



Statistical mooring devices data quality control and ocean model validation

Paolo Oliveri, Alessandro Grandi, and Simona Simoncelli
Istituto Nazionale di Geofisica e Vulcanologia, Italy (paolo.oliveri@ingv.it)

Marine data recorded by fixed observatories at key geographic sites represent a huge and extensive source of information on the real sea conditions. Nowadays these data are not completely reliable yet due to their functioning. Marine environment it's not the ideal condition where electronic can work long and optimally, causing sampling discontinuities and needing frequent maintenance and sensors' calibration. This requires a continuous control of the device state and of the sampled data quality both from the data provider and the Data Assembly Center. The goals of this work are the following:

- to have data of the best possible quality for the final users release (quality control);
- to use these data to calibrate and validate (Cal/Val) ocean circulation models both in real time and in delayed mode.

Data from 150 moorings located in the Mediterranean Sea have been downloaded from the European Copernicus Marine Environment Monitoring Service CMEMS (https://doi.org/10.25423/MEDSEA_ANALYSIS_FORECAST_PHYS_006_001) in situ TAC (Thematic Assembly Center) Med FTP server (<ftp://medinsitu.hcmr.gr/>), the data is accessible after free registration to Copernicus marine service portal) and analyzed to implement an automated secondary quality check procedure based on simple statistical properties (mean and standard deviation). Two model data from the RITMARE (<http://www.ritmare.it/>) Italian project and from CMEMS have been considered to test the Cal/Val procedure. The analyzed ocean variables are: temperature, salinity, sea level, 3D water speed.

A Python module has been set up to automatically assess the data quality of all moorings, which returns the best hourly and daily time series to be compared with any model data. In order to create a validation of the model in terms of RMSE (root mean square error) and statistical bias. The quality control procedure is divided in three phases:

1. in situ TAC quality flag application and discard of too sparse variables or stations;
2. spike removal (gross check);
3. redundant statistic quality check by computing standardized anomaly and the probability distribution (kernel density estimation).

Data are flagged as good if the standardized anomaly does not exceed a specific standard deviation threshold chosen during the calibration phase of the quality control procedure. The statistic check is then repeated iteratively in order to increase data quality. The results are promising, both daily and hourly moorings data are not anymore influenced by statistically improbable data and the model Cal/Val is now providing more reliable skill scores.