Deutsches Geodätisches Forschungsinstitut
TUM School of Engineering and Design
Technische Universität München

On the potential of mapping sea level anomalies from Copernicus Marine Service with Random Forest Regression

Marcello Passaro and Marie-Christin Juhl

Deutsches Geodätisches Forschungsinstitut, Technical University of Munich

Motivation and summary

The sea level observations from satellite altimetry are characterised by a sparse spatial and temporal coverage. For this reason, along-track data are routinely interpolated into daily grids provided by the Copernicus Marine Service (CMEMS). These are strongly smoothed in time and space and are generated using an optimal interpolation routine requiring several pre-processing steps and covariance characterisation.

In this study, we assess the potential of Random Forest Regression to estimate daily sea level anomalies. One-year-long records of along-track sea level are used to build a training dataset whose predictors are the neighbouring observations. The validation is based on the comparison against daily averages from tide gauges (from GESLA).

As an example, four time series estimated from satellite altimetry from this study (ML, blue) and CMEMS (orange) at the closest point to four tide gauges (green) are shown in Fig.3. The generated dataset is on average 10% more correlated to the tide gauge records than the commonly used product from Copernicus. The most remarkable result of the validation is that in almost all of the domain (29 tide gauges out of 32) our technique (called ML here) performs better than CMEMS. In more than half of the domain, there is at least a 5% improvement in both correlation and root mean square error (RMSE) considering the tide gauges as ground truth.

Moreover, improvements in the temporal characterisation of the sea level variability is shown by means of a coherence analysis to be spread over all subannual periods (Fig.5). While the current Copernicus daily sea level anomalies are more optimised for the detection of spatial mesoscales, we show how the methodology of this study can improve the characterisation of sea level variability, particularly in the coastal zone.

Our study fits into the use of Copernicus Marine Service data in the context of pan-European coastal zone monitoring, since this innovative machine-learning based technique is validated along the coast of the North Sea. A publication of this study is available in Ocean Dynamics at this link: https://doi.org/10.1007/s10236-023-01540-4

Methodology

The concept of our methodology is the use of along-track SLAs as truth to train the random forest regressor in the estimation of unknown SLAs (our target variable) on a set of grid points. As predictors, we use means, weighted means and standard deviations of the SLAs at different neighbourhoods in space and time. Furthermore, to better describe the evolution of the target variable in both space and time, the ratios among these predictors from the different neighbourhoods are also used as predictors, similarly to Leirvik and Yuan (2021). An example of the spatial and temporal neighbourhoods is provided in Fig.1, while Fig.2 shows example of four predictors related to different neighbourhoods.

Leirvik T., Yuan M (2021) A machine learning technique for spatial interpolation of solar radiation observations. Earth Space Sci 8(4):e2020EA001527. https://doi.org/10.1029/2020EA001527

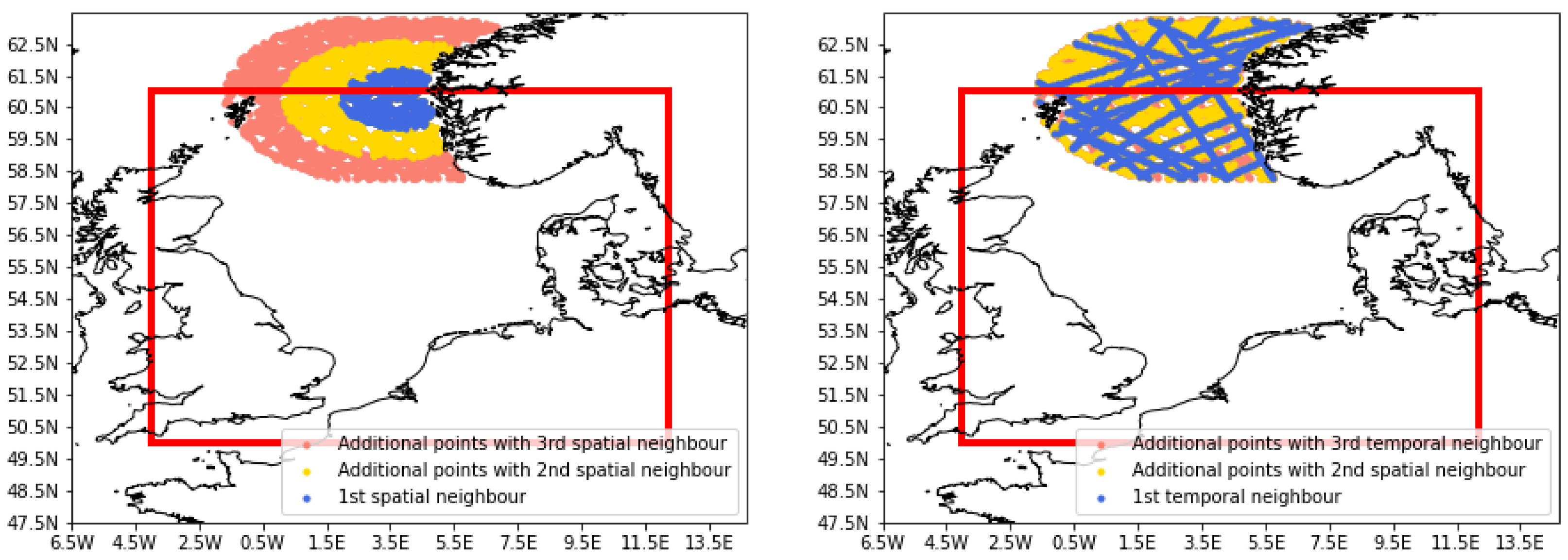

Fig. 1: Examples of along-track observations included in spatial (left) and temporal (right) neighbourhoods associated to one particular location. The red box indicates the area of study. The latter is extended in the search for neighbouring observations, in order to allow for the estimations at the domain's border.

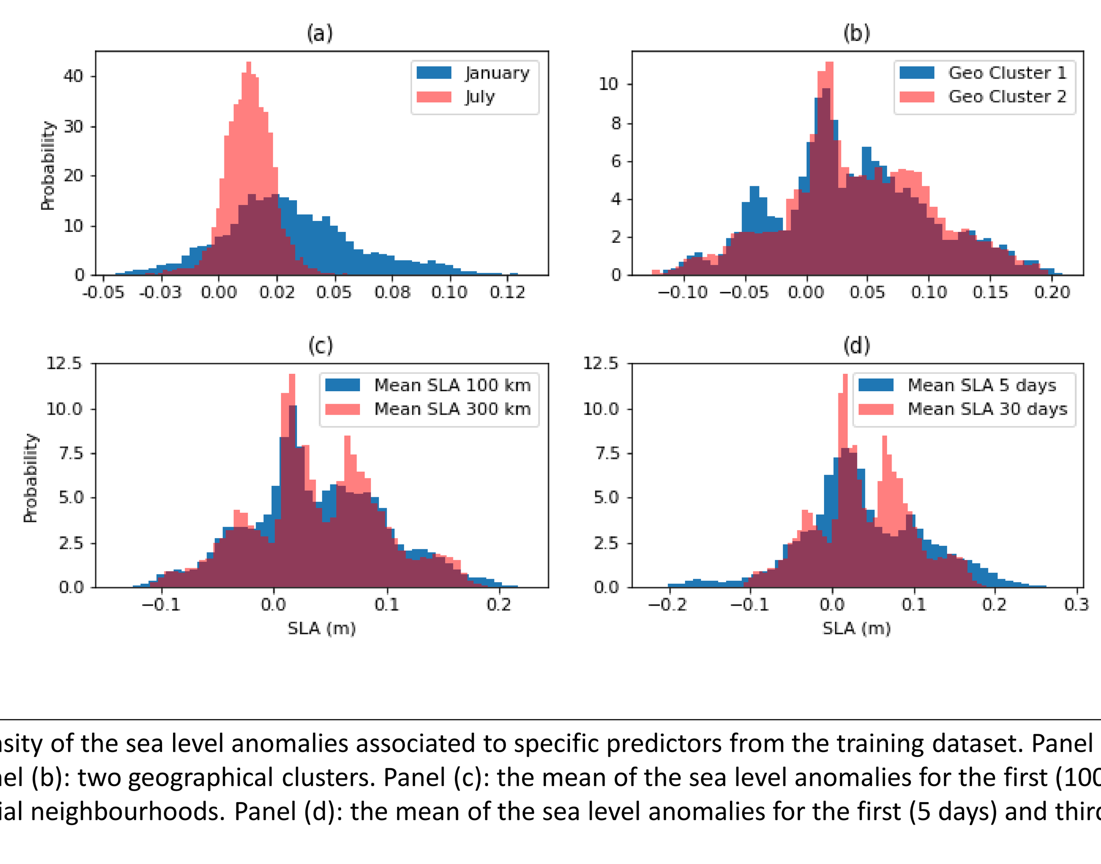

Fig.2: Probability density of the sea level anomalies associated to specific predictors from the training dataset. Panel (a): months of January and July. Panel (b): two geographical clusters. Panel (c): the mean of the sea level anomalies for the first (100-km radius) and third (300-km radius) spatial neighbourhoods. Panel (d): the mean of the sea level anomalies for the first (5 days) and third (30 days) time neighbourhoods.

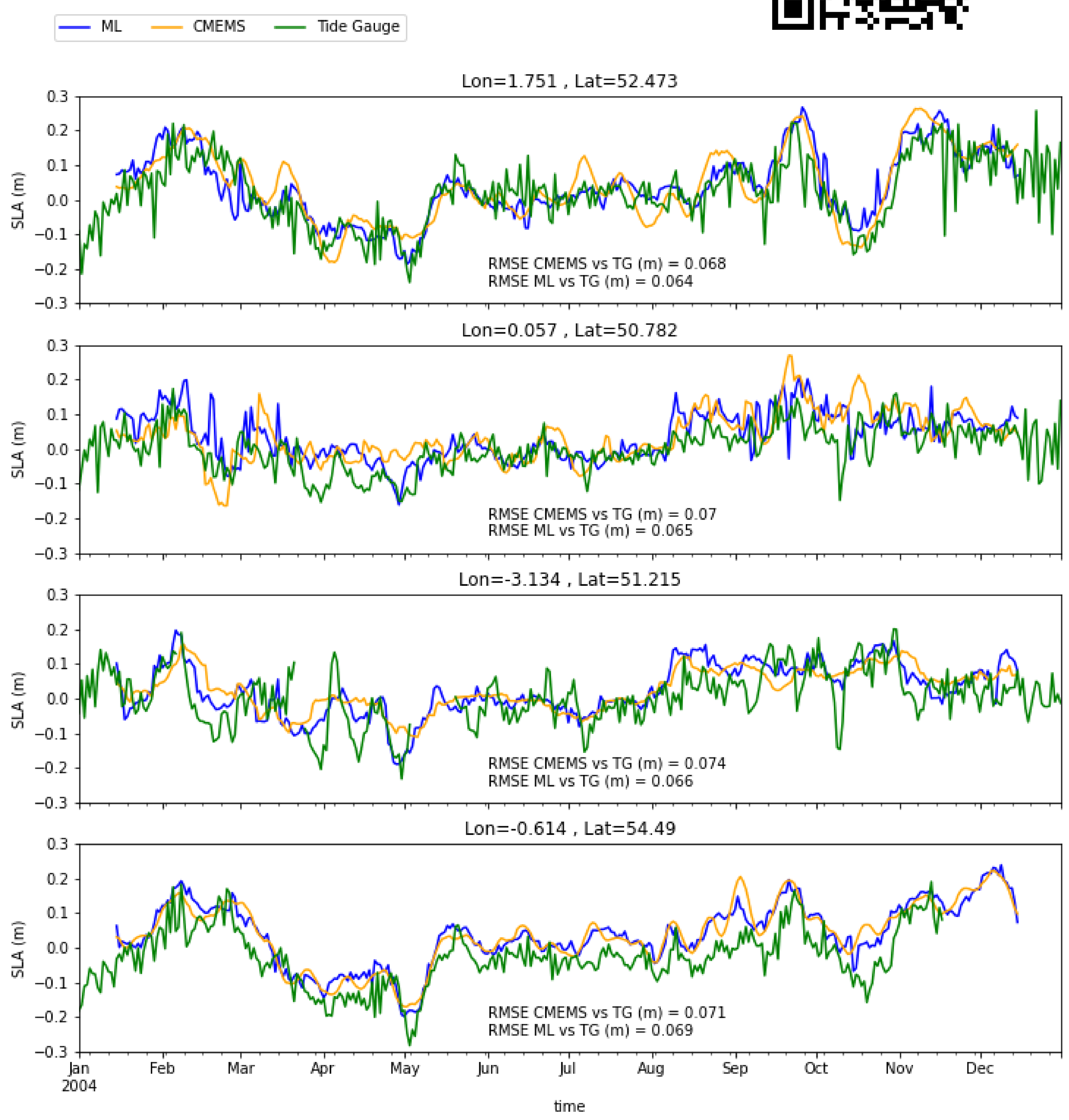

Fig.3: Time series estimated from satellite altimetry from this study (ML, blue) and CMEMS (orange) at the closest point to four tide gauges (green), whose coordinates are shown at the top of each panel. Also shown as text is the root mean square error (RMSE) of the altimetry dataset considering the tide gauges as ground truth.

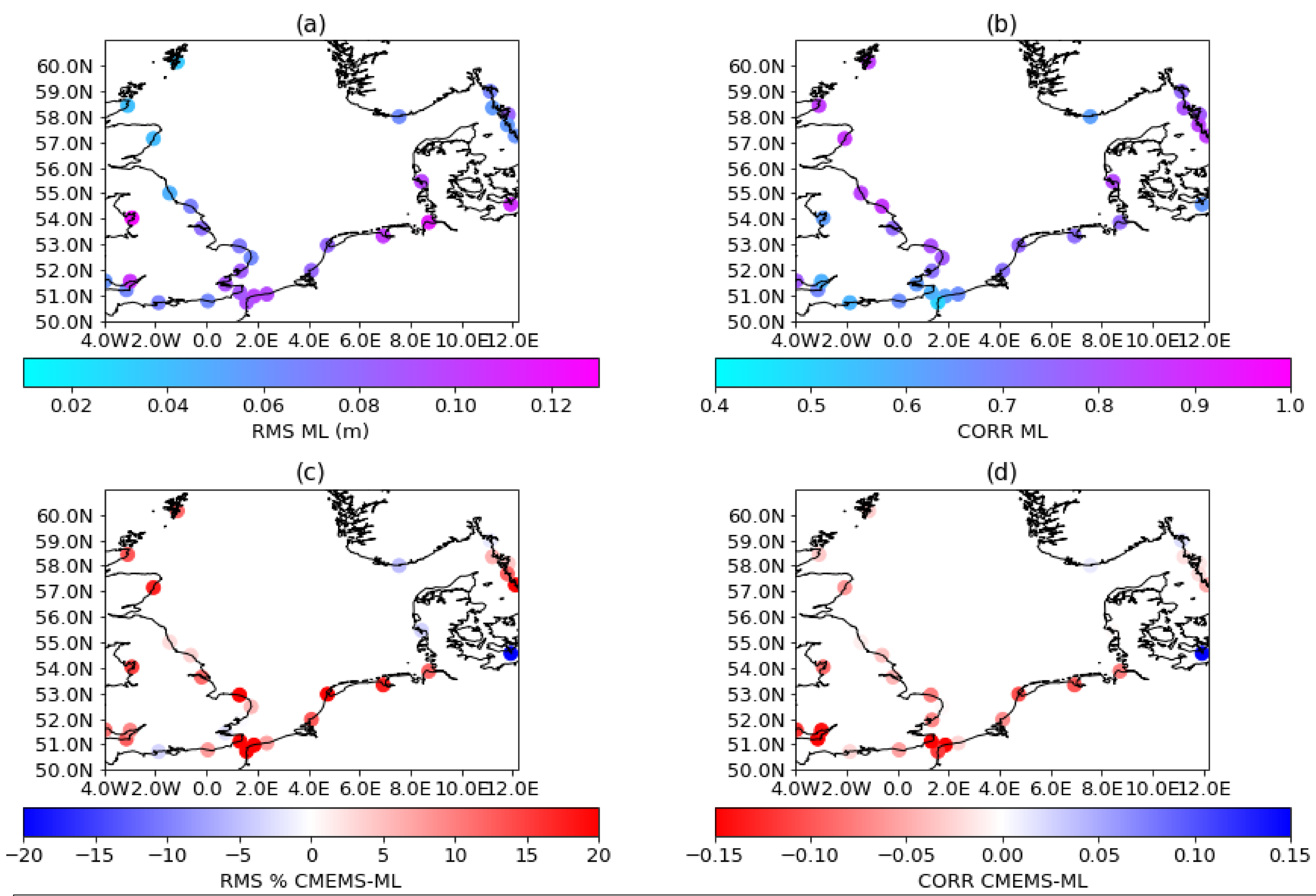

Fig. 4: Results of the validation of daily sea level anomaly maps coupled with tide gauges at the closest point. Root mean square (RMS, panel **a**) and Pearson's correlation coefficient (CORR, panel **b**) between the product of this study (ML) and the time series from the tide gauges (panel **a**). Panels **c** and **d**: Difference between these statistics and the equivalent computed using the CMEMS product, in which the red colour palette indicates an improvement using ML.

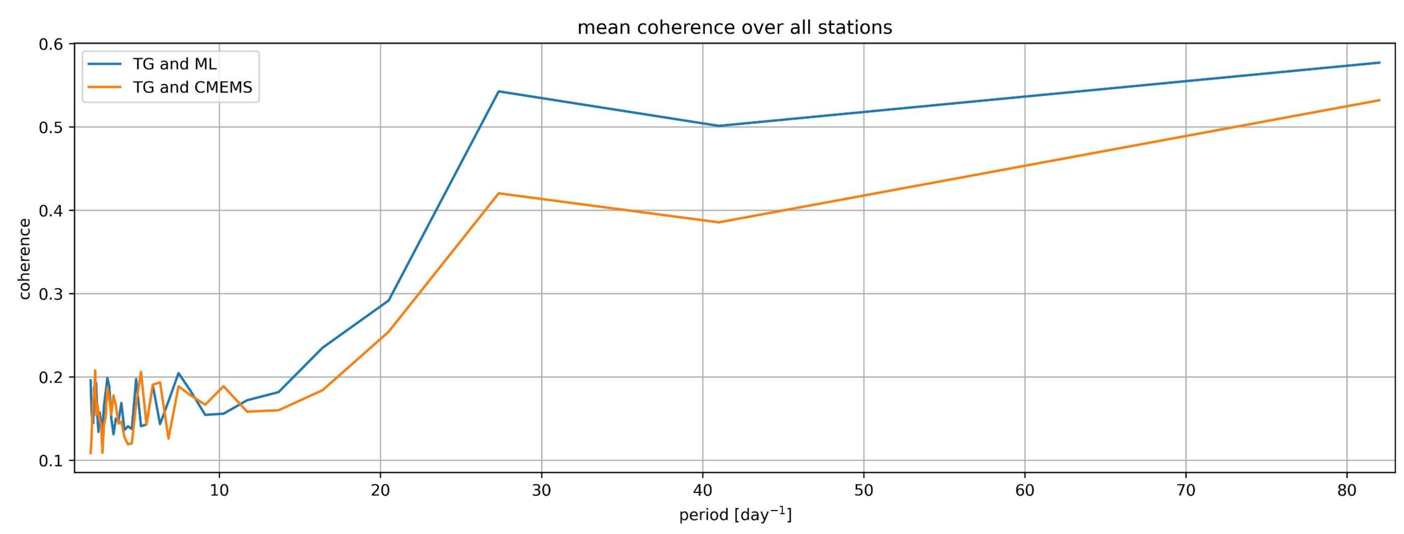

Fig. 5: Results of the validation of daily sea level anomaly maps coupled with tide gauges (TG) at the closest point: mean magnitude squared coherence for periods lower than 90 days expressing similarity between the time series from TG and CMEMS (orange), and from TG and ML (blue).

Concluding remarks

The method presented allows for a more realistic representation of the sea level variability, as verified by the comparison against coastal in situ data. Such comparison has been conducted using high-frequency tide gauges, which is in our opinion a much more realistic external validation than the use of monthly means, if the objective is to assess the capability of the altimetry constellation to observe sea level at short time scales. We have shown that there is a clear potential to establish this strategy as an operational Copernicus product on a regional scale.