Guidelines for triple collocation validation of significant wave height measurements: application to satellite altimeters, in-situ data, and the new NORA3 hindcast

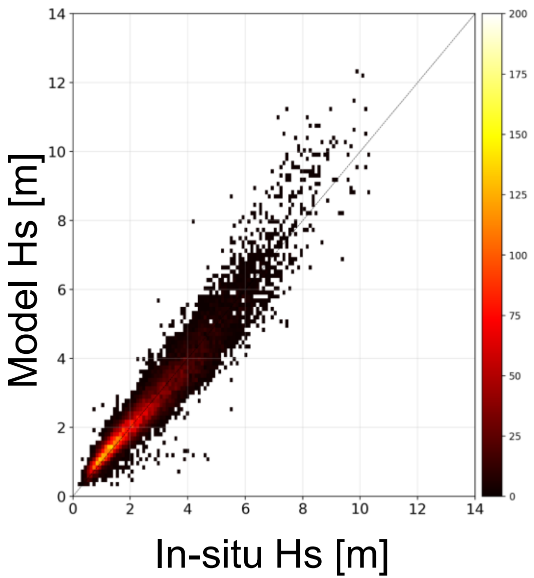
Fabien Collas[1,2], Patrik Bohlinger[1], Ad Stoffelen[3], Øyvind Breivik[1,2]

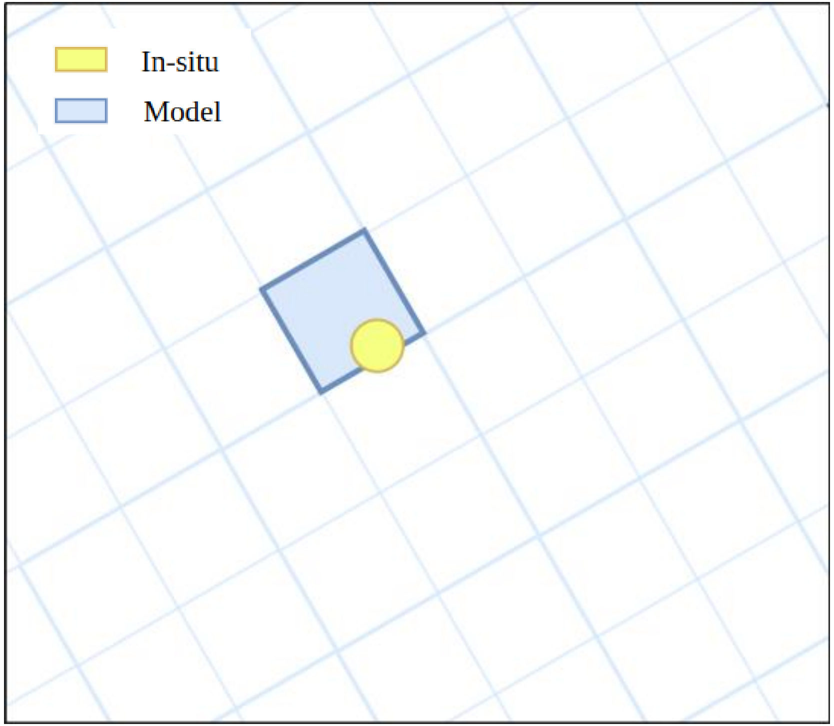
[1]Norwegian Meteorological Institute, Bergen, Norway, [2]University of Bergen, Bergen, Norway, [3]KNMI, De Bilt, Netherlands

Norwegian Meteorological Institute

Motivation

$$X_{Model} = X_{Insitu} + \epsilon$$

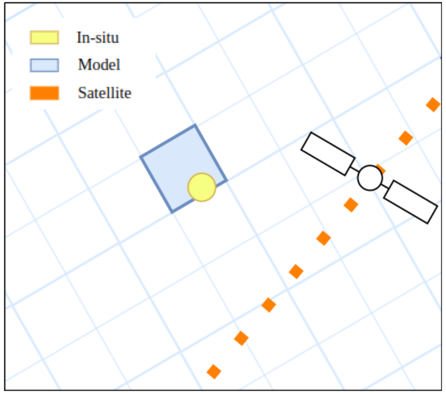




$$RMSE(X_{Model}) = \sqrt{< (X_{Model} - X_{Insitu})^2 >} = \sqrt{< \epsilon^2 >}$$

Triple collocation (Stoffelen, 1998)

- 3 independent collocated measurements (*x, y, z*)
- Common truth *t*



Error model

$$x = t + \epsilon_x$$
$$y = a_y t + b_y + \epsilon_y$$
$$z = a_z t + b_z + \epsilon_z$$

for $i = x, y, z$

Assumptions

- Unbiased errors:
  $$< \epsilon_i >= 0$$
- Errors independent of common truth:
  $$< \epsilon_i t >=< \epsilon_i >< t >= 0$$
- No cross-correlation for errors:
  $$< \epsilon_x \epsilon_y >=< \epsilon_x \epsilon_z >=< \epsilon_y \epsilon_z >= 0$$
- Error variances constant over range of values considered (homoscedasticity)

Result: variance of errors

$$< \epsilon_x^2 > = C_{xx} - \frac{C_{xy} C_{xz}}{C_{yz}}$$
$$< \epsilon_y^2 > = C_{yy} - \frac{C_{yx} C_{yz}}{C_{xz}}$$
$$< \epsilon_z^2 > = C_{zz} - \frac{C_{zx} C_{zy}}{C_{xy}}$$

C_{ij} : covariance between measurements *i* and *j*.

Using formulation as in Dodet et al., 2022

Calibration

$$x = t + \epsilon_x$$

$$y = \boxed{a_y} t + b_y + \epsilon_y$$

$$z = \boxed{a_z} t + b_z + \epsilon_z$$

$$\boxed{\begin{aligned} a_y &= \frac{C_{yz}}{C_{xz}} \\ a_z &= \frac{C_{yz}}{C_{xy}} \end{aligned}}$$

$$\boxed{\begin{aligned} y^x &= t + \frac{\epsilon_y}{a_y} \\ z^x &= t + \frac{\epsilon_z}{a_z} \end{aligned}}$$

$$RMSE(x) = \sqrt{< (x-t)^2 >} = \sqrt{< \epsilon_x^2 >}$$

$$RMSE(y^x) = \sqrt{< (y^x-t)^2 >} = \frac{\sqrt{< \epsilon_y^2 >}}{a_y}$$

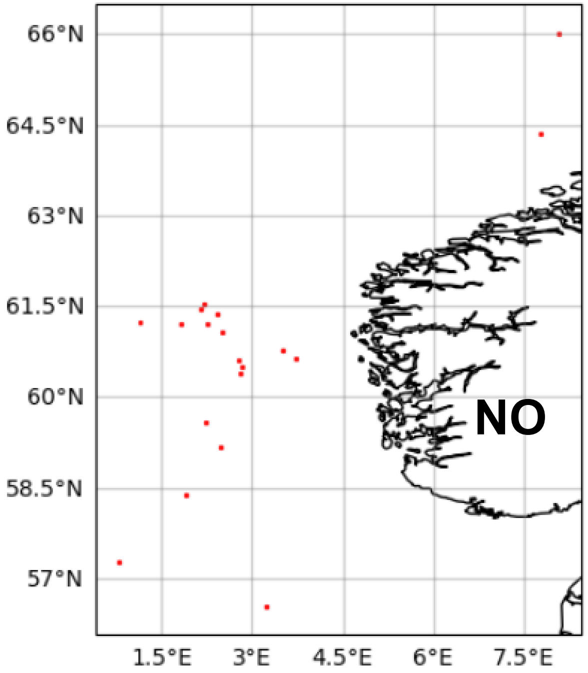
$$RMSE(z^x) = \sqrt{< (z^x-t)^2 >} = \frac{\sqrt{< \epsilon_z^2 >}}{a_z}$$

Significant wave height (Hs) measurements

- NORA3 wave model hindcast - 3 km resolution (Breivik et al., 2022)
  Regridded to 0.5°x0.5° resolution

- Altimeter data from the ESA Sea State CCI v3 dataset (Piollé et al., 2022)

- CMEMS In-situ TAC - Near Real Time data (http://www.marineinsitu.eu/)



In-situ selection (18 stations):

- Distance to coast > 50 km

- Depth > 50m (https://erddap.emodnet.eu/)

- 2014-2021

Error model and assumptions

$$x = \boxed{t} + \epsilon_x$$
$$y = a_y \boxed{t} + b_y + \epsilon_y$$
$$z = a_z \boxed{t} + b_z + \epsilon_z$$

Common truth t

for $i = x, y, z$

- Unbiased errors:

$$< \epsilon_i >= 0$$

- Errors independent of common truth:

$$< \epsilon_i t >=< \epsilon_i >< t >= 0$$

- No cross-correlation for errors:

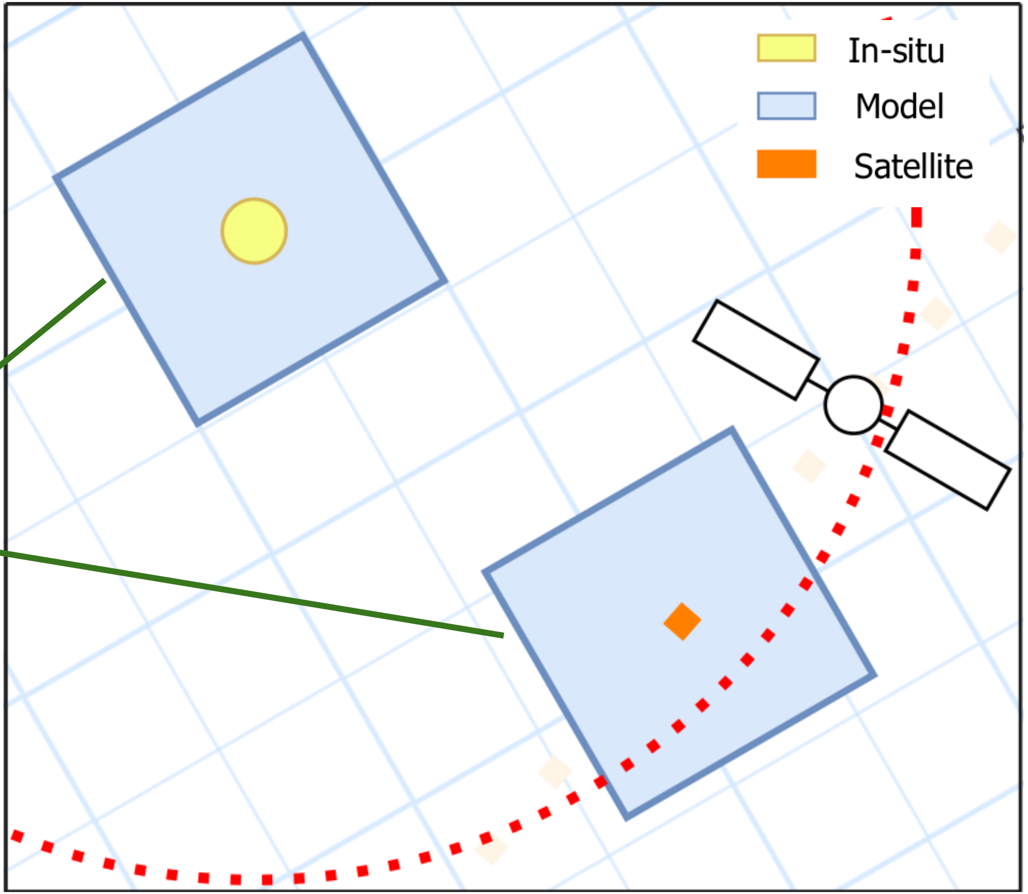
$$< \epsilon_x \epsilon_y >=< \epsilon_x \epsilon_z >=< \epsilon_y \epsilon_z >= 0$$

- Error variances constant over range of values considered (homoscedasticity)

Collocation

“Dynamic collocation”

Keep the triplets only if relative difference < 5%



Janssen et al., 2007, Dodet et al., 2025

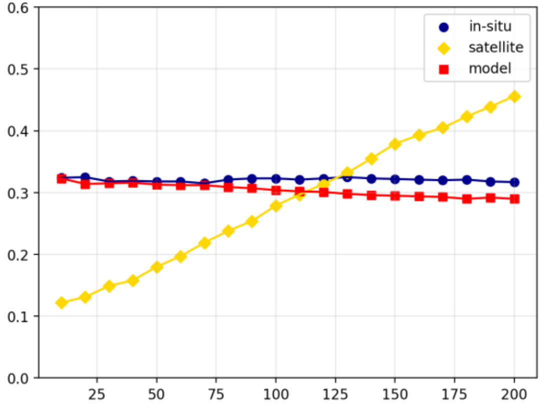
Collocation distance and filter

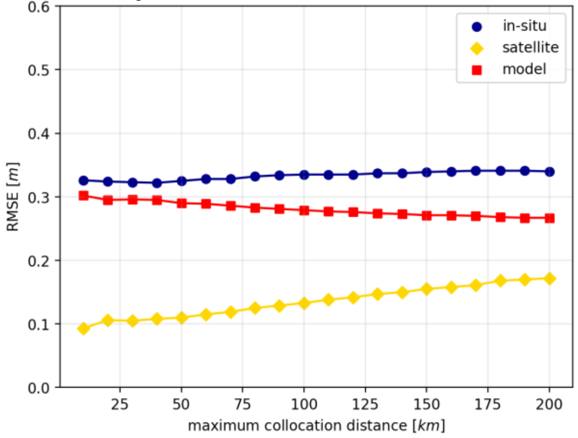
No filter

Dynamic collocation

Model at in-situ location

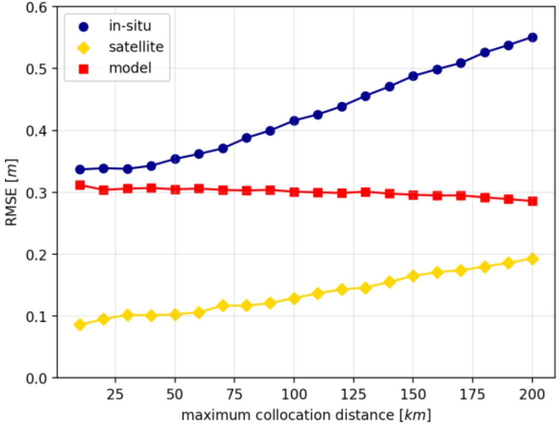
RMSE [m]

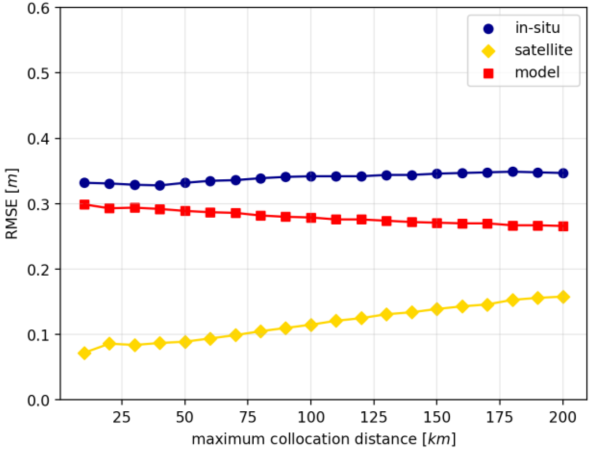




Maximum collocation distance [km]

Model at satellite location





Error model and assumptions

$$x = t + \epsilon_x$$

$$y = a_y t + b_y + \epsilon_y$$

$$z = a_z t + b_z + \epsilon_z$$

for $i = x, y, z$

- Unbiased errors:

$$< \epsilon_i >= 0$$

- Errors independent of common truth:

$$< \epsilon_i t >=< \epsilon_i >< t >= 0$$

- No cross-correlation for errors:

$$< \epsilon_x \epsilon_y >=< \epsilon_x \epsilon_z >=< \epsilon_y \epsilon_z >= 0$$

- Error variances constant over range of values considered (homoscedasticity)

Cross-correlation error

If $x, y >> z$:

$$\boxed{r^2} = < \epsilon_x, \epsilon_y > \neq 0$$

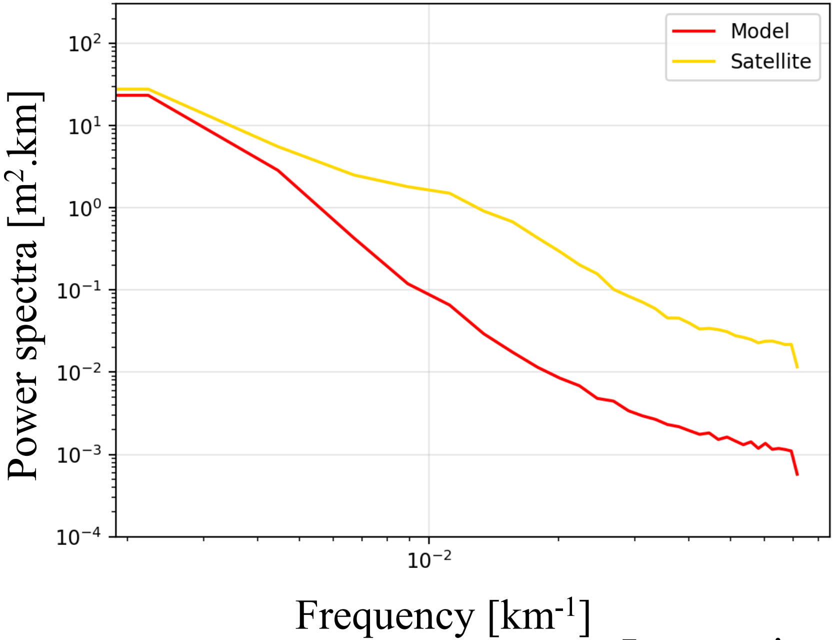
Solutions:

- Coarsen x and y - using “super observations” (Abdalla et al., 2011)
- Estimate the cross-correlation r^2 (Vogelzang et al., 2011)

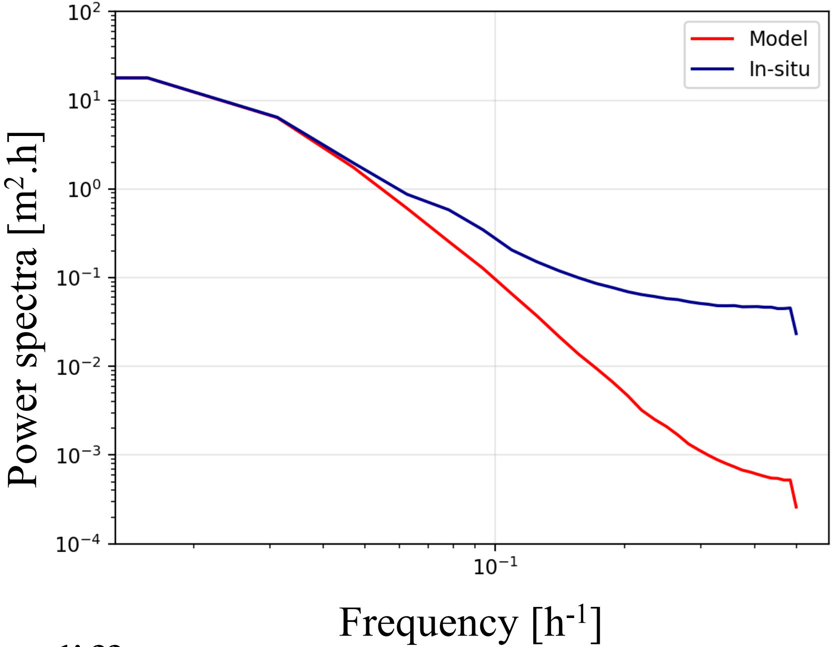
$$< \epsilon_x^2 > = C_{xx} - C_{xy} + \boxed{r^2}$$
$$< \epsilon_y^2 > = C_{yy} - C_{xy} + \boxed{r^2}$$
$$< \epsilon_z^2 > = C_{zz} - C_{xz}$$

Estimate $r^2 = <\varepsilon_{in\text{-}situ}, \varepsilon_{satellite}>$

Power spectra collocated satellite and model



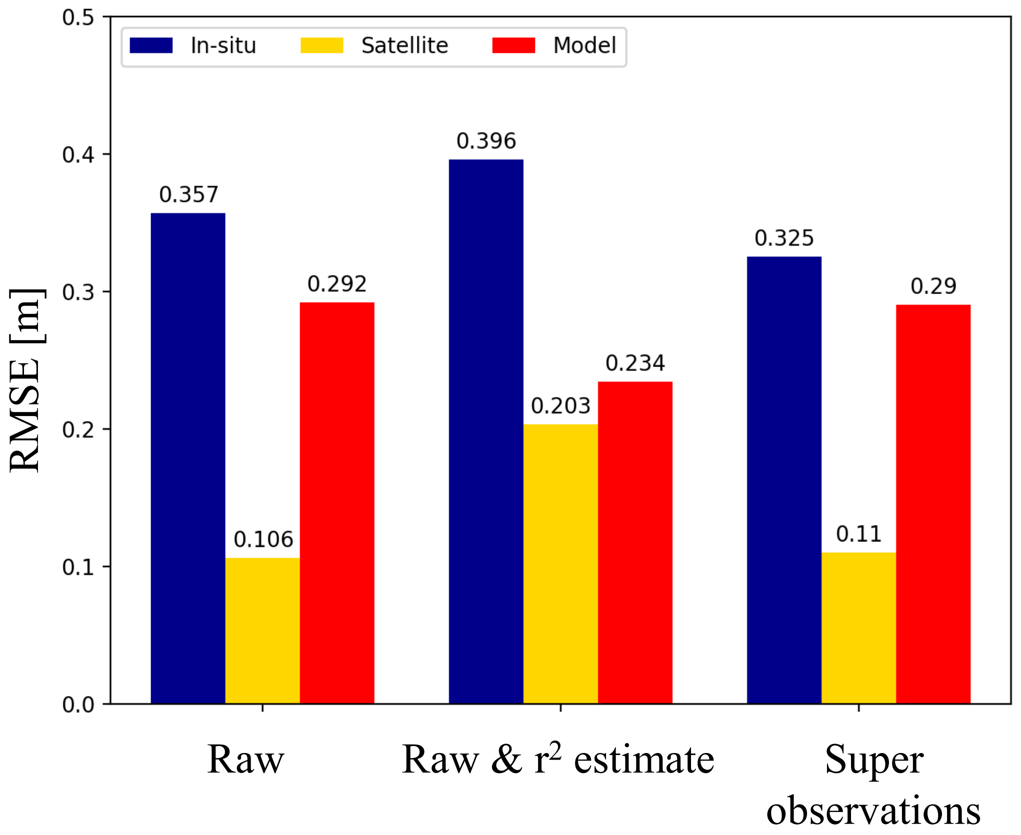
Power spectra collocated in-situ and model



Integrating the difference

$r^2 \sim 0.03\ m^2$

Differences in results with different approaches



Error model and assumptions

$$x = t + \epsilon_x$$

$$y = a_y t + b_y + \epsilon_y$$

$$z = a_z t + b_z + \epsilon_z$$

for $i = x, y, z$

- Unbiased errors:

$$< \epsilon_i >= 0$$

- Errors independent of common truth:

$$< \epsilon_i t >=< \epsilon_i >< t >= 0$$

- No cross-correlation for errors:

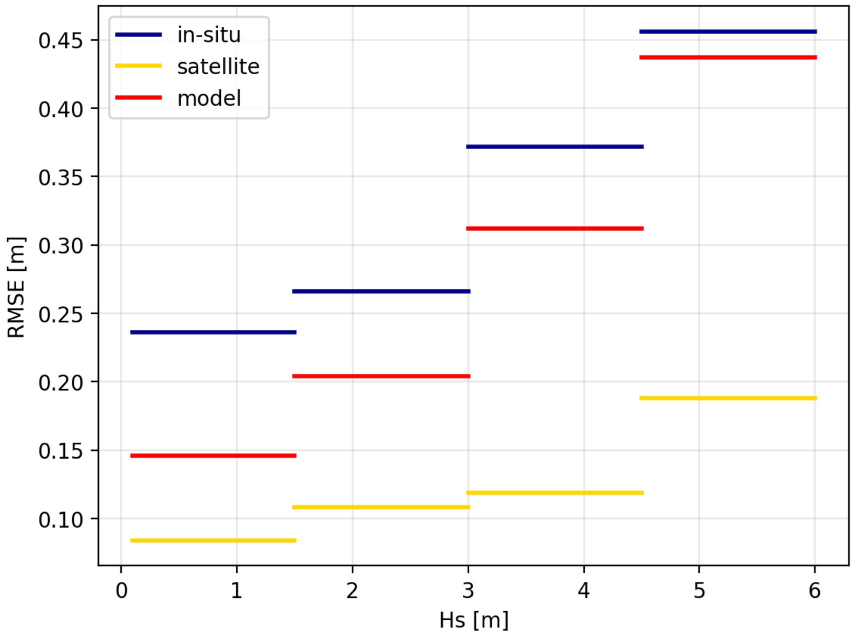
$$< \epsilon_x \epsilon_y >=< \epsilon_x \epsilon_z >=< \epsilon_y \epsilon_z >= 0$$

- Error variances constant over range of values considered (homoscedasticity)

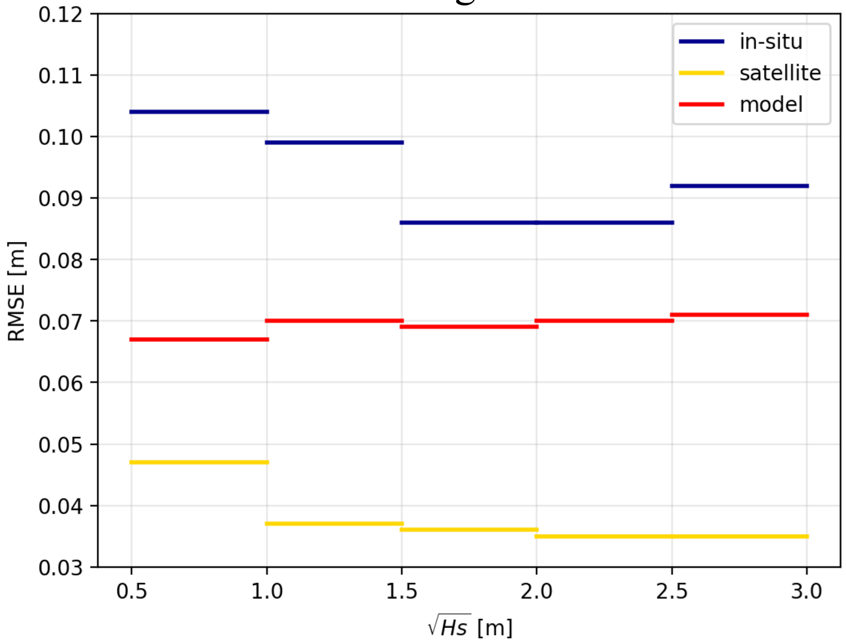
Independence with common truth *t* & constant variance

Solution? Apply triple collocation to √**Hs**?

Triple collocation results for different ranges of Hs values



Triple collocation on √Hs results for different ranges of √Hs values



See also Dodet et al., 2022

See Alemohammad et al., 2015

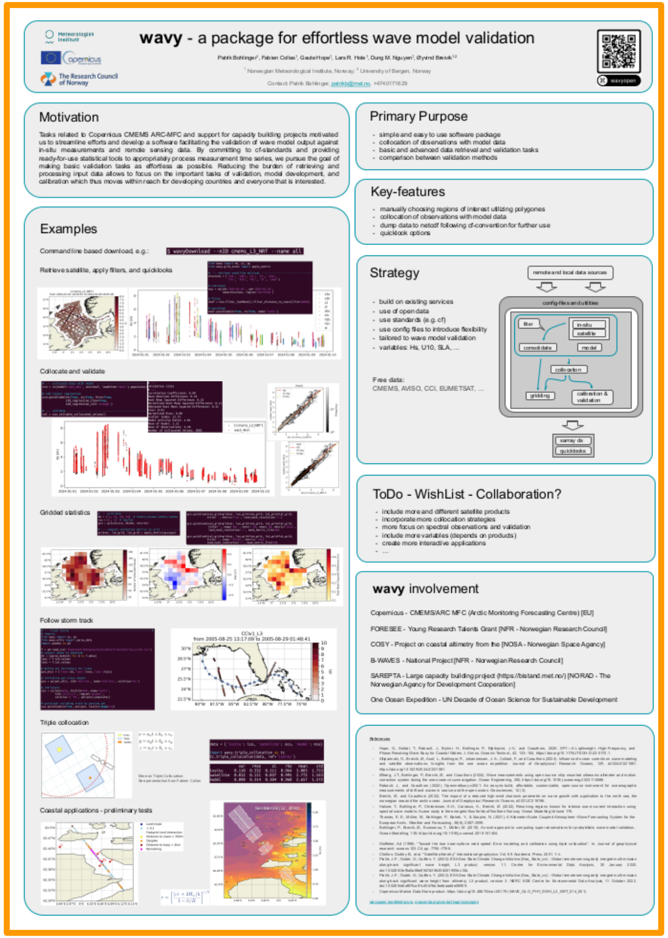
Wavy open-source Python package

Thursday afternoon - Room 1 - Patrik Bohlinger

```python
data = {
    'in-situ':insitu,
    'satellite':satellite,
    'model_is':model_is,
    'model_sat':model_sat
      }

data = tc.filter_collocation_distance(data,
                                      dist_max=50,
                                      name='satellite')
data = {k:data[k].vars.Hs.values for k in data.keys()}

data = tc.filter_dynamic_collocation(data, mod_1='model_is',
                                     mod_2='model_sat',
                                     max_rel_diff=0.05)

data = tc.calibration_triplets_tc(data, 'satellite')

tc.triple_collocation(data, ref='satellite')
```

	var	rmse	si	rho	mean	std
in-situ	0.106	0.326	0.125	0.954	2.604	1.514
satellite	0.012	0.111	0.042	0.994	2.604	1.483
model_is	0.083	0.288	0.111	0.963	2.604	1.506

wavyopen



Bohlinger et al., 2019

References

- Stoffelen, Ad. (1998). Error Modeling and Calibration; Towards the true surface wind speed. *J. Geophys. Res.*. 103. 7,755-7,766. 10.1029/97JC03180.
- Dodet, G., S. Abdalla, M. Alday, M. Accensi, J. Bidlot, and F. Ardhuin, 2022: Error Characterization off Significant Wave Heights in Multidecadal Satellite Altimeter Product, Model Hindcast, and In Situ Measurements Using the Triple Collocation Technique. *J. Atmos. Oceanic Technol.*, **39**, 887–901, https://doi.org/10.1175/JTECH-D-21-0179.1.
- Janssen, Peter AEM et al. (2007). “Error estimation of buoy, satellite, and model wave height data”. In: Journal of Atmospheric and Oceanic Technology 24.9, pp. 1665–1677.
- Zwieback, S., Scipal, K., Dorigo, W., and Wagner, W.: Structural and statistical properties of the collocation technique for error characterization, *Nonlin. Processes Geophys., 19, 69–80*, https://doi.org/10.5194/npg-19-69-2012, 2012.
- Breivik, Ø., Carrasco, A., Haakenstad, H., Aarnes, O. J., Behrens, A., Bidlot, J.-R., Björkqvist, J.-V., Bohlinger, P., Furevik, B. R., Staneva, J., and Reistad, M.(2022). The impact of a reduced high-wind Charnock parameter on wave growth with application to the North Sea, the Norwegian Sea, and the Arctic Ocean. *Journal of Geophysical Research: Oceans, 127, e2021JC018196*. https://doi.org/10.1029/2021JC018196
- Piollé, J.-F.; Dodet, G.; Quilfen, Y. (2022): ESA Sea State Climate Change Initiative (Sea_State_cci) : Global remote sensing daily merged multi-mission along-track significant wave height from altimetry, L3 product, version 3. NERC EDS Centre for Environmental Data Analysis, *11 October 2022*. doi:10.5285/e6af67fca81c40b7bb3eddaadde06909.
- Dodet, Guillaume, et al. "Impact of altimeter-buoy data-pairing methods on the validation of Sentinel-3A coastal significant wave heights." *Remote Sensing of Environment* 316 (2025): 114483.
- Bohlinger P., Breivik Ø., Economou T., Müller M., A novel approach to computing super observations for probabilistic wave model validation, Ocean Model., 139 (2019), Article 101404, https://doi.org/10.1016/j.ocemod.2019.101404
- Vogelzang, J. and A. Stoffelen (2012). Tech. Rep. NWPSAF-KN-TR-021: Triple Collocation. *Tech. rep. NWPSAF, 24 pp.*
- S. Abdalla, P. A. E. M. Janssen, J-R. Bidlot (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
- A. Gruber, C.-H. Su, S. Zwieback, W. Crow, W. Dorigo, W. Wagner (2016), Recent advances in (soil moisture) triple collocation analysis, *International Journal of Applied Earth Observation and Geoinformation Volume 45, Part B, March 2016, Pages 200-211,* https://doi.org/10.1016/j.jag.2015.09.002
- Alemohammad, Seyed Hamed, et al. "Characterization of precipitation product errors across the United States using multiplicative triple collocation." *Hydrology and Earth System Sciences* 19.8 (2015): 3489-3503.

Calibration

Using here formulation from Gruber et al., 2016

$$y^x = \frac{y - < y >}{a_y} + < x >$$
$$= t + \frac{\epsilon_y}{a_y}$$

$$z^x = \frac{z - < z >}{a_z} + < x >$$
$$= t + \frac{\epsilon_z}{a_z}$$

Where

$$\boxed{\begin{aligned} a_y &= \frac{C_{yz}}{C_{xz}} \\ a_z &= \frac{C_{yz}}{C_{xy}} \end{aligned}}$$

$$RMSE(x) = \sqrt{< (x-t)^2 >} = \sqrt{< \epsilon_x^2 >}$$

$$RMSE(y^x) = \sqrt{< (y^x-t)^2 >} = \frac{\sqrt{< \epsilon_y^2 >}}{a_y}$$

$$RMSE(z^x) = \sqrt{< (z^x-t)^2 >} = \frac{\sqrt{< \epsilon_z^2 >}}{a_z}$$

Super observations

- ***Altimeter data:*** approximately 6 km spatial resolution (*Piollé et al., 2022*)

  9 observations ~ 54 km

- ***In-situ data:*** 1 measure every 10 minutes

  Smooth over a time window according to average wave group velocity c_g *(Abdalla et al., 2011)*

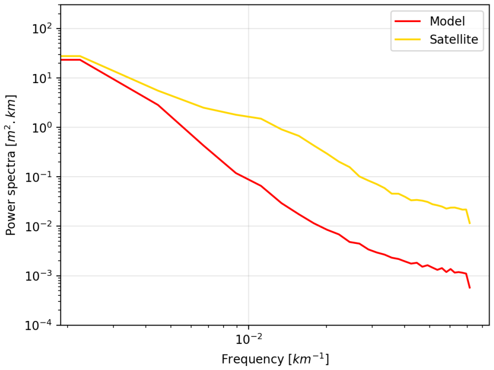
  17 observations/170 minutes ~ 50km
  with $c_g \simeq 4.9$ m.s^{-1}

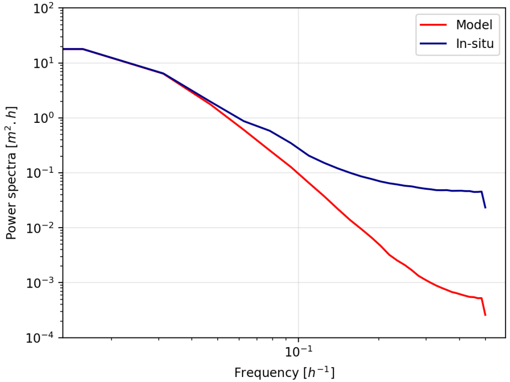
Meteorologisk institutt

Power spectra collocated satellite and model

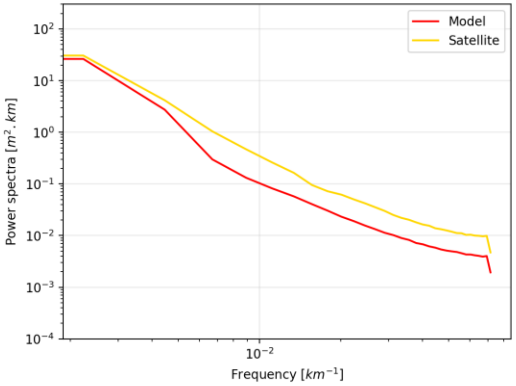
Power spectra collocated in-situ and model

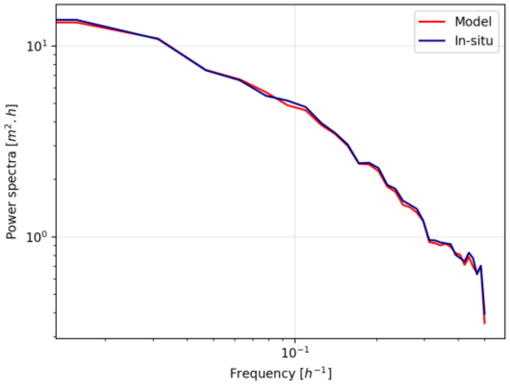
Raw data





super-observations



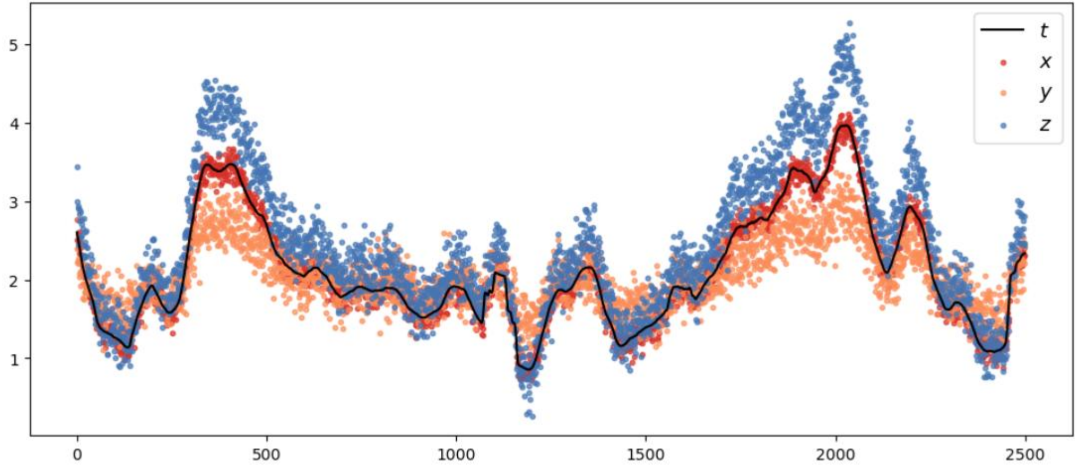


Implementation in wavy - example with simulated data

$$x = t + \epsilon_x$$
$$y = 0.5t + 1 + \epsilon_y$$
$$z = 1.3t - 0.3 + \epsilon_z$$

Where

$$\epsilon_x \sim \mathcal{N}(0, \boxed{0.01})$$
$$\epsilon_y \sim \mathcal{N}(0, \boxed{0.04})$$
$$\epsilon_z \sim \mathcal{N}(0, \boxed{0.04})$$



Wavy output:

	X	Y	Z
var est	0.0097	0.04	0.0405
RMSE	0.0984	0.2	0.2012
SI	4.6578	9.4645	9.5208
rho	0.9812	0.7592	0.9548
mean	2.1135	2.0539	2.4451
std	0.7173	0.4076	0.946

The reference for the SI is: X

Meteorologisk institutt

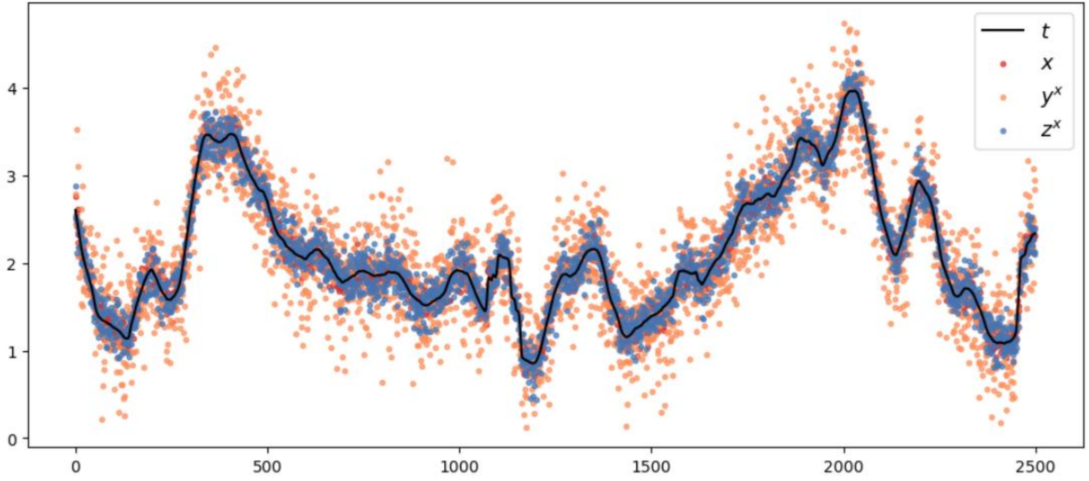
Implementation in wavy - example with simulated data

Now calibrating the data beforehand

$$< \epsilon_x^2 > = 0.1^2 = \boxed{0.01}$$

$$\frac{< \epsilon_y^2 >}{a_y^2} = \frac{0.2^2}{0.5^2} = \boxed{0.16}$$

$$\frac{< \epsilon_z^2 >}{a_z^2} = \frac{0.2^2}{1.3^2} = \boxed{0.024}$$



Wavy output:

	X	Y_X	Z_X
var_est	0.0097	0.1602	0.0239
RMSE	0.0984	0.4002	0.1547
SI	4.6578	18.9362	7.3186
rho	0.9812	0.7592	0.9548
mean	2.1135	2.1135	2.1135
std	0.7173	0.8155	0.7272

The reference for the SI is: X