Determination of Uncertainties of OC data

IOCCG Training

22-23 Apr. 2024

Frédéric Mélin

European Commission

Vocabulary

- **measurand**: well-defined physical quantity that is to be measured
- **uncertainty of a measurement**: a parameter, associated with the result of a measurement that characterizes the dispersion of the values that could reasonably be attributed to the measurand

dispersion => uncertainty expressed as a standard deviation (standard uncertainty)

- **error:** difference between the measurement and the true value of the measurand (or a reference quantity value, assumed to have negligible uncertainty)
- **accuracy:** closeness of agreement between a measured quantity value and the true value of the measurand

Validation

i.e., comparison with field data

European Commission

Validation (1): Comparison EO vs field data

measurand x_{SAT} x_F

uncertainty $u(x_{SAT})$ $u(x_F)$

NB: field data have their own uncertainties

- Different spatial scales
- Different times
- Definitional differences (λ)

representation error

Validation (2): Match-up Selection Protocol

- Select the closest pixel to the measurement location

- Extract a box of **N** pixels around that pixel (3x3, 5x5, …) with all relevant information:
  - geophysical outputs ($L_{WN}(\lambda)$, Chl*a*, …)
  - geometry (θ, θ_0, $\Delta\varphi$)
  - ancillary data (pressure, ozone, wind, …)

- Document :
  - **flags** (level of quality, geometry)
    -> fraction of valid values f_v
  - difference in time Δt
  - temporal variability ($CV_t = \sigma_t/\mu_t$) if available
  - spatial variability ($CV_s = \sigma_s/\mu_s$)
  - difference in wavelength
  - …

Bailey and Werdell 2006, Mélin & Franz 2014

Validation (3): Comparison statistics

$$|\psi| = 100\frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - x_i|}{x_i}$$

mean absolute relative difference

$$|\delta| = \frac{1}{N}\sum_{i=1}^{N}|y_i - x_i|$$

mean absolute difference

$$\psi = 100\frac{1}{N}\sum_{i=1}^{N}\frac{y_i - x_i}{x_i}$$

mean relative difference

$$\delta = \frac{1}{N}\sum_{i=1}^{N}(y_i - x_i)$$

mean difference (bias)

$$\Delta = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - x_i)^2}$$

root mean square difference

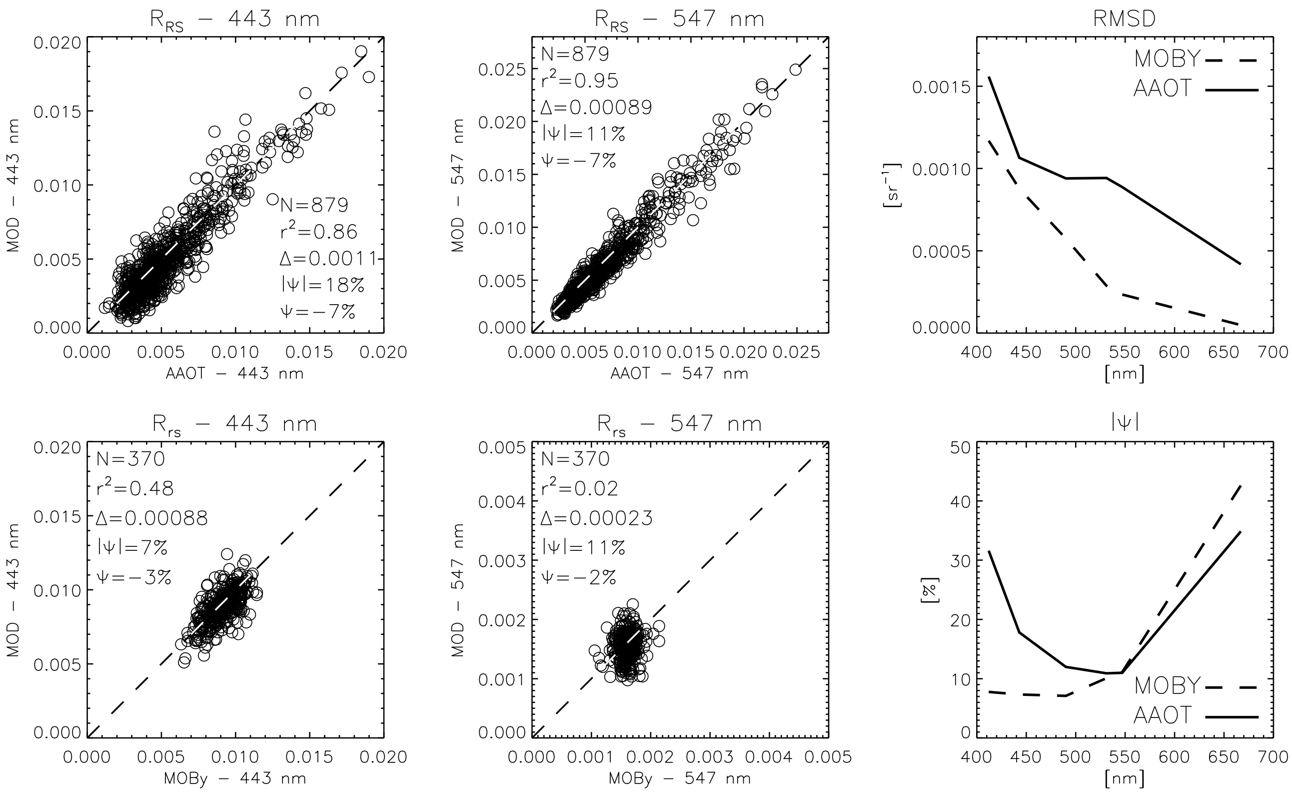
$$\Delta_c = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \overline{y} - x_i + \overline{x})^2} = \sqrt{\Delta^2 - \delta^2}$$

centered (unbiased) mean square difference

Validation (4): Example

MODIS-A

AAOT: Adriatic coastal site



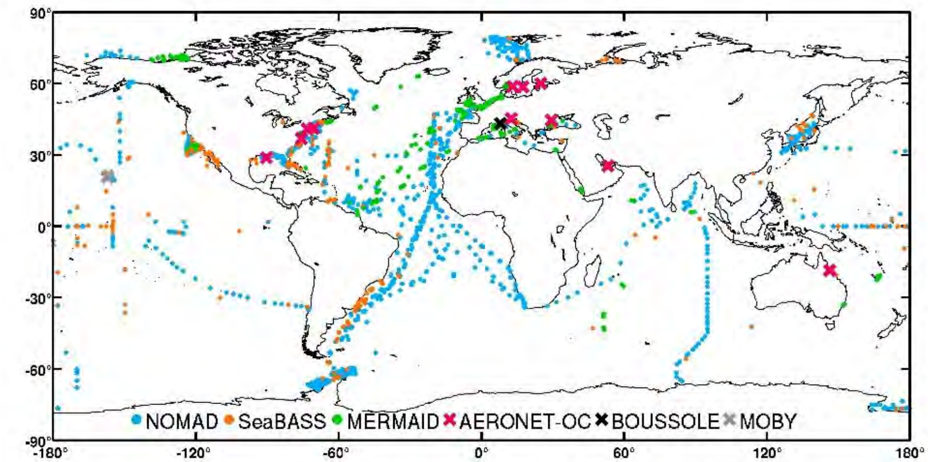
$$\Delta = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - x_i)^2}$$

$$|\psi| = 100.\frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - x_i|}{x_i}$$

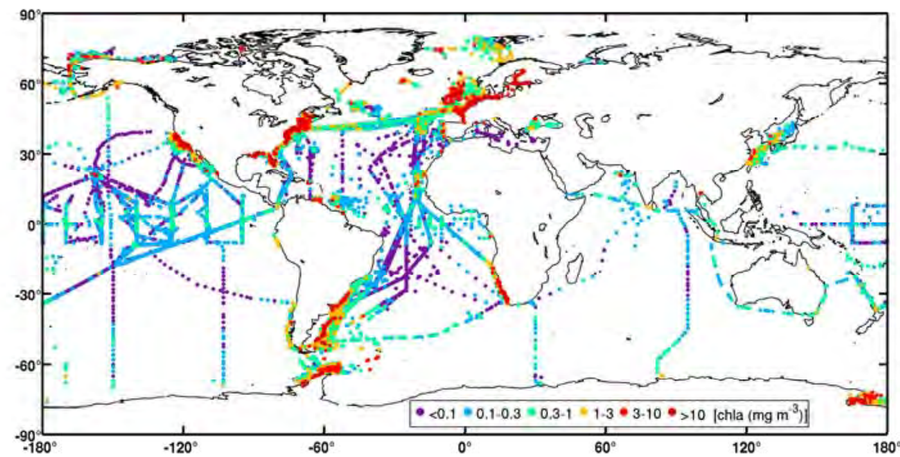
MOBY: Pacific oceanic site

Validation (5): Spatial distribution

R_{RS} data distribution



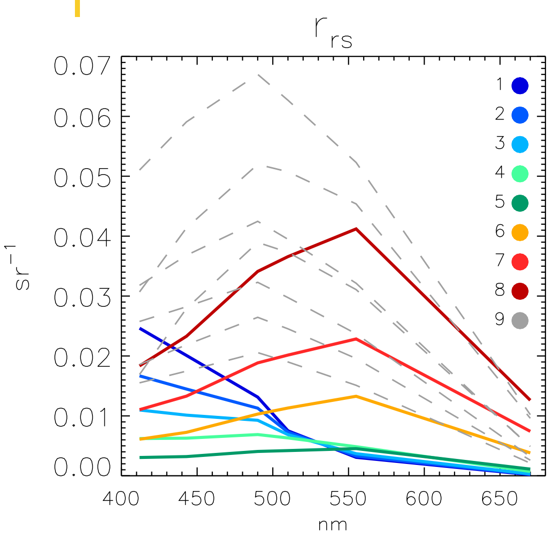
Chl-a data distribution



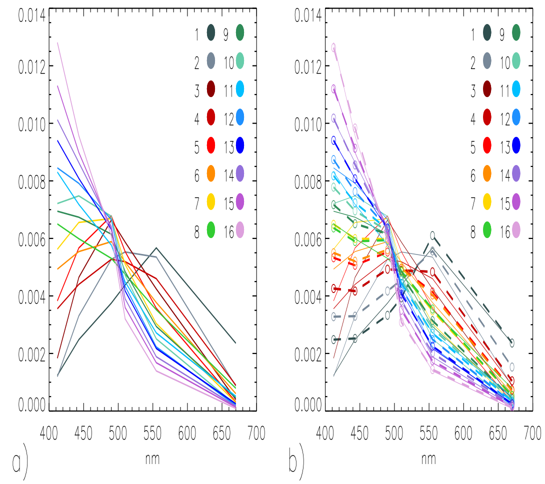
Valente et al., *ESSD* 2016

field data rather sparse

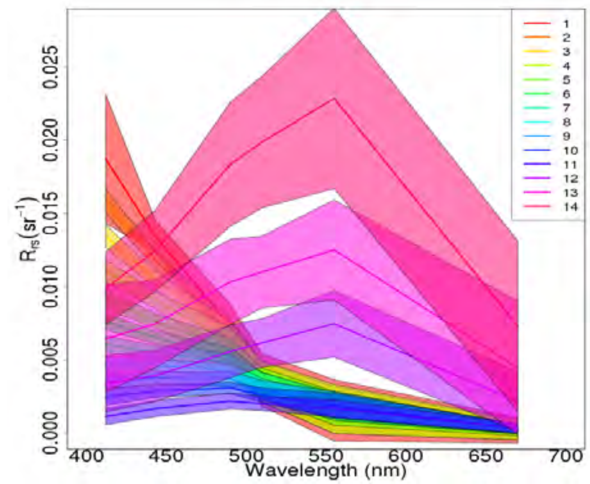
Validation (6a): Spatial extension with OWTs



Moore et al., *RSE* 2009



Mélin & Vantrepotte *RSE* 2015



Jackson et al., *RSE* 2017

Hypothesis: similar uncertainties for similar optical classes / optical water types

Validation (6b): Spatial extension with OWTs

Validation statistics for *M* OWTs

- Compute class membership $w_{k,i}$ for each class *k* and each match-up *i*

- Compute validation stats for each class *k*:

$$\Delta_k^{\ 2} = \frac{\sum_{i=1}^{N} w_{k,i}(x_{i,s} - x_{i,f})^2}{\sum_{i=1}^{N} w_{k,i}}$$ **RMSD**

$$\delta_k = \frac{\sum_{i=1}^{N} w_{k,i}(x_{i,s} - x_{i,f})}{\sum_{i=1}^{N} w_{k,i}}$$ **BIAS**

difference between satellite and field values

Validation statistics for pixel *p*

- Compute class membership $w_{k,p}$ for each class *k*

- Compute stats for pixel *p*:

$$\Delta_p^{\ 2} = \frac{\sum_{k=1}^{M} w_{k,p}\Delta_k^{\ 2}}{\sum_{p=1}^{M} w_{k,p}} \qquad \delta_p = \frac{\sum_{k=1}^{M} w_{k,p}\delta_k}{\sum_{k=1}^{M} w_{k,p}}$$

Uncertainty Propagation

Law of propagation of Uncertainties (1)

combined uncertainty of measurand *y*:

$$u_c^2(y) = \sum_{i=1}^{N}\left(\frac{\partial f}{\partial x_i}\right)^2 u^2\,(x_i) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\frac{\partial f}{\partial x_i}\frac{\partial f}{x_j}u(x_i, x_j)$$

$$u_c^2(y) = \sum_{i=1}^{N}\left(\frac{\partial f}{\partial x_i}\right)^2 u^2\,(x_i) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\frac{\partial f}{\partial x_i}\frac{\partial f}{x_j}r_{ij}u(x_i)u(x_j)$$

Law of propagation of Uncertainties (2): Example for band-ratio Chl-a algorithms

$$\log_{10} Chla = \sum_{i=0}^{N} A_i \left(\log_{10}\frac{R_g}{R_b}\right)^i$$

u: standard uncertainty for Chla, R_b and R_g

Taylor expansion

$$u_{Chl}^2 = \mathrm{Chla}^2 * \left(\sum_{i=1}^{N} A_i * i * \left(\log_{10}\frac{R_g}{R_b}\right)^{i-1}\right)^2 * \left(\frac{u_{R_b}^2}{R_b^2} - 2\,\frac{u_{R_bR_g}}{R_bR_g} + \frac{u_{R_g}^2}{R_g^2}\right)$$

if $u_{R_bR_g} = u_{R_b}u_{R_g}$ (error correlation = 1)

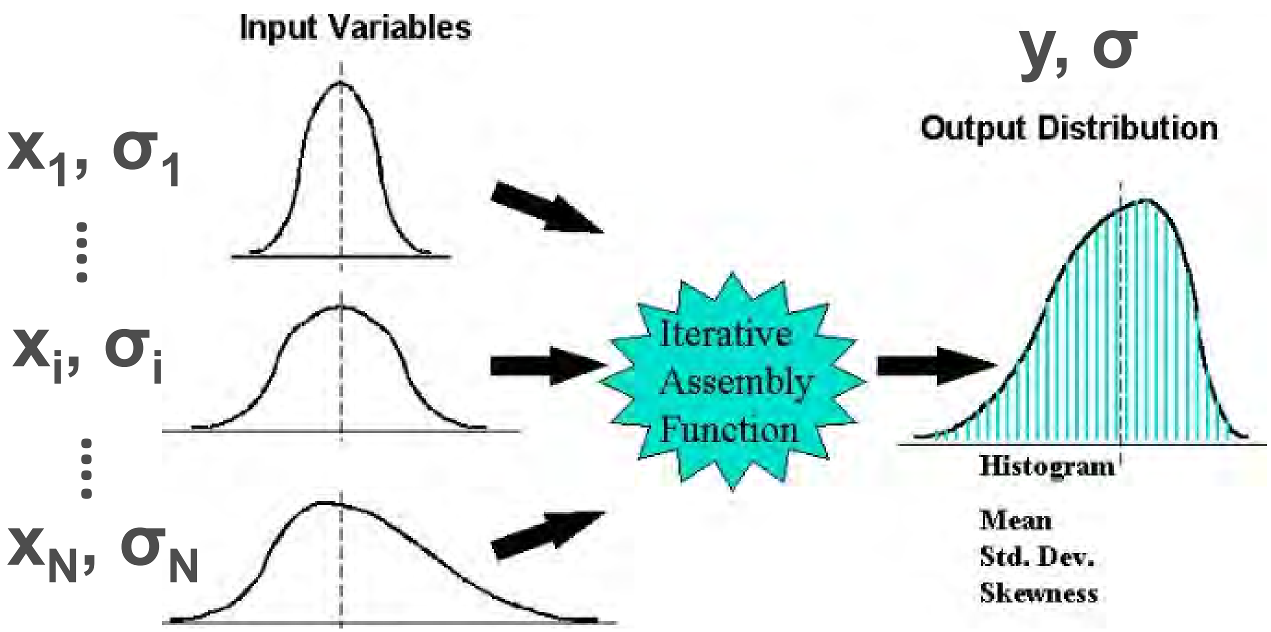
$$u_{Chl} = \mathrm{Chla}\left|\sum_{i=1}^{N} A_i * i * \left(log_{10}\frac{R_b}{R_g}\right)^{i-1}\right| * \left|\frac{u_{R_b}}{R_b} - \frac{u_{R_g}}{R_g}\right|$$

Notable reduction in uncertainty

See also McKinna et al., *FES* 2019

Monte Carlo simulations / Use of ensembles

$y = f(x_i)_{i=1,N}$



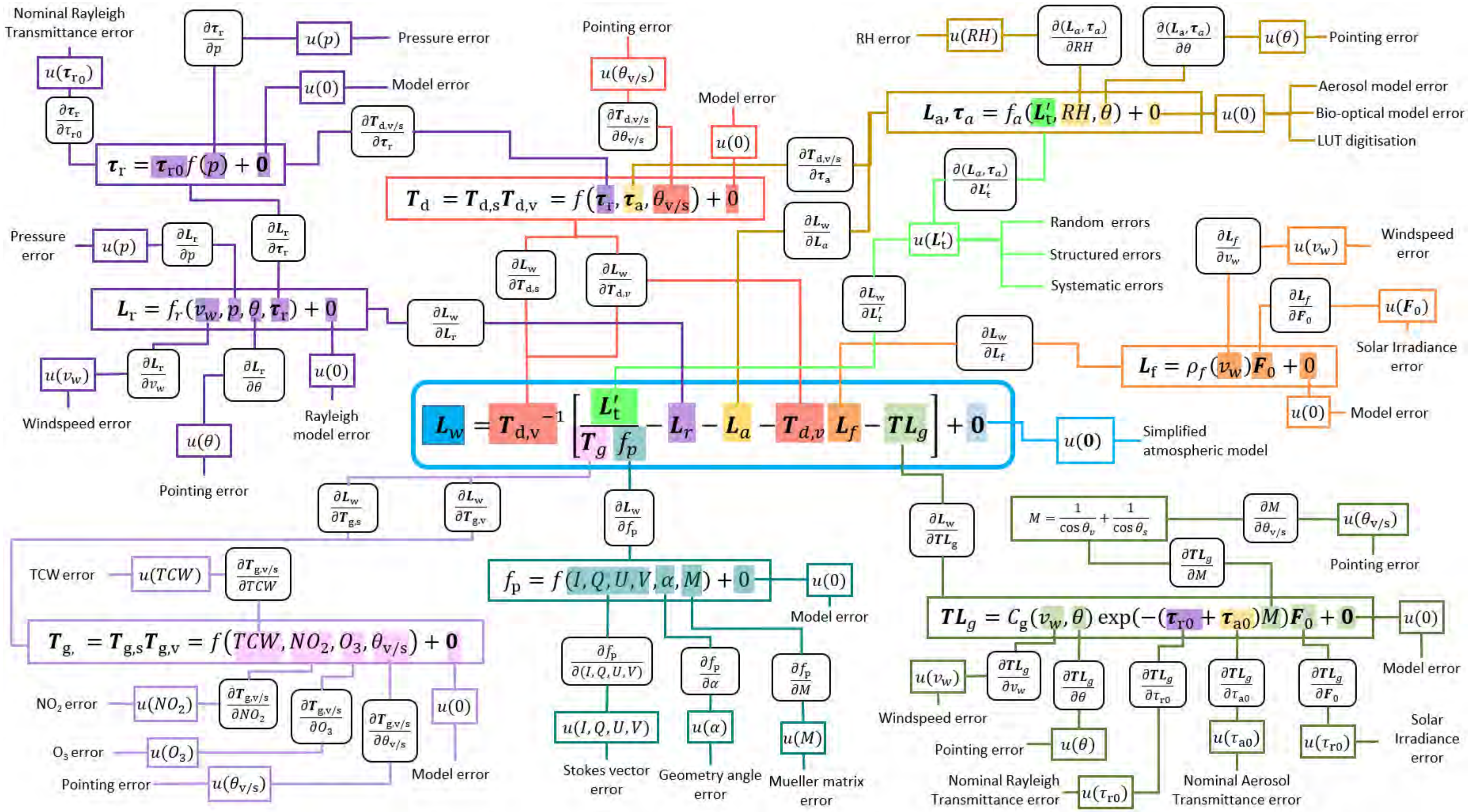
adcats.et.byu.edu

Use of ensembles: example on ancillary data

Question:
impact of the uncertainties associated with ancillary data on the uncertainty of R_{RS} through an atmospheric correction algorithm (*l2gen*)

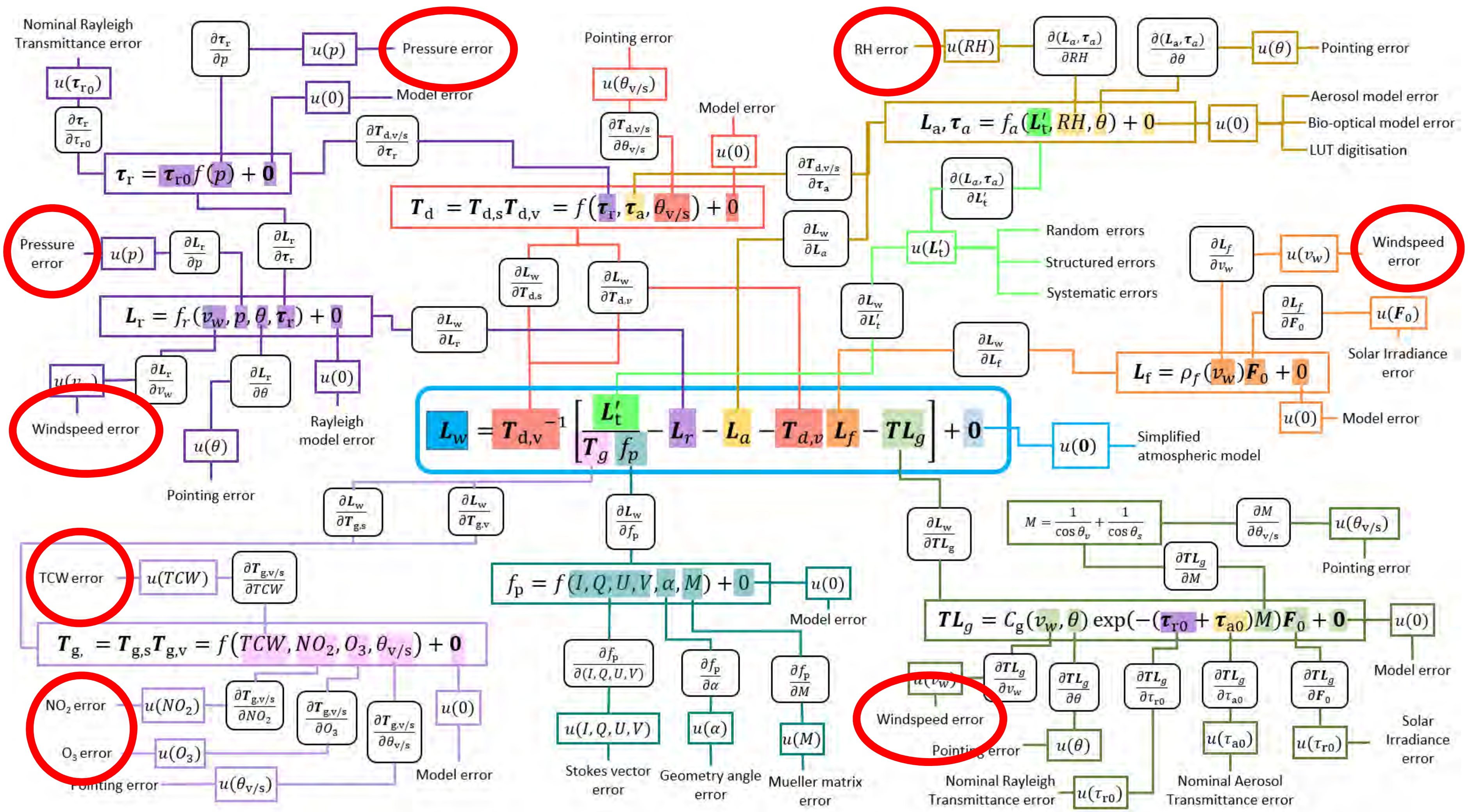
- **WS: wind speed**
- **SLP: sea level pressure**
- **PW: precipitable water**
- **RH: relative humidity**
- **O_3: ozone concentration**

l2gen uncertainty tree



Source: NPL; De Vis, Hunt & Mélin *RS* 2022

l2gen uncertainty tree



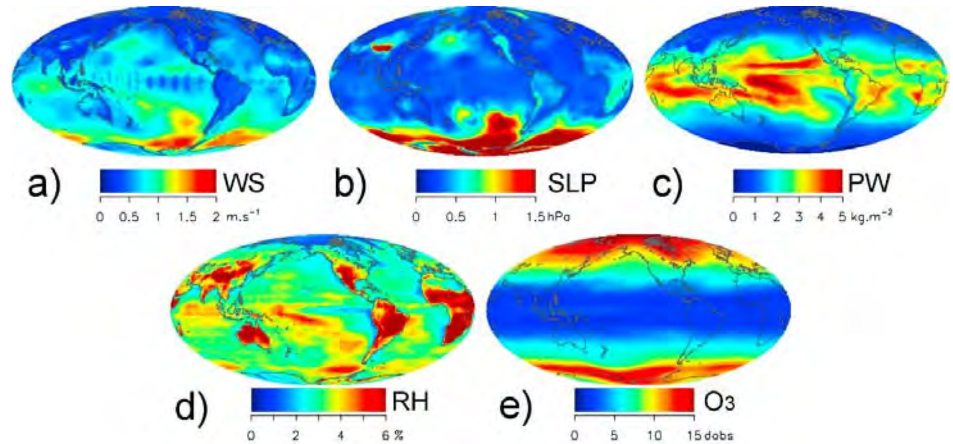
 Source: NPL; De Vis, Hunt & Mélin *RS* 2022

Use of ensembles: example (2)

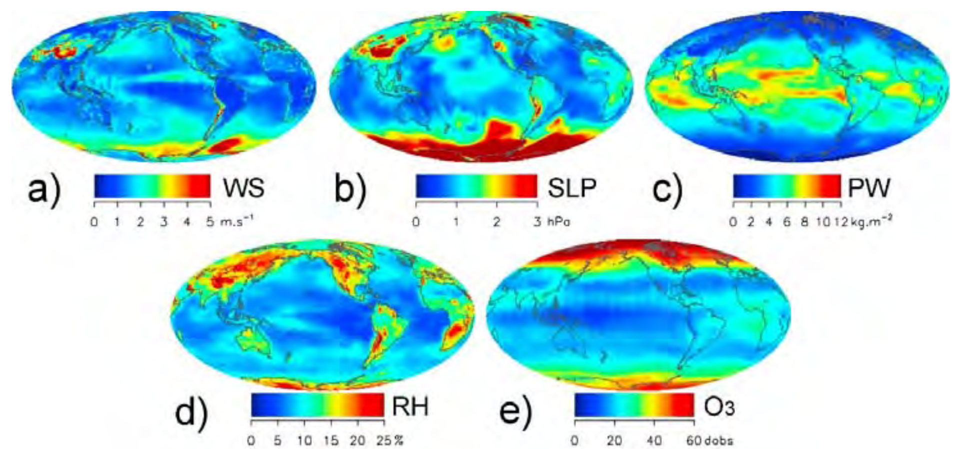
- Reference set: NCEP + TOMS
- CERA-20: Ensemble of 10 realizations of ERA5 (ECMWF)

Ensemble SeaWiFS global data (2003)

average s.d. over the ERA ensemble



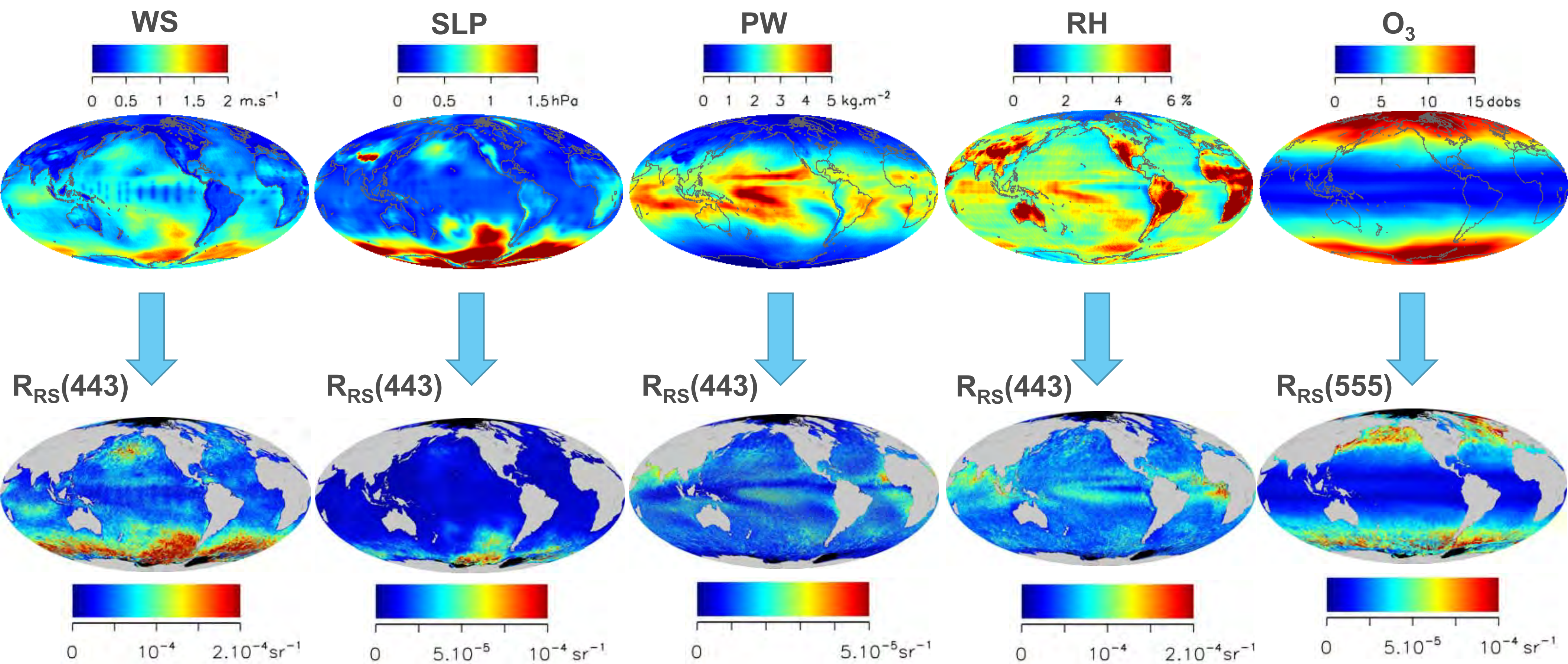
RMS difference ERA-NCEP/TOMS



De Vis, Hunt & Mélin *RS* 2022, Mélin et al. *IEEE* 2022

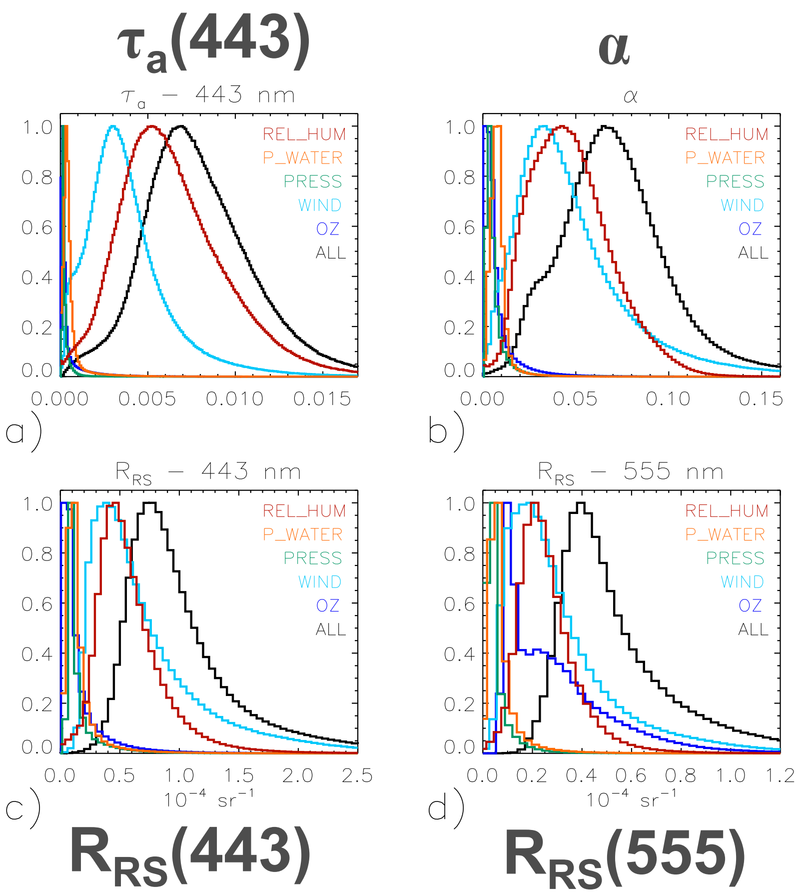
European Commission

Perturbation of 1 Ancillary Variable (s.d. over the ensemble):

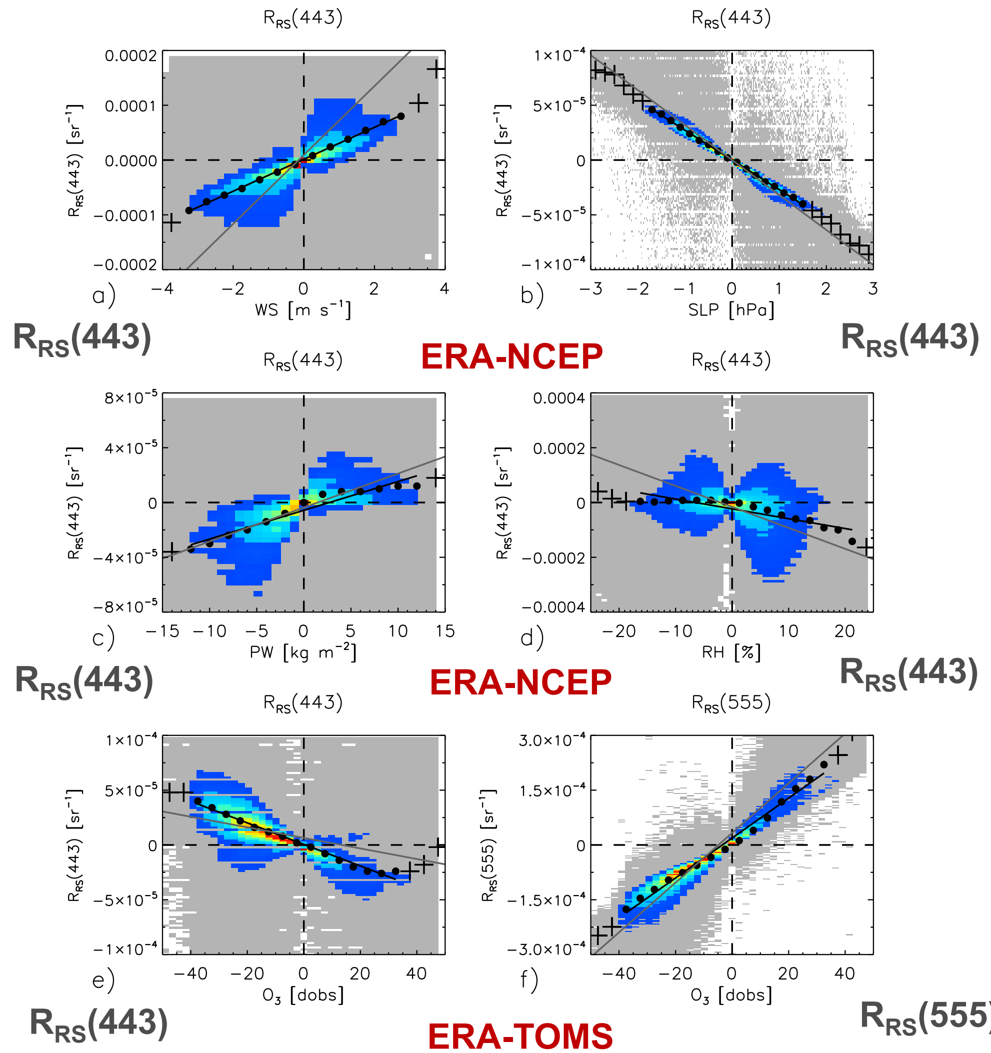


Mélin et al., *IEEE* 2022

Contribution of each Ancillary Variable:



Calculation of Sensitivity factors:



 Mélin et al., *IEEE* 2022

Collocation Statistics

Collocation (1): with Reference Data

For i=1,N validation points:

$$\begin{cases} x_i = r_i + \gamma_i \\ y_i = \beta\, r_i + \delta + \varepsilon_i \end{cases}$$

r: reference values ("truth")
x: in-situ measurements **with σ_γ known**
γ: associated random error term (mean=0, un-correlated)
y: satellite values, **with σ_ε unknown**
ε: associated random error term (mean=0, un-correlated)
δ: additive bias
β: multiplicative bias

$$\sigma_\varepsilon^2 = \sigma_y^2 - \frac{\sigma_{xy}^2}{\sigma_x^2 - \sigma_\gamma^2}$$

(δ and β can also be computed)

Mélin et al. *RSE* 2016; Mélin, *FMAS* 2021; Mélin *RSE* 2022

Collocation (2): with 2 Similar Data Sets

For i=1,N comparison points:

$$\begin{cases} x_i = r_i + \gamma_i \\ y_i = \beta\, r_i + \delta + \varepsilon_i \end{cases}$$

r: reference values ("truth")
x: in-situ measurements **with σ_γ unknown**
γ: associated random error term (mean=0, un-correlated)
y: satellite values, **with σ_ε unknown**
ε: associated random error term (mean=0, un-correlated)
δ: additive bias
β: multiplicative bias

$$\eta = \frac{\sigma_\varepsilon}{\sigma_\gamma}$$

$$\beta = \frac{-\eta^2\sigma_x^2 + \sigma_y^2 + \sqrt{(\sigma_y^2 - \eta^2\sigma_x^2)^2 + 4\sigma_{xy}^2\eta^2}}{2\sigma_{xy}}$$

Slope of Model II linear regression
Legendre & Legendre (1998)

$$\begin{cases} \sigma_\gamma^2 = \frac{1}{2}\left[\sigma_x^2 + \frac{\sigma_y^2}{\eta^2} - \sqrt{(\frac{\sigma_y^2}{\eta^2} - \sigma_x^2)^2 + 4\sigma_{xy}^2/\eta^2}\right] \\ \sigma_\varepsilon^2 = \frac{1}{2}\left[\eta^2\sigma_x^2 + \sigma_y^2 - \sqrt{(\sigma_y^2 - \eta^2\sigma_x^2)^2 + 4\sigma_{xy}^2\eta^2}\right] \end{cases}$$

Mélin, *FMAS* 2021; Mélin *RSE* 2022

Collocation (3): with 2 Similar Data Sets

By the way:

$$\Delta_c^2 = \sigma_x^2 + \sigma_y^2 - 2\sigma_{xy}$$

$$\Delta_c^2 = (\beta - 1)^2\sigma_x^2 + \beta(2-\beta)\sigma_\gamma^2 + \sigma_\varepsilon^2 \longrightarrow \sigma_\gamma^2 + \sigma_\varepsilon^2$$

if $\beta \longrightarrow 1$

Verification of Uncertainties

process providing objective evidence that stated uncertainty fulfills specific requirements

~

checking that uncertainty estimates are truthful

Use of Reference Data (1)

$(x_i, u^{est}(x_i))_{i=1,N}$ collocated with reference data $(x_{ref,i}, u(x_{ref,i}))_{i=1,N}$

estimated uncertainty

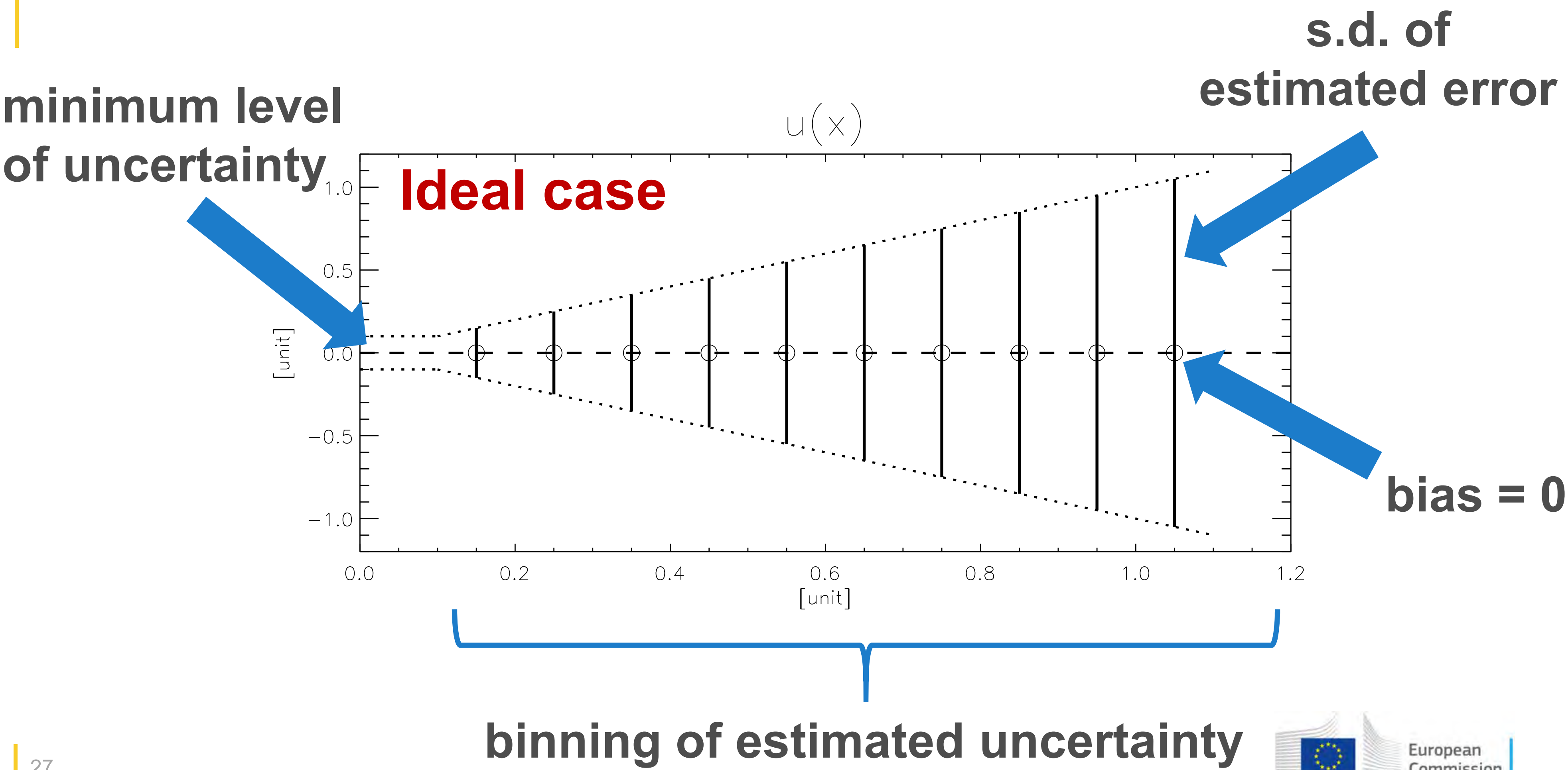
if $u(x_{ref,i}) << u(x_i)$: $x_i - x_{ref,i}$ ~ error

With normal hypothesis: $|X_i - X_{ref,i}| < k\sqrt{u^2(X_i) + u^2(X_{ref,i}) + \sigma^2_{RE}}$

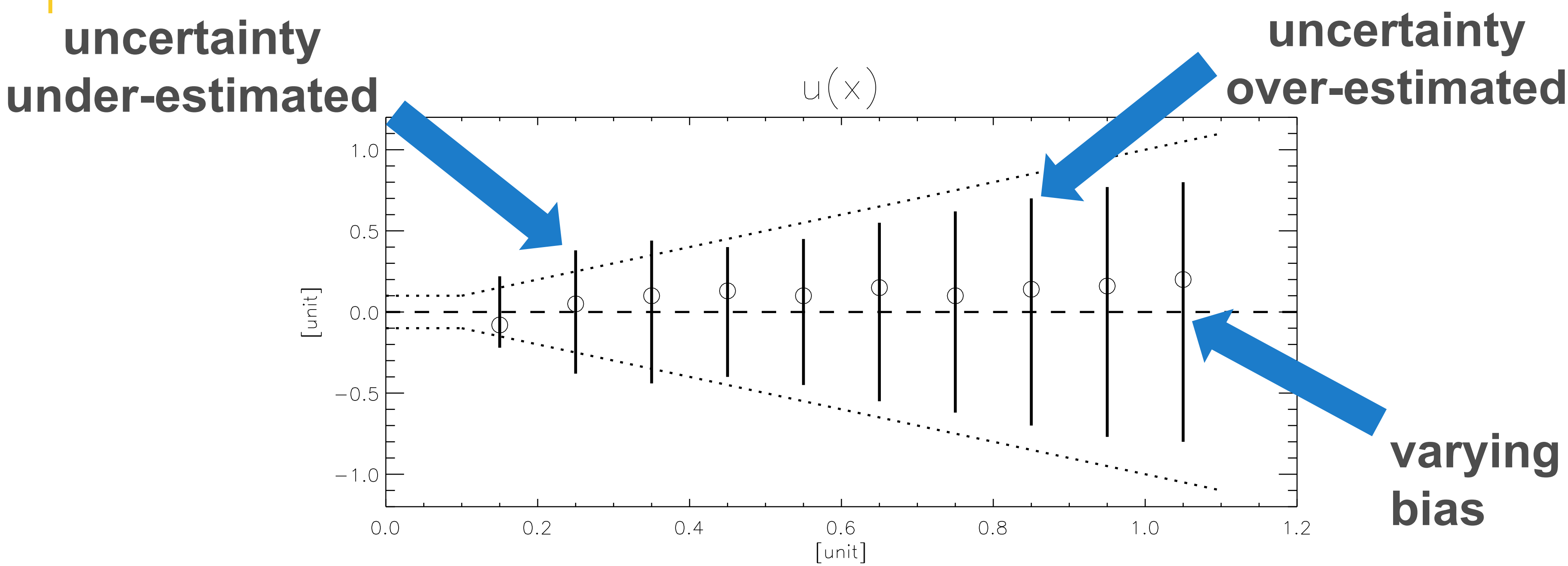
representation error

in 68% of cases (k=1)

Use of Reference Data (2): Cone Diagram



Use of Reference Data (3): Cone Diagram



Inter-comparison / Metrological Compatibility

$(x_{1,i}, u(x_{1,i}))_{i=1,N}$ collocated with data $(x_{2,i}, u(x_{2,i}))_{i=1,N}$

With normal hypothesis :
(and error correlation = 0
and no representation error)

$$|x_{2,i} - x_{1,i}| < k\sqrt{u^2(x_1) + u^2(x_2)}$$

in 68% of cases (*k*=1)

Comparison between Methods

IOCCG Report #18 (2019)
+ new publications

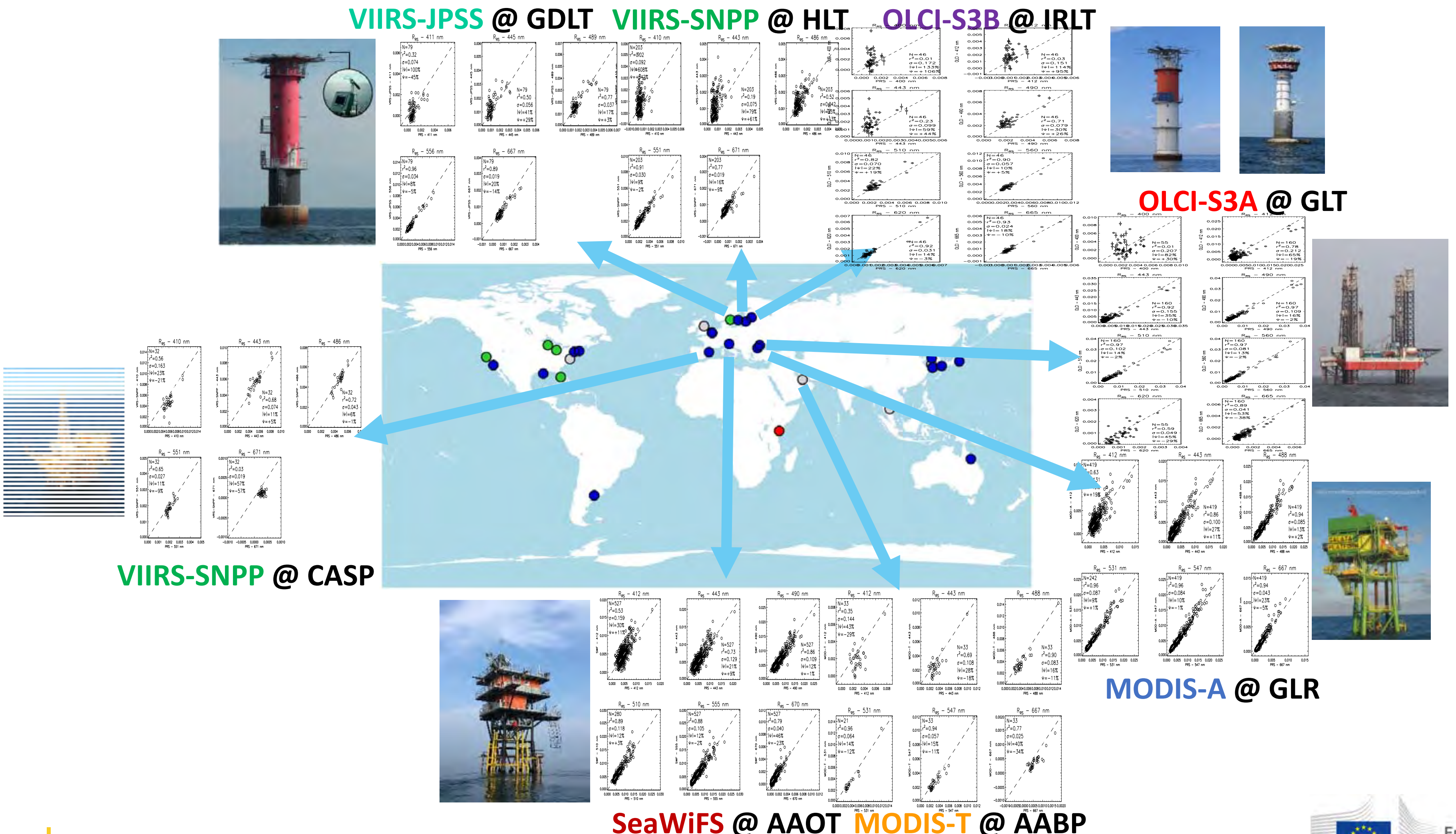
Various methods to derive uncertainty estimates have been proposed (none of them fully satisfactory)
Comparing their outputs is a fruitful exercise

Bonus on Validation

Some lessons on Validation of R_{RS} data

AERONET-OC

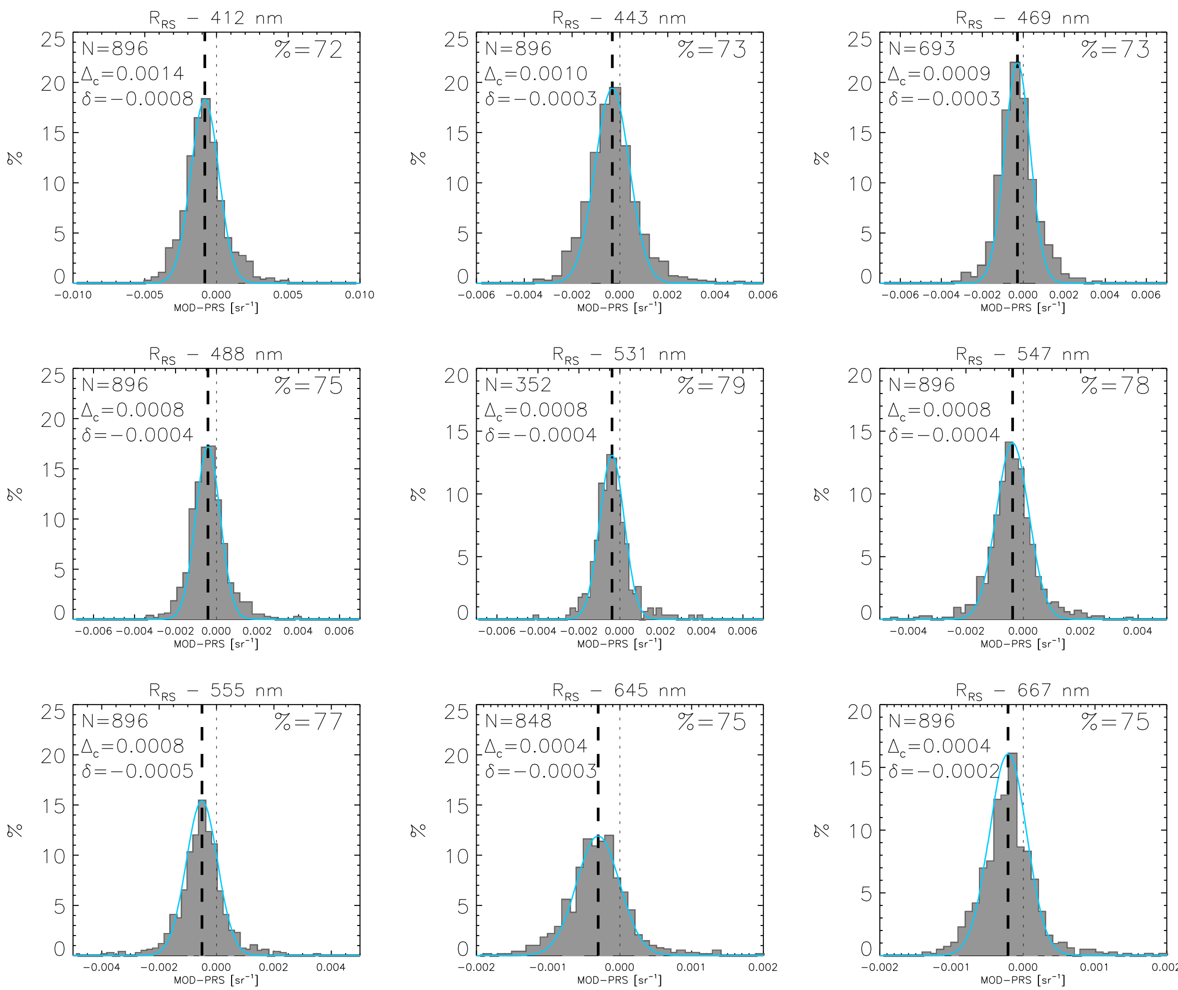
- SeaWiFS
- MODIS-Aqua
- MODIS-Terra
- VIIRS-SNPP
- VIIRS-JPSS
- OLCI-S3A
- OLCI-S3B



European Commission

❖ How do residuals look like? ➡ Fairly Gaussian

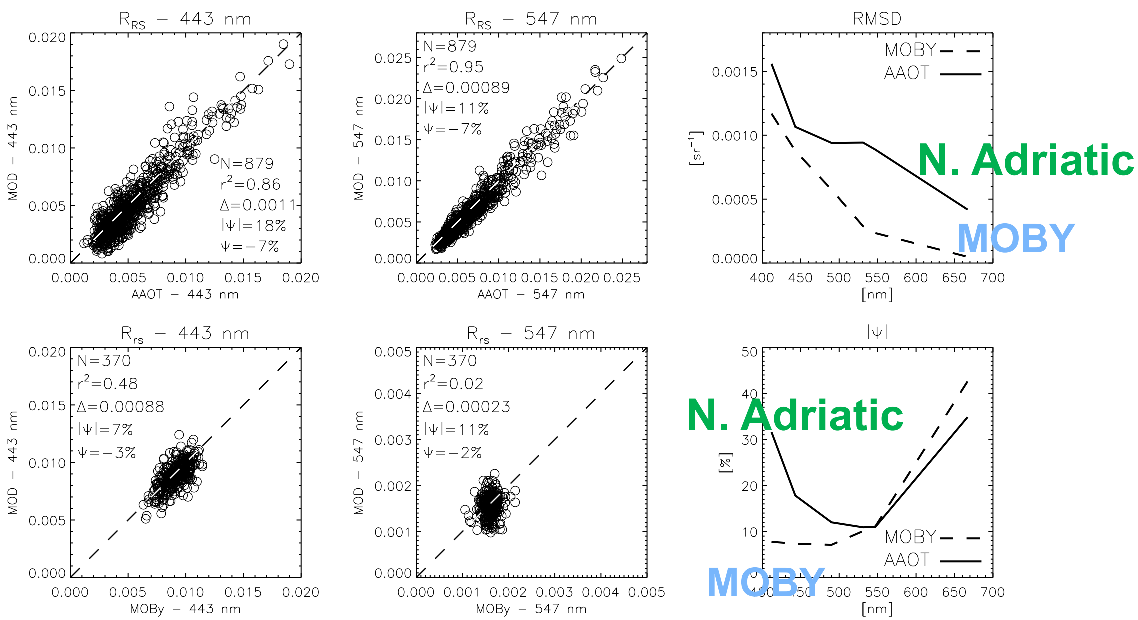
MODIS-A @ AAOT



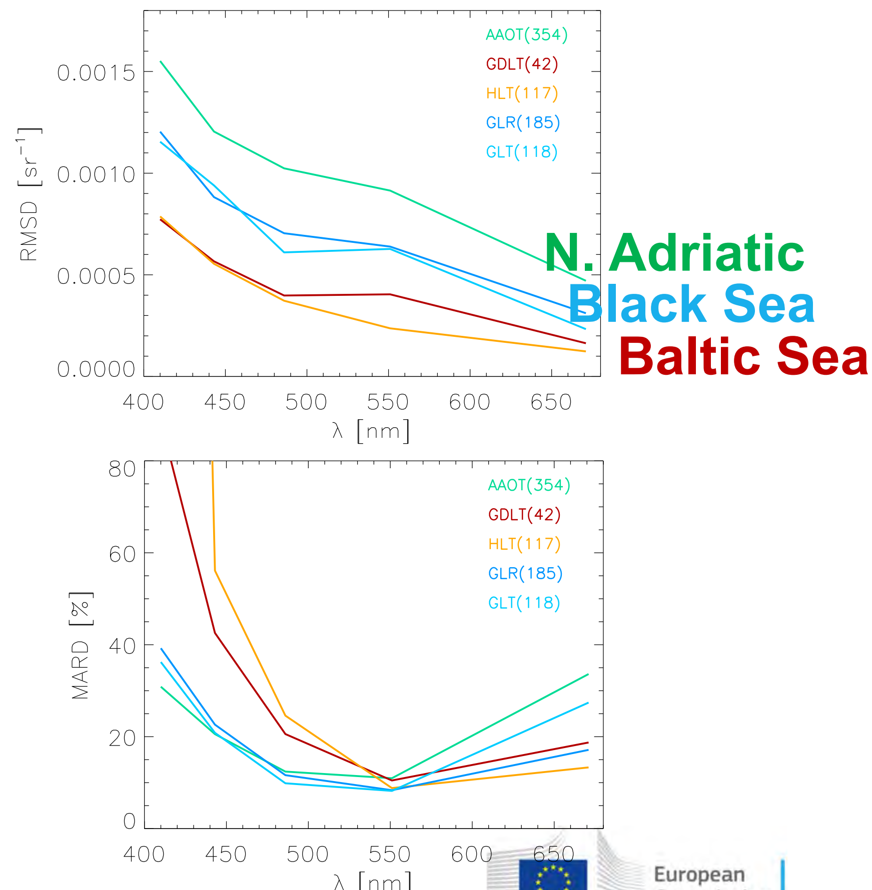
European Commission

❖ Do validation results change across sites?

MODIS-A

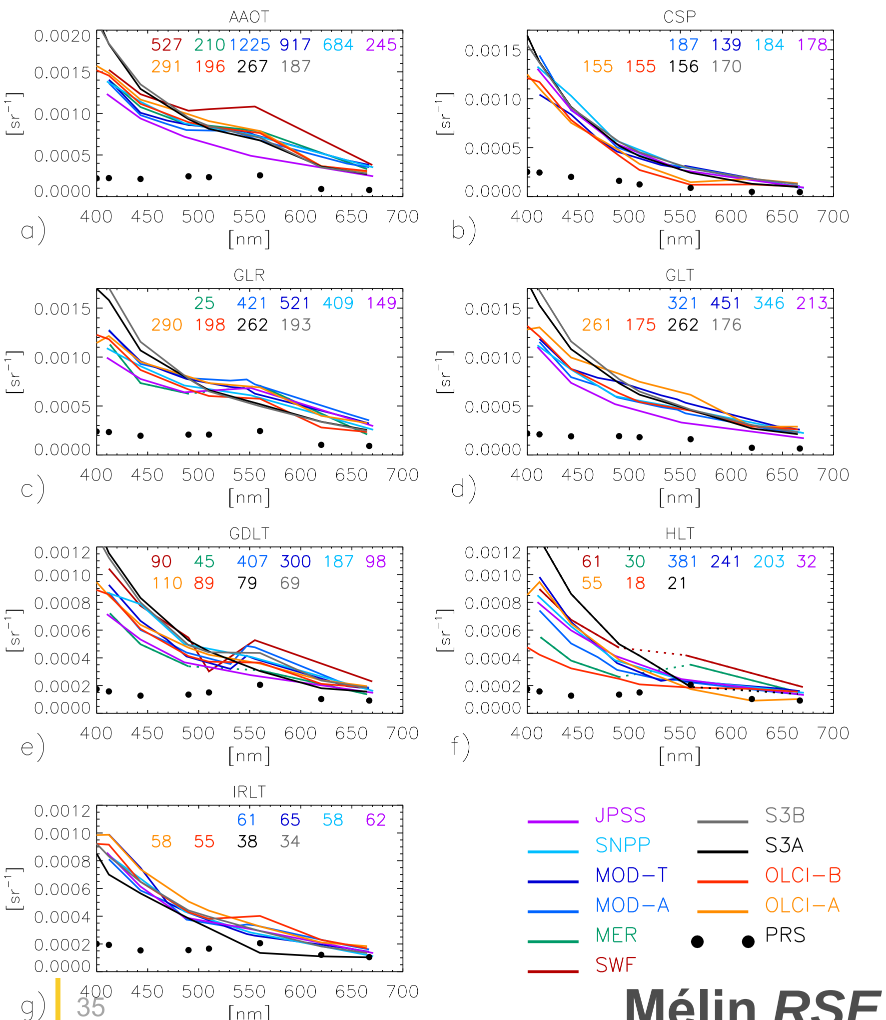


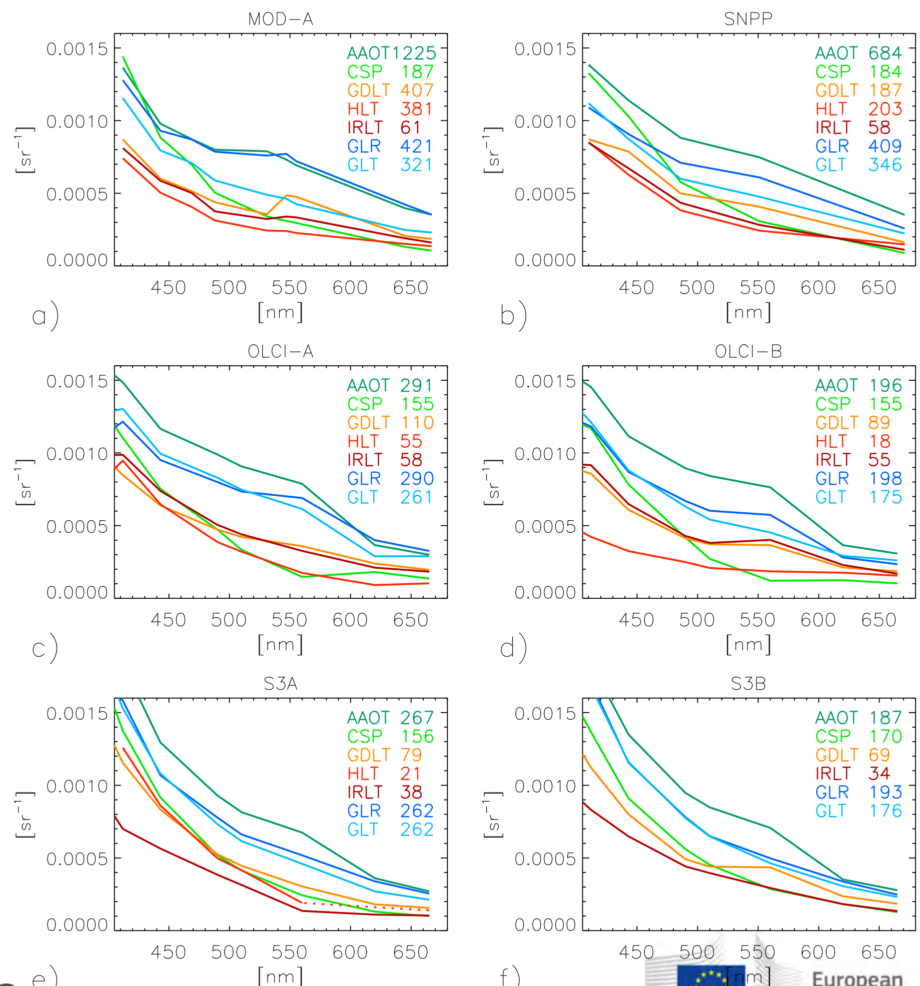
VIIRS



$$\Delta = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - x_i)^2} \qquad |\psi| = 100.\frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - x_i|}{x_i}$$

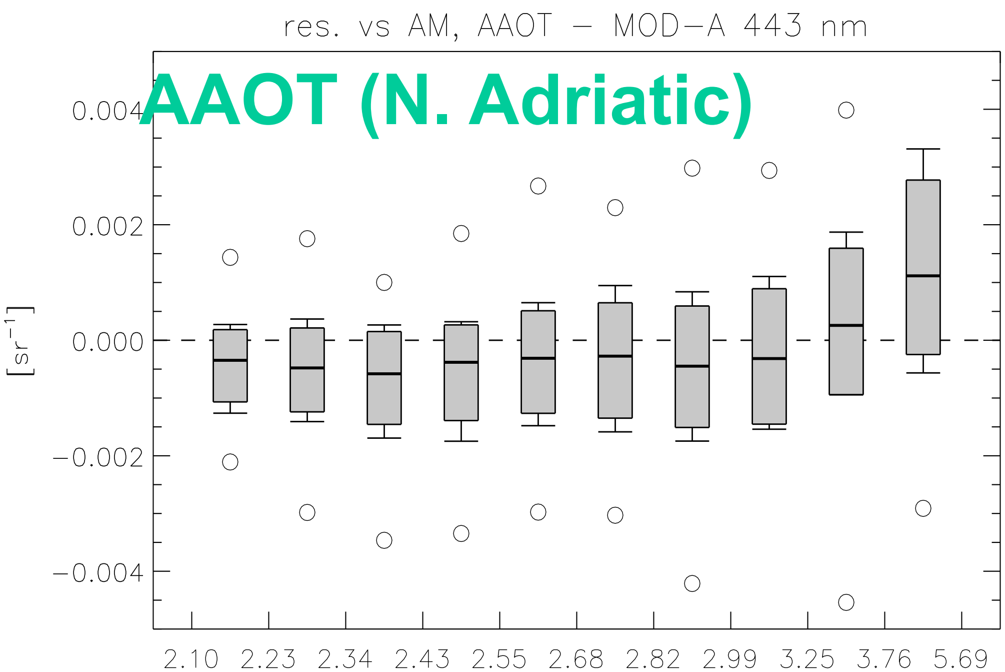
❖ Do validation results change across sites/missions?

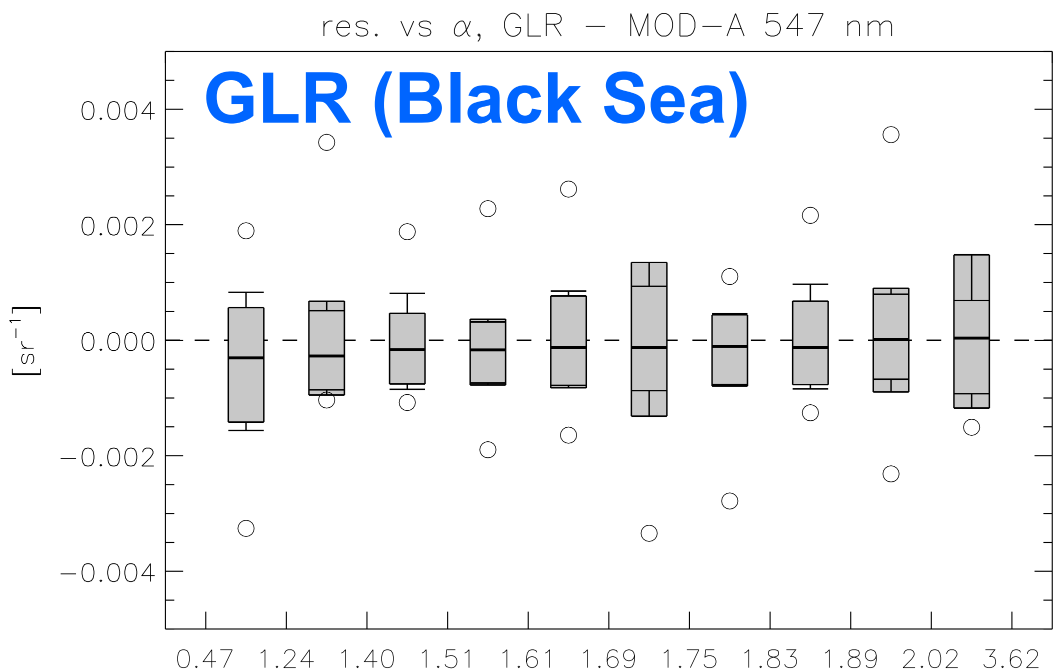




Mélin *RSE* 2022

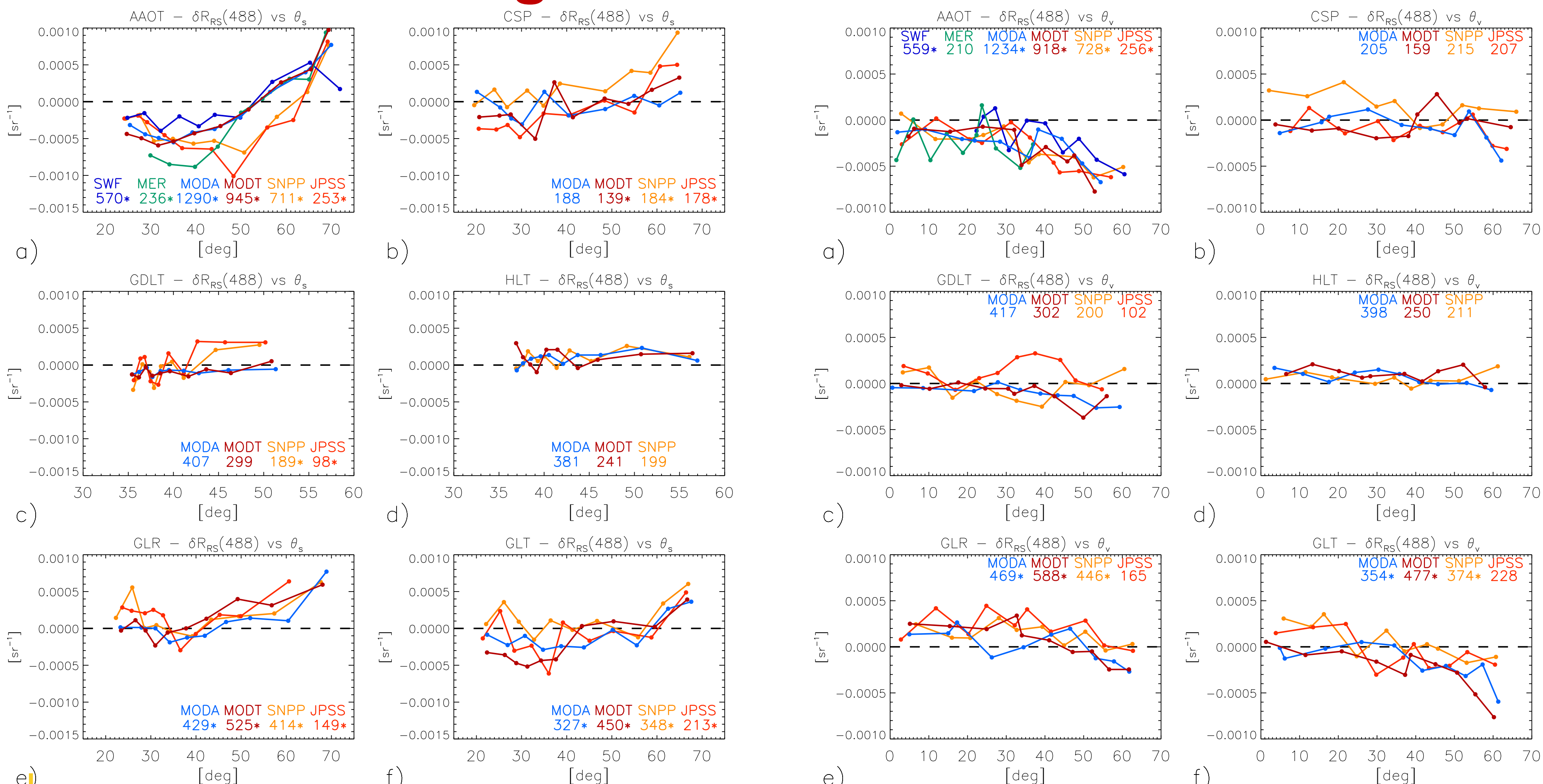
❖ Do results change with environmental conditions?





Residuals vs Air Mass **vs Ångström Exponent**

❖ Do results change with environmental conditions?



Mélin *RSE* 2022

❖ Spectral Correlation of Residuals

Inter-band correlation of the residuals

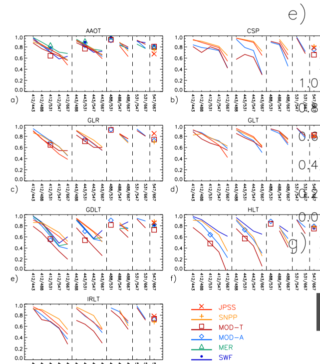
Implications for uncertainty propagation through bio-optical algorithms

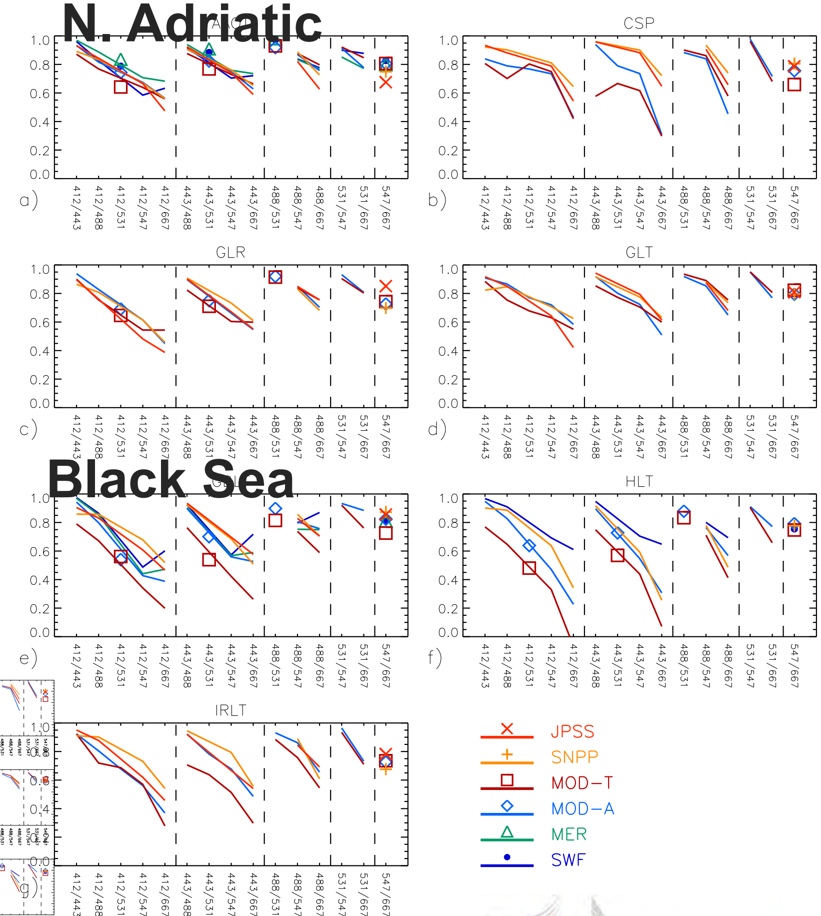
Constituents = function of R_{RS}

$$C = f(R_{RS}(\lambda_1), R_{RS}(\lambda_2))$$

Standard uncertainty for C, u(c):

$$u_c^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 u^2(x_1) + \left(\frac{\partial f}{\partial x_2}\right)^2 u^2(x_2) + 2\frac{\partial f}{x_1}\frac{\partial f}{x_2} u(x_1)u(x_2) r(\varepsilon_1, \varepsilon_2)$$





Mélin *RSE* 2022

European Commission

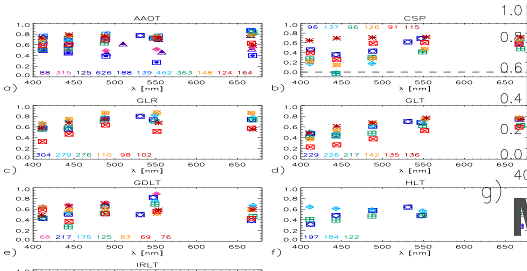
❖ Are residuals correlated across missions?

Implications for uncertainty propagation to multi-mission products

Merging by averaging of missions A and B:

Standard uncertainty for (A+B)/2

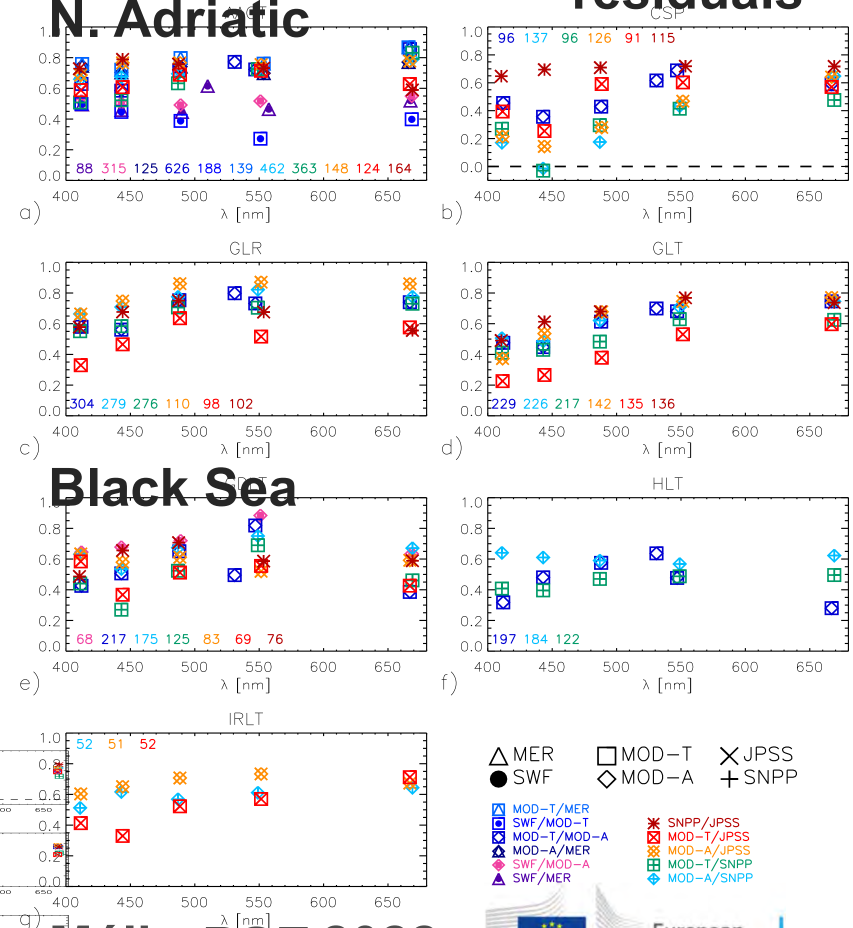
$$u_c^2 = \frac{1}{4}u^2(x_A) + \frac{1}{4}u^2(x_B) + \frac{1}{2}u(x_A)u(x_B)r(\varepsilon_A, \varepsilon_B)$$



Inter-mission correlation of residuals

N. Adriatic

Black Sea



Mélin *RSE* 2022

European Commission

Bonus on Collocation Statistics

European Commission

Collocation : with 2 Similar Data Sets, with error correlation

For i=1,N comparison points:

$$\begin{cases} x_i = r_i + \gamma_i \\ y_i = \beta\, r_i + \delta + \varepsilon_i \end{cases}$$

$$\eta = \frac{\sigma_\varepsilon}{\sigma_\gamma}$$

r: reference values ("truth")
x: in-situ measurements **with σ_γ unknown**
γ: associated random error term (mean=0, **correlated**)
y: satellite values, **with σ_ε unknown**
ε: associated random error term (mean=0, **correlated**)
δ: additive bias
β: multiplicative bias

$$\beta = \frac{\sigma_y^2 - \eta^2\sigma_x^2 + \sqrt{(\sigma_y^2 - \eta^2\sigma_x^2)^2 + 4(\sigma_{xy} - r_{\epsilon\gamma}\eta\sigma_x^2)(\eta^2\sigma_{xy} - r_{\varepsilon\gamma}\eta\sigma_y^2)}}{2(\sigma_{xy} - r_{\varepsilon\gamma}\eta\sigma_x^2)}$$

$r_{\varepsilon\gamma}$: Correlation between ε_i and γ_i

Mélin, *FRS* 2024

Collocation : with 2 Similar Data Sets, with error correlation

$$\begin{cases} \sigma_\gamma^2 = \dfrac{\beta\sigma_x^2 - \sigma_{xy}}{\beta - \eta r_{\varepsilon\gamma}} \\ \\ \sigma_\varepsilon^2 = \dfrac{\sigma_y^2 - \beta\sigma_{xy}}{1 - \beta r_{\varepsilon\gamma}/\eta} \end{cases}$$

$$\Delta_c^2 = (\beta - 1)^2 + \left[\beta(2 - \beta) + \eta^2 - 2\eta r_{\varepsilon\gamma}\right]\sigma_\gamma^2$$

$\longrightarrow \sigma_\gamma^2(1 + \eta^2 - 2\eta r_{\varepsilon\gamma})$ if $\beta \longrightarrow 1$

$\longrightarrow \sigma_\gamma^2(2 - 2r_{\varepsilon\gamma})$ if $\beta \longrightarrow 1$ and $\eta = 1$

Mélin, *FRS* 2024