SHORT TERM SCIENTIFIC MISSION (STSM) – SCIENTIFIC REPORT

The STSM applicant submits this report for approval to the STSM coordinator

Action number: ES1402
STSM title: Space and Time Interpolation of Remote Sensing Data in the Gulf of Trieste
STSM start and end date: 24/09/2018 to 05/10/2018
Grantee name:Matjaž Ličer

PURPOSE OF THE STSM/

Northern Adriatic shelf and coastal Adriatic regions in general are under growing environmental stress due to port, tourism and economical activities. Consequently, a better understanding of the state of the ocean in this region serves a very diverse set of purposes. Slovenia and Italy are therefore operating a common HF radar system and respective networks of moored sensors in the Gulf of Trieste. Regular monthly transects and CTD deployments in the Gulf are also performed. Satellite remote sensing data are furthermore getting more and more available and are increasing in quality. This data is however often sparse in space and time. Hence additional information could be gained from space-time interpolations of satellite and HF radar data on these scales. The purpose of the STSM was to use (and learn to use) interpolation tools developed at AGO-GHER (DINEOF) on available observational data with the purpose of creating a gridded representation of the observed state of the ocean.

Furthermore, since many rivers discharge into the Adriatic, stratification is often present, yielding first baroclinic Rossby radius of the order of several kilometers in the coastal regions and over the shelf, indicating strong current and density shear and other transient features near the coast. These features are however often missing from the reconstructed data. Further STSM purpose was therefore a feasibility study of (nonlinear) machine learning tools as a strategy to include satellite observed transient features into the reconstructed data.

Working on long timeseries is related to substantial numerical cost, hence a feasibility study of shared memory parallelization of the DINEOF code was to be performed.

DESCRIPTION OF WORK CARRIED OUT DURING THE STSMS

1. DINEOF code profiling was performed using GPROF and VALGRIND profilers. It turned out that functions related to most numerical cost (most notably `valsvd`) are called as part of an

iteration loop. This precludes straightforward `-openmp` parallelization of this segment of code. Other suitable loops where parallelized, but without much gain of performance.

2. Preprocessing tools for preparation of DINEOF input data were written in Matlab, employing filling of missing values with a CDO library. DINEOF reconstruction of SST data was performed for the Adriatic basin, employing satellite observations with QC4 and QC5 quality flags. Unfortunatelly this criterion precluded usage of SST data in the Gulf of Trieste. Consequently a subsetting script from generic satellite L3 grid to regular longitude latitude box was written in Matlab and employed to subset the 2017 SST data to Gargano peninsula. This region is of particular interest due to ubiquitous presence of baroclinic instabilities in the West Adriatic Current.

3. Convolutional Auto Encoder (CAE) network was then built in Python environment, using Keras machine learning library with a TensorFlow backend. It consisted of an encoder with 8 levels and a (8,8) convolutional bottleneck with ReLU activation. Decoder is a single layer with (3,3) bottleneck and sigmoid activation. The network was trained (70 epochs with `adadelta` gradient descent method) on the merged satellite + DINEOF data and tested on a complementary subset.

DESCRIPTION OF THE MAIN RESULTS OBTAINED

(max. 500 words)

1. DINEOF code seems possible to parallelize subject to reversing loop order between spatial and iteration loops or possibly using a different SVD library. In this case, distributed memory management like MPI might be beneficial.

2. I have learned how to use DINEOF and have a clear idea about how to put it in operational environment at my home institution.

3. CAE engine was shown to be promising for assimilating observed transient features into the reconstructed fields. It is able to learn and reproduce high-frequency features from given input data. However, the CAE we have employed tended to exhibit bias of the order of 1 K. This would most likely be solved by training a deeper network for more epochs. Additional feature that would have been of importance is a construction of STDEV dependent loss function for the CAE. This way one would be able to penalize observations with higher STDEV and hence give more weight to higher quality data points. This seems to be straightforward in low-level Tensorflow but a bit less elegant if using higher levels APIs like Keras.

FUTURE COLLABORATIONS (if applicable)

(max.500 words)

The work done shows promise and I'd be happy to continue this collaboration.