considered actual data available in different areas of the ocean. Doing so, the datasets are often not large enough to obtain a fine-tuning of the inverse method. In the present paper we propose to broaden the above approaches by associating the *S* profiles with all available surface parameters. This is done with the help of an automatic classification-based methodology consisting in Kohonen self-organizing maps (SOM). We first focus our study on model data that allow the large database necessary to implement the method. In order to validate the approach, we chose a particular area (the tropical Atlantic) in which we have in situ measurements.

This paper is organized as follows. Section 2 describes the characteristics, in terms of oceanic features, of the tropical Atlantic area where the experiences have been developed, together with the datasets we used. The methodology is presented in Section 3. Section 4 is devoted to the results and discussion, followed by the conclusion, in Section 5.

2. Regional area and data descriptions

2.1. Regional characteristics

The Atlantic Ocean plays a key role in the thermohaline circulation of the world ocean. It may be considered as the engine of the famous conveyor belt characterizing the above-mentioned thermohaline circulation. This "conveyor belt" results from the fact that the Atlantic Ocean is the only major ocean open to the north polar region (the Arctic) which allows significant cooling of the surface water which therefore sinks in order to reach its natural density level several hundred metres below the sea surface. The conveyor belt therefore constitutes the major oceanic contribution to the climate variability through the high-latitude convection occurring in the Norwegian Sea (Gordon, 1986). The variability of the tropical Atlantic impacts that of other regions: for example, northeastern Brazil, northwestern Africa, Central America and the Caribbean (Munoz et al., 2012). The conveyor belt provides the connection between the tropical Atlantic, the tropical Pacific (Losada et al., 2010) and the Indian Ocean (Kucharski et al., 2009). The tropical Atlantic Ocean is therefore an attractive area for oceanic climate studies.

The Atlantic Ocean has the highest *S* values (>37). These maximum-salinity water areas (MSW) occupy the subtropical areas, the *S* values in the equatorial/tropical zone are lower, at about 35°north and south of the equator and are maintained principally by evaporation. These MSW, advected poleward at the surface and subducted at depth and advected equatorward by the oceanic circulation, constitute key features of the conveyor belt dynamics.

Furthermore, the vertical variability of *S* in the tropical Atlantic consists mainly of two homogeneous layers separated by a strong salinity gradient, the so-called halocline.

The precise determination of the halocline depth is difficult, although it is a main contributor to local air–sea interactions through the barrier-layer phenomena. These barrier-layers can block local air–sea interactions and lead again to climate variations (Tanguy et al., 2010).

Even though the tropical Atlantic has been recognized as an important region in the Earths coupled ocean–atmosphere climate system, it has been challenging to model it satisfactorily by coupled climate models (Munoz et al., 2012), especially to simulate its *S* variability. Apart from the forcing uncertainties (river run-off, evaporation, precipitation, wind stress) this difficulty is mainly due to recurrent and unstable climatic features localized in some particular areas (Fig. 1) such as:

- the MSW zones around 15°S and 25°N;
- the region of the westward-flowing surface South Equatorial Current (SEC) and the eastward-flowing subsurface Equatorial UnderCurrent (EUC) around the equator;
- the Inter-Tropical Convergence Zone (ITCZ) and the eastward-flowing surface Northern Equatorial CounterCurrent (NECC) area around 5°N, southward of the westward-flowing Northern