Random Error Estimation of Sentinel-3 and Jason-3 Wind & Wave Data: Initial Efforts

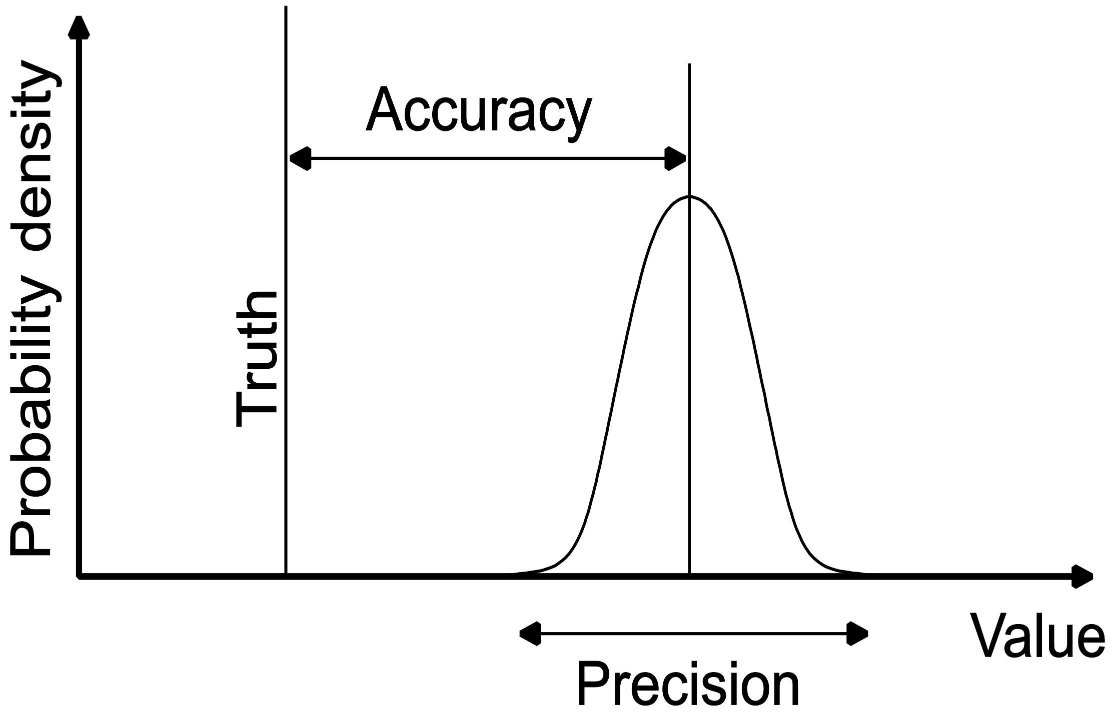
OSTST 2018, Ponta Delgada, Azores, Portugal, 27-29 Sep. 2018

Saleh Abdalla & Giovanna De Chiara

Saleh.Abdalla (AT) ecmwf.int

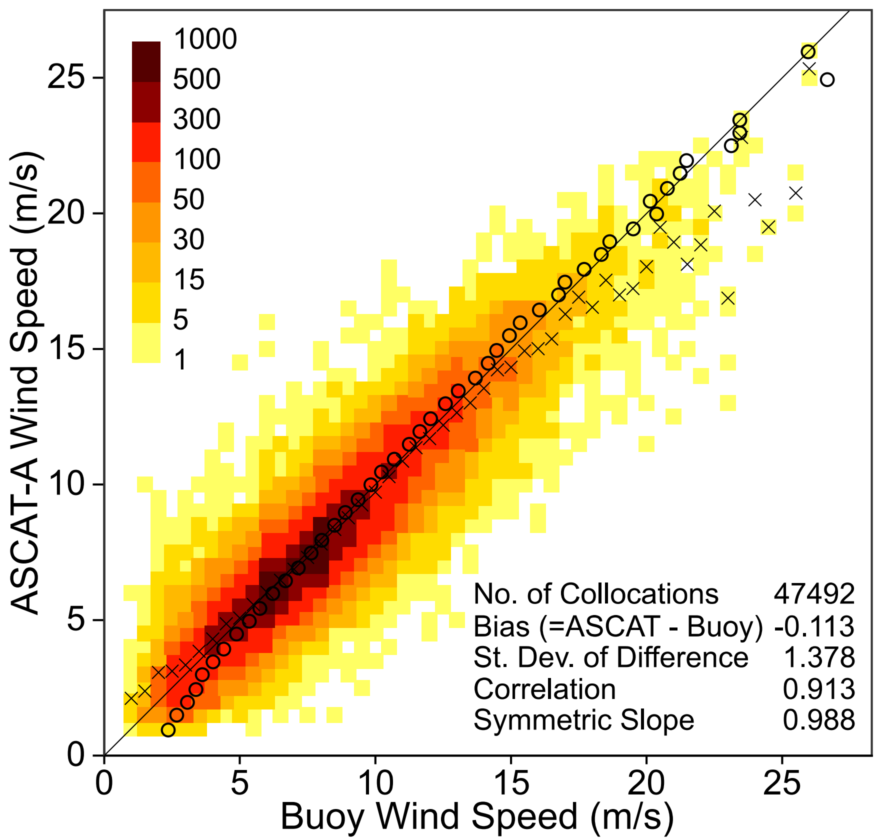
Introduction – Errors in the Measurements

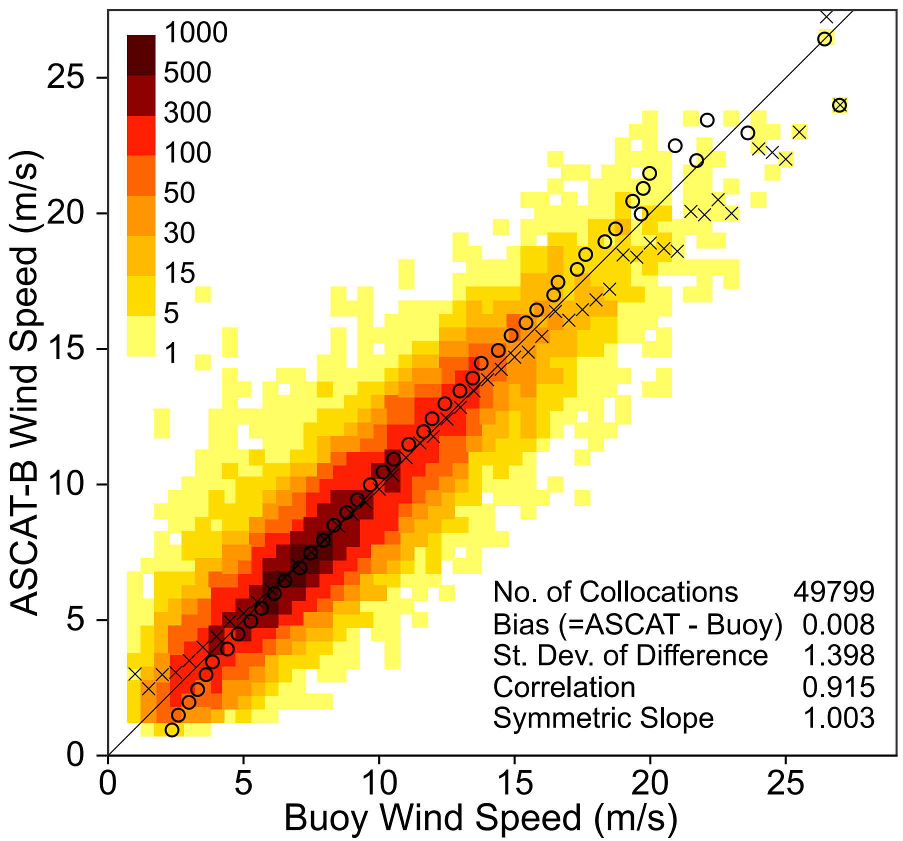
- Error = Measurement – Truth
- Truth is usually unknown.
- Statistical description:
  - Systematic error ➔ bias or mean difference.
  - Random error ➔ variance or standard deviation.
- Bias cannot be found in absolute sense.
  A reference is required.
  (will not be considered here.)



- Traditionally, estimation of the random error is done against a reference (not the truth).
  For example comparison of scatterometer wind speed against in-situ measurements.

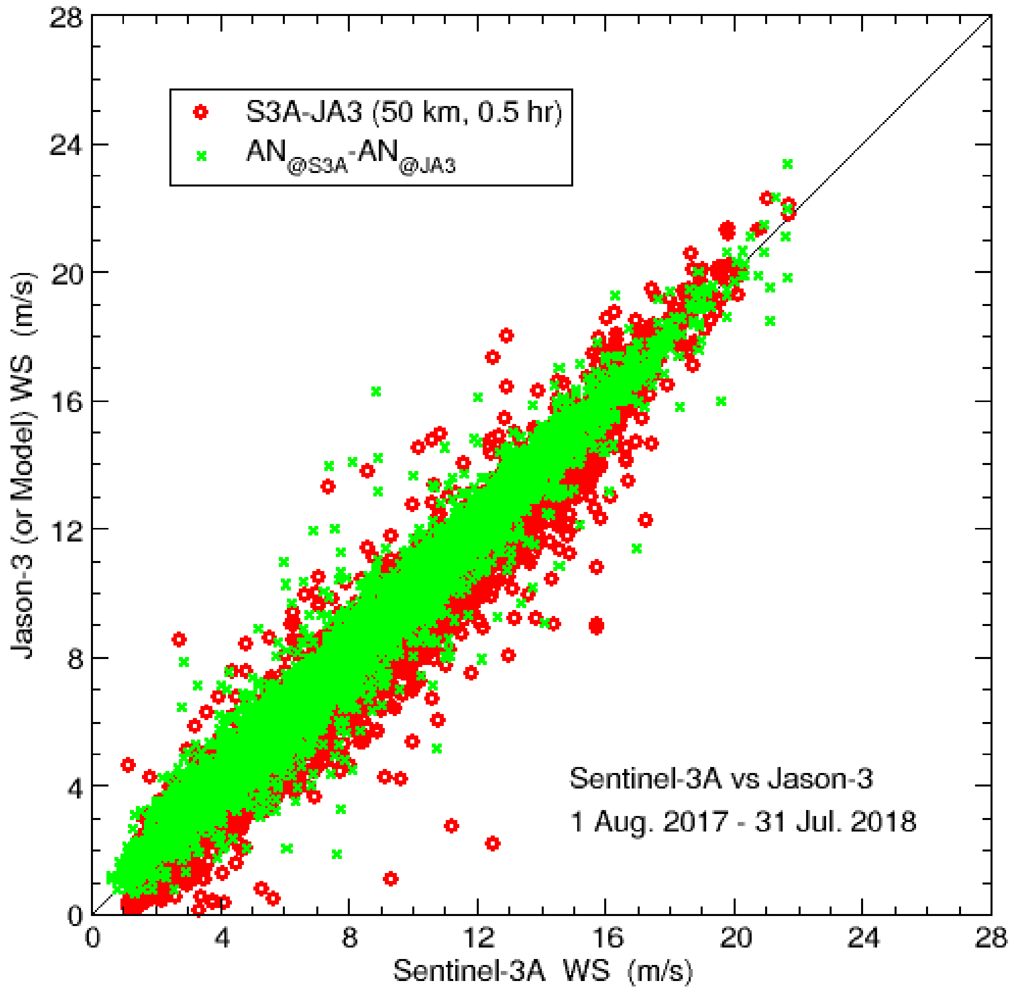
Comparison between ASCAT-A (left) and ASCAT-B (right) against in-situ surface wind speed measurements (1 August 2013 - 31 July 2015)

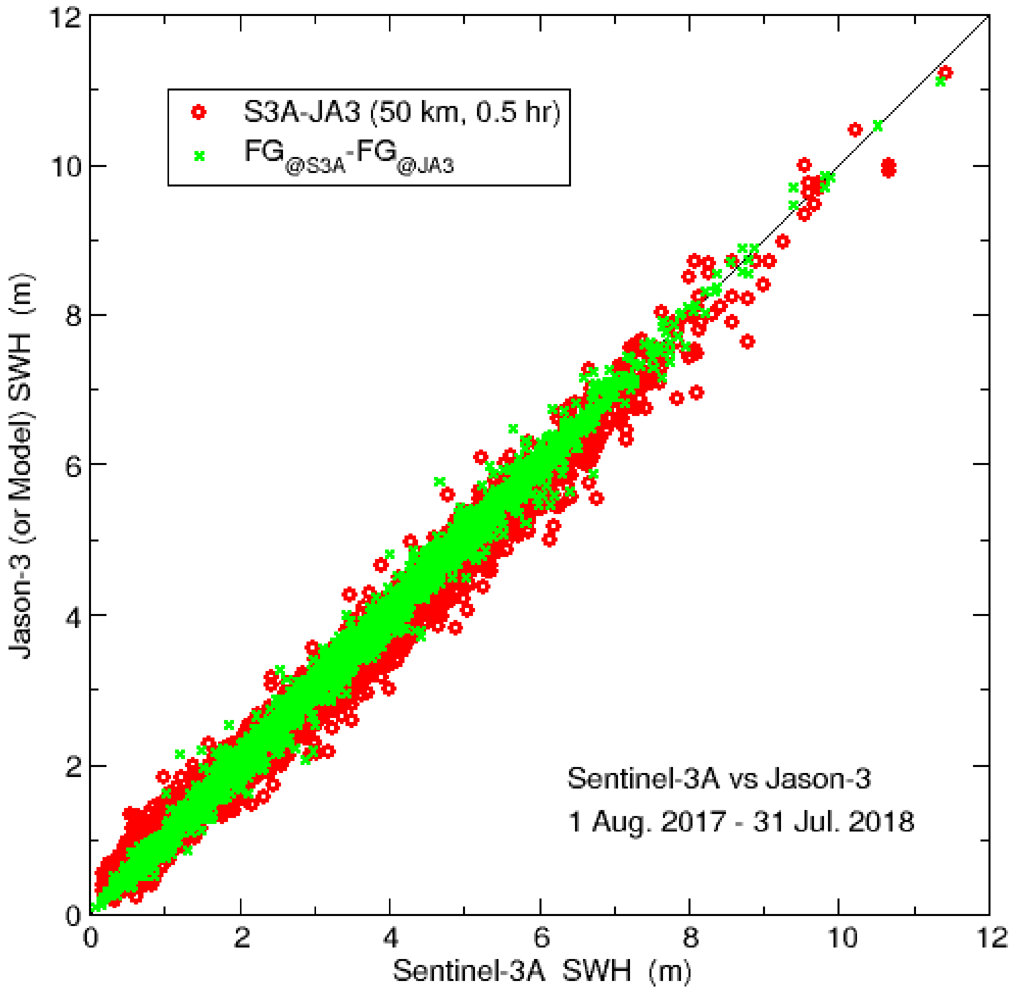




ECMWF

Comparison between Sentinel-3A & Jason-3 Wind Speed and SWH at cross overs





Error Estimation – Introduction

- For two systems (*X* and *Y*) measuring the same truth at the same location and time; it is assumed that:

$$\text{Error Variance} = N^{-1} \sum (X_i - Y_i)^2 - \text{Bias}^2$$

- But this is just the "*difference*" not the "*error*" unless system *Y* is "error-free" (which is highly unlikely).

- Using 3 (or more) systems instead of 2 solves this problem.
  ➔ "*Triple Collocation Technique*".

Triple Collocation Technique

Great!

But it makes several assumptions which are *sometimes* violated easily!

Error model

- Assume that the errors are linear additives to the true value (the “truth”).

- For any measurement, X_i, we assume that:

$$X_i = \alpha + \beta\, T_i + e_i$$

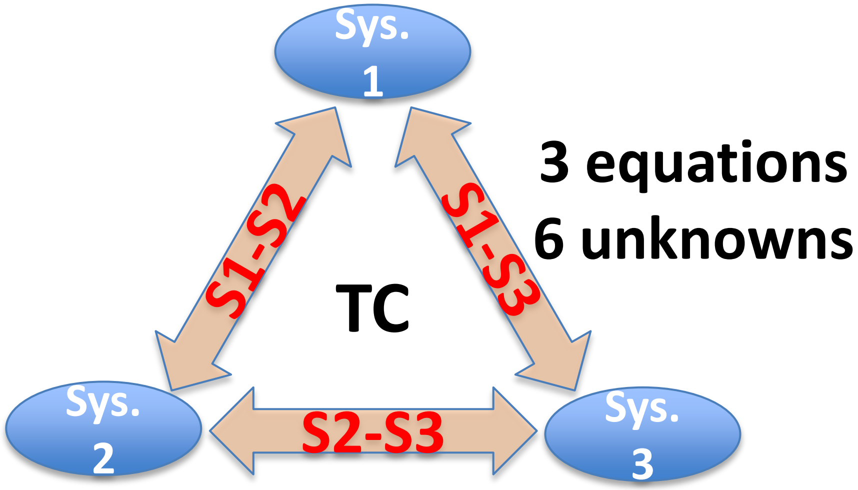
  here:
  - α is a fixed bias in the measurement system (accuracy).
  - β is a calibration constant of the measurement system (a bias that depends on the truth).
  - T_i is the truth.
  - e_i is the random error which is assumed to be of zero mean.

- Except for the measurement X_i, all the variables are unknown.

Triple Collocation

- Three measuring systems X, Y, Z (e.g. alt., scatt., model):

$$X_i = T_i + e_{x_i} \quad ; \quad Y_i = T_i + e_{y_i} \quad ; \quad Z_i = T_i + e_{z_i}$$



- Using the notation: $\langle X \rangle = N^{-1} \Sigma (X_i)$; … etc.

- Solving for the error variances and rearranging:

$$\langle e_x^2 \rangle = \langle X^2 \rangle - \langle X\,Y \rangle - \langle X\,Z \rangle + \langle Y\,Z \rangle + \langle e_x\,e_y \rangle + \langle e_x\,e_z \rangle - \langle e_y\,e_z \rangle$$

$$\langle e_y^2 \rangle = \langle Y^2 \rangle - \langle X\,Y \rangle - \langle Y\,Z \rangle + \langle X\,Z \rangle + \langle e_x\,e_y \rangle + \langle e_y\,e_z \rangle - \langle e_x\,e_z \rangle$$

$$\langle e_z^2 \rangle = \langle Z^2 \rangle - \langle Y\,Z \rangle - \langle X\,Z \rangle + \langle X\,Y \rangle + \langle e_y\,e_z \rangle + \langle e_x\,e_z \rangle - \langle e_x\,e_y \rangle$$

Note that we set:
- $\alpha_p = 0$ (all data sets have same bias);
- $\beta_p = 1$ (for the time being).

Note that:
- $\langle e_x^2 \rangle$, $\langle e_y^2 \rangle$, $\langle e_z^2 \rangle$ are the random error variances;
- $\langle e_x\,e_y \rangle$, $\langle e_x\,e_z \rangle$, $\langle e_y\,e_z \rangle$ are random error covariances.

ECMWF

Triple Collocation – Assumptions

- General Equations:

$$<e_x^2> = <X^2> - <X\,Y> - <X\,Z> + <Y\,Z> + < e_x\, e_y> + <e_x\, e_z> - < e_y\, e_z>$$

$$<e_y^2> = <Y^2> - <X\,Y> - <Y\,Z> + <X\,Z> + < e_x\, e_y> + <e_y\, e_z> - < e_x\, e_z>$$

$$<e_z^2> = <Z^2> - <Y\,Z> - <X\,Z> + <X\,Y> + < e_y\, e_z> + <e_x\, e_z> - < e_x\, e_y>$$

- 3 Equations with 6 unknowns (terms in black above)!

- **Assume no correlation between error pairs** ($< e_x\, e_y> = <e_x\, e_z> = < e_y\, e_z> = 0$)

$$<e_x^2> = <X^2> - <X\,Y> - <X\,Z> + <Y\,Z>$$

$$<e_y^2> = <Y^2> - <X\,Y> - <Y\,Z> + <X\,Z>$$

$$<e_z^2> = <Z^2> - <Y\,Z> - <X\,Z> + <X\,Y>$$

- But for the triplet SCAT-ALT-MODEL:
  - we know that: $<e_{scat}\, e_{model}> \neq 0$
  - we assumed that $<e_{scat}\; e_{alt}> \neq 0$ (?)
  - Also, we assumed that $<e_{alt}\, e_{model}> = 0$ (?)

ECMWF

Triple Collocation – Procedure

- We assume that one of the systems (say the first one, X) is calibrated (i.e. $\beta_x = 1$), and we calibrate the other two systems (i.e. To find β_y and β_z) accordingly.

- The neutral regression is used for that.
  (Conventional regression is not suitable as it assumes that one of the measurement systems is error-free).

- $\beta_y = [-B + (B^2 - 4\,A\,C)^{1/2}]\,/\,2\,A$

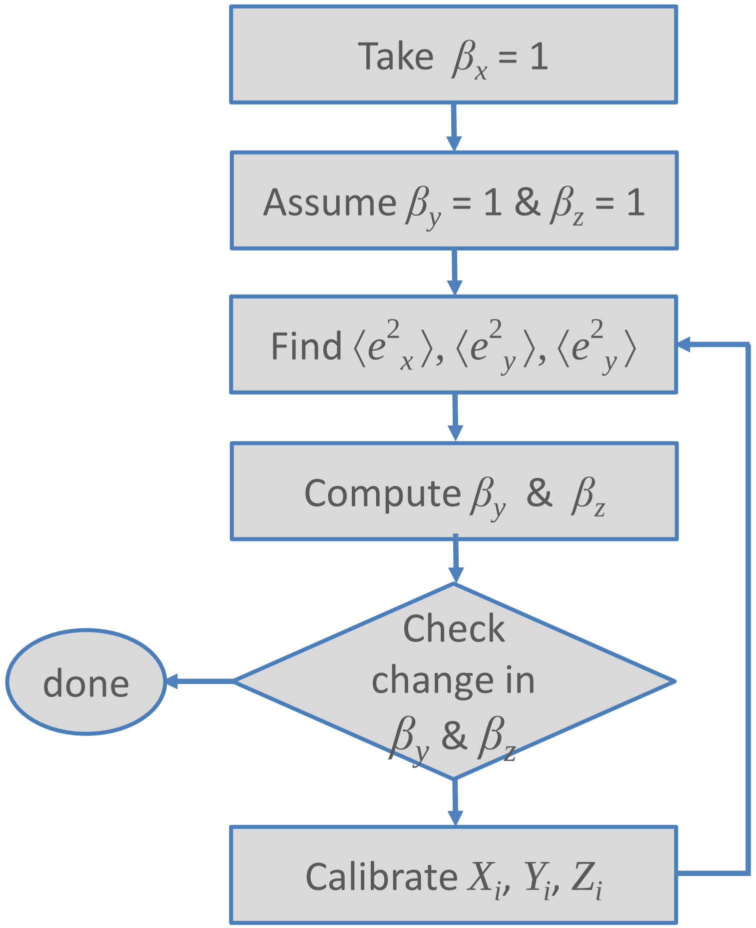
  where:

  $A = \gamma\langle X_i\,Y_i\rangle \quad ; \quad B = \langle X_i^2\rangle - \gamma\langle Y_i^2\rangle \quad ; \quad C = -\langle X_i\,Y_i\rangle$

  $\gamma = \langle e^2{}_x\rangle\,/\,\langle e^2{}_y\rangle$

- Similarly, β_z can be found by replacing Y above with Z.

- The calibration constants are found by iteration



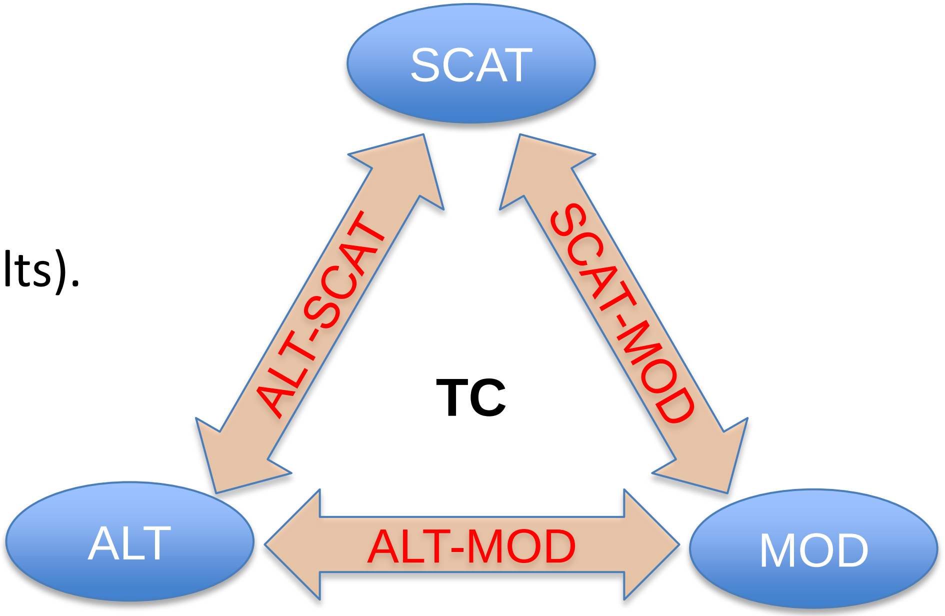
ECMWF

Results from an earlier study

Wind speed error estimation in a triplet of Altimeter (Jason-2), Scatterometer (ASCAT-B) & the ECMWF model

EARLIER STUDY: Altimeter – Model – Scatterometer Collocations

- Altimeter: Jason-2.
- Model: ECMWF IFS
- Scatterometer: ASCAT-B   (ASCAT-A provides same results).

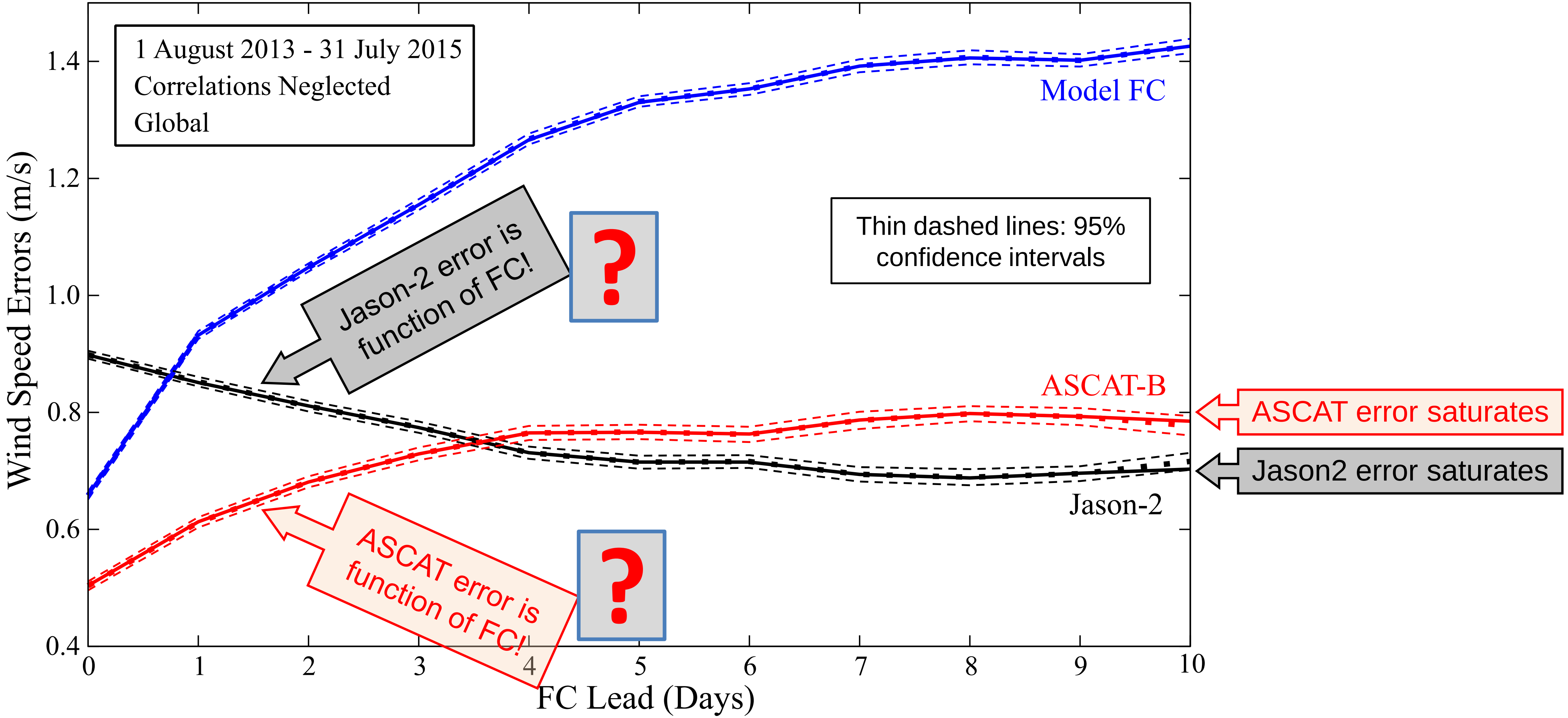


- Period: 1 August 2013 – 31 July 2015 (2 years).
- Only assimilated ASCAT data (good quality).

- Comparable scales by averaging altimeter data ➔ 70~100 km.

- Triple collocation (TC) assumes no (or known) correlation among data sets.
- The model assimilates ASCAT data ➔ correlations! (violation to the assumptions).

Surface wind speed errors estimated by ignoring the correlations



Altimeter – Model – Scatterometer Collocations (cont'd)

- Model error increases with increasing the FC lead time ➔ OK.
- Altimeter and scatterometer errors should not depend on FC lead time.
- The model assimilates ASCAT data ➔ correlations (which were ignored).
- Altimeter and scatterometer errors asymptote at long lead times (> ~7 days).
- Correlations between ASCAT and the model almost vanish.
- We can estimate the correlations (2 out of 3 only) and correct the error estimates.

Error correlations and error adjustment

- Change of error of model forecast as a function of the FC lead time is expectable.

- Change of altimeter and scatterometer errors as functions of FC lead time is due to the ignored error covariances.

- Tentatively; we expect the impact of correlations to vanish at long forecast leads (say beyond ~7 days) ➔ accept altimeter and scatterometer errors at those lead times are the correct error estimates and work back to find two of the three error covariances ($<e_{scat}\, e_{model}>$ & $<e_{scat}\, e_{alt}>$) and the model error variance $<e_{model}^2>$.

- We had still to assume that one of the covariances is zero ($<e_{alt}\, e_{model}>$).

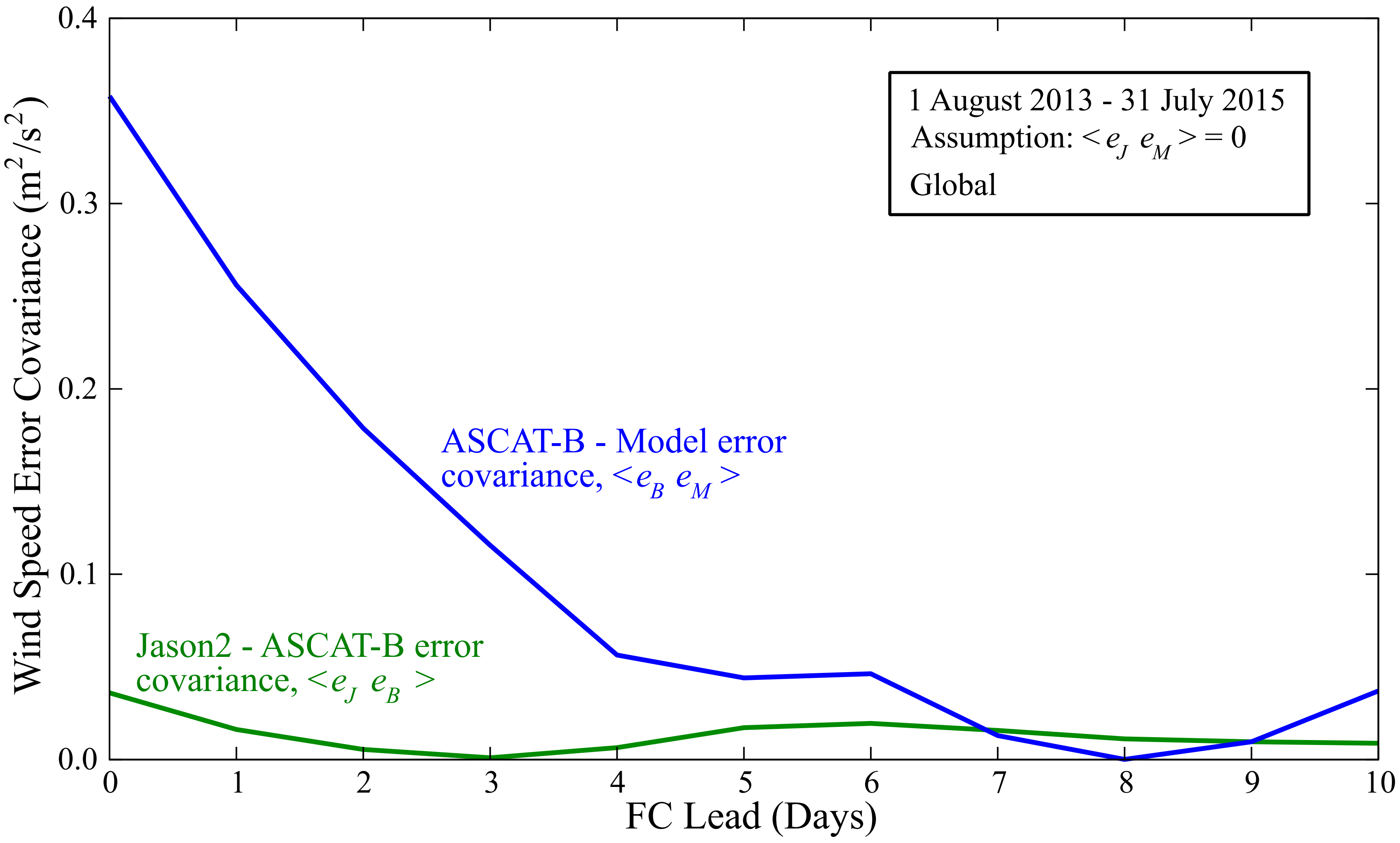
At FC lead times beyond 7 days $<e_x^2>$ (X=ALT), $<e_z^2>$ (Z=SCAT) represent the correct error variances: (Y=MODEL)

$$<e_x^2> = <X^2> - <X\,Y> - <X\,Z> + <Y\,Z> + 0 + <e_x\, e_z> - <e_y\, e_z>$$

$$<e_y^2> = <Y^2> - <X\,Y> - <Y\,Z> + <X\,Z> + 0 + <e_y\, e_z> - <e_x\, e_z>$$

$$<e_z^2> = <Z^2> - <Y\,Z> - <X\,Z> + <X\,Y> + <e_y\, e_z> + <e_x\, e_z> - 0$$

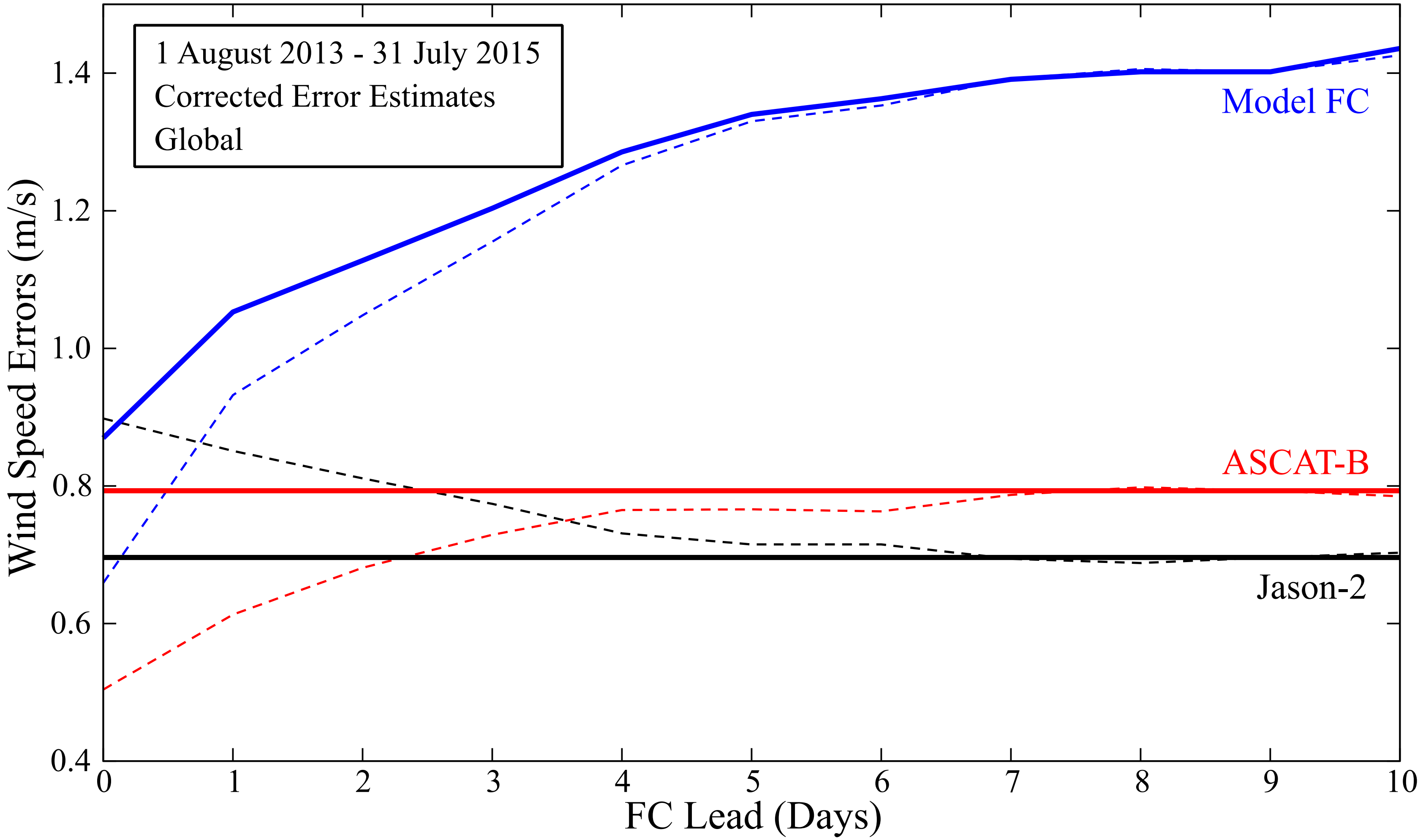
Estimation of surface wind speed error correlations (covariances)



Pearson correlation coefficient between the two data sets *x* & *y* is:

$r^2 = \langle x\, y\rangle^2 / (\langle x^2\rangle \langle y^2\rangle)$

Errors corrected for correlations



Concluding Remarks

- Triple collocation is a powerful technique for the estimation of random errors.
- Ocean surface wind speeds from altimeters and scatterometers have low random errors for scales in the order of 100 km (~0.7 & 0.8 m/s for Jason-2 and ASCAT-B; respectively, for the global ocean).
- The error in the model wind speed analysis is comparable with altimeters and scatterometers (better than 0.9 m/s for the global ocean).
- Some error correlations can be estimated by using model forecasts at 7-day lead time or beyond.

- **Work is in progress towards error estimation of Sentinel-3 and Jason-3 wind and wave data.**