



Des réseaux de neurones pour estimer l'humidité du sol et la biomasse depuis l'espace

Nemesio Rodríguez-Fernández

Centre d'Etudes Spatiales de la Biosphère (CESBIO),
Toulouse, France

nemesio.rodriguez@cesbio.cnes.fr

- Objectif: montrer un exemple d'utilisation des réseaux de neurones en géosciences en particulier pour l'inversion d'observations de télédétection spatiale
 - Inversion humidité du sol, estimation de la biomasse

Table de matières



- Contexte, problématique scientifique
 - Humidité du sol et contenu en eau de la végétation
- Concepts réseaux de neurones
 - *Feed-forward network*, Perceptron multi-couches
- Exemples d'applications
 - Synergie multi-capteur
 - Longues séries temporelles
 - Chaîne en temps quasi-réel de l'ESA (humidité SMOS)
 - Prétraitement de données avant assimilation dans le modèle de prévision météorologique du CEPMMT/ECMWF

Table de matières



- Contexte, problématique scientifique
 - Humidité du sol et contenu en eau de la végétation
- Concepts réseaux de neurones
 - *Feed-forward network*, Perceptron multi-couches
- Exemples d'applications
 - Synergie multi-capteur
 - Longues séries temporelles
 - Chaine en temps quasi-réel de l'ESA (humidité SMOS)
 - Prétraitement de données avant assimilation dans le modèle de prévision météorologique du CEPMMT/ECMWF

L'humidité du sol et la biomasse, deux variables climatiques essentielles



Humidité du sol:

- Agriculture
- Météorologie
- Climatologie
- Hydrologie
- Gestion de risques:
 - inondations
 - sécheresses



Biomasse:

- Déforestation
- Cycle du carbone

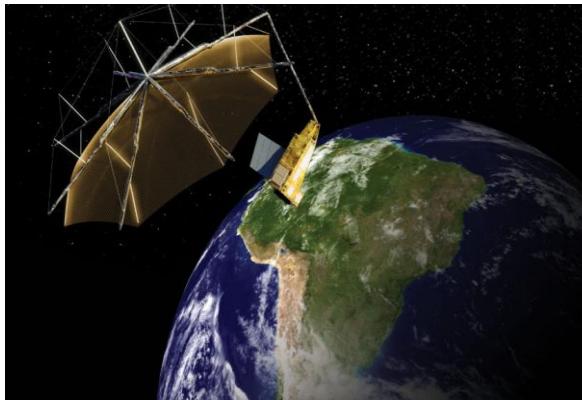


SMOS

Soil Moisture and Ocean Salinity

Launched in 2009 by ESA

Kerr et al. (2010, Proceedings IEEE)

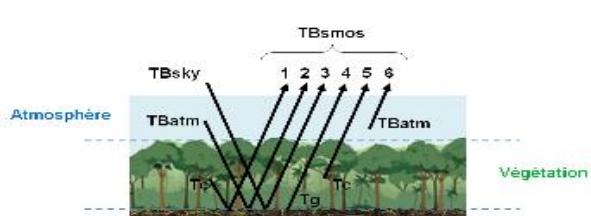
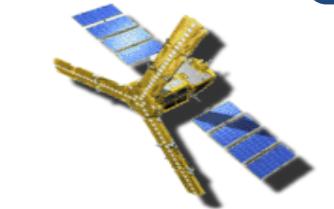


BIOMASS

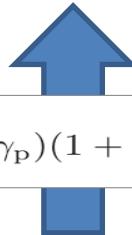
To be launched in ~2020 by ESA

Le Toan et al. (2011, Remote Sensing of Environment)

Observation: brightness temperatures



Geo-physical variables: soil moisture, vegetation optical depth, soil temperature, soil type, ...



$$TB_P = (1 - \omega_p)(1 - \gamma_p)(1 + \gamma_p r_{gp})T_c + (1 - r_{gp})\gamma_p T_g$$

Radiative transfer: tau-omega model

$$r_V = \frac{\epsilon \cos \theta - \sqrt{\epsilon - \sin^2 \theta}}{(\epsilon \cos \theta + \sqrt{\epsilon - \sin^2 \theta})}$$

$$\gamma_p = \exp(-\tau_p / \cos \theta)$$

Kerr et al. (2012, IEEE Tran. Geo. Remote Sensing)

Pourquoi utiliser des approches *data-driven*?



- Pour trouver une solution alternative à l'équation de transfert de rayonnement (par exemple en utilisant des mesures de terrain)
- Pour exploiter la synergie de différents capteurs pour lesquels on n'a pas forcément de modèle physique
- Pour avoir des résultats plus rapidement
- Pour utiliser un instrument comme référence et produire des données similaires avec un autre capteur
- Pour produire un jeu de données optimal avant de faire une assimilation de données avec un filtre de Kalman

Au commencement était la donnée ...



Un pré-requis indispensable

La *data science* est une démarche empirique qui se base sur des données pour apporter une réponse à des problèmes. Donc, avant toute chose, assurez-vous d'avoir des données...

La statistique algorithmique, s'intéresse aux processus sous-jacents aux données se concentrant à en extraire de l'information avec des modèles qui peuvent être des boîtes noires.

Mais cela marche mieux si grâce à notre compression du problème on arrive à faire des "boîtes grises" !

L'approche algorithmique connaît actuellement un formidable essor, notamment dans le champ du *machine learning* (apprentissage *automatique, machine, statistique* en français »).

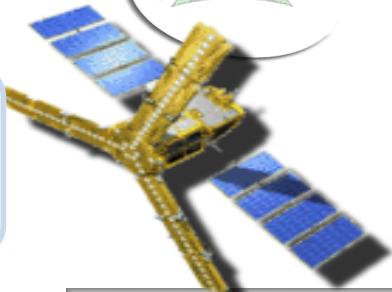
Biernat & Lutz, Data science: fondamentaux et études de cas, 2015, Ed. Eyrolles

Remote sensing observations



Soil Moisture and Ocean Salinity (SMOS)

- Synthetic aperture (equivalent to a 7 m dish). Resolution ~ 43 km
- Full polarization, Multi-angular (0-60°), L-Band (1.4 GHz)



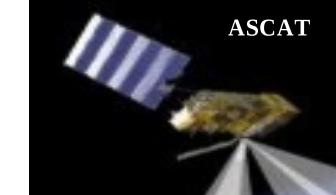
Advanced Microwave Scanning Radiometer- EOS (AMSR-E)

- Two polarizations, Single incidence angle
- 6 Bands (6.9 GHz, 18 GHz, 23.8 GHz, 36.5 GHz, 89.0 GHz)
- Resolution: 56 km @ 6.9 GHz



Advanced scatterometer (ASCAT)

- Active, C-band (5GHz) resolution ~ 50 km



MODIS

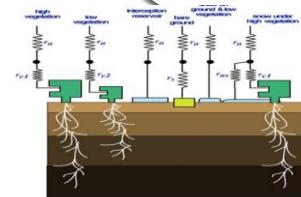
- visible and infrared bands, resolution 250 m – 1km



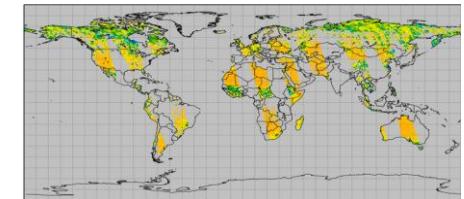
Models and in situ measurements



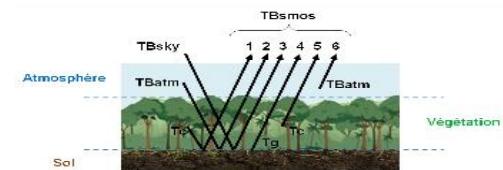
ECMWF land surface model H-TESSEL
(Balsamo et al. 2009)



SMOS Level 3 soil moisture (and opacity)
Al Bitar et al. (2017, ESSD)



Radiation transfer computations using L-MEB
(Wigneron et al. 2007)



In situ measurements compiled by the
ISMN (Dorigo et al. 2011, HESS)



Preparation des données



- En résumé, nous avons beaucoup de données de sources diverses pour tester des techniques d'apprentissage automatique
- D'abord il faut co-localiser les données dans le temps et dans l'espace: interpolations, agrégation (résolutions différentes), re-échantillonnages,... souvent cette étape demande plus de temps que la phase d'apprentissage elle-même
- Les données seront utilisées:
 - pour établir des critères de filtrage
 - comme données d'entrée
 - pour tester les sorties du modèle statistique

Table de matières

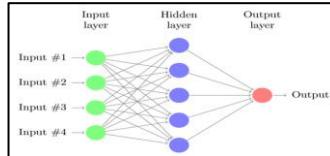
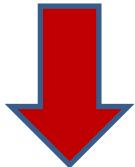


- Contexte, problématique scientifique
 - Humidité du sol et contenu en eau de la végétation
- Concepts réseaux de neurones
 - *Feed-forward network, Perceptron multi-couches*
- Exemples d'applications
 - Synergie multi-capteur
 - Longues séries temporelles
 - Chaîne en temps quasi-réel de l'ESA (humidité SMOS)
 - Prétraitement de données avant assimilation dans le modèle de prévision météorologique du CEPMMT/ECMWF

Statistical retrievals using Neural Networks



**Input data: SMOS TBs, AMSR-E TBs,
ASCAT σ , MODIS NDVI,...**



NN soil moisture



Training: comparison and new modeling step if needed

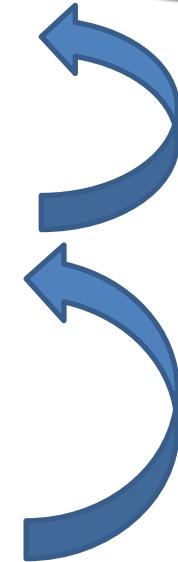
Soil moisture examples

- SMOS Level 2 SM
- Radiation transfer
- Surface models
- In situ measurements

Depends on the goal !

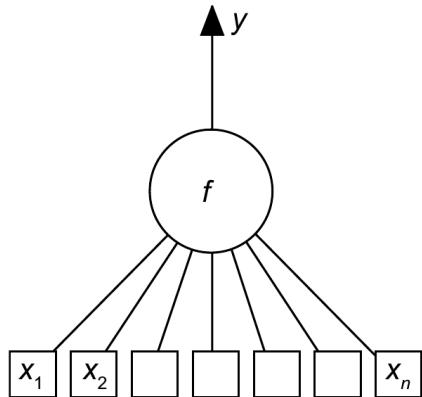
Test different
input data

Adapt NN
weights



La neurone

$$y = \text{th} \left[w_0 + \sum_{i=1}^n w_i x_i \right]$$



f est appelée *fonction d'activation*

potentiel v $v = w_0 + \sum_{i=1}^n w_i x_i$

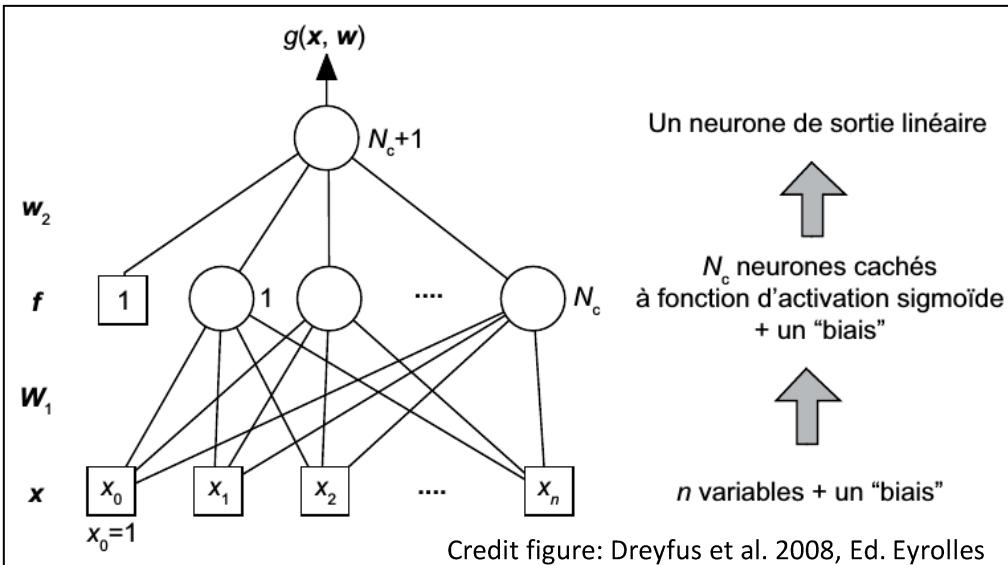
variables $\{x_i\}$ pondérées par les paramètres $\{w_i\}$

Poids synaptiques

w_0 terme constant ou « biais »

Credit figure: Dreyfus et al. 2008, Ed. Eyrolles

Le réseau de neurones non-bouclé (feed-forward)



$$g(\mathbf{x}, \mathbf{w}) = \sum_{i=1}^{N_c} \left[w_{N_c+1,i} \text{th} \left(\sum_{j=1}^n w_{ij} x_j + w_{i0} \right) \right] + w_{N_c+1,0}$$
$$= \mathbf{w}_2 \cdot f(\mathbf{W}_1 \mathbf{x})$$

On lit parfois en géosciences que les réseaux de neurones sont des modèles non-paramétriques par contraposition à des modèles basés sur une équation physique avec des paramètres à ajuster. Cependant, il s'agit bien évidemment de modèles paramétriques, mais en utilisant un réseau de fonctions sigmoïdes sans lien directe avec la physique du problème à résoudre.

- Exemple processeur humidité du sol SMOS en temps quasi-réel de l'Agence Spatiale Européenne (Rodriguez-Fernandez et al. 2017, Hydro. Earth Sys. Sci.)
 - Un vecteur d'entrée avec 13 éléments
 - Une couche cachée avec 5 neurones sigmoïdes
 - Une neurone linéaire dans la 2eme couche
 - Paramètres: $(13+1) \times 5 + (5+1) = 76$
 - ... pour un modèle valable pour tout le Globe !

Pour quoi utiliser ces réseaux de neurones



Les réseaux de neurones sont des approximateurs universels

Propriété

Toute fonction bornée suffisamment régulière peut être approchée uniformément, avec une précision arbitraire, dans un domaine fini de l'espace de ses variables, par un réseau de neurones comportant une couche de neurones cachés en nombre fini, possédant tous la même fonction d'activation, et un neurone de sortie linéaire [HORNIK 1989] [HORNIK 1990] [HORNIK 1991].

Les réseaux de neurones non linéaires par rapport à leurs paramètres sont des approximateurs parcimonieux

Propriété fondamentale

On montre [BARRON 1993] que, si l'approximation dépend des paramètres ajustables de manière non linéaire, elle est plus parcimonieuse que si elle dépend linéairement des paramètres.

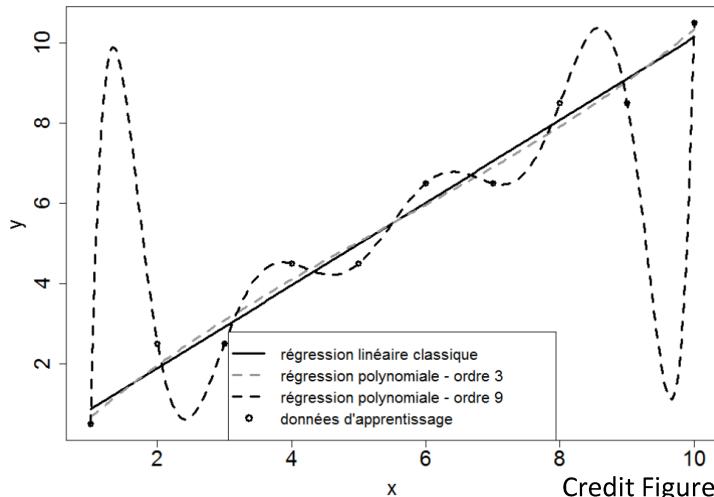
la complexité d'un modèle est liée au nombre de ses paramètres : contrôler le surajustement

Credit: Dreyfus et al. 2008, Ed. Eyrolles

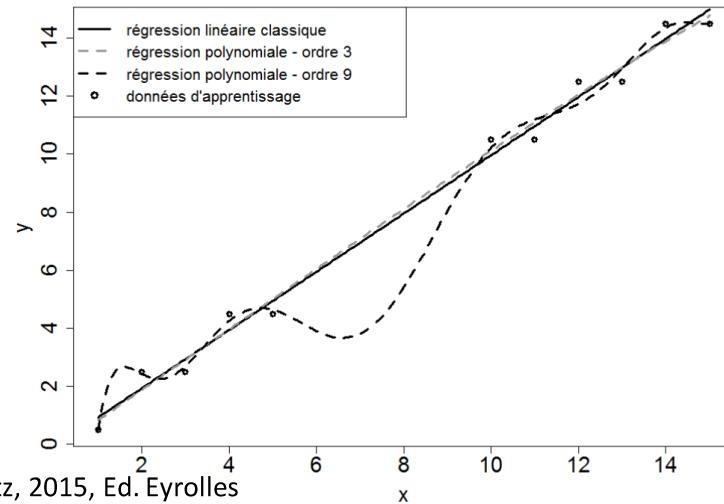
- Exemples d'apprentissage (humidité du sol dans notre cas)
- Minimisation fonction de couts
 - Levenberg-Marquardt, smooth conjugate gradient,...
- Eviter le sur-apprentissage !

Sur-apprentissage

Exemple avec des régressions linéaires et polynomiales

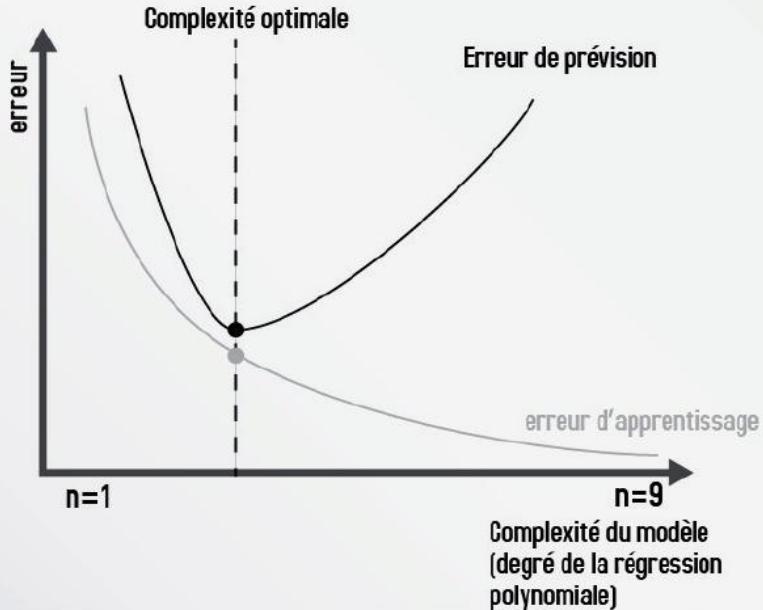


Credit Figures: Biernat & Lutz, 2015, Ed. Eyrolles



La régression polynomiale d'ordre 9 explique très bien les points utilisés pour l'apprentissage (faible biais) mais montre une très grande variance qui fait qu'elle soit instable et pas un bon *predictor* sur des nouveaux données d'entrée

Le compromis biais-variance



Solution:

1. Diviser les données dans une sous-ensemble d'apprentissage et un autre d'évaluation
2. Choisir le modèle que minimise l'erreur les données d'évaluation de la prévision et non pas l'erreur d'apprentissage

Credit Figure: Biernat & Lutz, 2015, Ed. Eyrolles

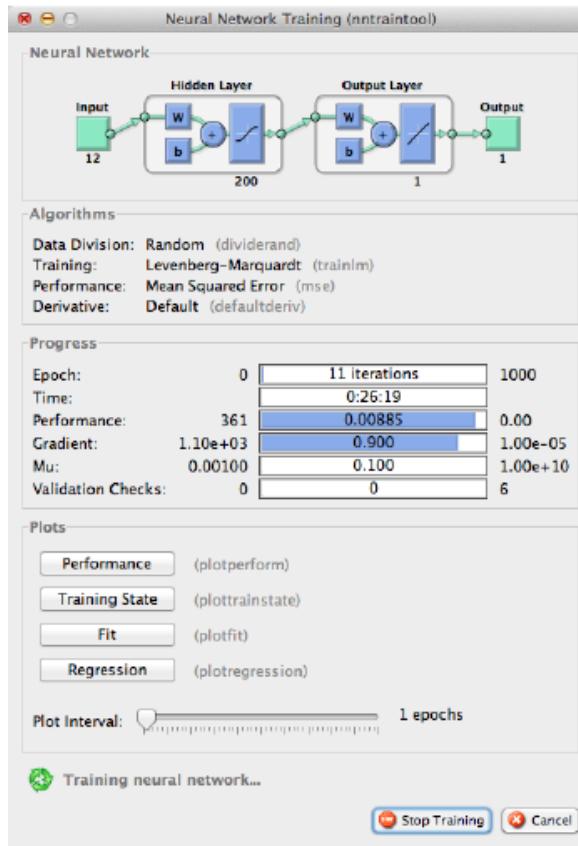
Comment diviser les données ?



- Il est possible de le faire de façon purement aléatoire
- Mais dans nos applications, avec 8 ans de données SMOS, nous utilisons l'approche suivante:
 - 2-3 ans pour apprentissage
 - Sélection d'un jour tous les 3-6 jours (conserve la climatologie)
 - Sélection d'un point de la grille tous les 2 ou 3 points en latitude et longitude (bonne représentativité spatiale)
 - Typiquement 10^5 échantillons : trois sous-ensembles aléatoirement
 - 60 % pour apprentissage
 - 20 % validation pendant apprentissage
 - 20 % test après apprentissage
 - Reste des données/années pour évaluation à posteriori

~ 8 Go de mémoire vive
(dépend de l'algorithme
de minimisation et de
l'architecture du réseau
de neurones)

Matlab neural network toolbox



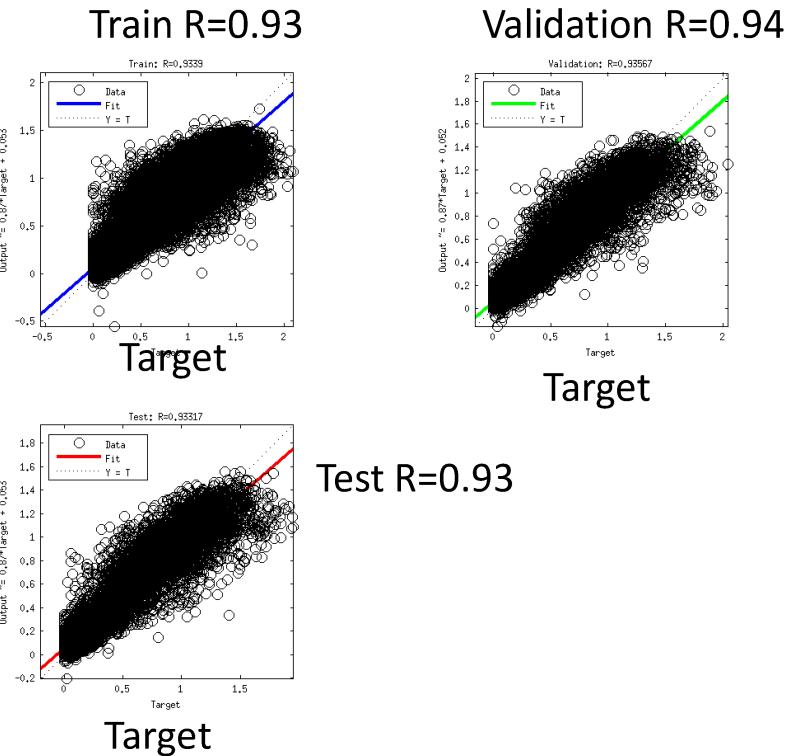
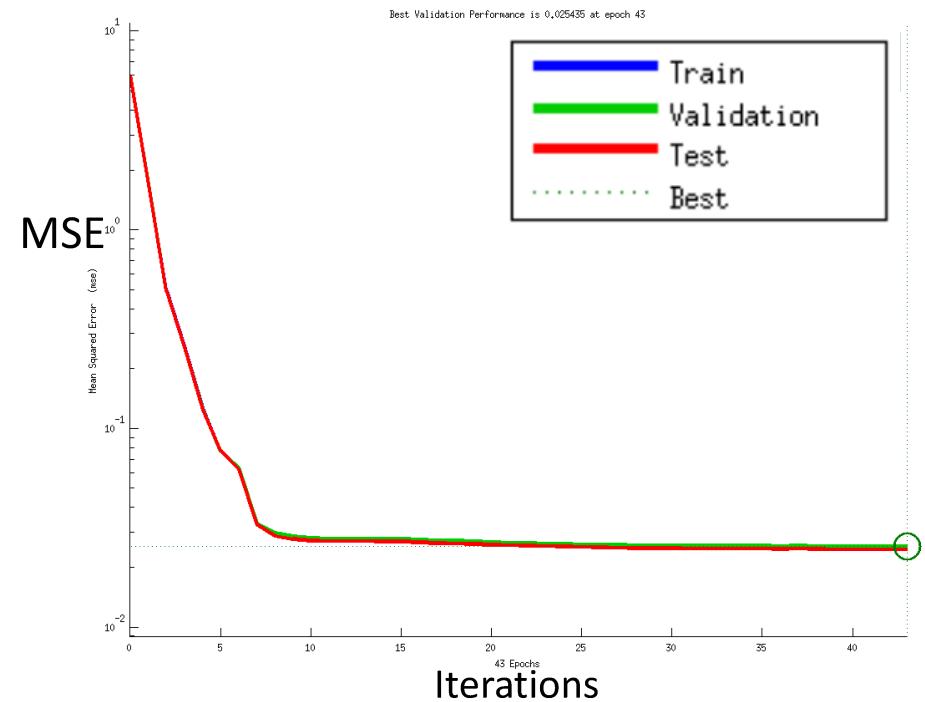
Métrique de performance pour l'apprentissage:
Mean Square Error (MSE)

$$\text{MSE} = \frac{1}{N} \times \sum (\text{SM}_{\text{nn}} - \text{SM}_{\text{ref}})^2$$

Métriques d'évaluation à posteriori:

- Corrélation Pearson
- RMSE
- STD
- Bias

Early stopping



Régularisation



- Pour éviter le sur-apprentissage il est possible aussi de faire une régularisation avec un terme qui pénalise des poids élevés

$$\text{MSE}_{\text{REG}} = \alpha \text{ MSE} + \beta \frac{1}{n} \sum_{j=1}^n \omega_j^2$$

- Utile s'il est nécessaire de faire l'apprentissage avec des données synthétiques avant le lancement d'un nouveau capteur. Les performances du réseau de neurones appliqué à des données réels sont meilleures dans ce cas (Rodriguez-Fernandez et al, 2017, IEEE Symp. Geo. And Rem. Sensing)

Feature normalization or standardization



- Permet de donner la même importance relative à tous les éléments du vecteur d'entrée pour la minimisation
- Exemple: normalisation [-1,1]

$$v_i^{\text{norm}} = -1 + 2 \frac{v_i - v_i^{\min}}{v_i^{\max} - v_i^{\min}}, \quad \forall i = 1 \dots n_{\text{in}}$$

$$v_j^{\text{L1}} = \tanh \left(\sum_{i=1}^{n_{\text{in}}} W_{\text{L1}}^{ij} v_i^{\text{norm}} + B_{\text{L1}}^j \right), \quad \forall j = 1 \dots n_{\text{L1}} \qquad \qquad v^{\text{L2}} = \sum_{j=1}^{n_{\text{L1}}} W_{\text{L2}}^j v_j^{\text{L1}} + B_{\text{L2}}$$

$$v^{\text{out}} = v_{\text{newMin}}^{\text{L2}} + \frac{v_{\text{newMax}}^{\text{L2}} - v_{\text{newMin}}^{\text{L2}}}{v_{\text{oldMax}}^{\text{L2}} - v_{\text{oldMin}}^{\text{L2}}} \left(v^{\text{L2}} - v_{\text{oldMin}}^{\text{L2}} \right)$$

Feature extraction



- *Building derived values (features) intended to be informative and non-redundant, facilitating the subsequent learning and generalization steps*
- *In our case, normalized indexes giving information on the local dynamic ranges of brightness temperatures and soil moisture*

$$I_{\lambda\phi}(t) = \text{SM}_{\lambda\phi}^{T_b^{\min}} + \left[\text{SM}_{\lambda\phi}^{T_b^{\max}} - \text{SM}_{\lambda\phi}^{T_b^{\min}} \right] \times I_{1_{\lambda\phi}}(t)$$

$$I_{1_{\lambda\phi}}(t) = \frac{T_{b_{\lambda\phi}}(t) - T_{b_{\lambda\phi}}^{\min}}{T_{b_{\lambda\phi}}^{\max} - T_{b_{\lambda\phi}}^{\min}}$$

Rodriguez-Fernandez et al. 2015 (IEEE Trans. Geo. And Rem. Sensing)

Errors



- Output errors computed from the input errors (SMOS Near-Real-Time soil moisture, Rodriguez-Fernandez et al. 2017)

$$\begin{aligned}\Delta I_{\lambda\phi}^2(t) = & \left[\text{SM}_{\lambda\phi}^{T_b^{\max}} - \text{SM}_{\lambda\phi}^{T_b^{\min}} \right]^2 (\Delta I_{1_{\lambda\phi}}(t))^2 \\ & + [1 - I_{1_{\lambda\phi}}(t)]^2 \left(\Delta \text{SM}_{\lambda\phi}^{T_b^{\min}} \right)^2 \\ & + [I_{1_{\lambda\phi}}(t)]^2 \left(\Delta \text{SM}_{\lambda\phi}^{T_b^{\max}} \right)^2,\end{aligned}$$

$$\begin{aligned}\Delta I_{1_{\lambda\phi}}^2(t) = & \frac{\Delta T_{b_{\lambda\phi}}(t)^2}{T_{D_{\lambda\phi}}^2} + \frac{\left(\Delta T_{b_{\lambda\phi}}^{\max} \right)^2}{T_{D_{\lambda\phi}}^2} \left(\frac{T_{m_{\lambda\phi}}(t)}{T_{D_{\lambda\phi}}} \right)^2 \\ & + \frac{\left(\Delta T_{b_{\lambda\phi}}^{\min} \right)^2}{T_{D_{\lambda\phi}}^2} \left(-1 + \frac{T_{m_{\lambda\phi}}(t)}{T_{D_{\lambda\phi}}} \right)^2,\end{aligned}$$

$$(\Delta v^{\text{L2}})^2 = \sum_{i=1}^{n_{\text{in}}} \left\{ (\Delta \boldsymbol{v}_i^{\text{norm}})^2 \left(\sum_{j=1}^{n_{\text{L1}}} W_{\text{L2}}^j W_{\text{L1}}^{ij} \sigma^j \right)^2 \right\}$$

$$\sigma^j = 1 - \tanh^2 \left(\sum_{i=1}^{n_{\text{in}}} W_{\text{L1}}^{ij} \boldsymbol{v}_i^{\text{norm}} + B_{\text{L1}}^j \right), \forall j = 1 \dots n_{\text{L1}}$$

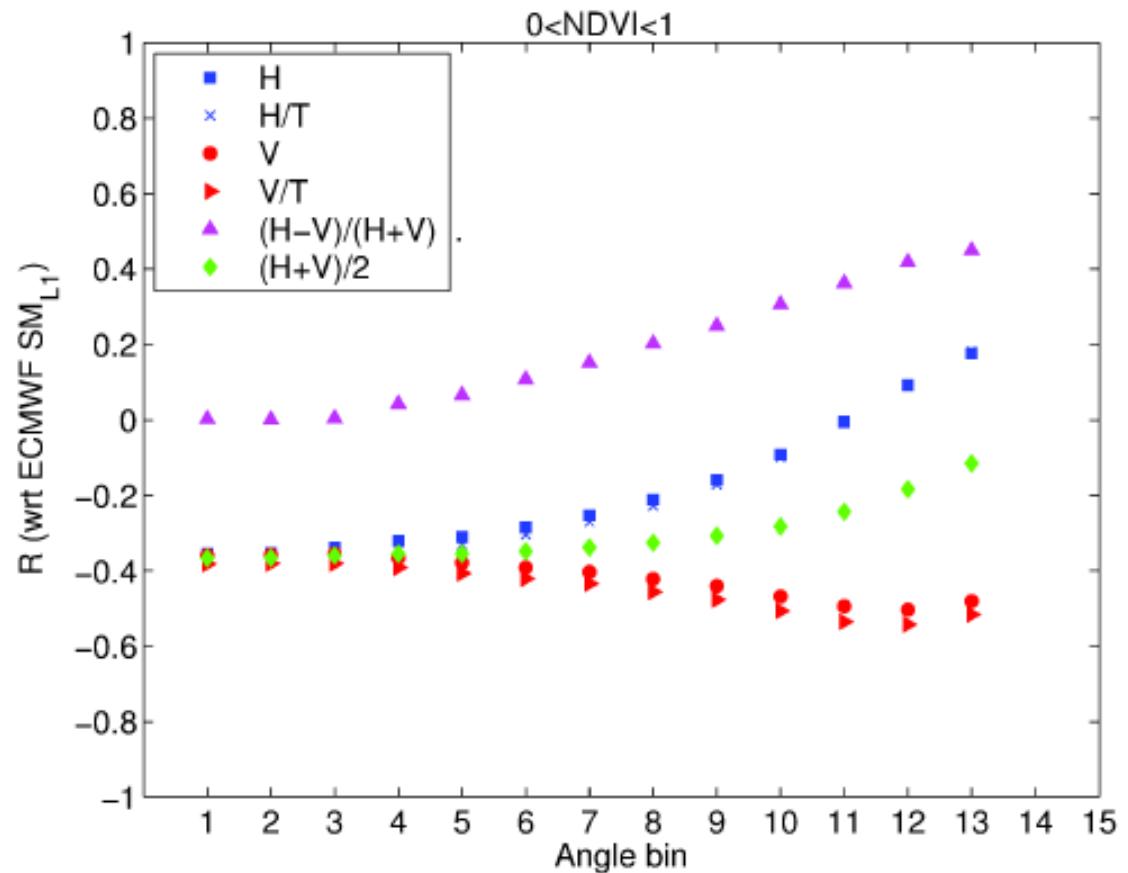
- It is also possible to estimate the errors in the weights... But it is much more complicated and time consuming (Aires et al. 2005)

Table de matières



- Contexte, problématique scientifique
 - Humidité du sol et contenu en eau de la végétation
- Concepts réseaux de neurones
 - *Feed-forward network*, Perceptron multi-couches
- Exemples d'applications
 - Synergie multi-capteur
 - Longes séries temporelles
 - Chaine en temps quasi-réel de l'ESA (humidité SMOS)
 - Prétraitement de données avant assimilation dans le modèle de prévision météorologique du CEPMMT/ECMWF

Analyse de sensibilité



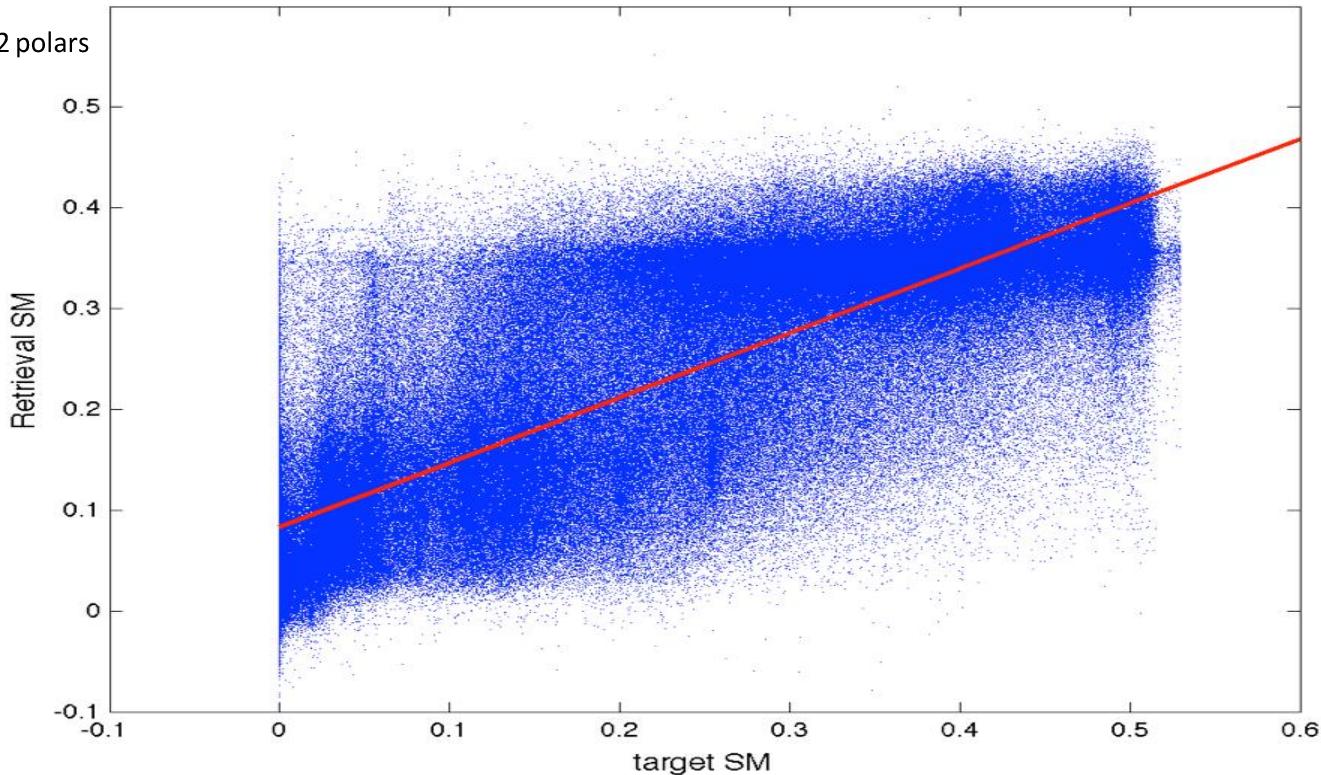


SMOS NN SM vs ECMWF SM



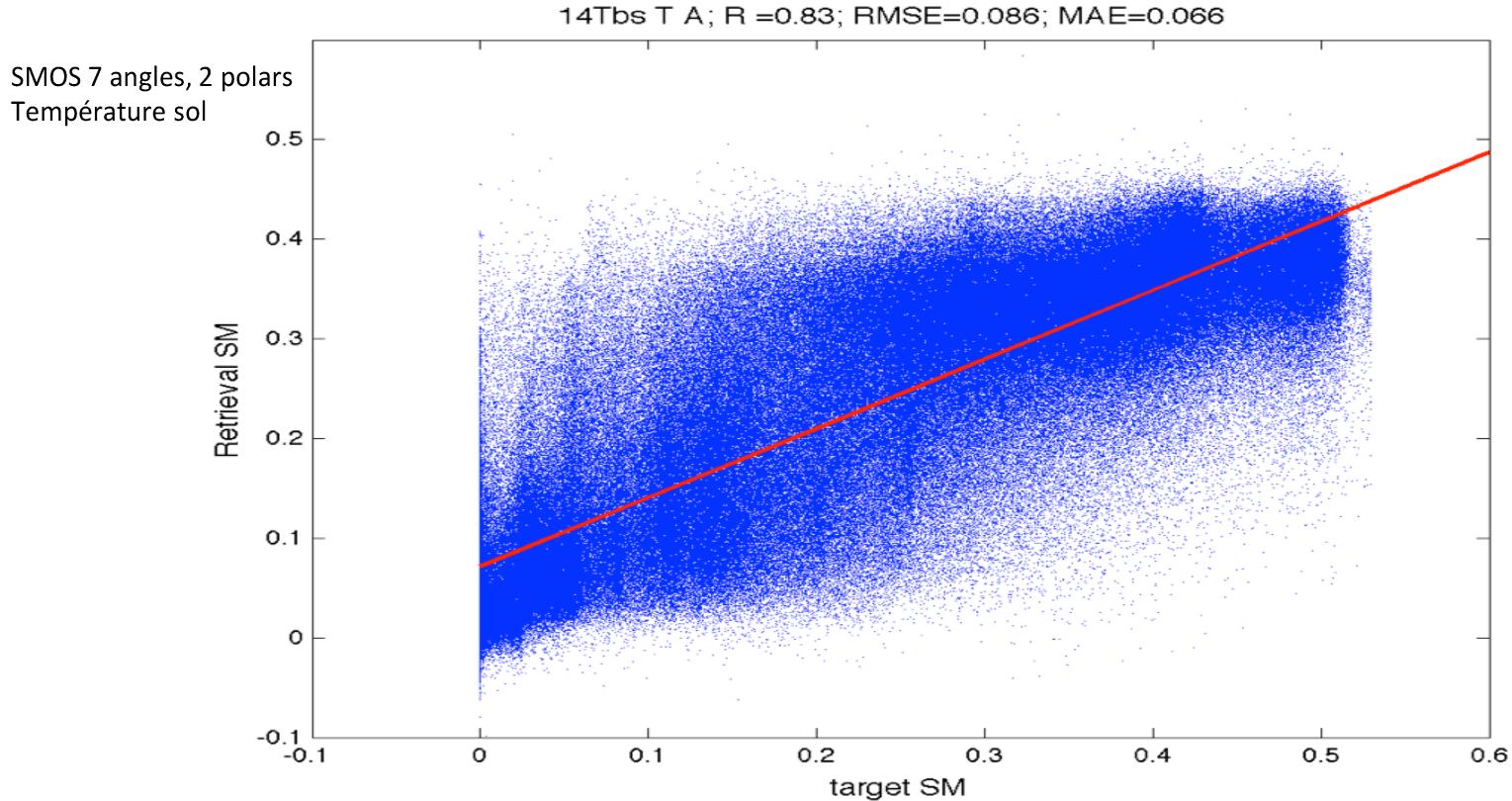
SMOS 7 angles, 2 polars

14Tbs A; R = 0.8; RMSE=0.093; MAE=0.072





SMOS NN SM vs ECMWF SM



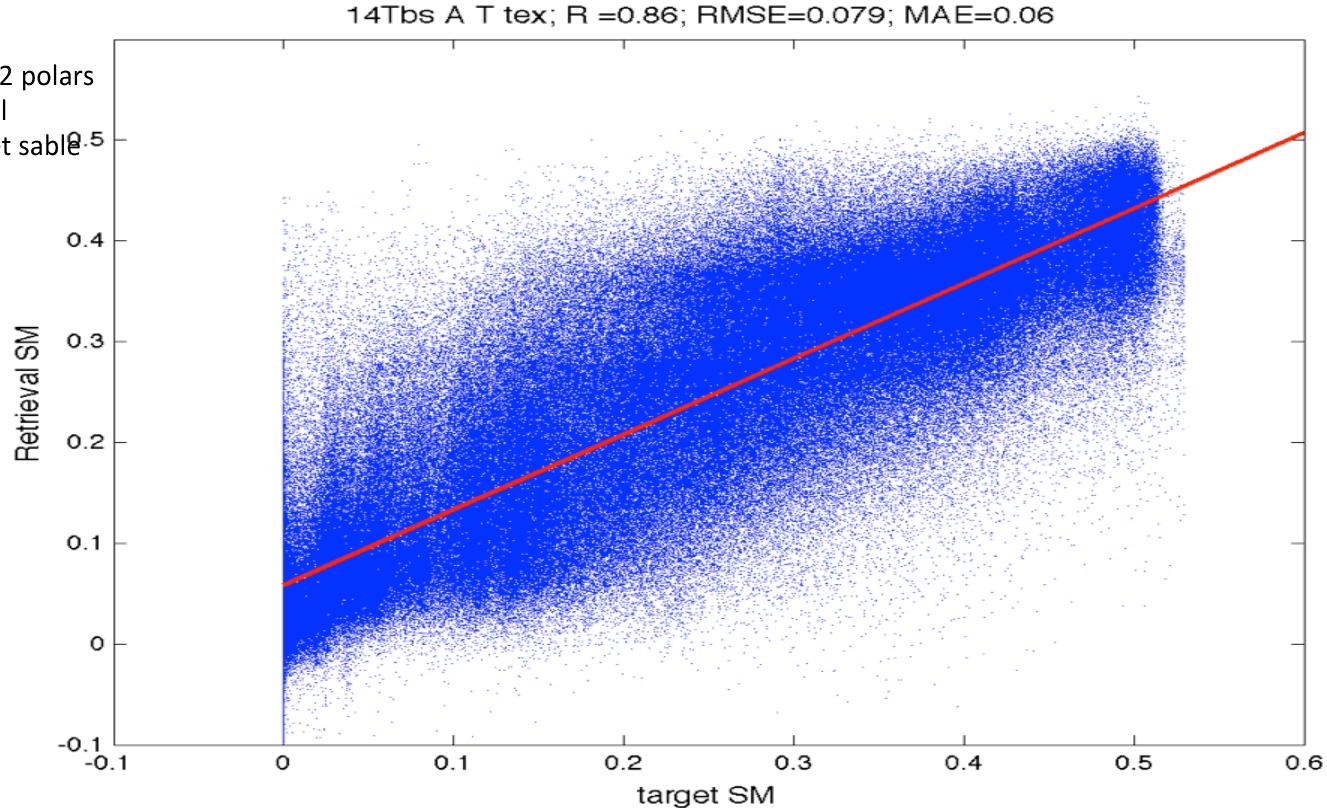
32



SMOS NN SM vs ECMWF SM



SMOS 7 angles, 2 polars
Température sol
Fraction argile et sable



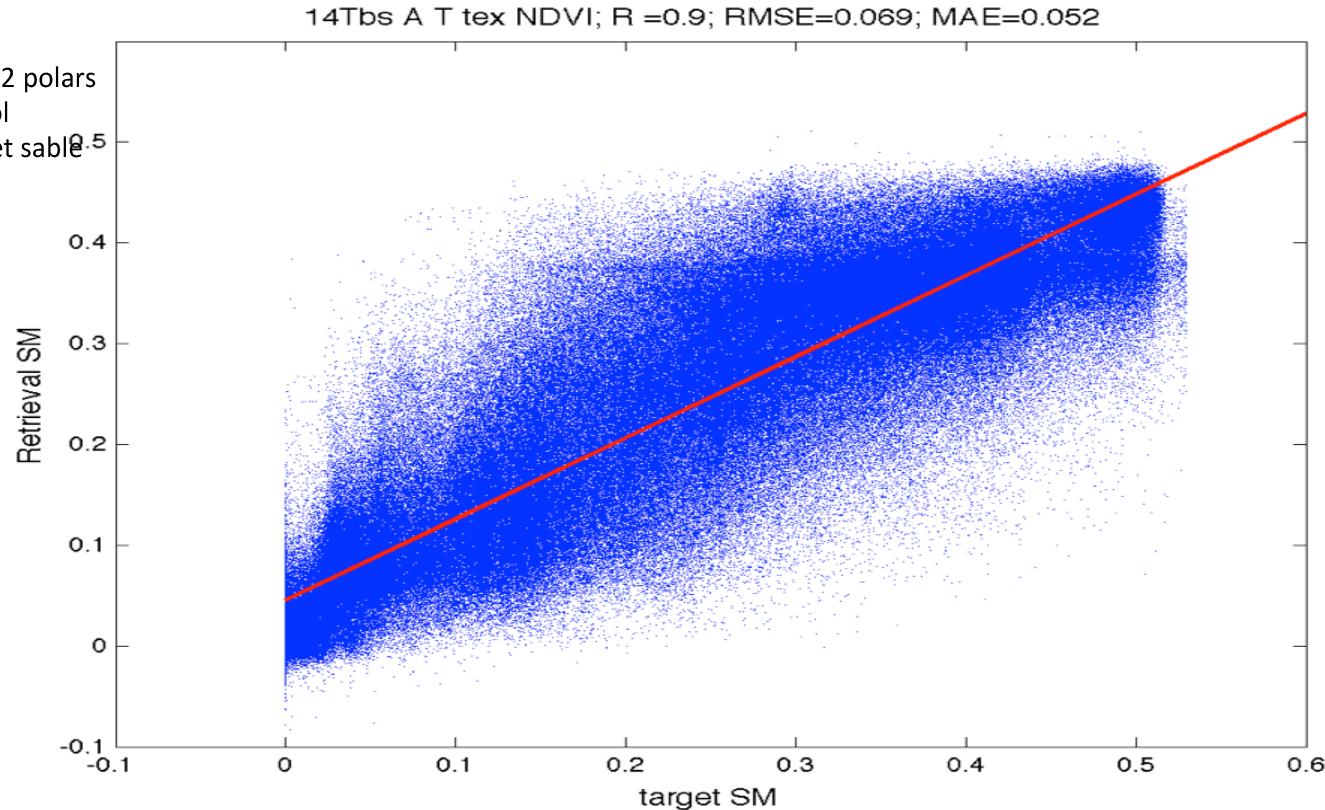
33



SMOS NN SM vs ECMWF SM

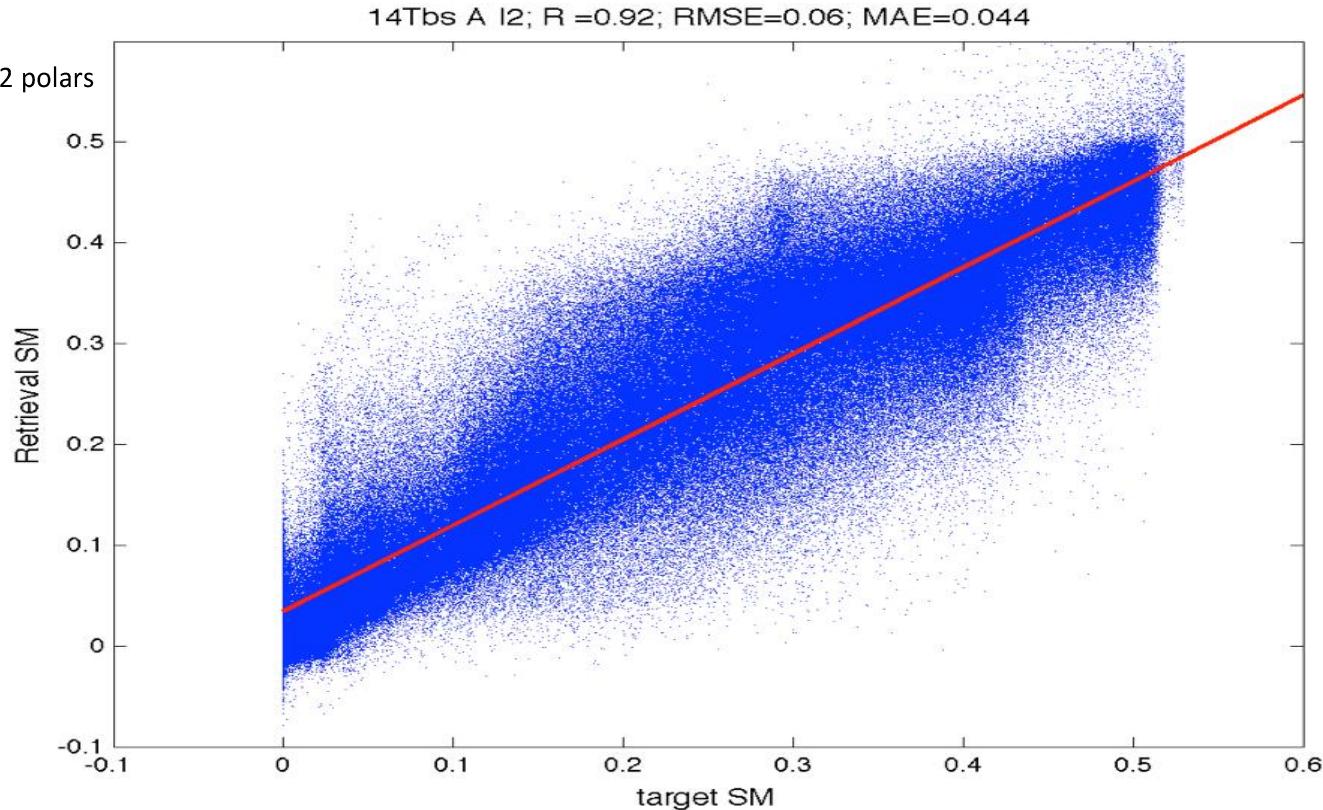


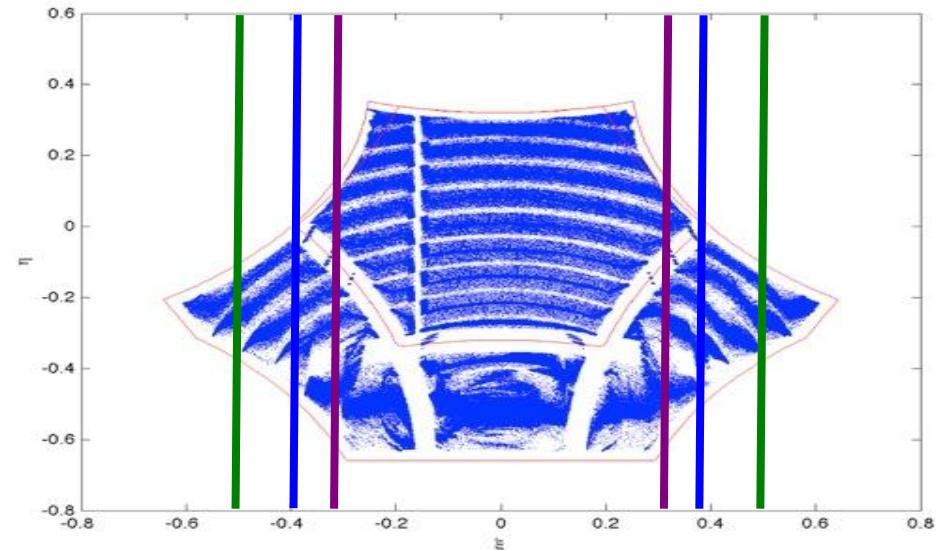
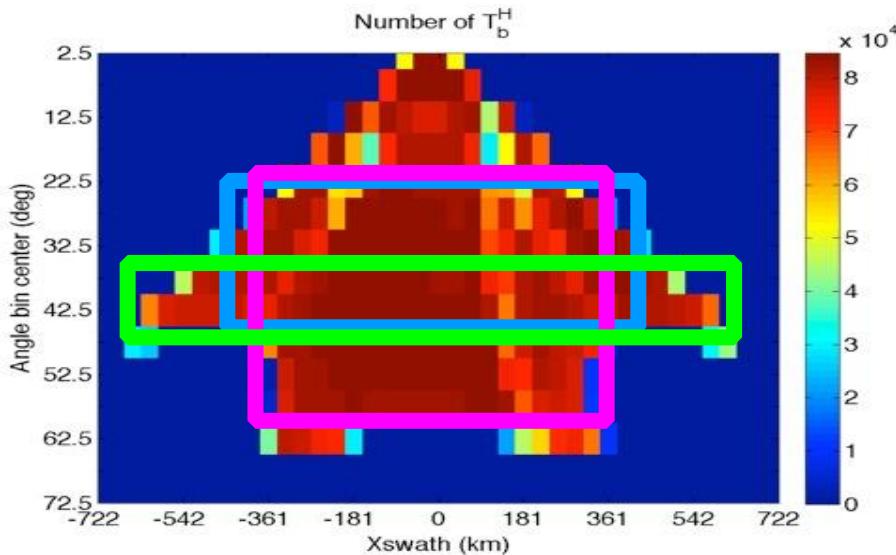
SMOS 7 angles, 2 polars
Température sol
Fraction argile et sable
NDVI



Using local SM information

SMOS 7 angles, 2 polars
Indices locaux





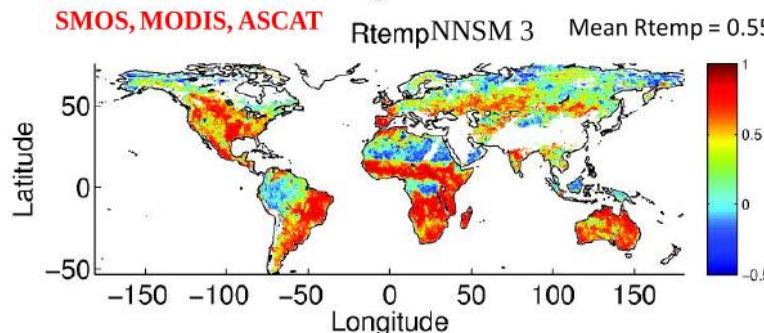
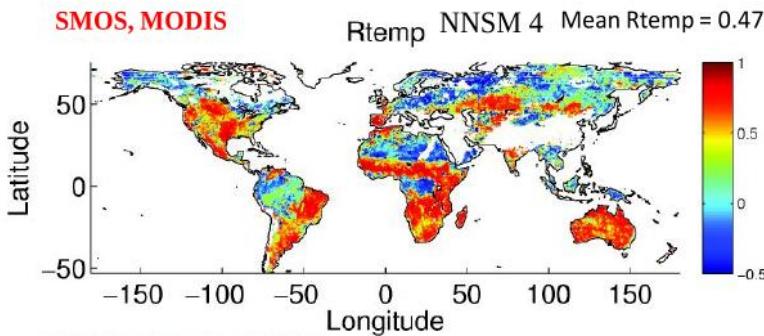
	R	RMSE	MAE	swath
filt1 net14I2 27-57	0.93	0.058	0.042	+ 335
filt2 net8I2_27-42	0.92	0.06	0.043	+ 380
filt3 net4I2_37-42	0.91	0.064	0.046	+ 540

- Une bonne compréhension des données, des instruments, facilite énormément la construction d'un algorithme neuronal

Multi-sensor synergy



- Training on ECMWF models
- Active and passive microwaves



- MODIS NDVI improve the retrievals
- Clay and soil fractions from ECOCLIMAP as well
- Active microwaves improve the NN ability to capture the time dynamics

A multi-sensor retrieval algorithm becomes possible even when a multi-wavelength physical algorithm is not available

Rodríguez-Fernandez et al. (2015, TGARS)

Neural networks as an un-biased observation operator for data assimilation



$$x_a^t = x_b^t + K(y_0^t - \mathcal{H}[x_b^t])$$

Kalman gains: $K = BH^T(HBH^T+R)^{-1}$

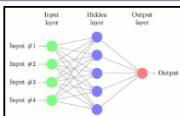
Innovations (first guess departures): $y_0^t - \mathcal{H}[x_b^t]$

The observations should be unbiased wrt the model

NN SM for the DA project

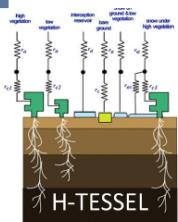


- SMOS L3 Tbs, polarization H & V, angles 30°-45°
- Normalization with local extreme SM



NN soil moisture and associated uncertainties

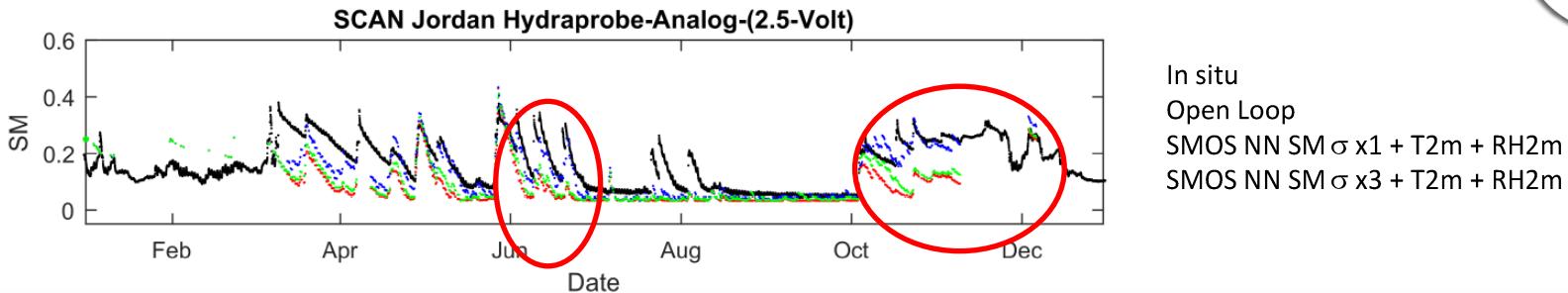
Ready for DA!



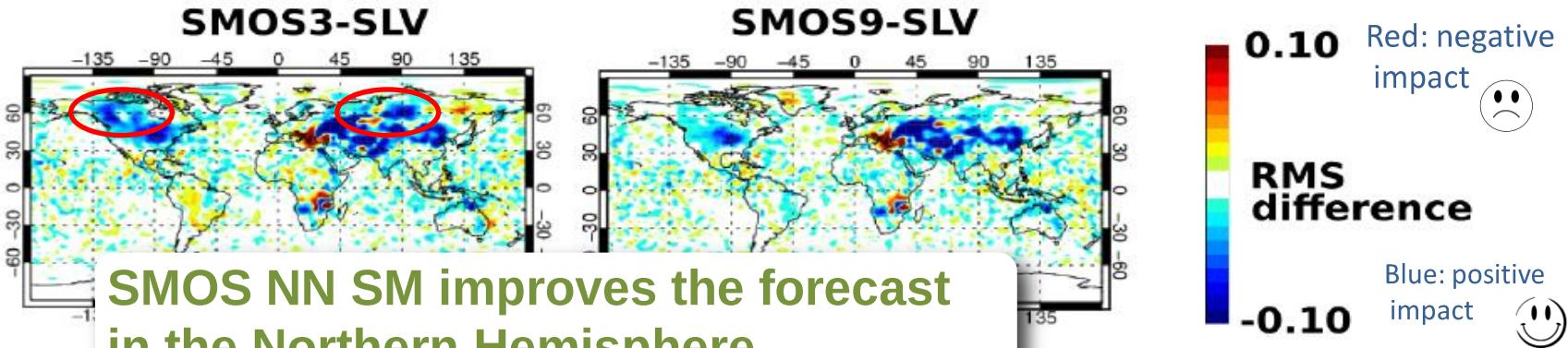
IFS (0-7cm) soil moisture

- ESA SMOS auxiliary files
- Spatial averaging to SMOS resolution (~43 km)
- Temporal interpolation to the time of SMOS acquisitions

Data assimilation results



Results: On average, for more than 400 in situ sites, the performances of the analysed soil moisture fields are close (within 2-3 %) to those of the open loop experiment



ESA neural net near-real-time soil moisture

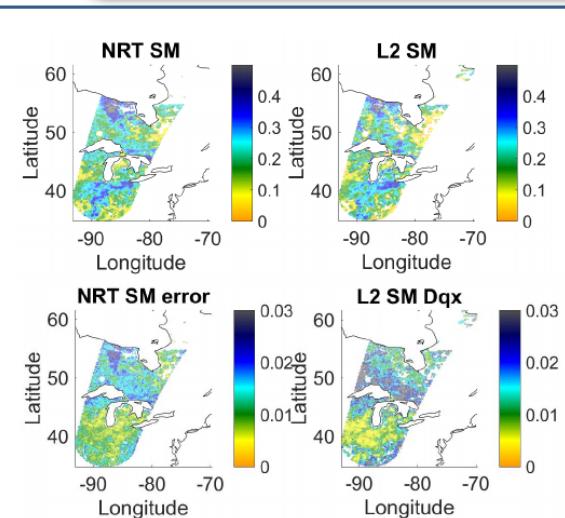


- Input NRT TBs
- Training on SMOS Level 2 SM
- Available in less than 3.5 hours after sensing since January 2016

Rodriguez-Fernandez et al. (2016, CESBIO ESA report)

Muñoz-Sabater et al. (2016, ECMWF ESA report)

Rodriguez-Fernandez et al. (2017, HESS)



Implemented by :



With support by :



Delivered to :



Disseminated by:



And ESA SMOS
Dissemination site

NRT flow chart

Plan:

- Definition of best Neural Network configuration for a NRT algorithm (CESBIO with Feedback from ESA and ECMWF)

- Matlab breadboard
(CESBIO-ECMWF)

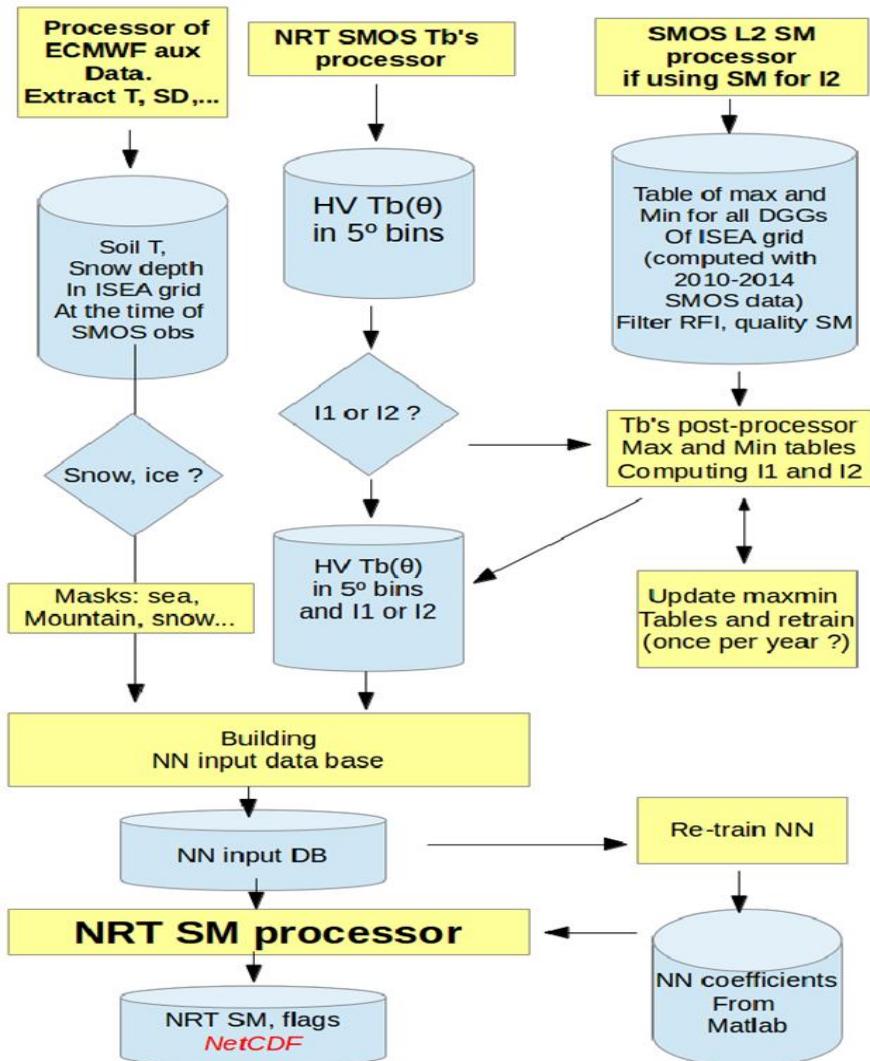
- BUFR reader
- XY2HV
- Binning

- ECMWF implementation
(Fortran)

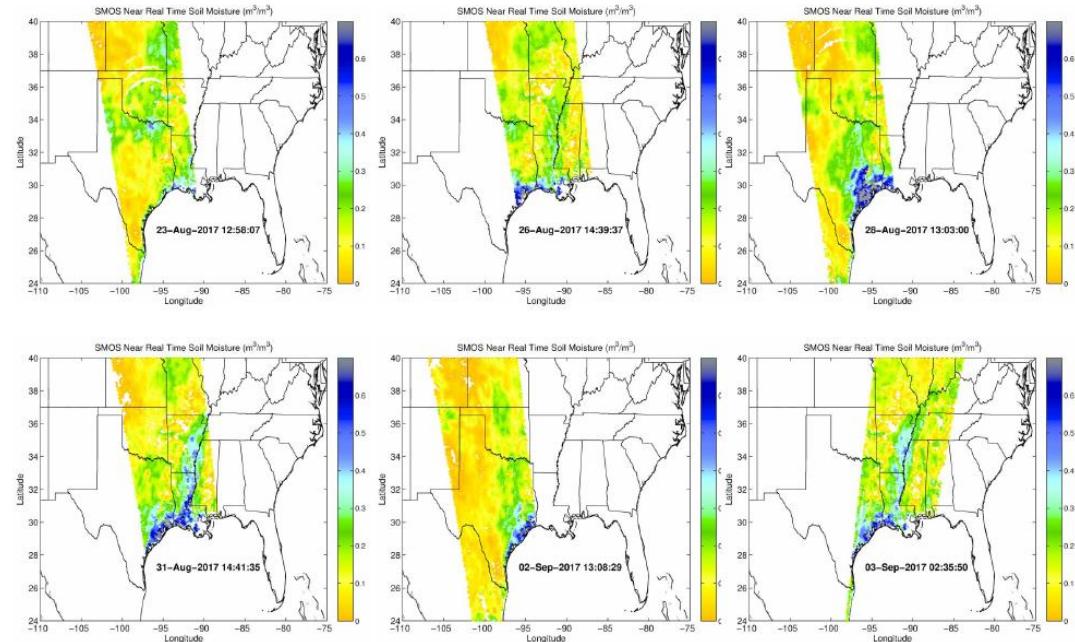
- Evaluation of tests data sets by CESBIO

- Training Neural Networks
(CESBIO)

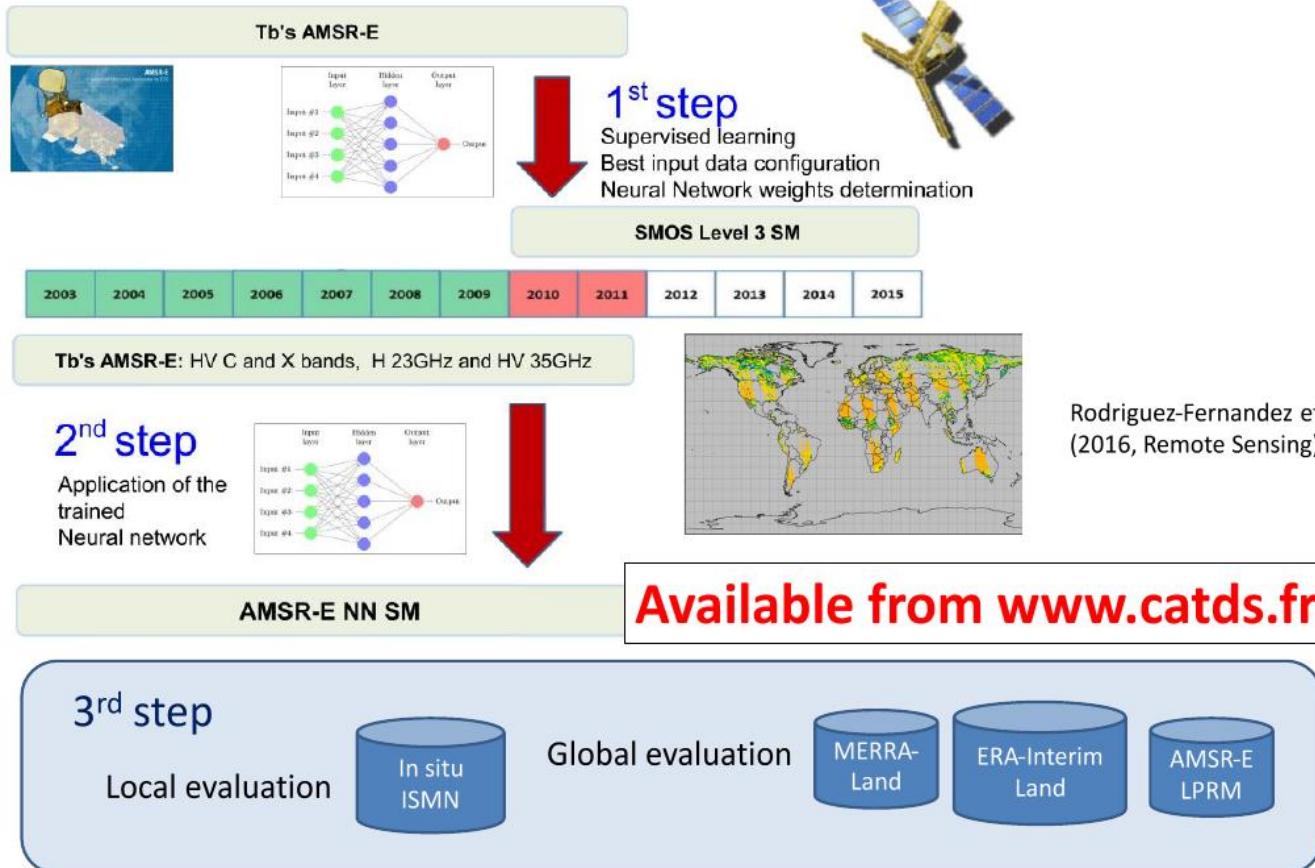
- Testing results (ECMWF &CESBIO)



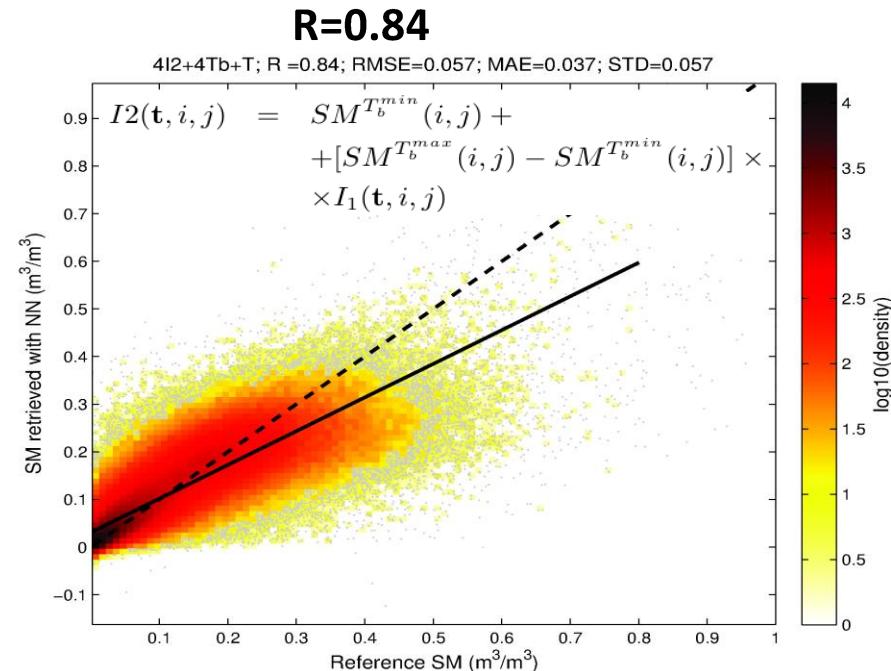
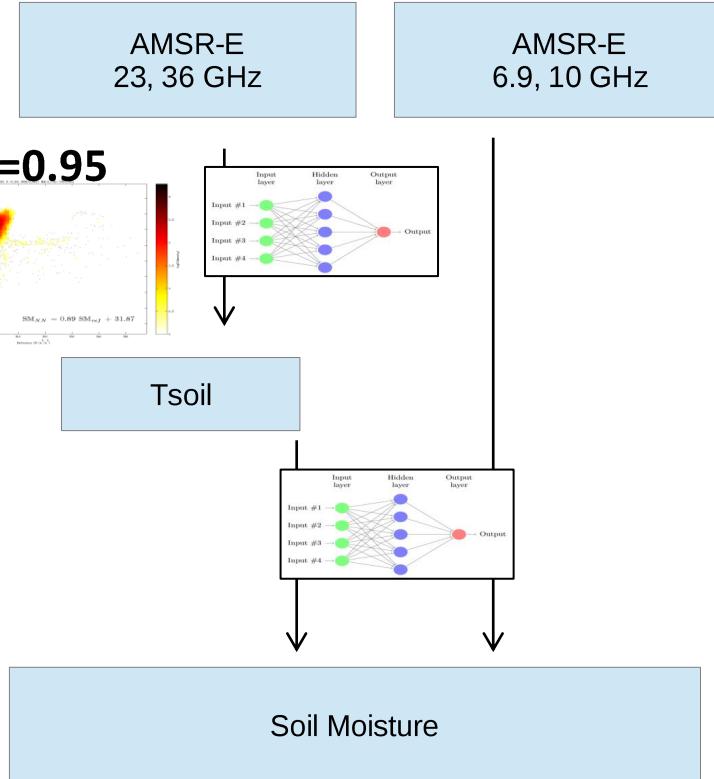
- The NRT SM has captured well an extreme event such as the Harvey floods in Texas



Long time series from AMSR-E and SMOS



AMSR-E SM Neural Network retrieval using SMOS L3 SM as reference



- Pure data-driven retrieval using neural networks
- Input: SMOS brightness temperatures
- Reference for the training: in situ measurements

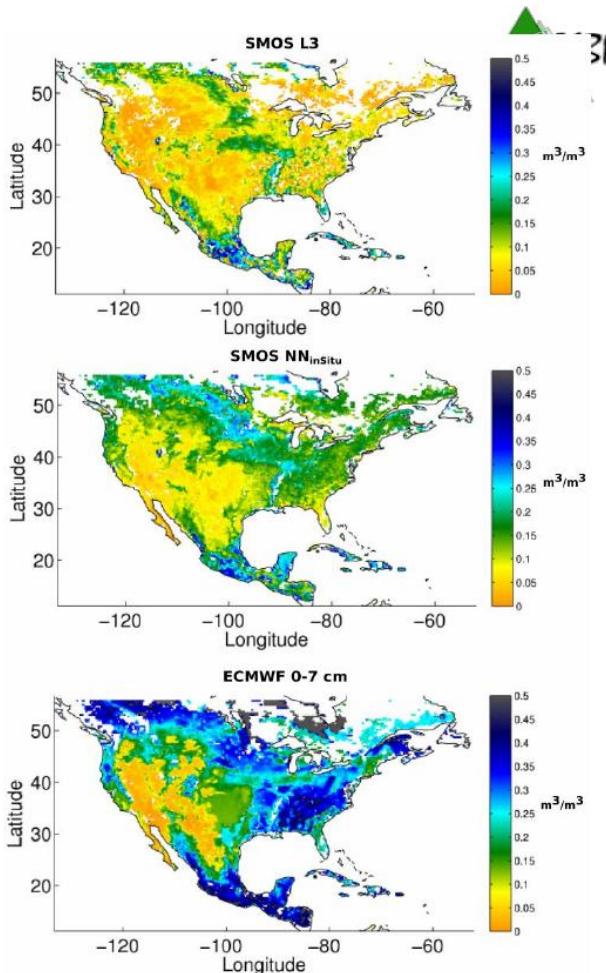


Using in situ measurements for the training

Analysis of International Soil Moisture Network : SCAN, SNOTEL, USCRN



Rodríguez-Fernández et al (2017a, IGARSS)

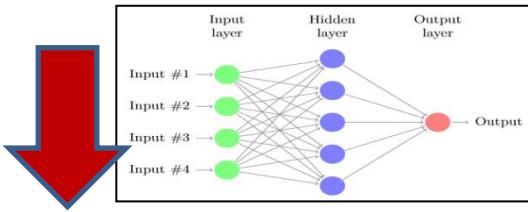


Training the NN



**Year averages: Tb's, SM,
tau, Pol ratios, NDVI ...**

Test different
input data



Adapt NN
weights

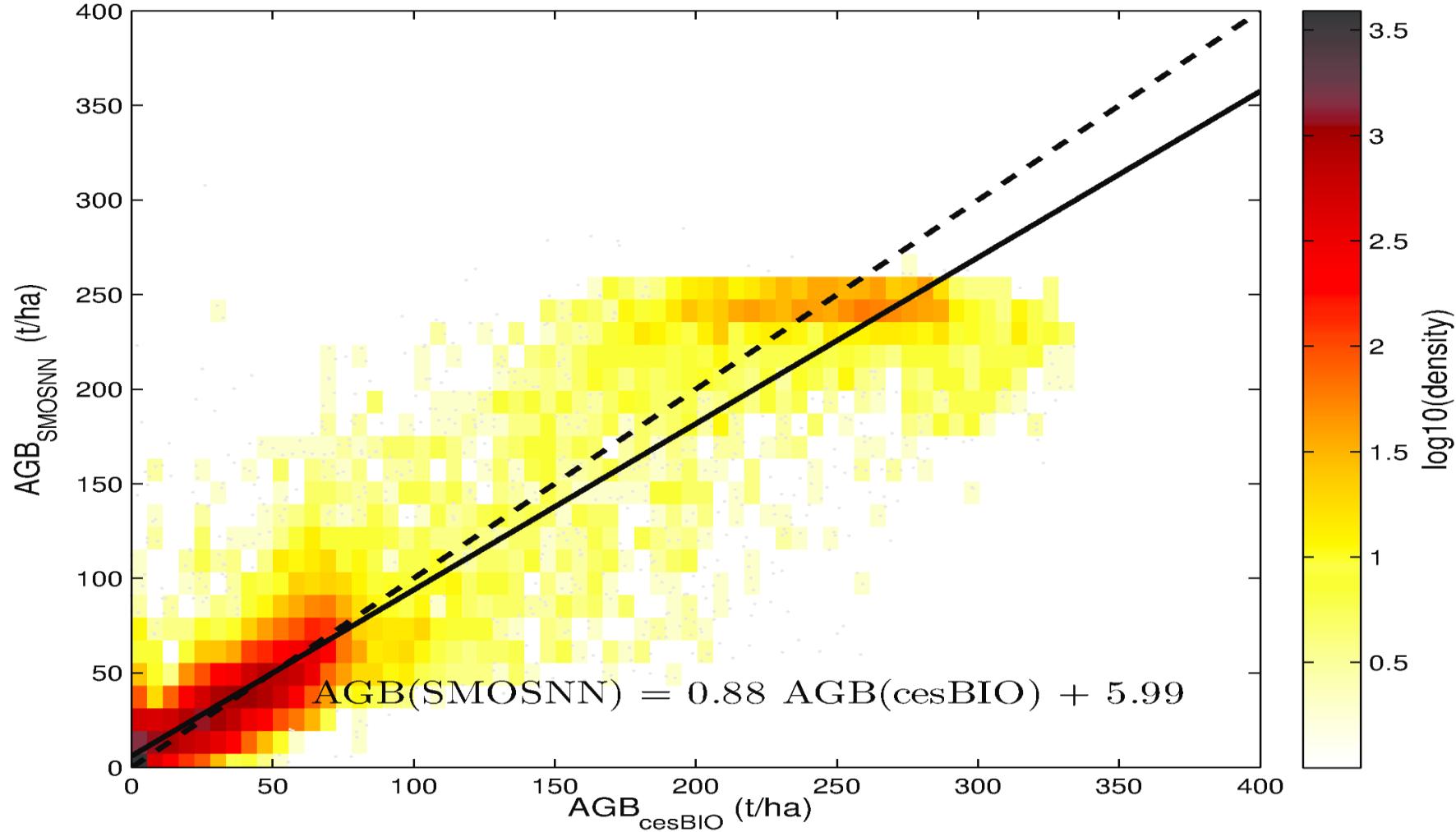
NN AGB

Training: comparison and new modeling step if needed

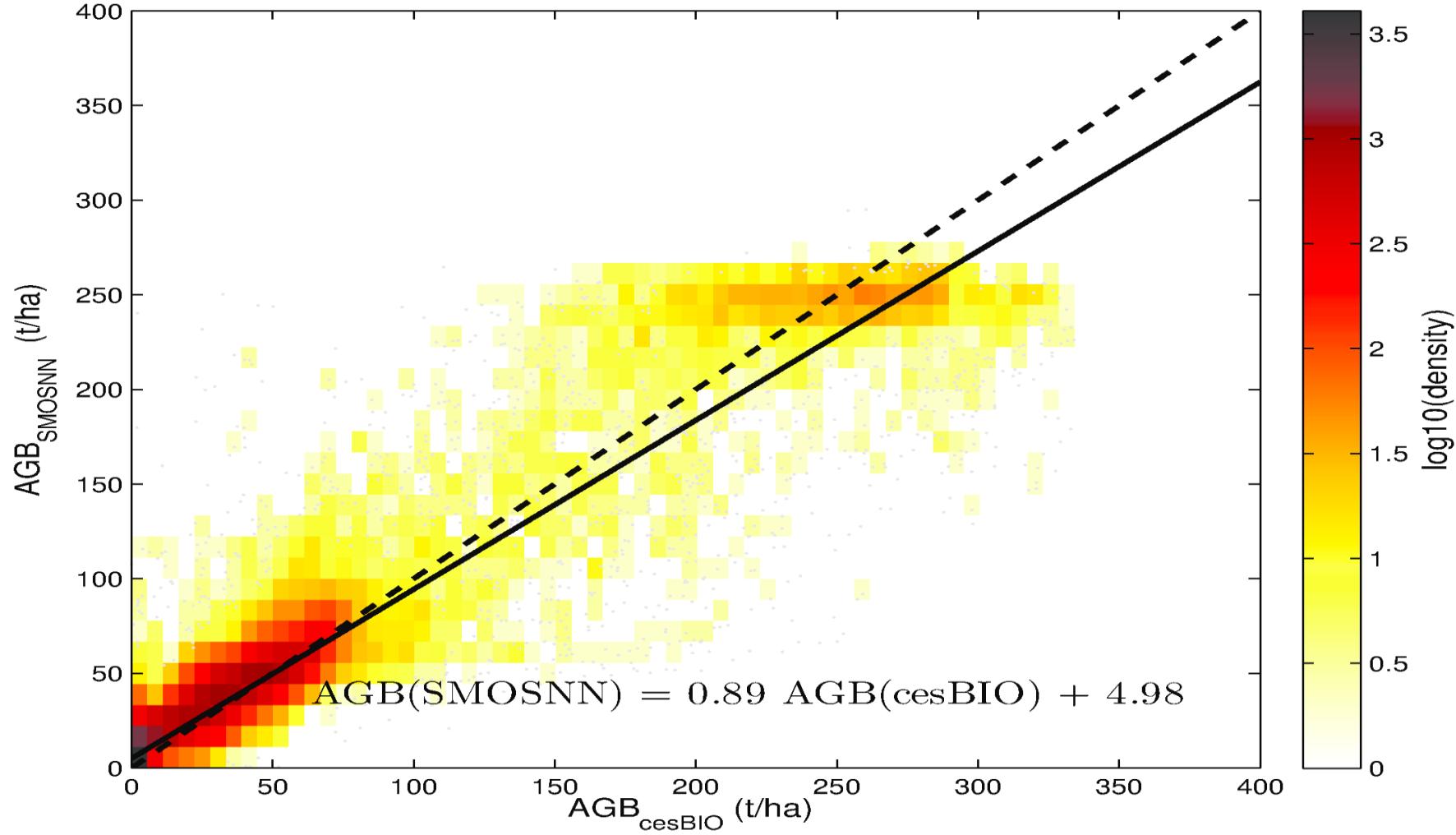
AGB examples: CESBIO AGB



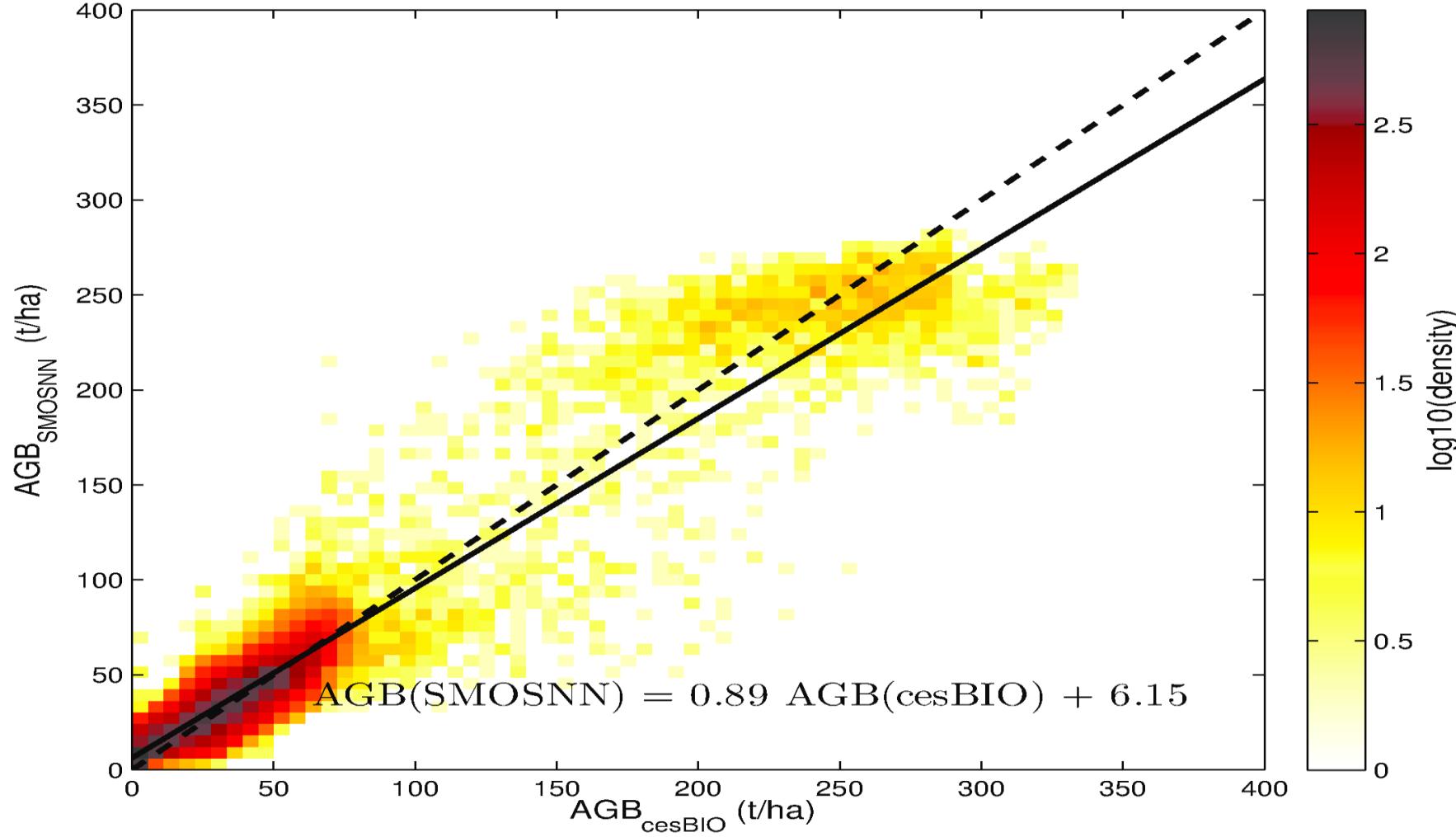
tb tarcesBIO; R =0.94; RMSE=22.161; MAE=12.667; STD=22.161

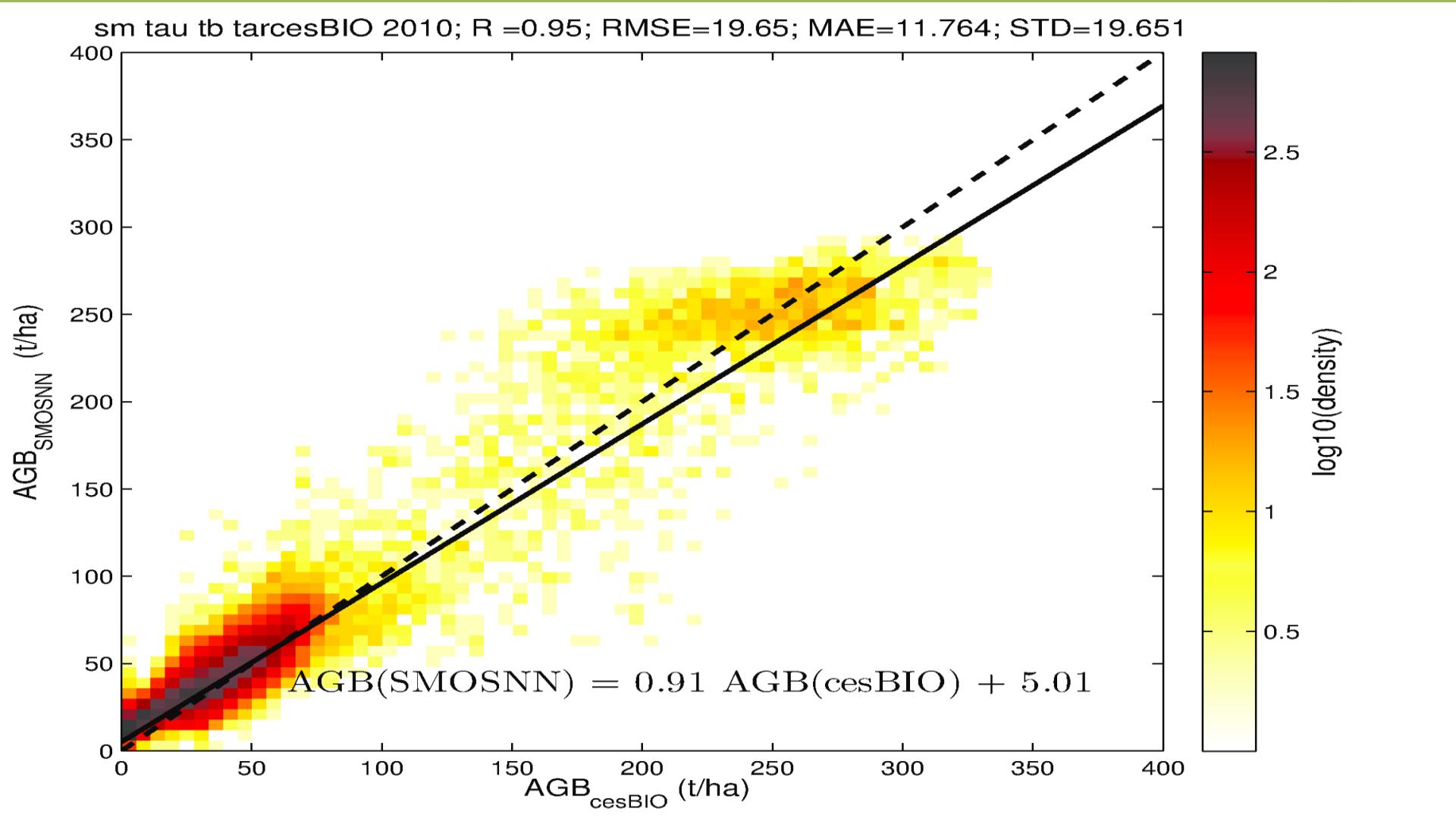


sm tb tarcesBIO; R = 0.94; RMSE=20.509; MAE=11.39; STD=20.509



tau tb tarcesBIO; R =0.94; RMSE=21.229; MAE=12.544; STD=21.228





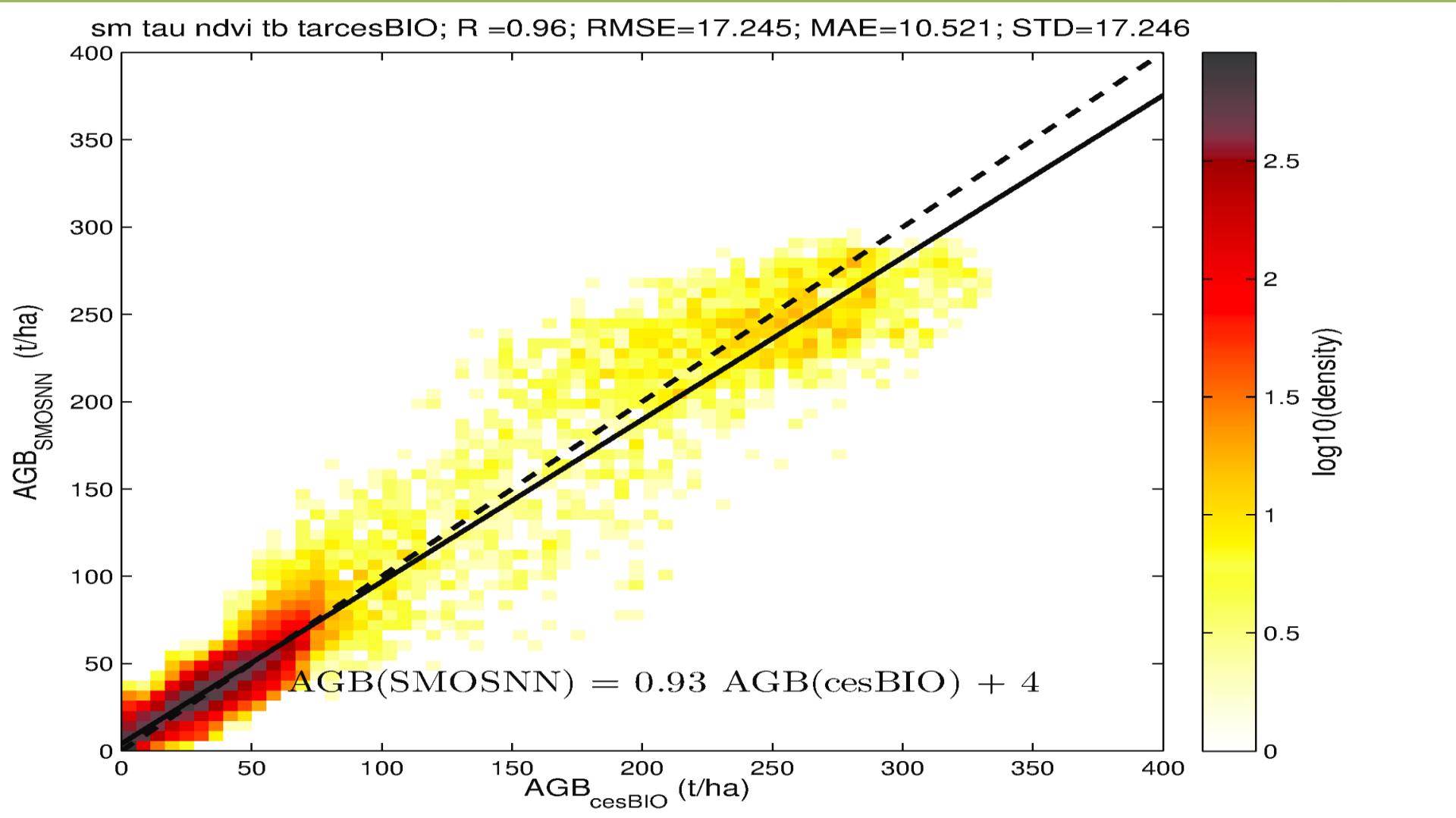


Table de matières



- Contexte, problématique scientifique
 - Humidité du sol et contenu en eau de la végétation
- Concepts réseaux de neurones
 - *Feed-forward network*, Perceptron multi-couches
- Exemples d'applications
 - Synergie multi-capteur
 - Longues séries temporelles
 - Chaîne en temps quasi-réel de l'ESA (humidité SMOS)
 - Prétraitement de données avant assimilation dans le modèle de prévision météorologique du CEPMMT/ECMWF

Thanks for your attention !



- More information



A screenshot of a computer screen displaying the SMOS blog website. The header features the CESBIO logo, a satellite in space, and the text "Soil Moisture and Ocean Salinity Satellite". Below the header, the title "SMOS blog" is displayed. A navigation bar includes links for Home, General Information, Data, Projects Pages, Drought Monitor, Team, FAQ, and Memo Conf. The main content area shows a post titled "SMOS is on twitter!" with a link to their Twitter page. Another post titled "LEWIS on the move: Part 2" is shown, dated 11-11-2014, posted by Yann KERR. This post includes a map of the world showing soil moisture levels and a map of RFI probability. On the right side of the blog, there are two small images: "Latest Soil Moisture map from SMOS CATDS" and "Latest RFI probability maps from SMOS".



@SMOS_satellite