IEEE
GEOSCIENCE AND REMOTE
SENSING LETTERS

Reconstruction of Sub-Surface Velocities from Satellite Observations Using Iterative Self-Organizing Maps

Christopher Chapman and Anastase Alexandre Charantonis

***Abstract*—A new method based on modified self-organizing maps is presented for the reconstruction of deep ocean current velocities from surface information provided by satellites. This method takes advantage of local correlations in the data-space to improve the accuracy of the reconstructed deep velocities. No assumptions regarding the structure of the water column, nor the underlying dynamics of the flow field, are made. Using satellite observations of surface velocity, sea-surface height and sea-surface temperature, as well as observations of the deep current velocity from autonomous Argo floats to train the map, we are able to reconstruct realistic high–resolution velocity fields at a depth of 1000m. Validation reveals promising results, with a speed root mean squared error of $\sim$2.8cm.s^{-1}, more than a factor of two smaller than competing methods, and direction errors consistently smaller than 30°. Finally, we discuss the merits and shortcomings of this methodology.**

***Index Terms*—Oceans,Remote sensing,Self-organizing feature maps**

I. Introduction

SUBSURFACE observations of the world's ocean, particularly of climatically interesting fields such as the velocity of ocean currents, are generally sparse both temporally and spatially. Despite recent attempts to improve ocean observing networks, our ability to directly measure oceanic properties at depth is still limited. The lack of long-term data with broad spatio–temporal coverage impedes our ability to make robust inferences about changes in the climate system.

In contrast, since the early 1980s, quasi-global measurements from satellites have enabled near continuous measurement of the global ocean's surface. In particular, observations of the sea-surface height anomaly (which enables direct measurement of the surface geostrophic velocity) from altimeters, and observations of the sea-surface temperature from radiometers and infrared sensors have revolutionized the understanding of the ocean's dynamics. These observations have revealed that the ocean is rich in flow features of varying spatial and temporal scales [1]. We illustrate the difference in the spatial coverage between surface and deep measurements in Fig. 1, which shows the velocity from satellite altimetry (shaded contours/black vectors) and at approximately 1000m depth from Argo floats (red vectors) in the South Indian Ocean on the 17th of April, 2009. It is clear that the satellite data provide broad spatial coverage of the region while the measurements at depth are scattered. This gap in coverage has lead to numerous efforts to reconstruct sub-surface quantities from high-resolution satellite data.

Attempts to reconstruct the deep flow from surface observations fall into two different categories: 'statistical" or "dynamical" methods. "Statistical" methods take advantage of empirical relationships between surface and subsurface quantities to reconstruct the subsurface fields, subject to the assumptions that these relationships are static in time and that the vertical structure of the water column can be represented as a simple function of depth [2], [3]. "Dynamical" methodologies combine the equations of fluid motion with surface information from satellites to estimate the sub-surface fields [4], [5]. However, dynamical reconstructions require the basic stratification of the ocean to be slowly varying in space and tend to smooth out important small scale flow structures.

Recently, machine-learning methods have been proposed for similar problems [6], [7], [8]. These

C. Chapman is with the LOCEAN-IPSL at the Université de Pierre et Marie Curie, Paris 75005 France e-mail: chris.chapman.28@gmail.com.

A. A. Charantonis is with SAMOVAR, Télécom SudParis, CNRS, Université Paris-Saclay, 9 rue Charles Fourier 91011 EVRY

Manuscript received

Digital Object Identifier: 10.1109/LGRS.2017.2665603

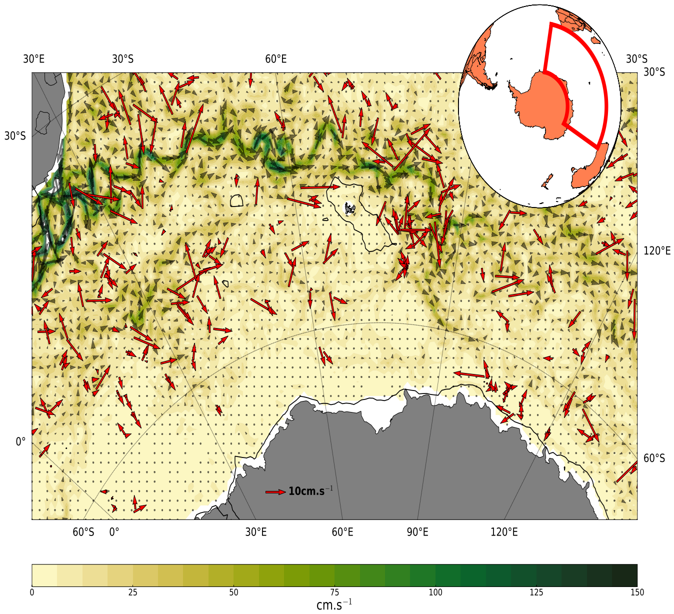

Fig. 1: Current velocity at the surface from the AVISO satellite product (shaded contours and black vectors) and near 1000m depth from the Argo floats (solid red vectors) in the South Indian Ocean on the 17th of April, 2009. The location of this region is indicated in the inset box. The solid contour line indicates the 1000m depth contour.

methods have several advantages over the traditional methods described above: the relationships between surface and sub-surface quantities can vary in space and time; it is not necessary to make any assumptions about the vertical structure of the water column; and non-linear relationships can be extracted from the data.

In this letter, we tackle the problem of reconstructing sub-surface velocities from surface data using a self-organizing maps, based on previous work by Charontonis *et al.* [8]. We restrict our attention to the Southern Ocean, the region that encircles the Antarctic continent, between 65°S and 35°S, as it hosts a complex flow field that presents a challenge for reconstruction schemes, and as it one of most data sparse ocean basins. Hence, a robust reconstruction of the deep flow from satellite data in this region could be of immense benefit to the oceanographic community.

II. Data and Methodology

A. Data

The data used in this study consists of *surface data* that are used as predictors, and *sub-surface* measurements of current speed that are used to train the SOM and for validation.

1) Surface Data: The altimeteric data used in this study is obtained from the Archiving, Validation, and Interpretation of Satellite Oceanographic data (AVISO) daily gridded absolute dynamic topography (ADT) (http://www.aviso.altimetry.fr/), a “level 4” product. We use ADT for the 5 year period 2005–2011, mapped to a 1/4 degree Mercator grid using optimal interpolation of along-track data series. This dataset provides estimates of both the sea-surface height and velocity. An example of the surface velocity from this dataset is shown in Fig. 1.

The sea–surface temperature (SST) data used are daily averages of Version 2 of the NOAA combined AVHRR-AMSR optimally interpolated SST product (https://www.ncdc.noaa.gov/oisst)[9]. The combined use of infrared and microwave instruments in cloud-free regions reduces systematic biases as the errors of each sensor are independent.

2) Sub-Surface Velocity Data: In order to estimate the ocean current velocity at depth, we use the velocity data provided by autonomous Lagrangian drifters called Argo floats [10]. After deployment, Argo floats descend to a pre–programmed “parking” depth (generally 1000m) where they drift with the current for approximately 10 days, then ascend to the surface (taking a profile of temperature and salinity) and transmit their location via satellite. The floats then re-descend to their parking depth and repeat the cycle. With knowledge of the time between each resurfacing, and the surfacing locations, the parking depth velocity can be estimated.

In this study, we use the ANDRO data set (http://www.umr-lops.fr/Donnees/ANDRO) [11]. This dataset provides estimates of the current velocity at the float parking depth between 2005 and 2011 and covers the entire ocean north of about 65°S. There are 122,174 independent data records in the dataset and errors due to the delay between the float surfacing and the satellite location fix and vertical shear in the water column are estimated to be small. We use only floats with parking depths within 50m of 1000m, as there are sufficient floats at this depth to enable broad geographical coverage. The number of floats at parking depths different from 1000m is much more limited.

B. Methodology

Self-Organizing Maps are a neuronal network classification algorithm that incorporate a topological structure on a 2D lattice [12]. Each class is represented by an index, c, and a reference vector, $\mathbf{ref}^c \in \mathbb{R}^D$, in the data space. $\mathbf{ref}^c$ is (approximately) the mean of all training data assigned to that class. The reference vectors of two neighboring classes on the 2D lattice are, by construction, close in data space. A practical guide to the application of SOM in oceanography is given by Liu et al. 2006 [13].

After the initial training, the SOM map can be used to reconstruct "missing" data (in our case, the deep current velocities) from available data. This is generally accomplished by matching to input data (with missing values) to a class by finding the closest reference vector in a Euclidian sense, i.e has the smallest distance:

$$d_T = \left[\sum_{i \in \text{avail.}}^{D} (x_i - ref_i^c)^2 \right]^{1/2} \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^D$ is input data vector with missing values and d_T is the *truncated distance*, as the sum is performed only over the dimensions with available data. The missing values are then "filled" by extracting the corresponding dimensions from the reference vector of the best matching class. However, due to local correlations in the data space, the pertinence of each parameter to the retrieval of a missing parameter's value varies throughout the data space. To improve the reconstruction, we follow Charantonis *et al*. [8] and introduce a similarity function, $\mathbf{s}^c$, that weights the truncated distance according to the correlation between the missing data variable (here the deep velocity) and available (surface) data. The modified distance is given by:

$$\begin{aligned} d_E^c(\mathbf{x}, \mathbf{ref}^c) &= \sum_{i \in avail.} \left(1 + \sum_{j \in missing} \left(cor_{ij}^c \right)^2 \right) \dots \\ &\times \left(x_i - ref_i^c \right)^2, \end{aligned} \tag{2}$$

where $cor_{i,j}^c$ is the correlation matrix between the missing and available variables, computed over the all data attributed to the class c during training. The first term is the similarity function that weights the distance by the correlations between the variables in the input data vector, while the second term is simply the truncated distance. As an example, imagine for a particular class, the "missing" deep currents were found to be highly correlated with the ADT and uncorrelated with SST. In this case, the ADT would be more influential when finding the class that best matches the input data. We find that the similarity function significantly improve the classification of the input data, and hence the reconstruction of the deep velocities.

A Matlab implementation our method is available as free software on the author's GitHub account.

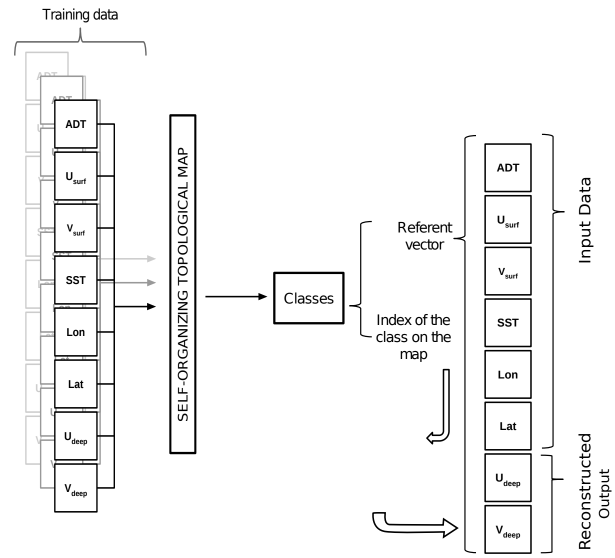

Fig. 2: Schematic of our methodology. On the left, we the training the SOM, which maps the training data to a discreet set of classes. Each class contains a reference vector that approximates the mean value of the data assigned to it. On the right, the reconstruction: the available (surface) components are projected via a similarity function onto the SOM and the missing (deep) values are extracted from reference vector of the best-matching class.

III. Reconstruction of Deep Velocities from Satellite Data

A. Validation and Errors

The SOM methodology is now applied to the problem of reconstructing velocity fields at depth from the satellite observations described in section II. To train the map, we use the altimetrically derived surface velocity and dynamic height, the SST, and the deep velocity obtained from the Argo floats as well as the latitude and longitude of each deep observation. Surface data is co–located at the subsurface data locations by linear interpolation. 80%

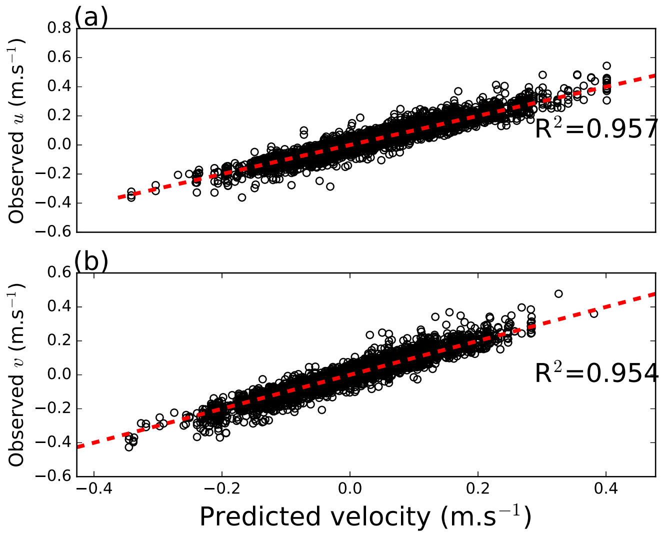

Fig. 3: The reconstructed versus observed zonal (a) and meridional (b) velocities over the Southern Ocean at 1000m. The red dashed line indicates the "perfect" reconstruction. R^2 values for each series are indicated.

of this data-set, (97,739 data records), is selected by random sampling and used to train the SOM. The remaining 20% (24,435 records) are retained for validation. The results of any SOM computation are sensitive to parameters in the training phase, such as the number of classes and the neighborhood radius [13]. After testing numerous parameter values, we have settled on using 2500 classes and initial and final neighborhood radii of 6 and 0.1, respectively. These values provide a good compromise between performance and computational expense.

The validation of the method is shown in Fig. 3. It is evident from this figure that our results are very promising. The coefficient of determination, R^2, for each velocity component is above 0.95, and the speed RMSE is 2.8cm.s^{-1}, more than a factor of 2 smaller than those obtained by Meijers *et al.* [3] and a factor of 3 smaller than those obtained from dynamical methods (eg. Isern–Fontenet *et al* [4]). For comparison with the internal variably, we note that the standard deviation of the deep current speed over the study domain is 4.8cm/s. To ensure that our method is not subject to overfitting, that is that our method 'fits' random noise in the training data and thus is unable to generalize to new data, we also validate against the training data. We obtain a speed RMSE of 2.6cm.s^{-1}, sufficiently close to the RMSE obtained from the validation data that we can rule out overfitting. We note that using the SOM methodology on its own, without the correction described by Eqn. 2, gives inferior results, with RMSEs of $\sim$7.5cm.s^{-1} and R^2 =$\sim$0.6.

We investigate the spatial distribution of the errors in both the current speed and direction, the latter quantified by the bearing angle $\theta = \tan^{-1}(v/u)$, by determining the error at the location of each deep velocity observation in the validation dataset, then interpolating the results to a regular latitude/longitude grid (not shown). Similarly to Meijers *et al* [3], we find elevated errors in current speed downstream of large sub–surface topography. Additionally, ϵ_{speed} correlates with regions of both high variance of both ADT and deep currents, suggestive of more intense meso-scale turbulence fields. The increasing error in highly turbulent regions indicates a decoupling of the surface and deep flow that may limit the effectiveness of the reconstruction. In contrast, errors in the bearing angle appear to be distributed quasi–randomly. We note that more than 85% of bearing angle errors are less that 30°, with a mean absolute error of 18° and no clear directional bias.

B. Reconstruction of Deep Southern Ocean Currents

We now apply our method to each of the daily output maps in the AVISO and OISST databases, between 2005 and 2011 to obtain 5 years of daily current velocities (1826 snapshots) at 1000m on a regular latitude/longitude grid with 1/4° grid spacing. Grid points with fewer than 10 deep velocity observations within 150km (typically south of 65°S) are masked. These data are freely available from the Dryad data repository (datadryad.org).

Fig. 4 shows the reconstructed time mean current speed for the year 2009. Our reconstructed currents display realistic behavior such as self-organisation into complicated small–scale ($\mathcal{O}$(10–20km)) "jets" (e.g. south of Africa between 20°E and 50°E), steering by subsurface topography (e.g. south of New Zealand between 170°E and 170°W) and western boundary currents (e.g. along the east coast of Australia and South America).

Our reconstructed deep velocity maps show similar features to the satGEM reconstruction of Meijers *et al.* [3]. However, there are some notable qualitative differences between the two reconstructions.

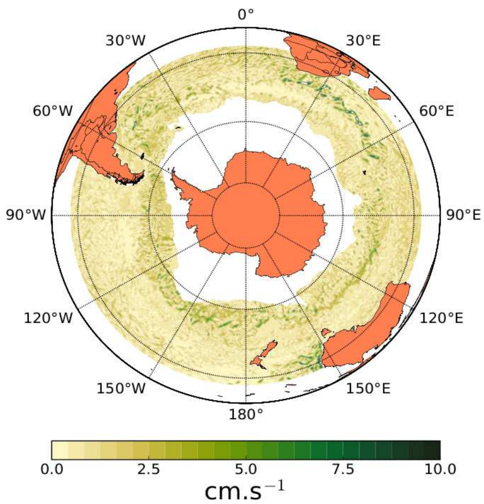

Fig. 4: The time mean reconstructed current speed at 1000m depth forthe year 2009.

Most notably, we find that our reconstructed currents are generally slower than the satGEM product, locally by as much 20%.

IV. Discussion and Conclusions

In this letter, we have used a machine-learning technique to reconstruct the velocity of ocean currents at 1000m depth from satellite observations. Our results yield errors that are 2 to 3 times smaller than competing methods. We are able to use this method to reconstruct realistic maps of deep currents with high temporal and spatial resolution.

Our methodology has several shortcomings. Most notably, to train the SOM we require velocity information at depth. While we have been able to exploit the near global coverage provided by Argo floats at 1000m, velocity data are more limited at other depths [11], which reduces our ability to apply this method more generally. The satGEM dataset of Meijers *et al.* [3] and dynamical methods [4] are not limited by the availability of deep velocity data and can provide reconstructions at any depth. However, despite these shortcomings, our method has numerous potential applications beyond the obvious extension to other quantities, such as temperature or salinity. This method could be used for real-time for data-assimilation into predictive ocean models, or for validating numerical models in data-sparse regions.

Acknowledgment

C.C. is supported by a National Science Foundation Ocean Sciences Directorate Postdoctoral Fellowship, #1521508. This work is supported by the Center for Data Science, funded by the IDEX Paris-Saclay, ANR-11-IDEX-0003-02.

References

[1] R. Morrow and P.-Y. Le Traon, "Recent advances in observing mesoscale ocean dynamics with satellite altimetry," *Advances in Space Research*, vol. 50, no. 8, pp. 1062–1076, 2012.

[2] J.-H. Park, D. R. Watts, K. L. Tracey, and D. A. Mitchell, "A multi-index gem technique and its application to the southwestern japan/east sea," *Journal of Atmospheric and Oceanic Technology*, vol. 22, no. 8, pp. 1282–1293, 2005.

[3] A. Meijers, N. Bindoff, and S. Rintoul, "Estimating the four-dimensional structure of the southern ocean using satellite altimetry," *Journal of Atmospheric and Oceanic Technology*, vol. 28, no. 4, pp. 548–568, 2011.

[4] J. Isern-Fontanet, G. Lapeyre, P. Klein, B. Chapron, and M. Hecht, "Three-dimensional reconstruction of oceanic mesoscale currents from surface information," *Journal of Geophysical Research: Oceans (1978–2012)*, vol. 113, no. C09, 2008.

[5] S. R. Keating and K. S. Smith, "Upper ocean flow statistics estimated from superresolved sea-surface temperature images," *Journal of Geophysical Research: Oceans*, vol. 120, no. 2, pp. 1197–1214, 2015.

[6] Y. Liu, R. H. Weisberg, and L. K. Shay, "Current patterns on the west florida shelf from joint self-organizing map analyses of hf radar and adcp data," *Journal of Atmospheric and Oceanic Technology*, vol. 24, no. 4, pp. 702–712, 2007.

[7] A. Charantonis, F. Badran, and S. Thiria, "Retrieving the evolution of vertical profiles of chlorophyll-a from satellite observations using hidden markov models and self-organizing topological maps," *Remote Sensing of Environment*, vol. 163, pp. 229–239, 2015.

[8] A. A. Charantonis, P. Testor, L. Mortier, F. D'Ortenzio, and S. Thiria, "Completion of a sparse glider database using multi-iterative self-organizing maps (itcomp som)," *Procedia Computer Science*, vol. 51, pp. 2198–2206, 2015.

[9] R. W. Reynolds, T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, and M. G. Schlax, "Daily high-resolution-blended analyses for sea surface temperature," *Journal of Climate*, vol. 20, no. 22, pp. 5473–5496, 2007.

[10] S. C. Riser, H. J. Freeland, D. Roemmich, S. Wijffels, A. Troisi, M. Belbéoch, D. Gilbert, J. Xu, S. Pouliquen, A. Thresher *et al.*, "Fifteen years of ocean observations with the global argo array," *Nature Climate Change*, vol. 6, no. 2, pp. 145–153, 2016.

[11] M. Ollitrault and J.-P. Rannou, "Andro: An argo-based deep displacement dataset," *Journal of Atmospheric and Oceanic Technology*, vol. 30, no. 4, pp. 759–788, 2013.

[12] T. Kohonen, *Self-Organizing Maps*, 3rd ed. Berlin: Springer, 2001, vol. 30.

[13] Y. Liu, R. H. Weisberg, and C. N. K. Mooers, "Performance evaluation of the self-organizing map for feature extraction," *Journal of Geophysical Research: Oceans*, vol. 111, no. C5, 2006, c05018.