A novel approach to computing super observations for probabilistic wave model validation

Patrik Bohlinger[a], Øyvind Breivik[a,c], Theodoros Economou[b], Malte Müller[a]

Contact: patrikb@met.no

Motivation

In conventional approaches, super observations of significant wave height (SWH) are based on block averages. The size of the blocks is determined by a fixed length scale motivated by physical arguments. However, the choice of the appropriate length scale can be ambiguous where errors might be propagated to the subsequent validation metrics without any information on uncertainty. Moreover, for high resolution wave modelling the block size might be too small to compute reliable averages and it is not clear how to proceed.

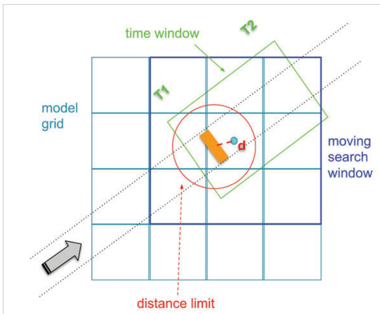
Data

➢ SWH from satellite Sentinel 3A (altimeter L3, 1 Hz): WAVE_GLO_WAV_L3_SWH_NRT_OBSERVATIONS _014_001

➢ SWH from wave model WAM (resolution 6.25 km): ARCTIC_ANALYSIS_FORECAST_WAV_002_010

➢ Collocation of satellite and wave model with constraints similar to Stopa et al. (2016).

- distance limit = 6 km
- time window = ± 30 min



➢ Collocation results in two time series that can be compared

Conventional approach

First the length scale is computed to make the represented physical scales from the wave model and the observations comparable, e.g. Zieger et al. (2009) or Abdalla et al. (2011).

For example, a model resolution of 9 km results in a length scale of 27-54 km. Knowing the distance between observations of 6.7 km, we can say that the scale of physical processes captured in the wave model corresponds to 4-8 consecutive footprints.

A super observation is then computed by averaging e.g. 7 consecutive footprints. Outliers can be identified based on the spread around this mean.

References

Abdalla, S., Janssen, P. A. E. M., Bidlot, J.-R., 2011. Altimeter near real time wind and wave products: Random error estimation. Marine Geodesy 34 (3-4), 393-406, doi: 10.1080/01490419.2011.585113.

Rasmussen, C. E., Williams, C. K., 2006. Gaussian processes for machine learning. 2006. The MIT Press, Cambridge, MA, USA 38, 715-719.

Stopa, J. E., Ardhuin, F., Girard-Ardhuin, F., 2016.Wave climate in the Arctic 1992-2014: seasonality and trends. Cryosphere 10 (4), doi: 10.5194/tc-10-1605-2016.

Zieger, S., Vinoth, J., Young, I., 2009. Joint calibration of multiplatform altimeter measurements of wind speed and wave height over the past 20 years. Journal of Atmospheric and Oceanic Technology 26 (12), 2549-2564, doi: 10.1175/2009JTECHA1303.1.

Novel approach

We formulate a Gaussian Process (GP) model (Rasmussen and Williams, 2006) to model the observational time series. Observations can thus be seen as samples from a multivariate Gaussian distribution where the covariance matrix of this distribution can be parameterized by a kernel function. The hyperparameters of the kernel function can be optimized with maximum likelihood.

By performing a parametric bootstrap we can produce distributions of synthetic observations drawn from our GP-model and compute distributions of the associated super observations. With these distributions we can express the uncertainty of having to estimate the mean function and the three hyperparameters from a data sample.

This uncertainty can be propagated to subsequent computation of validation metrics like correlation coefficient. The result is a full probabilistic statement of the model quality regarding the chosen metric.

The distribution of synthetic observation samples can further be used to see each observation in the context of its distribution and its probability to occur. Based on this information outliers can be detected following arbitrary non-stationarities in the underlying sea state caused by e.g. strong gradients due to sheltering effects.

Conclusion

The new approach is:

- ➢ highly flexible
- ➢ fully probabilistic
- ➢ no error dependence
- ➢ interpolation possible

$$
\begin{aligned}
y_t|f_t &= f_t + \epsilon_t \\
\epsilon_t &\sim N(0, \sigma_n^2) \\
y_t|f_t &\sim N(f_t, \sigma_n^2) \\
f_t &\sim N(0, \sigma_s^2) \\
\mathrm{cor}(f_t, f_{t'}) &= \exp\left(-\frac{(t-t')^2}{2l^2}\right)
\end{aligned}
$$

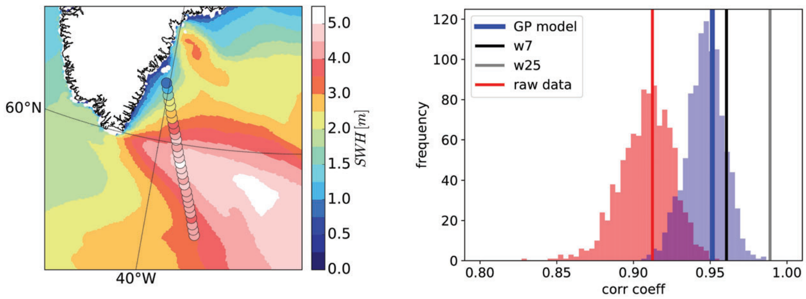
Hyperparameters

$$
\begin{aligned}
y &= (y_1, \ldots, y_n) \\
y &\sim N(0, \Sigma) \\
\Sigma &= K + \sigma_n^2 I \\
y_t, y_{t'} &= \sigma_s^2 \exp\left(-\frac{(t-t')^2}{2l^2}\right)
\end{aligned}
$$

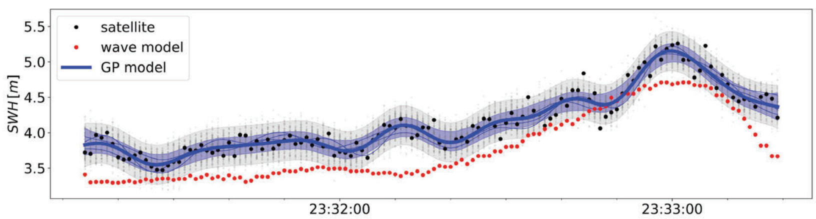
Multivariate Normal

Showcase

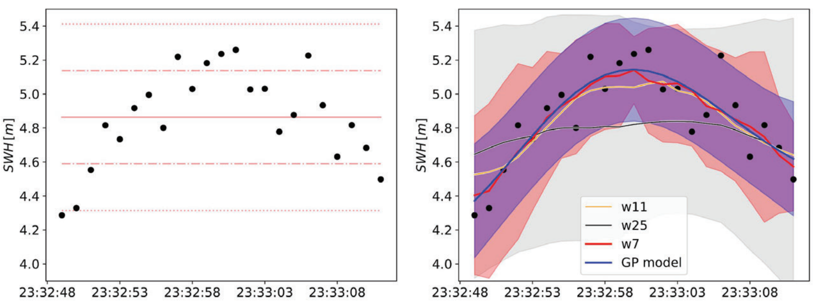
2018-05-02 00:00:00 UTC



b) Latent mean function, f-samples, y-samples, and uncertainty bounds



a) Block Estimate **b) Various Smoothing**



Meteorologisk institutt

[a]Norwegian Meteorological Institute, Norway
[b]University of Exeter, UK
[c]University of Bergen, Norway