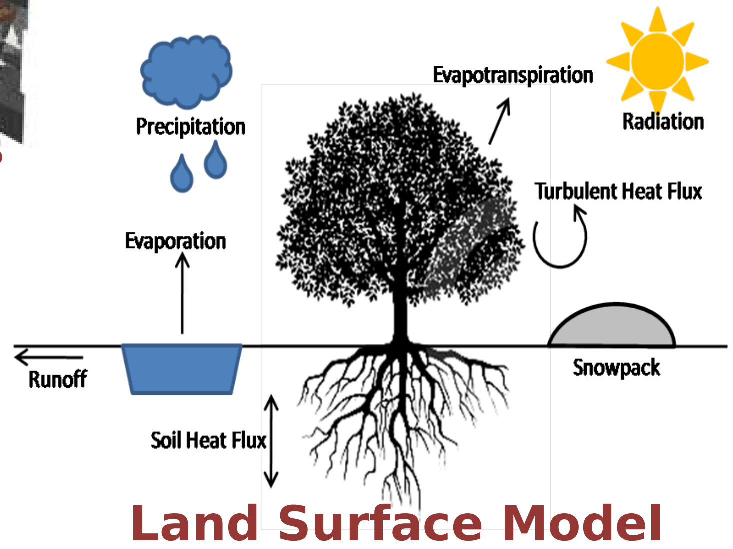
SAC-D/Aquarius Soil Moisture product development and evaluation for Pampas Plains (Argentina)

Cintia Alicia Bruscantini*, Francisco Grings, Federico Carballo, Matias Barber, Pablo Perna, Haydee Karszenbaum.

Remote Sensing Lab, , Institute of Astronomy and Space Physic (IAFE-CONICET-UBA). Buenos Aires, Argentina.

*E-mail: cintiab@iafe.uba.ar

Quebec, Canada – IGARSS 2014



Goal: Develop soil moisture (*sm*) retrieval algorithms for Aquarius/SAC-D, compare them with existing retrieval algorithms and with available *sm* products (SMOS, Aquarius, GLDAS).

Test theoretical and statistical approaches that uses satellite based data for the retrieval, using as benchmark a product derived from a land surface model.

SM retrievals

Theoretical approaches (ω-τ model)

- **SMOS** — *Multiple-incidence-angle dual channel algorithm*
- **Aquarius**
  - SCA
    - H
    - V
  - MPDA
  - Bayes
    - Mean
    - MAP
  - ANN

Single channel algorithms

Dual channel algorithms

Statistical approach → ANN

Process model → **GLDAS**

- **GLDAS** — *Land Surface Model (NOAH 3 hourly 1°x1°) Version 1*

SM retrievals

- **SMOS** ← *Multiple-incidence-angle dual channel algorithm*
- **Aquarius**
  - SCA
    - H
    - V
  - MPDA
  - Bayes
    - Mean
    - MAP
  - ANN
- **GLDAS** ← *Land Surface Model (NOAH 3 hourly 1°x1°)*

Theoretical approaches (ω-τ model): SMOS, Aquarius SCA, MPDA, Bayes

Statistical approach → ANN

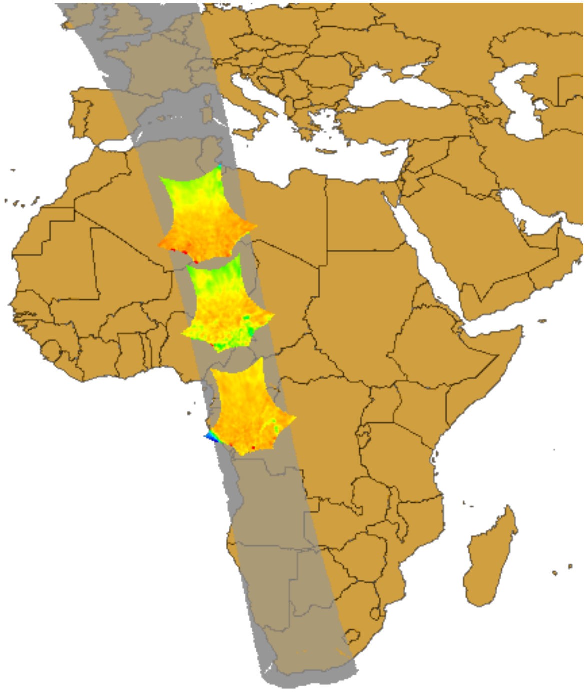
Process model → GLDAS

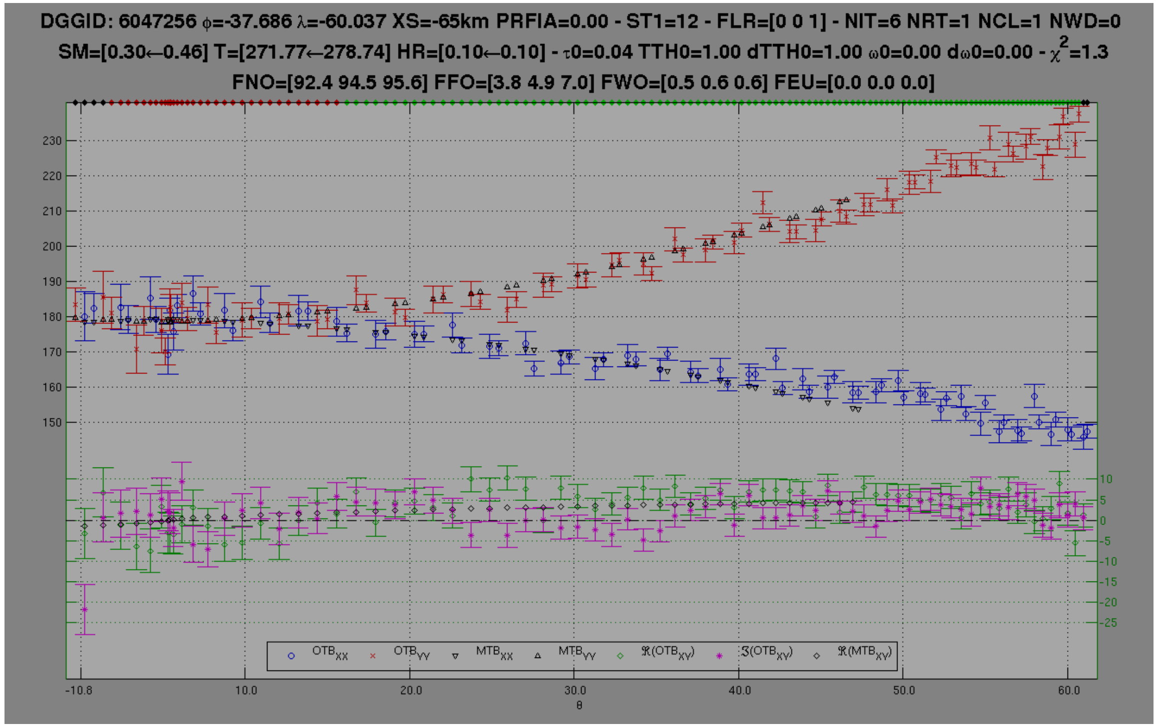
Single channel algorithms: SCA (H, V)

Dual channel algorithms: MPDA, Bayes (Mean, MAP), ANN

SMOS

SMOS provides multi-angular L-band observations. The retrieval algorithm is based on the minimization of the difference between measured and simulated brightness temperature.
Temporal resolution: 3 days
Spatial resolution: 35-60 km
Grid: EASE 25 km
Version 5.5.1





SM retrievals

- **SMOS** — *Multiple-incidence-angle dual channel algorithm*
- **Aquarius**
  - SCA
    - H
    - V
  - MPDA
  - Bayes
    - Mean
    - MAP
  - ANN
- **GLDAS** — *Land Surface Model (NOAH 3 hourly 1°x1°)*

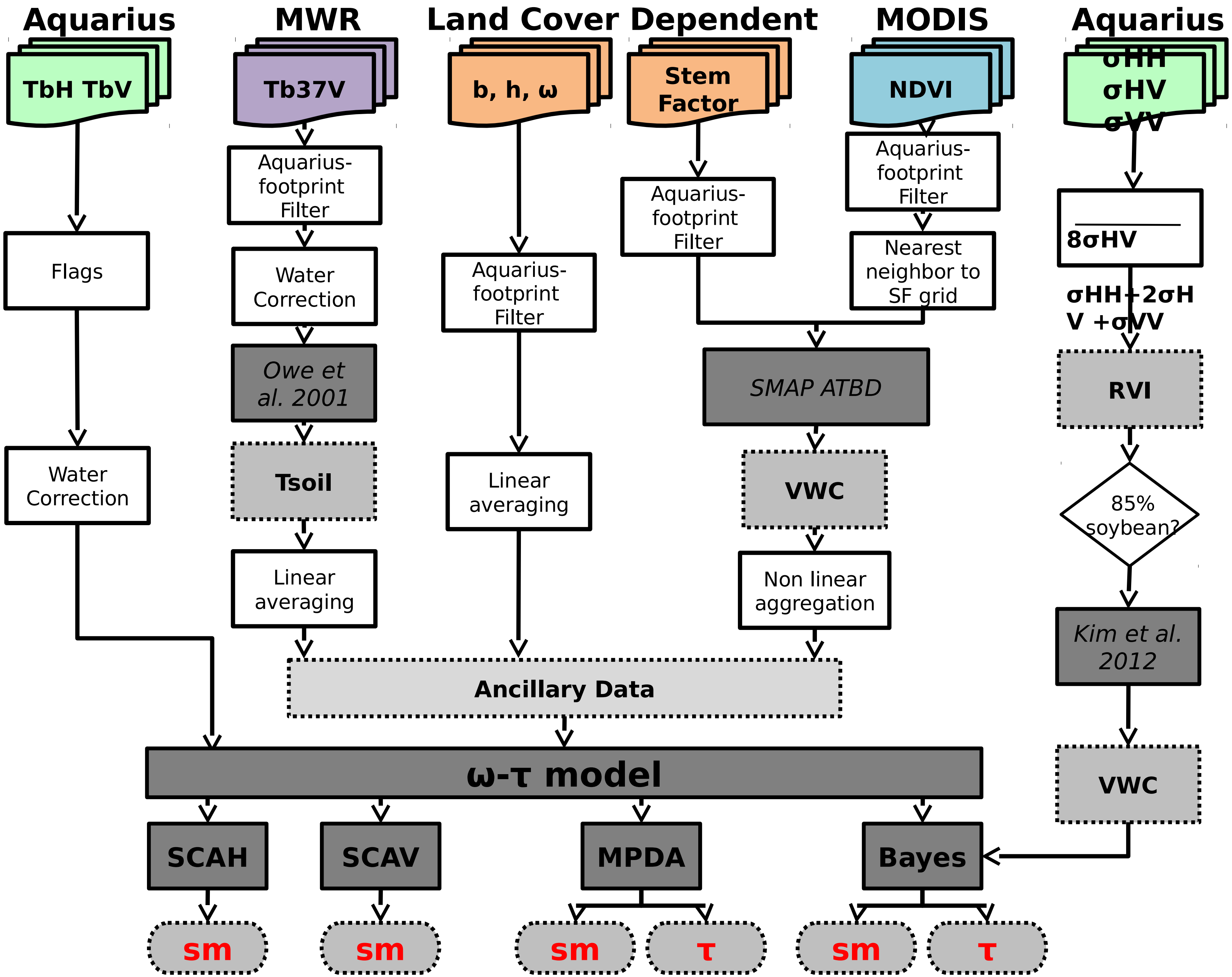
Single channel algorithms

Dual channel algorithms

Theoretical approaches (ω-τ model)

Statistical approach

Process model



Aquarius
MWR
Land Cover Dependent
MODIS
Aquarius
TbH TbV
Tb37V
b, h, ω
Stem Factor
NDVI
σHH
σHV
σVV
Aquarius-footprint Filter
Aquarius-footprint Filter
Aquarius-footprint Filter
Aquarius-footprint Filter
Flags
Water Correction
Nearest neighbor to SF grid
8σHV
σHH+2σHV +σVV
Owe et al. 2001
SMAP ATBD
RVI
Tsoil
VWC
Linear averaging
85% soybean?
Water Correction
Linear averaging
Non linear aggregation
Kim et al. 2012
Ancillary Data
ω-τ model
VWC
SCAH
SCAV
MPDA
Bayes
sm
sm
sm
τ
sm
τ

SM retrievals

Theoretical approaches (ω-τ model)

- **SMOS** — *Multiple-incidence-angle dual channel algorithm*
- **Aquarius**
  - SCA
    - H
    - V
  - MPDA
  - Bayes
    - Mean
    - MAP
  - ANN

Single channel algorithms (SCA H, V)

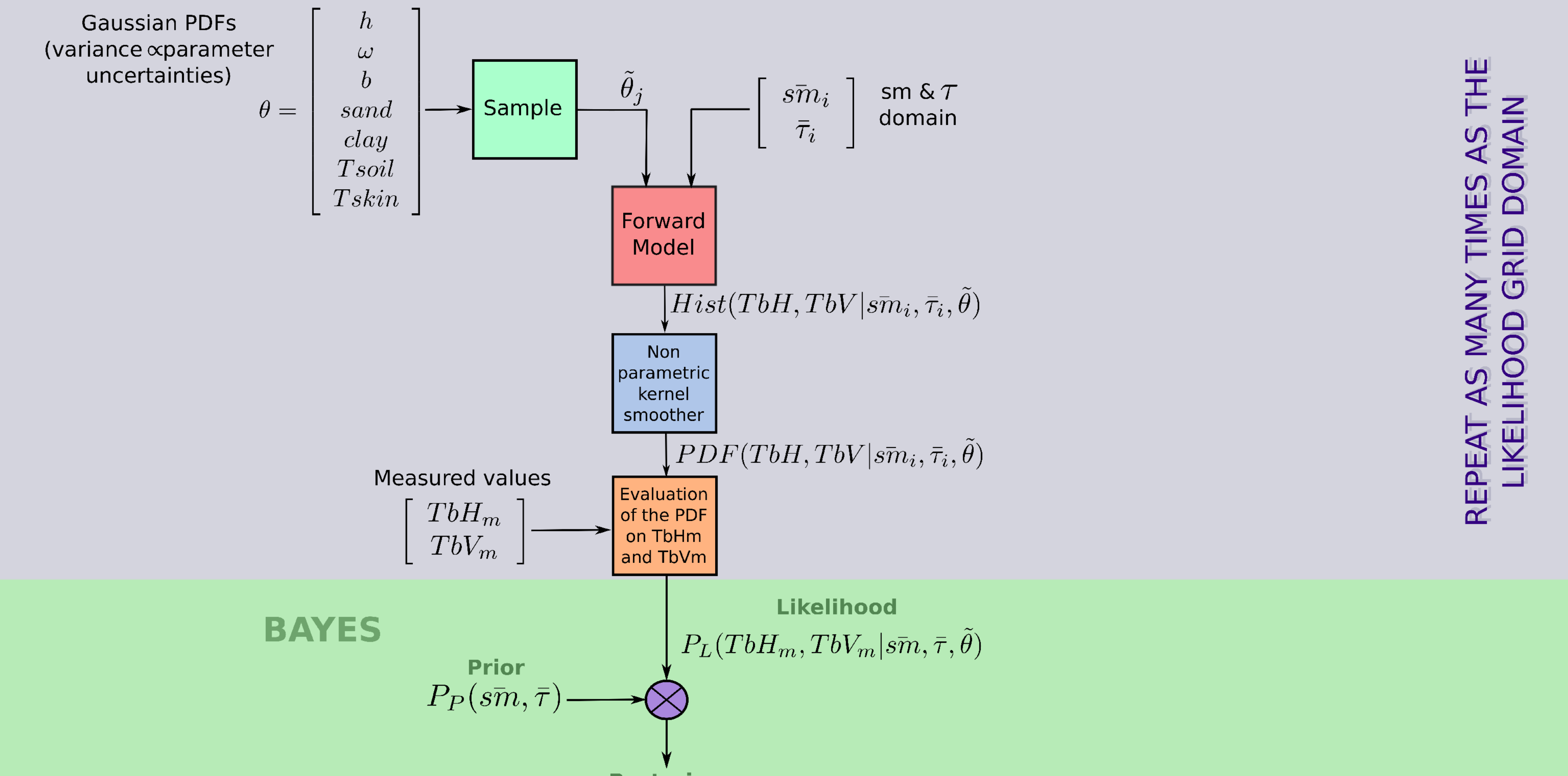
Dual channel algorithms (MPDA, Bayes, ANN)

Statistical approach → ANN

Process model →

- **GLDAS** — *Land Surface Model (NOAH 3 hourly 1°x1°)*

Bayesian Algorithm



$$P_Z(\bar{sm}, \bar{\tau}|TbH_m, TbV_m, \tilde{\theta}) = \frac{P_L(TbH_m, TbV_m|\bar{sm}, \bar{\tau}, \tilde{\theta})P_P(\bar{sm}, \bar{\tau})}{\int\int P_L(TbH_m, TbV_m|\bar{sm}, \bar{\tau}, \tilde{\theta})P_P(\bar{sm}, \bar{\tau})dsm\, d\tau}$$

ESTIMATORS

MAP

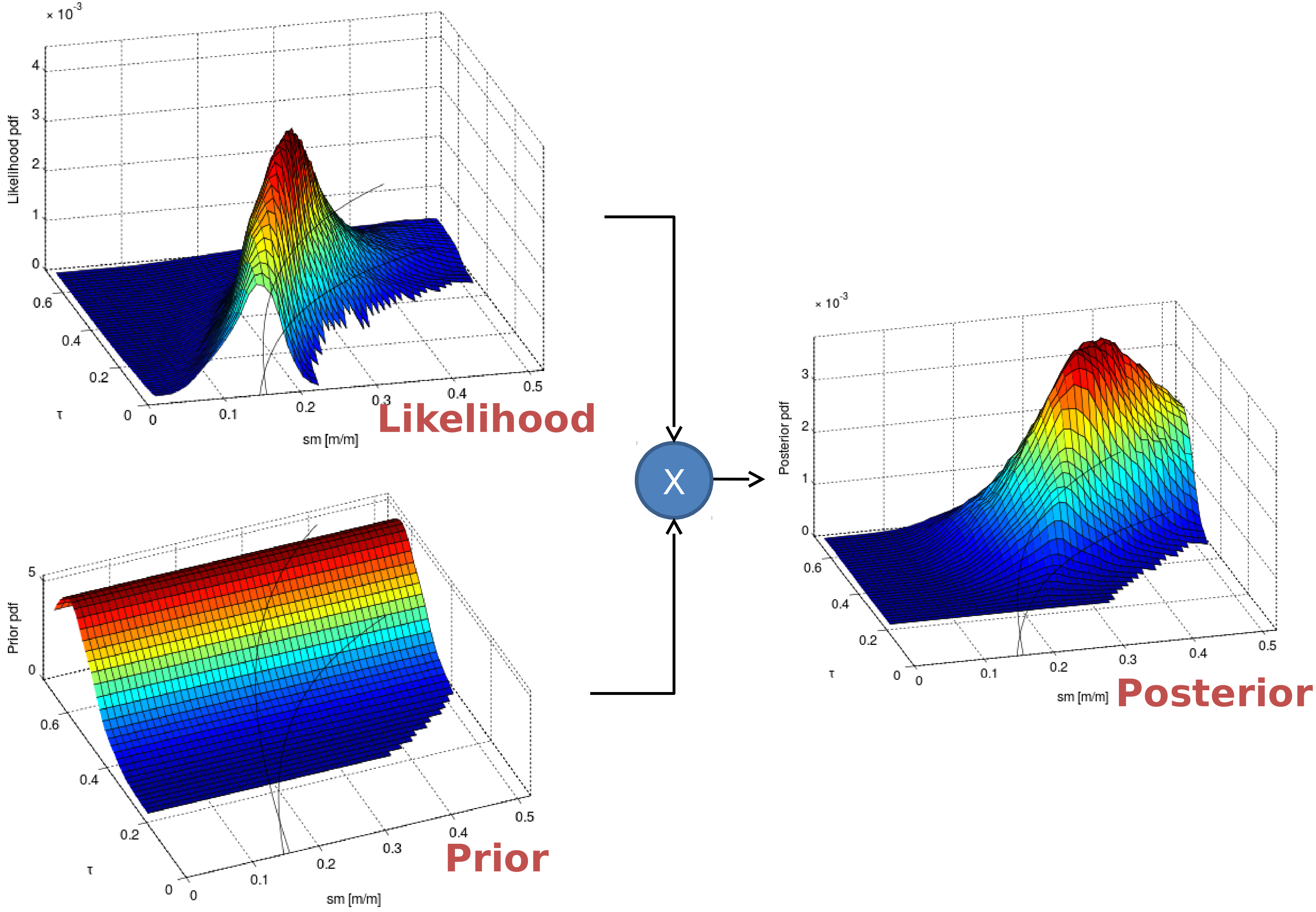
$$\hat{sm}_{map} = \arg\max_{sm} P_Z(\bar{sm}, \bar{\tau}|TbH_m, TbV_m, \tilde{\theta})$$

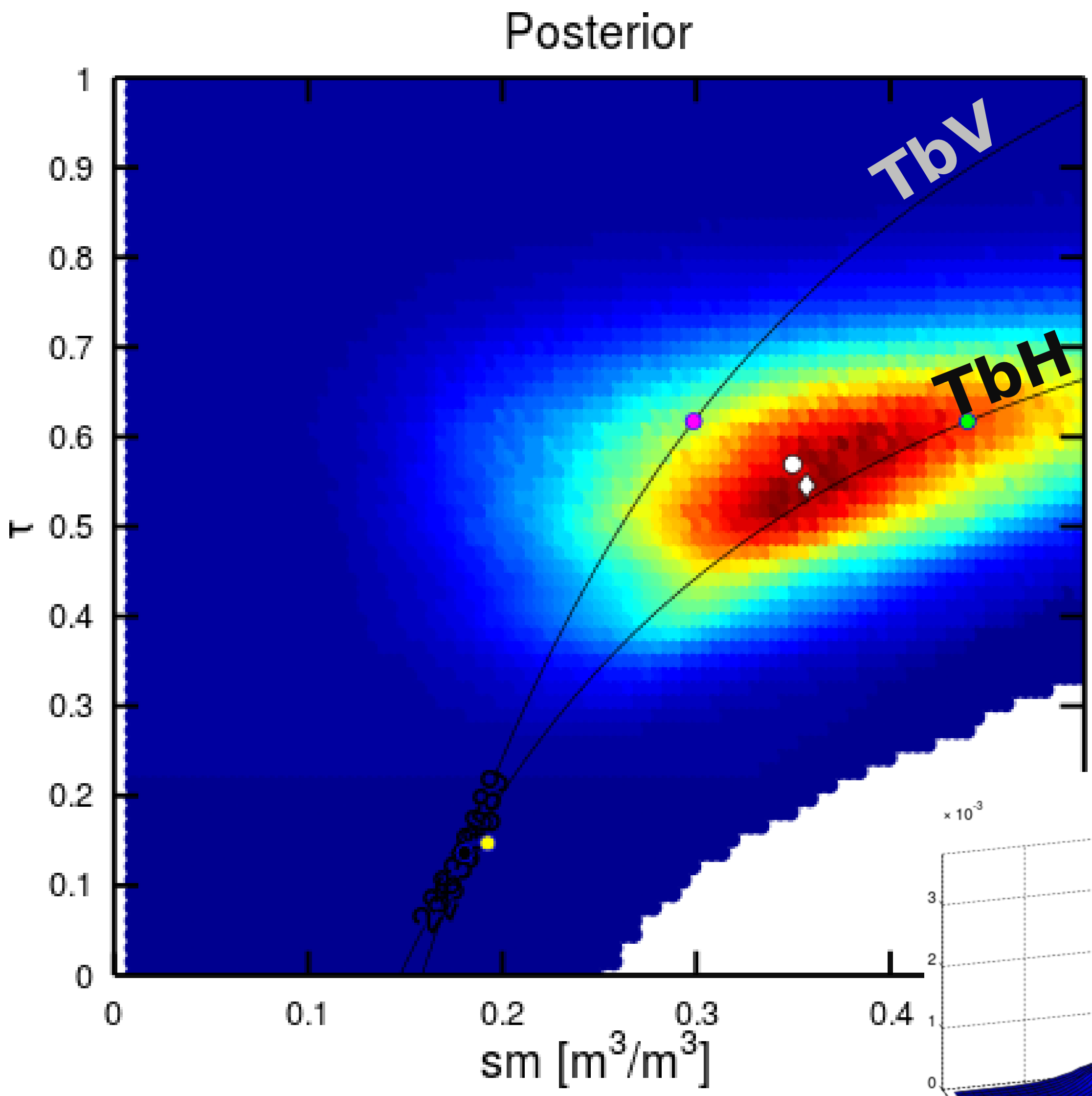
$$\sigma^2_{\hat{sm}_{map}} = \sigma^2_{\hat{sm}_{mean}} + (\hat{sm}_{mean} - \hat{sm}_{map})^2$$

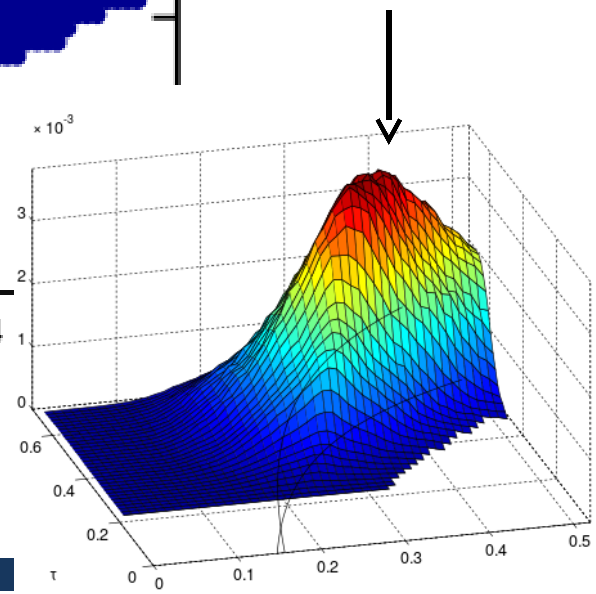
MEAN

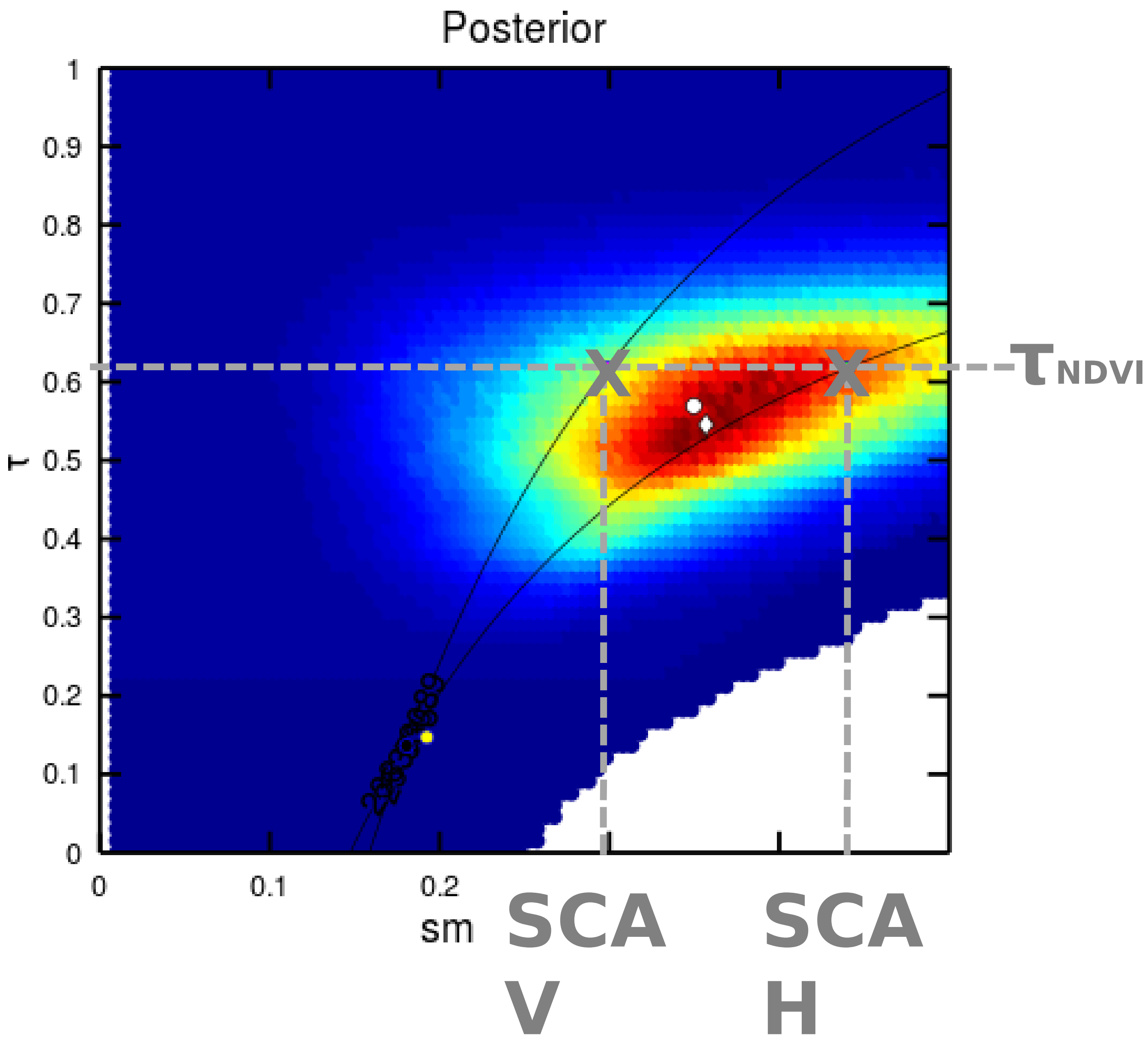
$$\hat{sm}_{mean} = \int\int sm P_Z(\bar{sm}, \bar{\tau}|TbH_m, TbV_m, \tilde{\theta})dsm\, d\tau$$

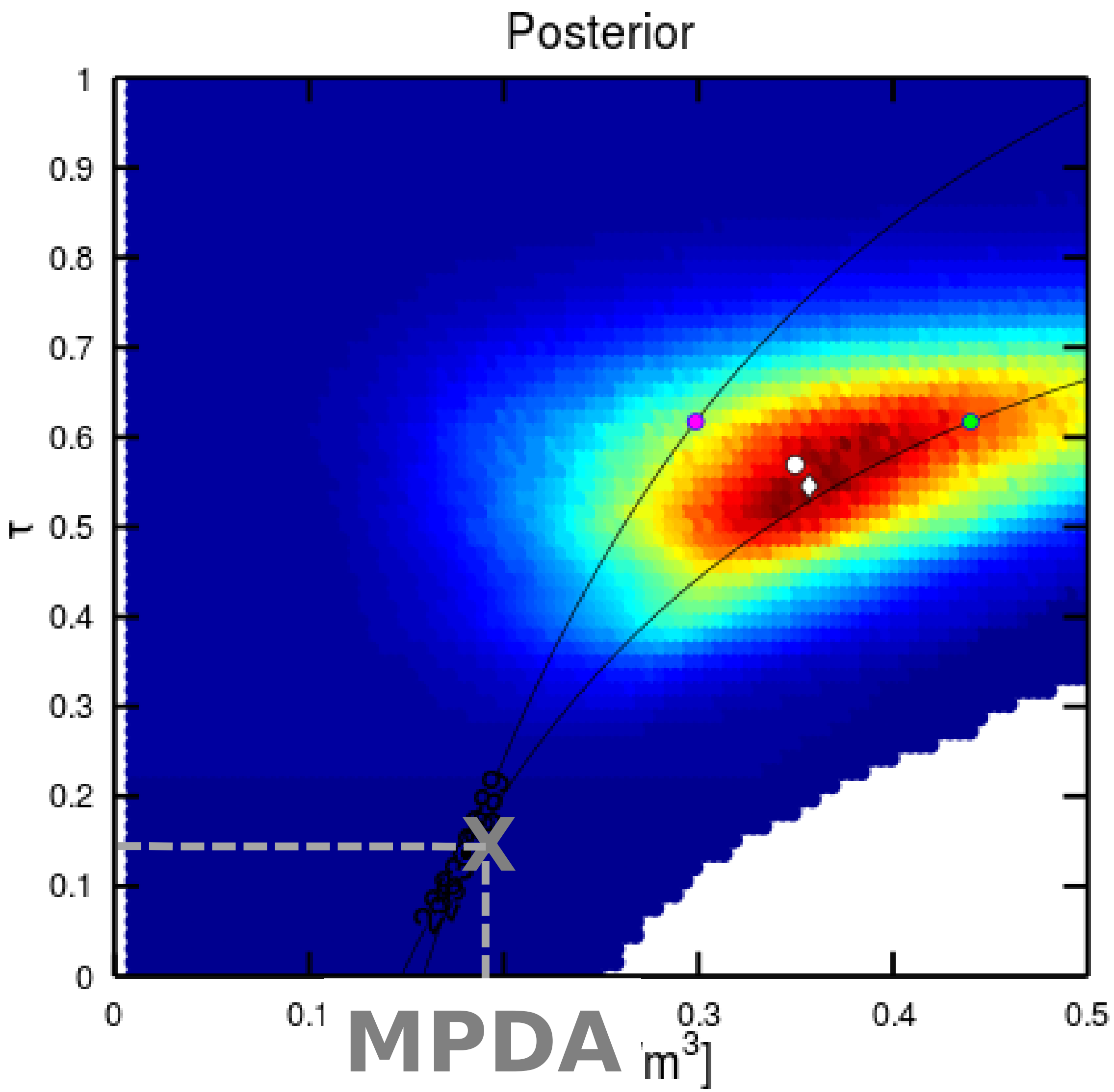
$$\sigma^2_{\hat{sm}_{mean}} = \int\int (sm - \hat{sm}_{mean})^2 P_Z(\bar{sm}, \bar{\tau}|TbH_m, TbV_m, \tilde{\theta})dsm\, d\tau$$

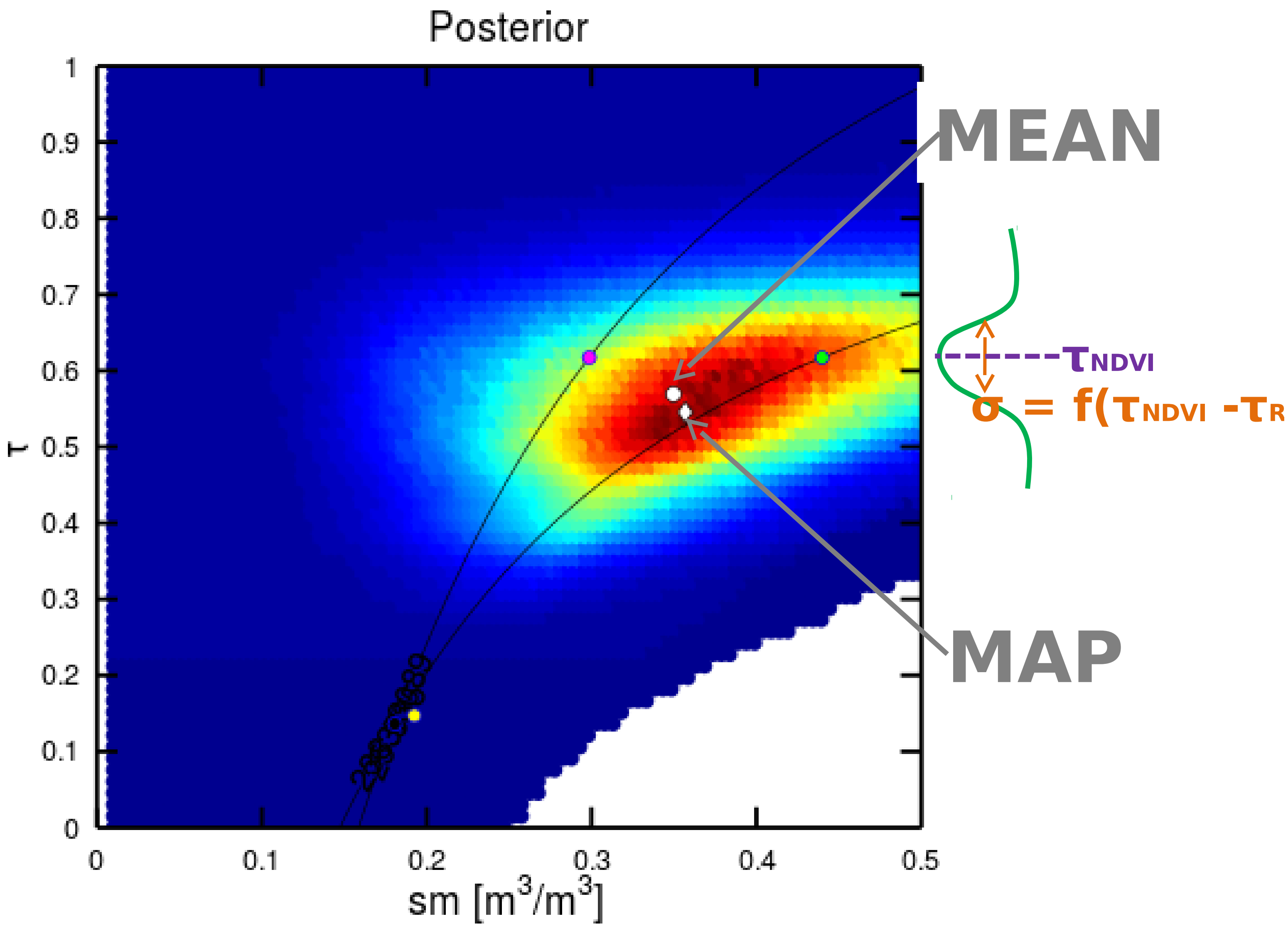

× 10-3
Likelihood pdf
τ
sm [m/m]
Likelihood
X
Prior pdf
τ
sm [m/m]
Prior
× 10-3
Posterior pdf
τ
sm [m/m]
Posterior

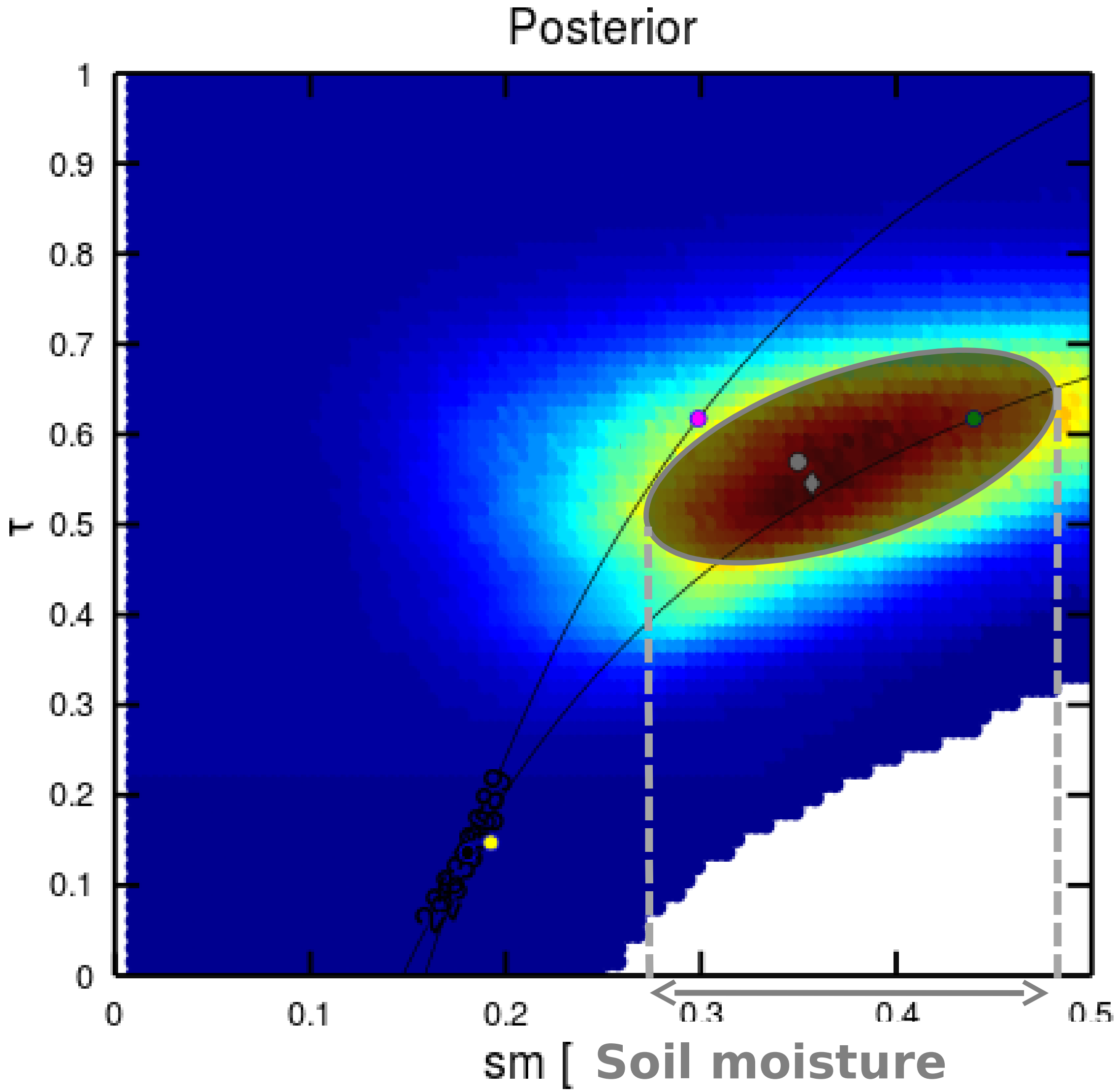

Posterior
TbV
TbH
τ
sm [m^3/m^3]




Posterior
τ
sm
τNDVI
SCA V
SCA H


Posterior
τ
MPDA
X
m³]


Posterior
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
τ
0
0.1
0.2
0.3
0.4
0.5
sm [m^3/m^3]
MEAN
MAP
τNDVI
σ = f(τNDVI -τR



Contour: confidence level (1-α)%

Soil moisture interval for the selected confidence level

The *advantages* of **Bayesian Approach** are:

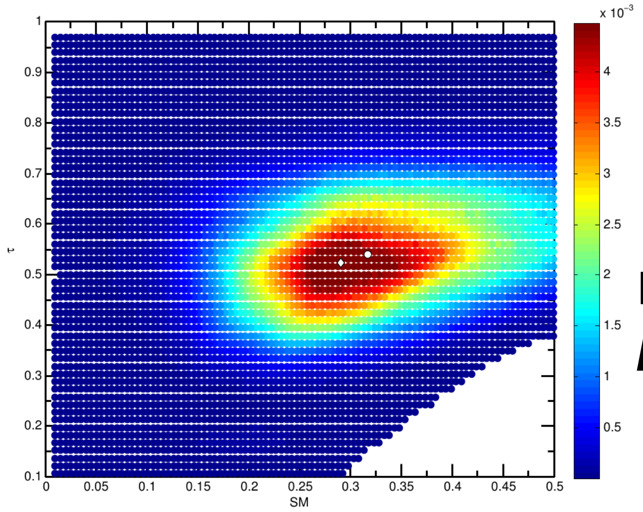
i) **Errors** on the retrieved variables can be estimated in an univocal way,

ii) It gives the possibility to use **prior** information about the retrieved variables (provided by other sensors or in situ historical data),

iii) It can handle **uncertainties** on the ancillary parameters.

The main *disadvantage* of **Bayesian Approach** is:

TIME PERFORMANCE!

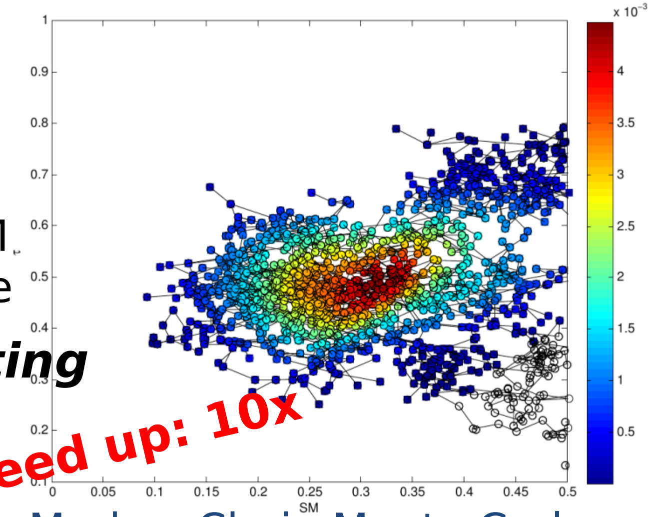


So much processing takes a lot of time, it won't be possible to make a global SM product, so how can we solve this?

h Performance Computing

MPI

Speed up: 10x



Markov Chain Monte Carlo

with chains running in parallel

SM retrievals

- **SMOS** ← *Multiple-incidence-angle dual channel algorithm*
- **Aquarius**
  - SCA
    - H
    - V
  - MPDA
  - Bayes
    - Mean
    - MAP
  - ANN
- **GLDAS** ← *Land Surface Model (NOAH 3 hourly 1°x1°)*

Theoretical approaches (ω-τ model): SMOS, Aquarius (SCA, MPDA, Bayes)

Statistical approach → ANN

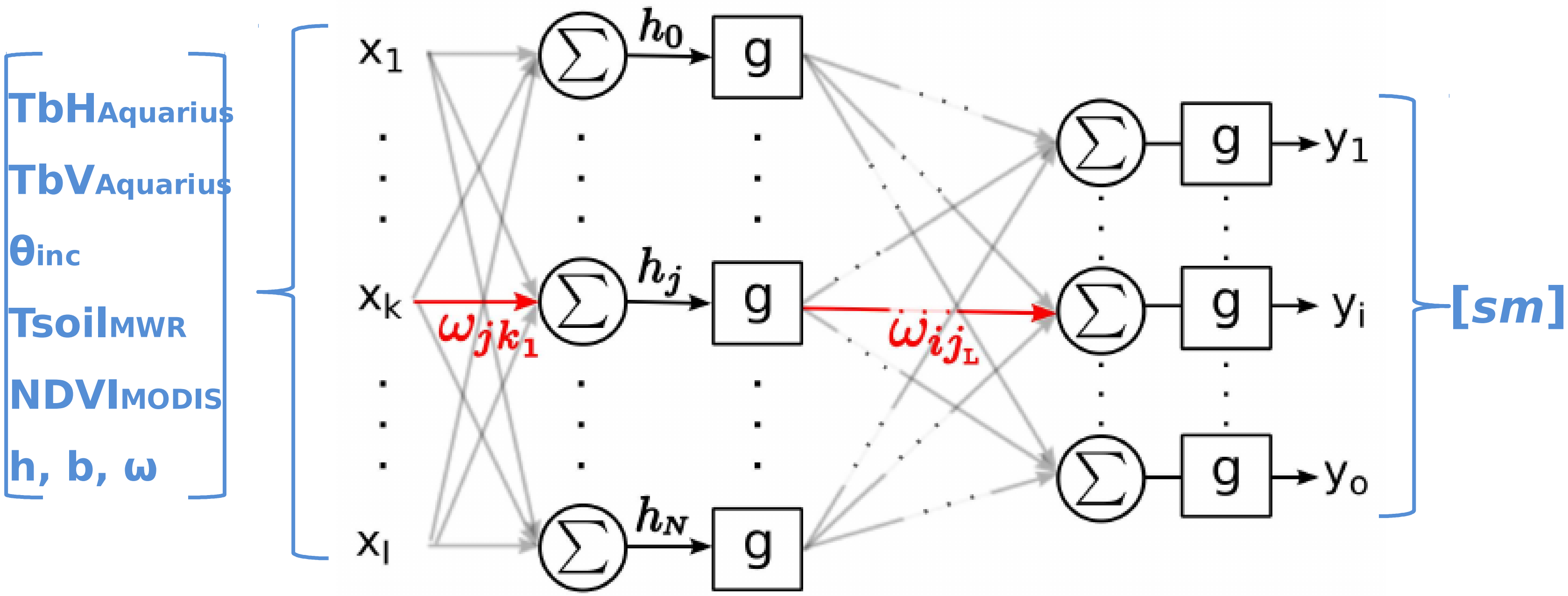
Process model → GLDAS

Single channel algorithms: SCA (H, V)

Dual channel algorithms: MPDA, Bayes (Mean, MAP), ANN

Artificial Neural Network Algorithm

Objective: provide an Aquarius *sm* product consistent to SMOS available Level 2 *sm* product.



Multi-Layer Perceptron Topology

Training Phase

Output target: SMOS L2 *sm*

Training period: January 1st, 2012 to May 1st, 2013 (excluding testing period)

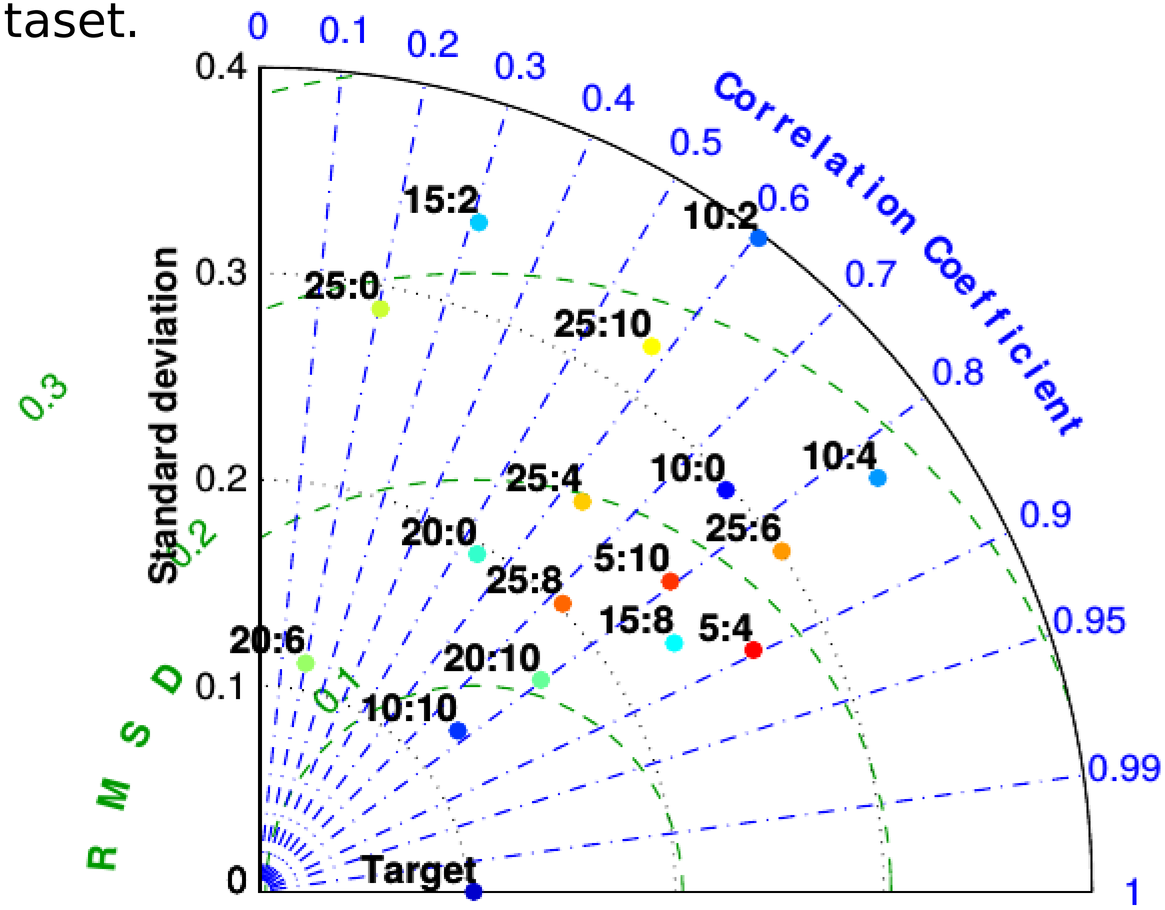
 # training samples used: ~ 7000. *# validation samples used*: ~ 3000.

Learning algorithm: Levenberg-Marquad backpropagation

(input and output datasets are normalized and randomize)

Artificial Neural Network Algorithm

In order to find the optimum ANN topology for Aquarius *sm* retrieval, several ANNs were trained and tested varying the number of hidden layers (one or two) and the number of neurons in each layer (2, 4, 6, 8, 10, 15, 20 and 25). Performance metrics were derived for a testing dataset.



SM retrievals

- **SMOS** ← *Multiple-incidence-angle dual channel algorithm*
- **Aquarius**
  - SCA
    - H
    - V
  - MPDA
  - Bayes
    - Mean
    - MAP
  - ANN
- **GLDAS** ← *Land Surface Model (NOAH 3 hourly 1°x1°)*

Theoretical approaches (ω-τ model): SMOS, Aquarius SCA, MPDA, Bayes

Statistical approach → ANN

Process model → GLDAS

Single channel algorithms: SCA (H, V)

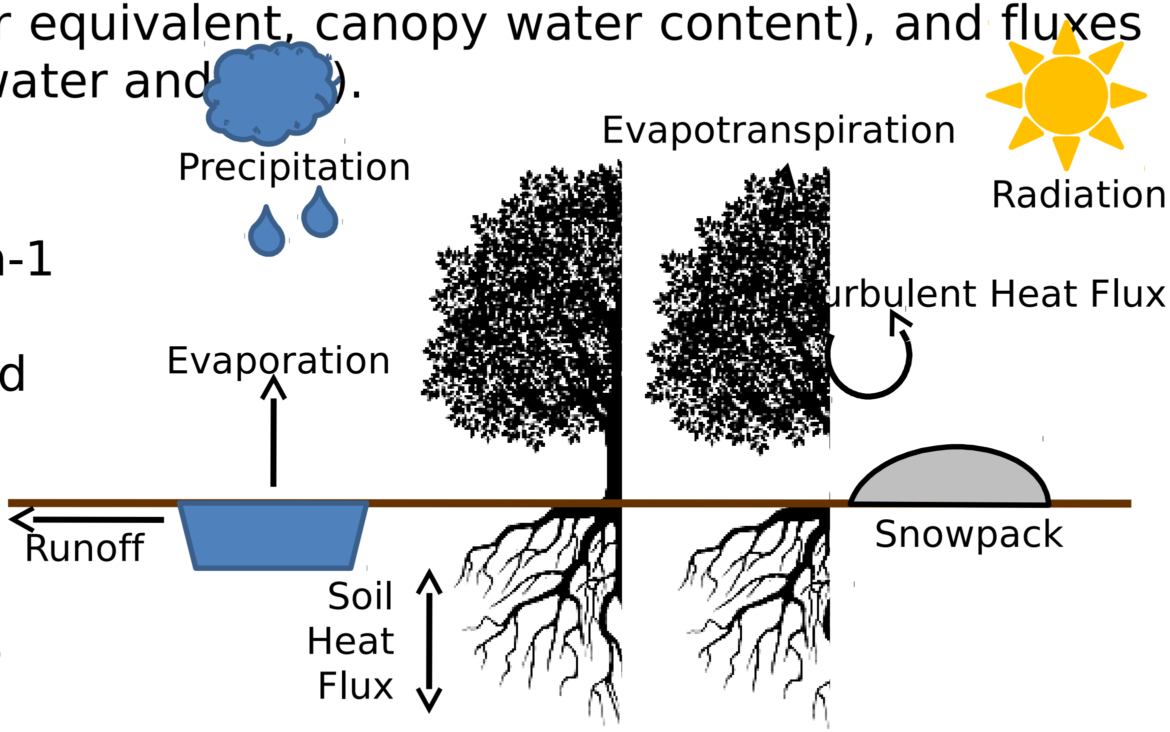
Dual channel algorithms: MPDA, Bayes (Mean, MAP), ANN

GLDAS

The Global Land Data Assimilation System (developed by NASA and NOAA) is a global, offline (uncoupled to the atmosphere) terrestrial modeling system that uses both ground and satellite observations as forcing of advanced land surface models and integrated to data assimilation techniques in order to generate optimal fields of land surface states (soil moisture, both liquid water and ice content, soil temperature, skin temperature, snow depth, snow water equivalent, canopy water content), and fluxes (surface energy, water and).

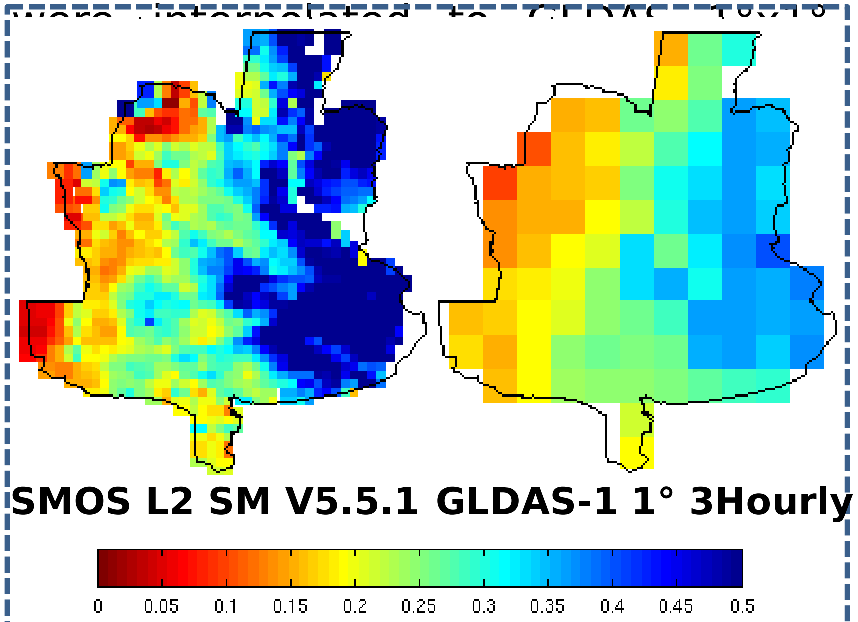
SM data product used:

- GLDAS version-1
- Noah LSM
- 1.0 Degree grid
- 3-Hourly Data (closest to Aquarius overpass)
- 0-10 cm layer

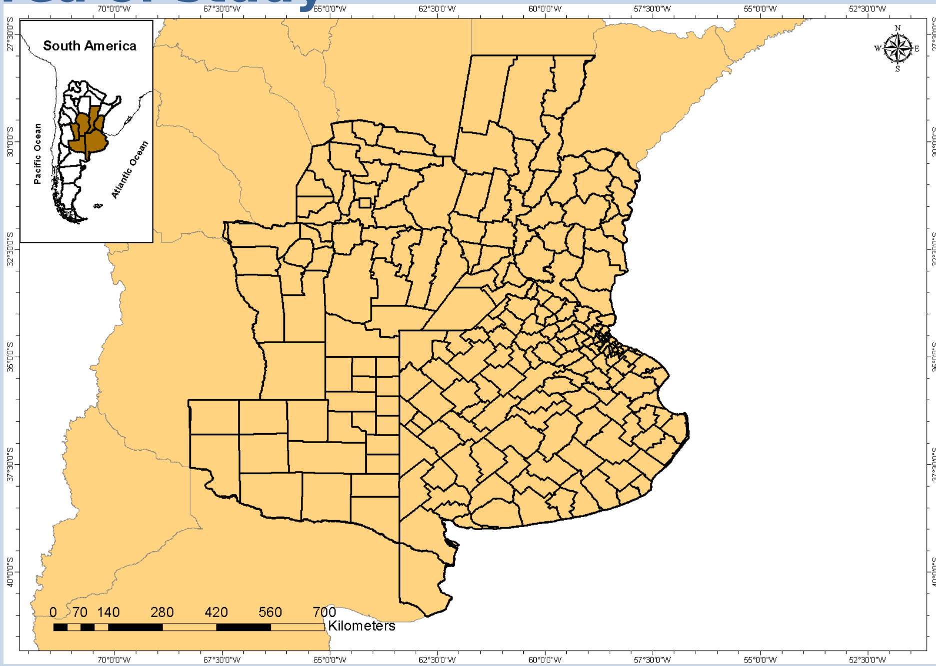


Results

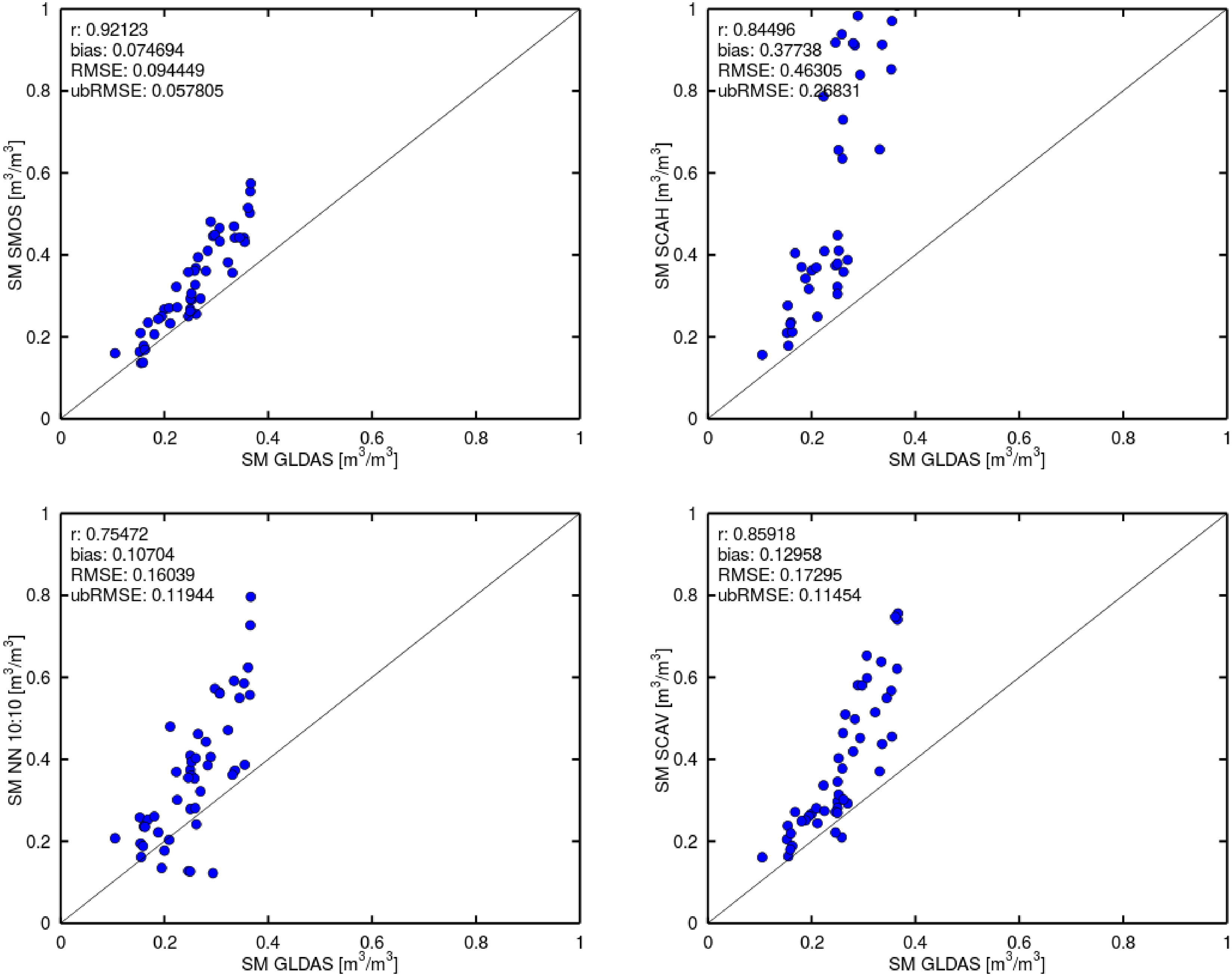
Sm products derived from the Bayesian approach (Mean and MAP), SCAH, SCAV, MPDA, ANN, SMOS and USDA were evaluated through several performance metrics (correlation, bias, root mean square error RMSE, unbiased RMSE) for a day in August 2012 (austral winter, low vegetation, marked dry-wet soil conditions). GLDAS sm was considered as benchmark product. All SM products were interpolated to GLDAS 1°x1°

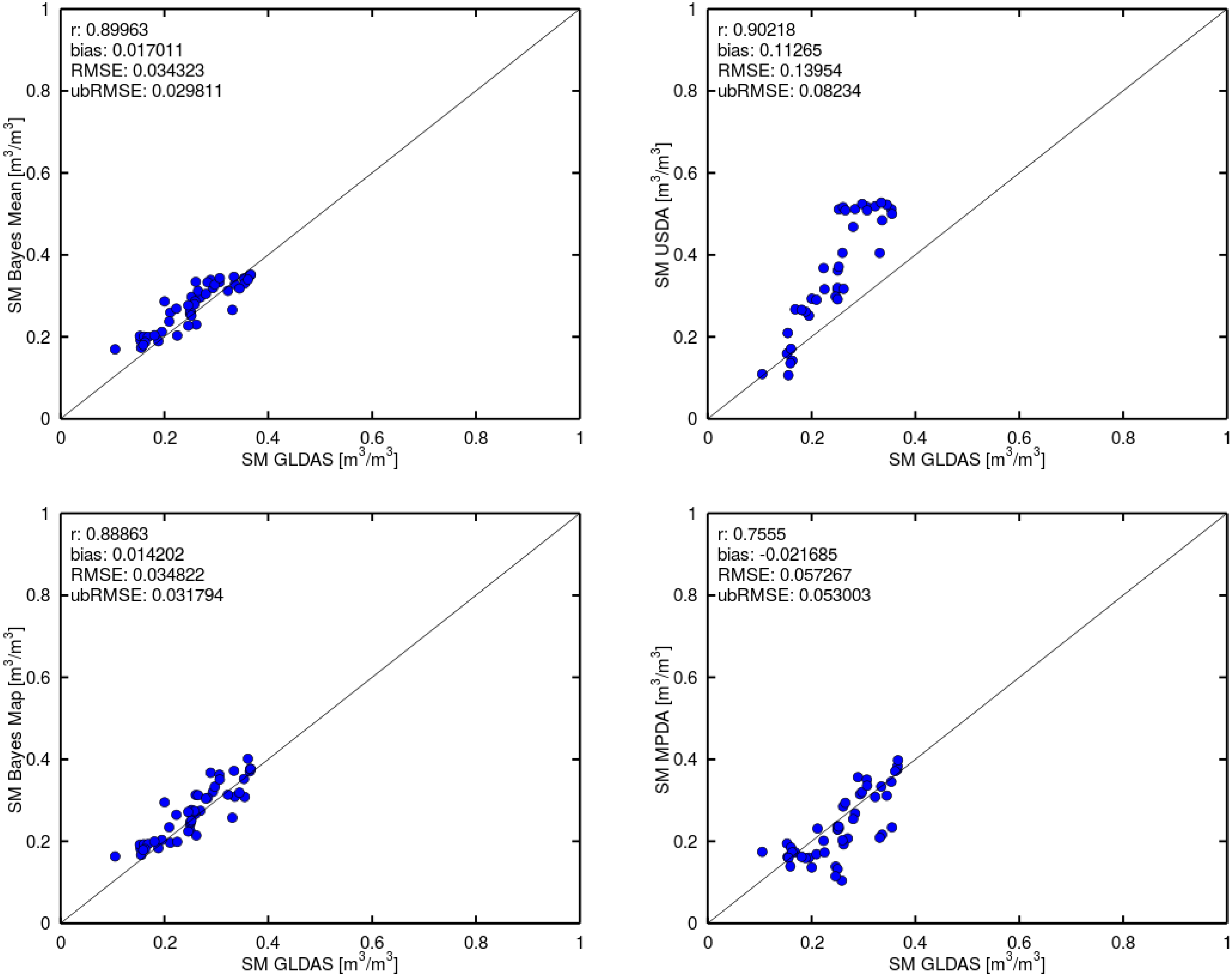


Area of study



The Argentina's Pampas region is a wide plain located in the center-east of Argentina where the main agricultural activities are cereal production and cattle-raising. It extends approximately 50 million hectares of fertile lands and accounts for more than **90% of the national grain production**. Soybean, wheat, maize and sunflower are the main crops. Weather is among the most important and uncontrollable elements affecting agriculture in this region. Most of the Pampas region is significantly affected by **cyclical drought and flood episodes** that impact both crop and cattle production. In general, along the region, the area is **drier in the west and becomes**


r: 0.92123
bias: 0.074694
RMSE: 0.094449
ubRMSE: 0.057805
SM SMOS [m^3/m^3]
SM GLDAS [m^3/m^3]
r: 0.84496
bias: 0.37738
RMSE: 0.46305
ubRMSE: 0.26831
SM SCAH [m^3/m^3]
SM GLDAS [m^3/m^3]
r: 0.75472
bias: 0.10704
RMSE: 0.16039
ubRMSE: 0.11944
SM NN 10:10 [m^3/m^3]
SM GLDAS [m^3/m^3]
r: 0.85918
bias: 0.12958
RMSE: 0.17295
ubRMSE: 0.11454
SM SCAV [m^3/m^3]
SM GLDAS [m^3/m^3]


r: 0.89963
bias: 0.017011
RMSE: 0.034323
ubRMSE: 0.029811
SM Bayes Mean [m3/m3]
SM GLDAS [m3/m3]
r: 0.90218
bias: 0.11265
RMSE: 0.13954
ubRMSE: 0.08234
SM USDA [m3/m3]
SM GLDAS [m3/m3]
r: 0.88863
bias: 0.014202
RMSE: 0.034822
ubRMSE: 0.031794
SM Bayes Map [m3/m3]
SM GLDAS [m3/m3]
r: 0.7555
bias: -0.021685
RMSE: 0.057267
ubRMSE: 0.053003
SM MPDA [m3/m3]
SM GLDAS [m3/m3]

Performance Metrics

	Mean	Map	MPDA	USDA	SCAH	SCAV	ANN	SMOS
r	0.900	0.890	0.756	0.902	0.845	0.859	0.755	0.921
bias	0.017	0.014	-0.022	0.113	0.377	0.130	0.107	0.075
RMSE	0.034	0.035	0.057	0.140	0.463	0.173	0.160	0.094
ubRMSE	0.030	0.032	0.053	0.082	0.268	0.115	0.119	0.058

Metrics Computation

$$r = \frac{E[(sm_{est} - E[sm_{est}])(sm_{GLDAS} - E[sm_{GLDAS}])}{\sigma_{est}\sigma_{GLDAS}}$$

$$bias = E[sm_{est}] - E[sm_{GLDAS}]$$

$$RMSE = \sqrt{E[(sm_{est} - sm_{GLDAS})^2]}$$

$$ubRMSE = \sqrt{E\{[(sm_{est} - E[sm_{est}]) - (sm_{GLDAS} - E[sm_{GLDAS}])]^2\}}$$

Summary

A **new retrieval** model based on a Bayesian approach (**BRA**) was proposed for Aquarius soil moisture product. Existing soil moisture models were also implemented. The performances of all the implemented models were evaluated using GLDAS sm as benchmark. Results were analyzed in terms of standard metrics.

Conclusions

✓The proposed methodology showed encouraging results. Of all the tested algorithms, the **BRA approach exhibited the best performance** with the lowest bias and ubRMSE.

✓The high SM values obtained on SCA (H & V) are due to **overestimation of optical depth**. Such overestimation indicates the presence of a bias on VWC and/or *b* vegetation parameter. Either way, *b* value and the methodology to derive VWC from NDVI follow the SMAP ATBD and should be considered for SMAP passive SM retrieval. Furthermore, SCAV showed a better performance, thus results might improve by using *b*H≠*b*V (***b*H<*b*V**).

✓<u>SM dynamic range</u>:

- BRA (Mean & MAP), MPDA and GLDAS display SM values as high as 0.4 m3/m3. BRA highest SM values were conditioned by the selected prior (0.5 m3/m3).
- USDA SCAH saturate SM at field capacity *by design* (depending on soil texture, ~0.55 m3/m3).
- SMOS shows values as high as 0.6 m3/m3 at GLDAS grid.
- SCAH, SCAV & ANN displayed values higher than 0.7 m3/m3.

✓<u>SM biases</u>: USDA, SCAH, SCAV, SMOS, MPDA and ANN exhibit high positive

BACK UP SLIDES

Why uncertainties?

Gaussian PDFs
(variance ∝ parameter
uncertainties)

$$\theta = \begin{bmatrix} h \\ \omega \\ b \\ sand \\ clay \\ Tsoil \\ Tskin \end{bmatrix}$$

Sample

$\tilde{\theta}_j$

Source	λ (cm)	f (GHz)	b (m²/kg)	Canopy Type	Comments/ Incidence Angle
Wang et al. [32], [33]	6	5	1.77	Short grass	10°
	21	1.4	0.3	,,	,,
Wang et al. [32], [33]	6	5	1.99	Tall grass	10°
	21	1.4	0.72	,,	,,
Wang et al. [31]	21	1.4	0.15	,,	10°
2nd part					
Pardé et al. [20]	21	1.4	0.26 ± 0.04	Corn	10° to 60°
Wigneron et al. [36]	6	5.0	0.41 ± 0.04	Soybean	38°
	21	1.4	0.19 ± 0.01	,,	,,
Haboudane et al. [8]	6	5.0	0.48 ± 0.05	,,	40°
	21	1.4	0.28 ± 0.03	,,	,,
Burke et al. [4]	21	1.4	0.114	,,	10° (τ-ω with ω = 0)
	21	1.4	0.122	,,	10° (discrete model)
Pardé et al. [20]	21	1.4	0.30 ± 0.02	,,	10° to 60°
Van de Griend et al. [28]	6	5	1.53	Short Wheat	10° to 60°
	21	1.4	0.27	,,	,,
Wigneron et al. [36]	6	5.0	0.48 ± 0.08	Wheat	38°
	21	1.4	0.12 ± 0.01	,,	,,
Haboudane et al. [8]	6	5.0	0.32 ± 0.04	,,	40°
	21	1.4	0.13 ± 0.01	,,	,,
Pardé et al. [20]	21	1.4	0.11 ± 0.01	,,	10° to 60°
Pardé et al. [20]	21	1.4	0.54 ± 0.02	Alfalfa	10° to 60°
Pardé et al. [20]	21	1.4	0.56 ± 0.05	Tall grass	10° to 60°
Mätzler [16]	6.1	4.9	0.72	Oat	55° to 80°
	2.88	10.4	3.7	,,	,,
	1.43	21.0	6.1	,,	,,
	0.86	34.8	6.1	,,	,,

Source	λ (cm)	f (GHz)	b (m²/kg)	Canopy Type	Comments/ Incidence Angle
1st part					
Jackson and O'Neill [9]	6	5	0.15	Corn	10° and 20°
	21	1.4	0.115	,,	,,
Pamapl. and Pal. [19]	0.8	37.5	0.6	,,	(-)
	3.1	9.7	0.34	,,	,,
Ulaby et al. [24]	6	5	0.186	,,	0° and 40°
	21	1.4	0.113	,,	,,
Jackson et al. [11]	6	5	0.162	,,	10° and 40°
	21	1.4	0.133	,,	,,
O'Neill et al. [18]	6	5	0.131	,,	10°
	21	1.4	0.102	,,	,,
Jackson and O'Neill [9]	6	5	0.288	Soybean	10° and 20°
	21	1.4	0.086	,,	,,
Ulaby and Wilson [26]	2.8	10.7	1.34	,,	(*)
	6.5	4.6	0.44	,,	,,
	18.2	1.6	0.1	,,	,,
Brunf. and Ulaby [2], [3]	5.9	5.1	0.366	,,	0° to 50° (**)
	11.1	2.7	0.264	,,	,,
Jackson et al. [11]	6	5	0.24	,,	10° and 40°
	21	1.4	0.087	,,	,,
Ulaby and Wilson [26]	2.8	10.7	0.38	Wheat	(*)
	6.5	4.6	0.15	,,	,,
	18.2	1.6	0.05	,,	,,
Kirdiashev et al. [14]	2.25	13.3	0.442	Winter rye	(-)
	10	3	0.229	,,	,,
	20	1.5	0.114	,,	,,
	30	1	0.043	,,	,,
O'Neill et al. [18]	21	1.4	0.105	Sorghum	10°
Chukhl. and Shutko [6]	18	1.7	0.142	Cereals	(-)
Pampal. and Pal. [19]	0.8	37.5	1.85	Alfalfa	(-)
	3.1	9.7	0.93	,,	,,
Chukhl. and Shutko [6]	18	1.7	0.182	,,	,,
O'Neill et al. [18]	6	5	0.138	Sweet	10°
Shutko [22]	3	10	0.475	Broadleaf	(-)
	20	1.5	0.075	,,	,,
Vyas [30]	19.3	1.6	0.092	,,	35°

Why uncertainties?

Gaussian PDFs
(variance ∝parameter
uncertainties)

$$\theta = \begin{bmatrix} h \\ \omega \\ b \\ sand \\ clay \\ Tsoil \\ Tskin \end{bmatrix}$$

Sample

$\tilde{\theta}_j$

$\begin{bmatrix} \bar{sm}_i \\ \bar{\tau}_i \end{bmatrix}$ sm & $\mathcal{T}$ domain

Forward Model

$Hist(TbH, TbV|\bar{sm}_i, \bar{\tau}_i, \tilde{\theta})$

Non parametric kernel smoother

Measured values

$\begin{bmatrix} TbH_m \\ TbV_m \end{bmatrix}$

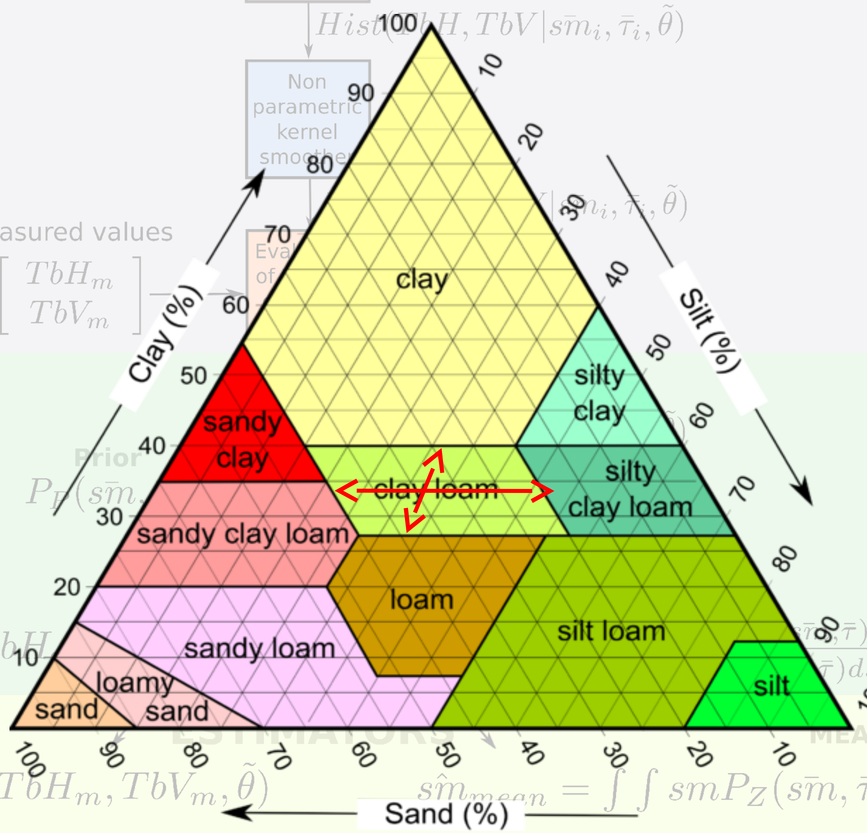
BAYES

Prior

$P_P(\bar{sm}$

$P_Z(\bar{sm}, \bar{\tau}|TbH$

REPEAT AS MANY TIMES AS THE LIKELIHOOD GRID DOMAIN



MAP

ESTIMATORS

MEAN

$$\hat{sm}_{map} = \arg\max_{sm} P_Z(\bar{sm}, \bar{\tau}|TbH_m, TbV_m, \tilde{\theta})$$

$$\hat{sm}_{mean} = \int\int sm P_Z(\bar{sm}, \bar{\tau}|TbH_m, TbV_m, \tilde{\theta}) dsm\, d\tau$$

$$\sigma^2_{\hat{sm}_{map}} = \sigma^2_{\hat{sm}_{mean}} + (\hat{sm}_{mean} - \hat{sm}_{map})^2$$

$$\sigma^2_{\hat{sm}_{mean}} = \int\int (sm - \hat{sm}_{mean})^2 P_Z(\bar{sm}, \bar{\tau}|TbH_m, TbV_m, \tilde{\theta}) dsm\, d\tau$$

Why uncertainties?

Gaussian PDFs
(variance ∝parameter
uncertainties)

$$\theta = \begin{bmatrix} h \\ \omega \\ b \\ sand \\ clay \\ Tsoil \\ Tskin \end{bmatrix}$$

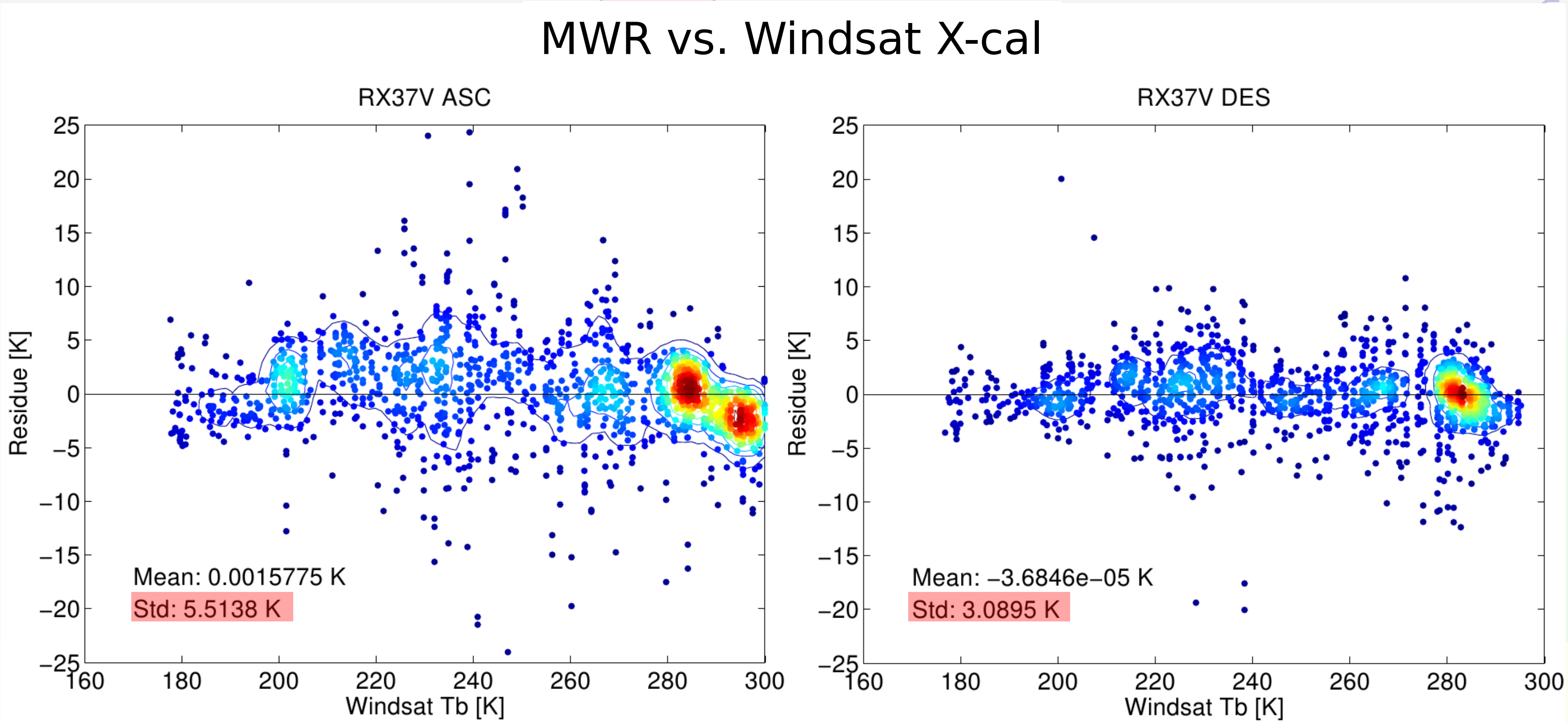
Sample

$\tilde{\theta}_j$

$\begin{bmatrix} s\bar{m}_i \\ \bar{\tau}_i \end{bmatrix}$ sm & τ domain

Forward Model

Y TIMES AS THE LIKELIHOOD GRID DOMAIN



$\hat{sm}_{map} = \arg\max_{sm} P_Z(sm, \tau | TbH_m, TbV_m, \theta)$ $\quad$ $\hat{sm}_{mean} = \int\int sm P_Z(sm, \tau | TbH_m, TbV_m, \theta) dsm\, d\tau$

$\sigma^2_{\hat{sm}_{map}} = \sigma^2_{\hat{sm}_{mean}} + (\hat{sm}_{mean} - \hat{sm}_{map})^2$ $\quad$ $\sigma^2_{\hat{sm}_{mean}} = \int\int (sm - \hat{sm}_{mean})^2 P_Z(s\bar{m}, \bar{\tau} | TbH_m, TbV_m, \tilde{\theta}) dsm\, d\tau$

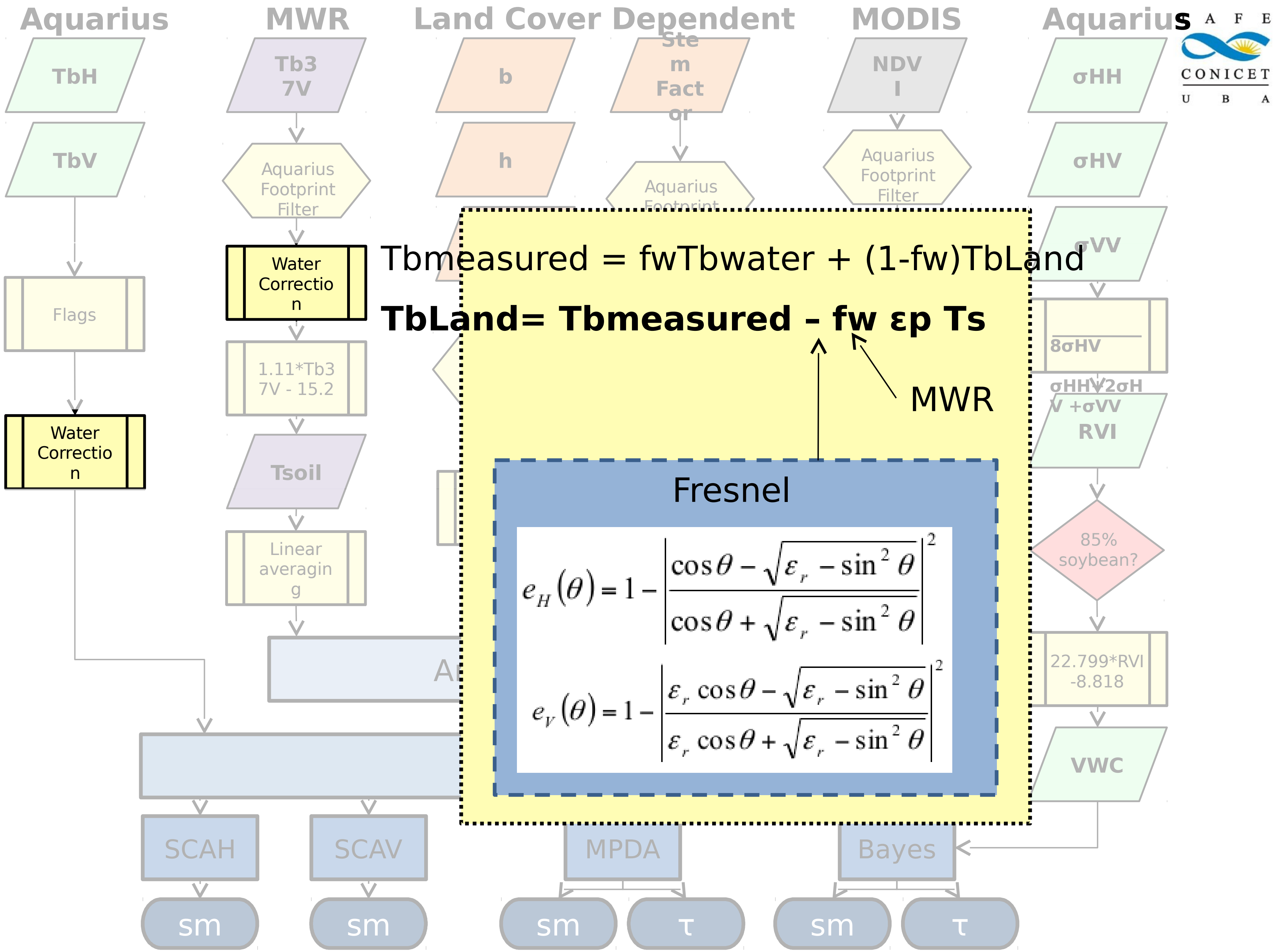
Land cover-dependent parameters

ID	MODIS IGBP land classification	*s*	*h*	*b*	*ω*	*Stem factor*
0	Water Bodies	--	-9999	-9999	-9999	--
1	Evergreen Needleleaf Forests	1.60	0.160	0.100	0.120	15.96
2	Evergreen Broadleaf Forests	1.60	0.160	0.100	0.120	19.15
3	Deciduous Needleleaf Forests	1.60	0.160	0.120	0.120	7.98
4	Deciduous Broadleaf Forests	1.60	0.160	0.120	0.120	12.77
5	Mixed Forests	1.60	0.160	0.110	0.120	12.77
6	Closed Shrublands	1.00	0.110	0.110	0.050	3.00
7	Open Shrublands	1.10	0.110	0.110	0.050	1.50
8	Woody Savannas	1.00	0.125	0.110	0.120	4.00
9	Savannas	1.00	0.156	0.110	0.080	3.00
10	Grasslands	1.56	0.156	0.130	0.050	1.50
11	Permanent Wetlands	1.00	-9999	-9999	-9999	4.00
12	Croplands - Average	1.08	0.108	0.110	0.050	3.50
	- Wheat	0.83	0.083	TBD	TBD	TBD
	- Mixed (Wheat, Barley, Oats)	1.08	0.108	TBD	TBD	TBD
	- Corn	0.94	0.094	TBD	TBD	TBD
	- Soybean	1.48	0.148	TBD	TBD	TBD
13	Urban and Built-up Lands	--	0	0.100	0.030	6.49
14	Crop-land/Natural Vegetation Mosaics	1.30	0.130	0.110	0.065	3.25
15	Snow and Ice	--	0	0	0	0
16	Barren	1.50	0.150	0	0	0

SCAH USDA & SCAH IAFE

Why the differences?

	SCAH USDA	SCAH IAFE
Soil temperature	NCEP	MWR (36.5V)
VWC	MODIS NDVI climatology	MODIS NDVI
b	0.8	land cover dependent
h	0.1	land cover dependent
ω	0.05	land cover dependent
Dielectric mixing model	Wang and Schmugge (1980)	Hallikainen (1985)



Tbmeasured = fwTbwater + (1-fw)TbLand

TbLand= Tbmeasured – fw εp Ts

MWR

Fresnel

$$e_H(\theta) = 1 - \left| \frac{\cos\theta - \sqrt{\varepsilon_r - \sin^2\theta}}{\cos\theta + \sqrt{\varepsilon_r - \sin^2\theta}} \right|^2$$

$$e_V(\theta) = 1 - \left| \frac{\varepsilon_r \cos\theta - \sqrt{\varepsilon_r - \sin^2\theta}}{\varepsilon_r \cos\theta + \sqrt{\varepsilon_r - \sin^2\theta}} \right|^2$$