Integration of novel ground-based and SAR-based soil moisture observation systems:

A first step towards automated monitoring of hydrometeorology using Artificial Intelligence

NGUYEN HOANG HAI,
Ph.D.
Sejong Rain Co., Ltd.
In-house Venture of K-Water

세종강우

CONTENTS

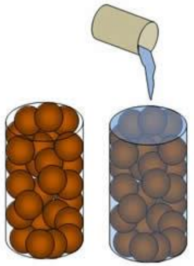
I. INTEGRATION OF NOVEL FIELD-SCALE SOIL MOISTURE OBSERVATION SYSTEMS

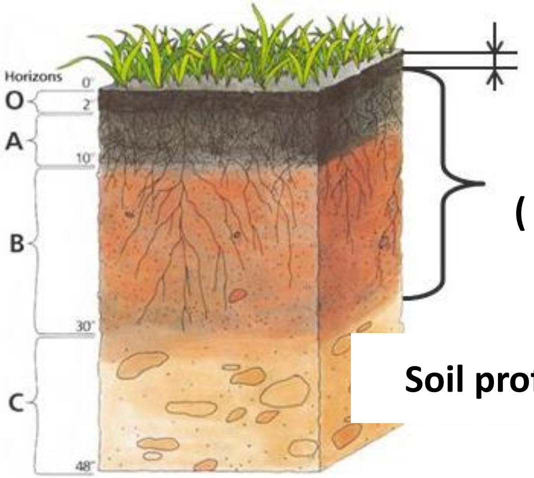
1. Introduction and Objectives

1.1. Theoretical background

Soil moisture and its applications

Soil moisture is the quantity of water contained in the soil



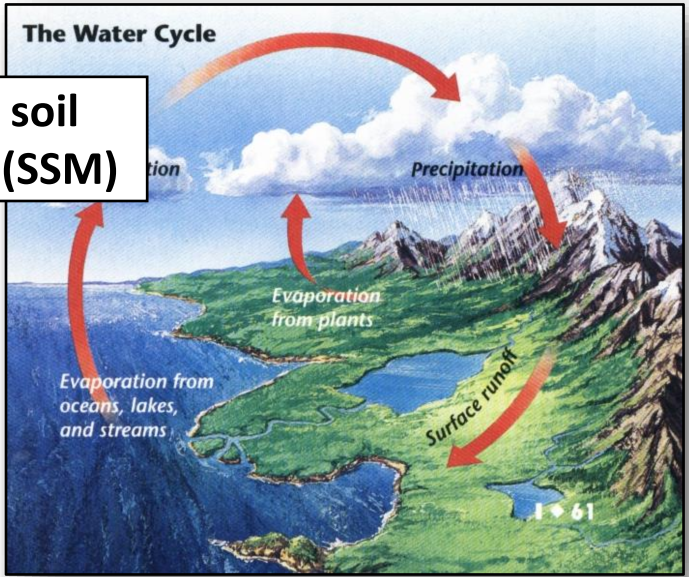


Soil profile

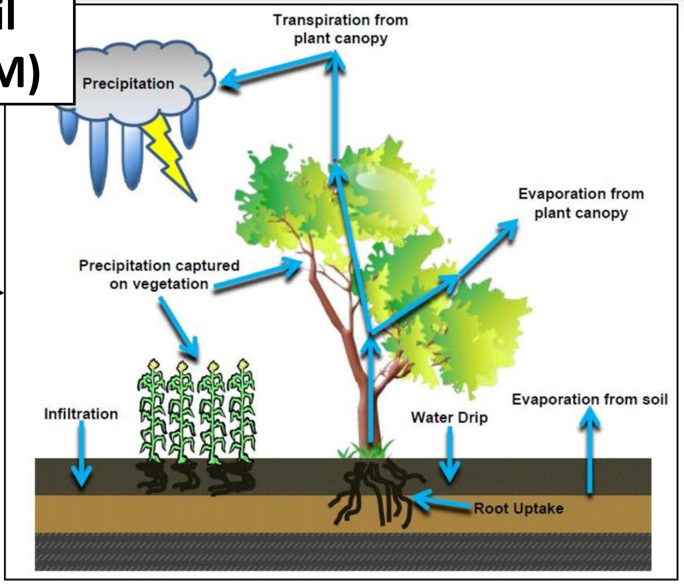
Surface layer
(~ 5 cm depth)

Root-zone layer
(up to 100 cm depth)

Surface soil moisture (SSM)



Root-zone soil moisture (RZSM)



Applications

Rainfall & Flood Forecasting

Drought Monitoring

Irrigation Scheduling

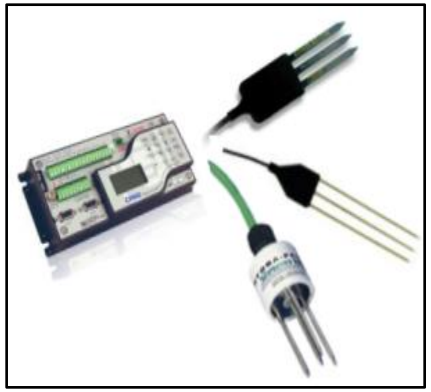
Sand Dust Outbreaks Risk Assessment

1.1. Theoretical background

Soil moisture measurements

In-situ measurements
Point scale

TDR (Time Domain Reflectometry)



FDR (Frequency Domain Reflectometry)

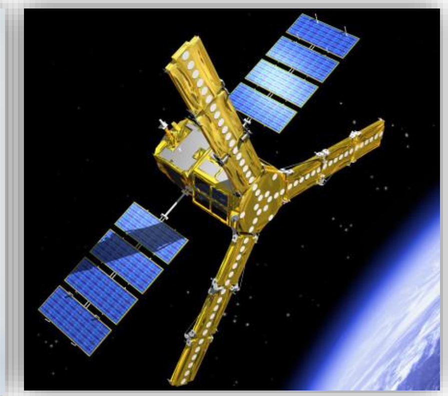
- ❖ **Advantages:**
- Higher temporal resolution
- Can provide RZSM at deeper layers
- ❖ **Disadvantages:**
- Point scale measurement (low spatial coverage)
- Require dense network for field-scale monitoring

Scale mismatch

Cosmic-ray Neutron Probe (CRNP)
Intermediate scale

Remote sensing measurements
Large scale

Airborne Remote Sensing



Satellite Remote Sensing

- ❖ **Advantages:**
- Large scale (regional, continental, global scale) measurement (high spatial coverage)
- ❖ **Disadvantages:**
- Lower temporal resolution
- Only provide shallow SSM
- Coarse spatial resolution cannot provide detailed information for field-scale monitoring

1.1. Theoretical background

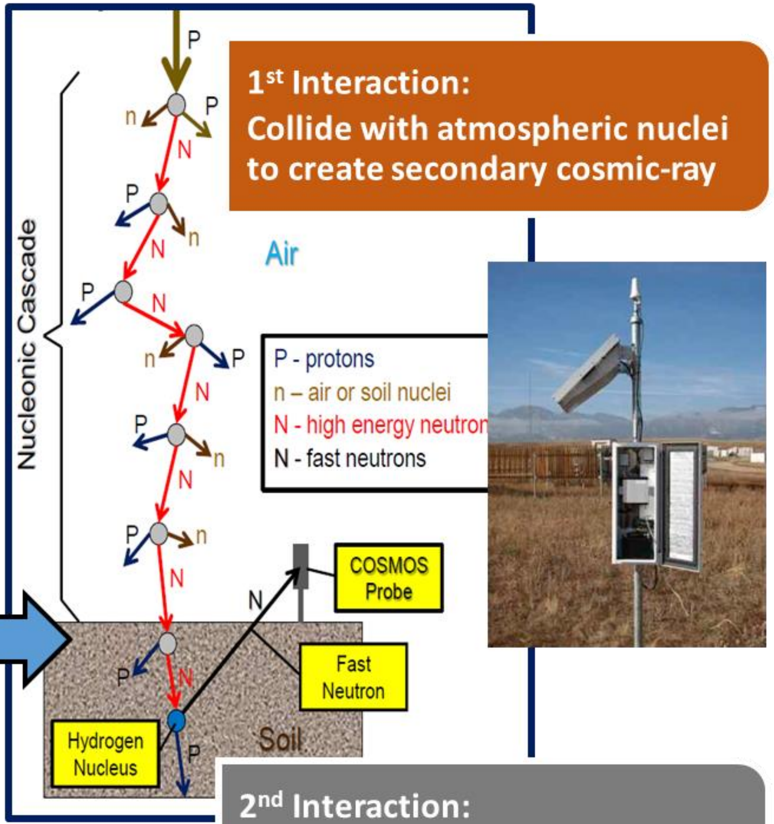
Cosmic-ray soil moisture theory

supernovae

Primary cosmic rays (high energy neutrons)

High energy radiation originated from outside the Solar system

Interaction of primary cosmic rays on the Earth

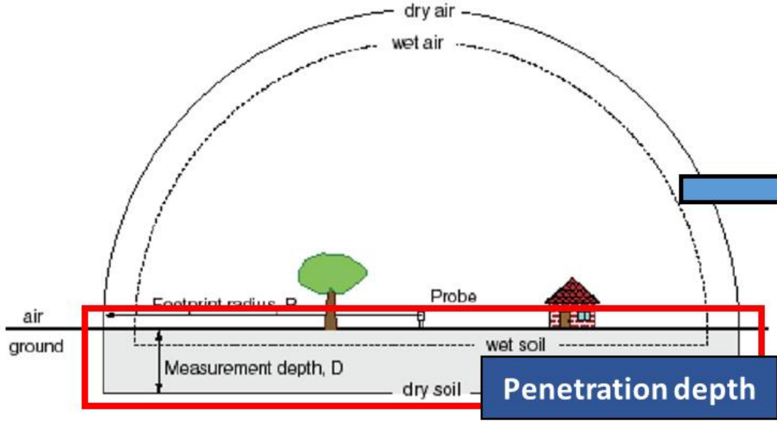


1st Interaction:
Collide with atmospheric nuclei to create secondary cosmic-ray

2nd Interaction:
Collide with soil nuclei and create low-energy fast neutrons

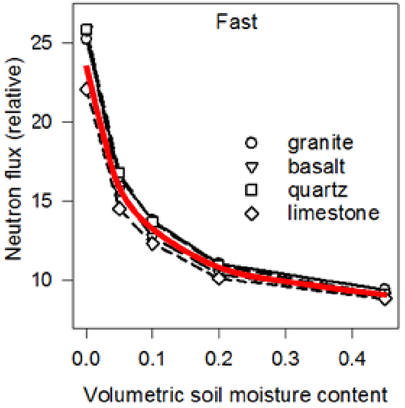
Soil moisture measurement from CRNP

Supporting volume



Horizontal footprint

Traditional Calibration



$$\theta_{vwc} = \left(\frac{a_0}{\frac{N_{corr}}{N_0} - a_1} - a_2 \right) \rho_{bd}$$

$a_0 = 0.0808$
$a_1 = 0.372$
$a_2 = 0.115$
ρ_{bd} = soil bulk density
N_0 (cph) = neutron intensity over dry soil (calibration parameter)

Holding the potential for measuring both field-scale SSM and RZSM

1.2. Motivation

(1) Uncertainties in CRNP Traditional Calibration

Traditional Calibration

$$\theta_{vwc} = \left(\frac{a_0}{\frac{N_{corr}}{N_0} - a_1} - a_2 \right) \rho_{bd}$$

Controlled by 1 parameter N_0

N_0 depends on calibration datasets

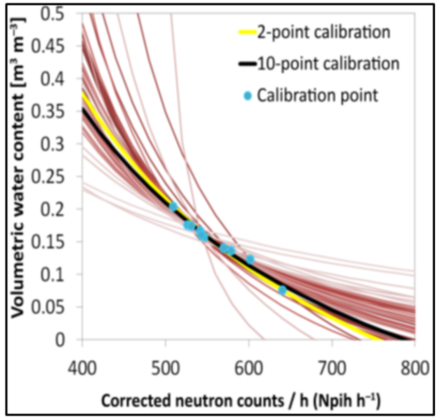
1-point calibration is not sufficient to define the shape

Multi-point calibrations (Heidbuchel et al., 2016)

Select preferred soil wetness conditions can limit the number of calibration points (Iwema et al., 2015)

Test the calibration considering soil wetness conditions over vegetated areas

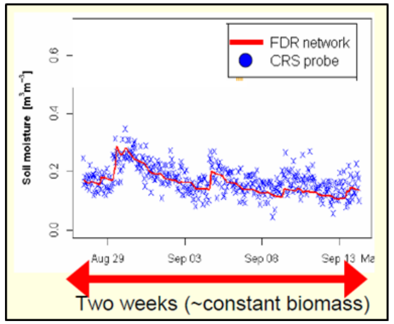
1-point vs. Multi-point calibrations



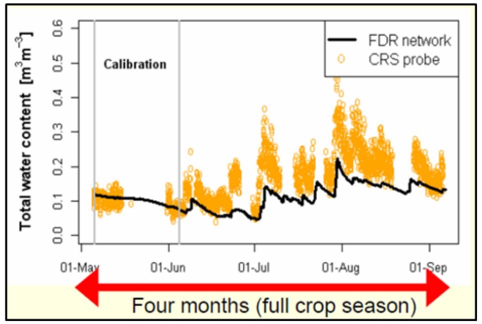
Can vegetation affect to determination of the calibration shape ?

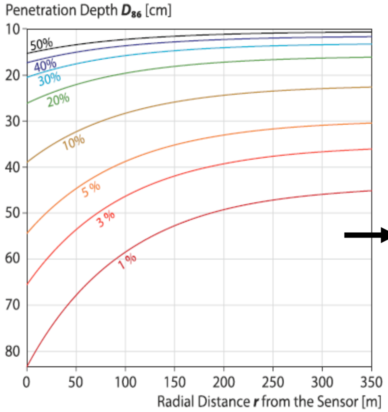
(2) Limitation in CRNP Penetration Depth

Constant biomass
(Rivera Villareyes et al., 2011)



Crop growth
(Baroni and Oswalds, 2015)





Total soil water content is inversely correlated with CRNP penetration depth

Increase of total soil water content can make decrease of CRNP penetration depth

Over vegetation growth, biomass water increase -> decrease in penetration depth

CRNP captures both biomass water and SM but in-situ sensors capture only SM

Combination of CRNP and In-situ sensors for improving field-scale RZSM estimation over vegetated areas

Limited in penetration depth

Field-scale measurement

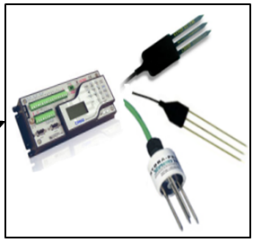
Traditional Calibration

$$\theta_{total} = \left(\frac{a_0}{\frac{N_{corr}}{N_0} - a_1} - a_2 \right) \rho_{bd}$$

$$\theta_{total} = \theta_{vwc} \pm \theta_{biomass}$$

Provide RZSM

In-situ sensors



Limited in point-scale measurements

1.2. Motivation

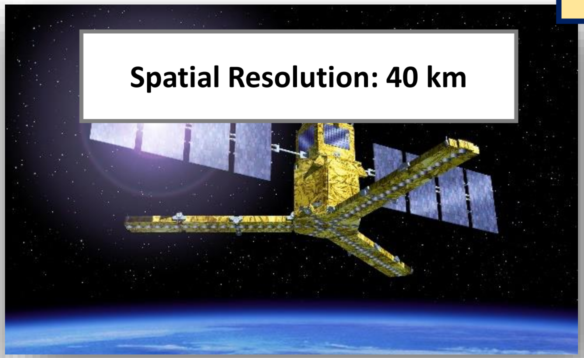
(3) Applications for supporting remote sensing products

Advanced Scatterometer (ASCAT)

Advanced Microwave Scanning Radiometer 2 (AMSR2)

Too coarse resolution

Soil Moisture Active Passive (SMAP)



Soil Moisture Ocean Salinity (SMOS)

CRNP

Footprint Diameter ~ 500 m

~

Synthetic Aperture Radar (SAR) Sentinel-1

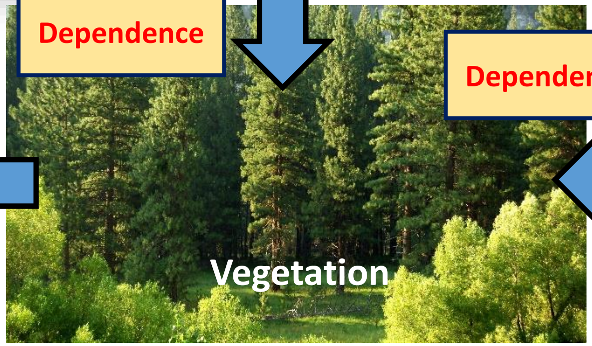
- **Only provide backscattering measurements**

❖ **Advantages:**
- High spatial resolution: 20 m
- Field-scale to catchment scale measurement
- Medium temporal resolution: 6-12 days
- Sensitive to soil moisture

❖ **Disadvantages:**
- Require retrieval methods to generate soil moisture products
- Highly affected by **vegetation**

Dependence

Dependence



Can CRNP be an useful tool for SAR Sentinel-1 soil moisture retrievals across different vegetation conditions?

Can we characterize the interactions of (SAR backscattering – vegetation – cosmic-ray soil moisture) at field-scale through modelling their interdependences?

1.2. Objectives

	Objective	Questions
1	Improvement of CRNP calibration for field-scale SSM estimation	• Which wetness conditions can produce the most reliable cosmic-ray soil moisture? • How does vegetation covers affected the calibration accuracy?
2	Integration of CRNP & In-situ Sensors for field-scale RZSM estimation	• How to combine CRNP and a representative In-situ soil moisture measurements for improving field-scale RZSM estimation? • How does field-scale merged RZSM vary across vegetation covers and its relationship to parent products quality?
3	Application of CRNP for field-scale SAR Sentinel-1 SM retrieval	• How are inter-dependence structures of field-scale radar backscatter, vegetation, and cosmic-ray soil moisture over vegetation covers? • Can SM be probabilistically retrieved based on the coupling CRNP-SAR? and how does its uncertainties vary across vegetation covers?

I. INTEGRATION OF NOVEL FIELD-SCALE SOIL MOISTURE OBSERVATION SYSTEMS

2. Integration of CRNP and In-situ Sensors for field-scale SSM and RZSM estimation

2.1. Study areas and Datasets

Tonzi Ranch & Kendall

Cosmic-ray Neutron Probe (COSMOS Network)

- hourly neutron intensity
- Study period: 1 year (2016)

In-situ soil moisture (SOILSCAPE Network)

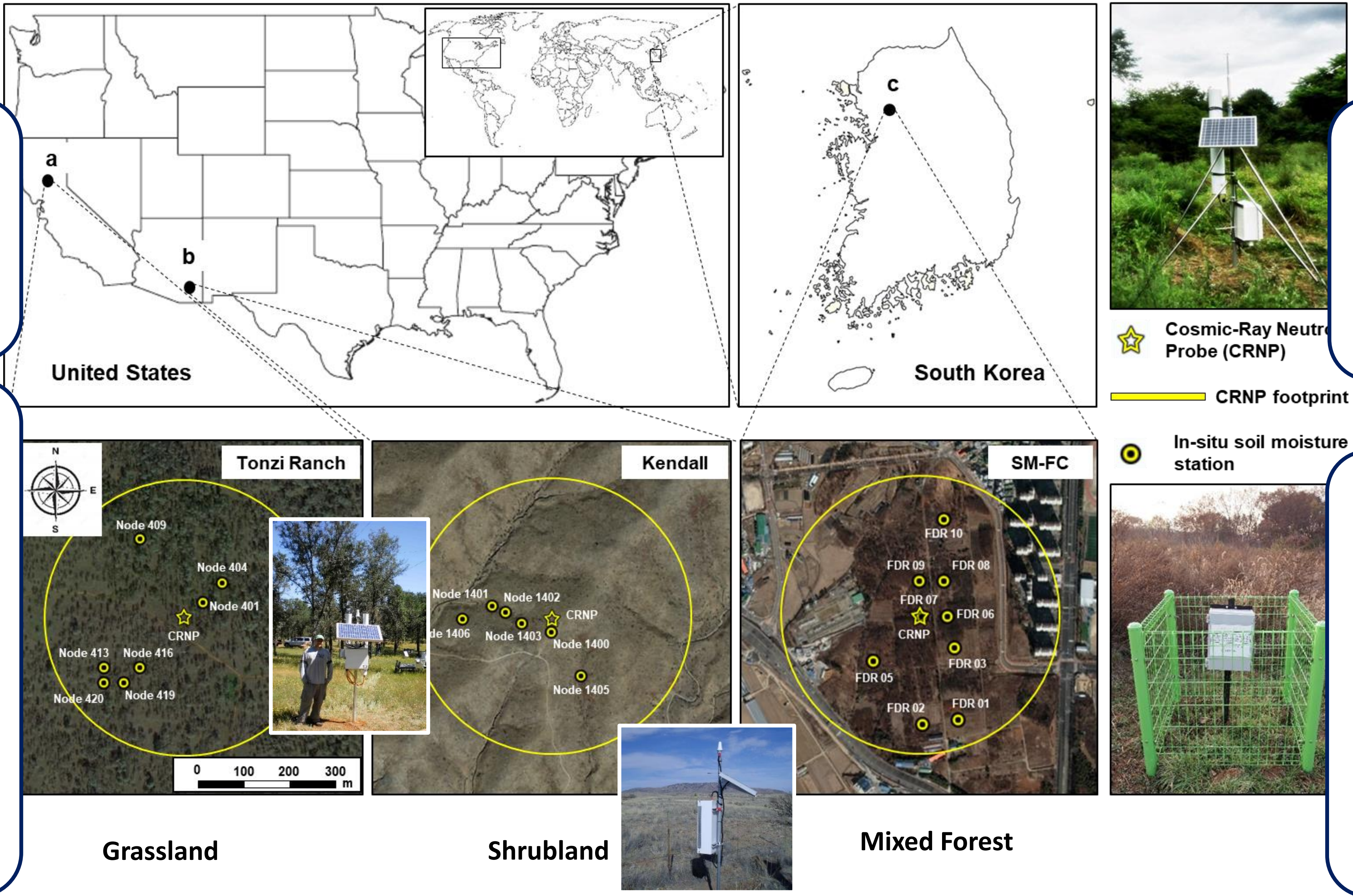
- hourly volumetric soil moisture

❖ **Tonzi Ranch:**

- 7 stations
- 3 measurement depths: 5, 20, 50 cm
- Study period: 1 year (2016)

❖ **Kendall:**

- 6 stations
- 3 measurement depths: 5, 15, 30 cm
- Study period: 1 year (2016)



SM-FC

Cosmic-ray Neutron Probe (CRNP)

- hourly neutron intensity
- Study period: (Sep. – Nov. 2015) & (Apr. – Jun. 2016)

In-situ soil moisture Frequency Domain Reflectometry (FDR) Network

- hourly volumetric soil moisture
- 9 stations
- 3 measurement depths: 10, 20, 30 cm
- Study period: (Sep. – Nov. 2015) & (Apr. – Jun. 2016)

2.1. Study areas and Datasets

United States → COSMOS + SoilSCAPE Network → Tonzi Ranch (Grassland); Kendall (Shrubland)

South Korea → SM-FC (Mixed Forest)

Tonzi Ranch (Grassland)

Site ID	Location		Distance r (m)	Soil texture
	Lat	Lon		
CRNP	38.432 N	120.966 W	0	• **0 – 5cm:** Loam • **5 – 20cm:** Loam • **20 – 50cm:** Clay Loam
Node 401	38.432 N	120.965 W	55	
Node 403	38.432 N	120.965 W	81	
Node 412	38.431 N	120.967 W	99	
Node 415	38.431 N	120.967 W	127	
Node 416	38.431 N	120.967 W	149	
Node 417	38.431 N	120.967 W	155	
Node 418	38.431 N	120.968 W	211	

Kendall (Shrubland)

Site ID	Location		Distance r (m)	Soil texture
	Lat	Lon		
CRNP	31.737 N	109. 942 W	0	• **0 – 5cm:** Loam • **5 – 15cm:** Loam • **15 – 30cm:** Loam
Node 1400	31.736 N	109. 942 W	36	
Node 1401	31.737 N	109. 943 W	157	
Node 1402	31.737 N	109. 943 W	102	
Node 1403	31.737 N	109. 943 W	78	
Node 1405	31.735 N	109. 941 W	164	
Node 1406	31.737 N	109. 944 W	226	

SM-FC (Mixed Forest)

Site ID	Location		Distance r (m)	Soil texture
	Lat	Lon		
CRNP	37.292 N	126.966 E	0	Sandy Loam
FDR 01	37.290 N	126.967 E	265	Sandy Loam
FDR 02	37.289 N	126.966 E	258	Loamy Sand
FDR 03	37.291 N	126.967 E	107	Sandy Loam
FDR 05	37.291 N	126.964 E	157	Loamy Sand
FDR 06	37.292 N	126.966 E	68	Sandy Loam
FDR 07	37.292 N	126.966 E	0	Sandy Loam
FDR 08	37.293 N	126.966 E	113	Loamy Sand
FDR 09	37.293 N	126.966 E	94	Loamy Sand
FDR 10	37.294 N	126.966 E	251	Loamy Sand

2.2. Methodology

Correction for neutron intensity

Correction for Atmospheric Pressure

Correction factor for atmospheric pressure (f_p):

$$f_p = \exp\left[\beta\left(P - P_{ref}\right)\right]$$

P = atmospheric pressure at specific site (mbar)
P_{ref} = reference atmospheric pressure (mbar)
β = atmospheric attenuation coefficient ($mbar^{-1}$) (~ 0.0077 $mbar^{-1}$)

Correction for Atmospheric Water Vapor

Correction factor for atmospheric water vapor (f_{wv}):

$$f_{wv} = 1 + 0.0054\left(\rho_{v0} - \rho_{v0}^{ref}\right)$$

ρ_{v0} = absolute humidity at the time measurement ($g.m^{-3}$)
ρ_{v0}^{ref} = reference absolute humidity ($g.m^{-3}$) (~ 0 $g.m^{-3}$ for dry air)

Correction for Incoming Neutron Intensity

Correction factor for incoming neutron intensity (f_i):

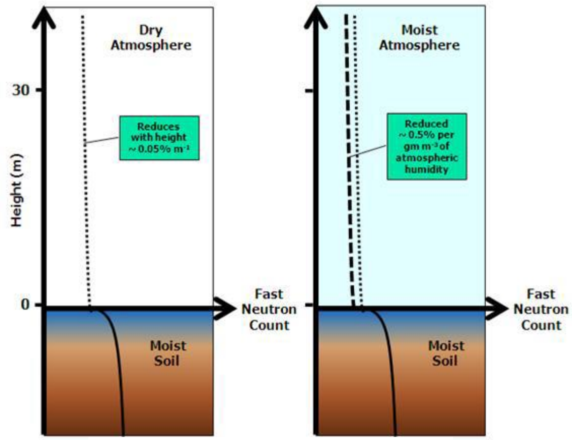
$$f_i = \frac{I_m}{I_0}$$

I_m = neutron monitor intensity at time of measurement
I_0 = reference neutron monitor intensity

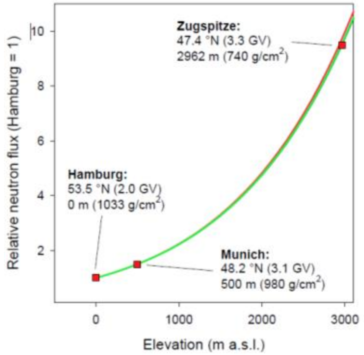
Variation of Cosmic-ray Neutron Intensity

- Vary with altitude -> change in barometric pressure
- Vary with atmospheric water vapor (atmospheric moist contributes to hydrogen sources that is sensitive to neutrons)
- Vary with latitude and longitude (geomagnetic cutoff rigidity)

Variation with atmospheric water vapor



Variation with altitude (pressure)



Correction for Neutron Intensity

Correction for Neutron Intensity:

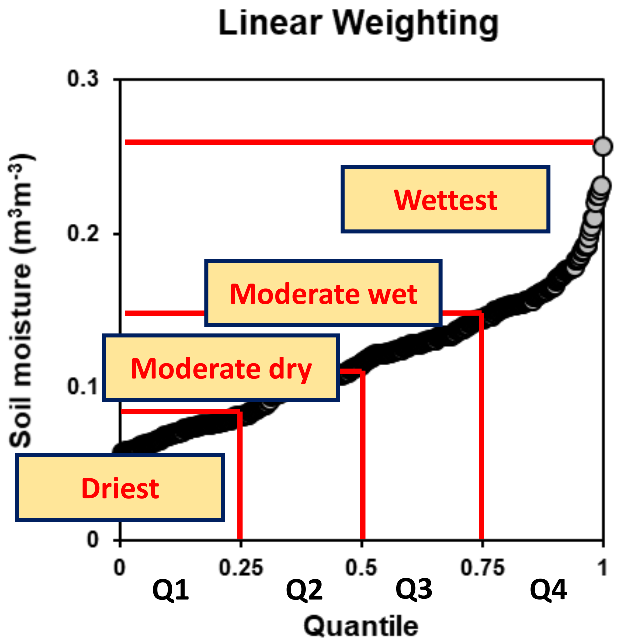
$$\boxed{N_{corr} = N_{raw} \cdot \frac{f_p \cdot f_{wv}}{f_i}}$$

N_{raw} = raw neutron intensity measured at CRNP
N_{corr} = corrected neutron intensity

2.2. Methodology

Calibration scenarios

In-situ point measurements → Weighting In-situ measurements (using nonlinear methods) → Weighted average SM →



Separation of wetness levels based on soil moisture Quartiles

→

$$\theta_{weighted} = \left(\frac{a_0}{\frac{N_{corr}}{N_0} - a_1} - a_2 \right) \rho_{bd}$$

Traditional Calibration

Metrics for performance evaluation

- Pearson Correlation (R)
- Root-Mean-Square Error (RMSE)
- Bias

Kling-Gupta Efficiency (KGE)

- Euclidian Distance:

$$ED = \sqrt{[s_R.(R-1)]^2 + [s_\gamma(\gamma-1)]^2 + [s_\varepsilon(\varepsilon-1)]^2}$$

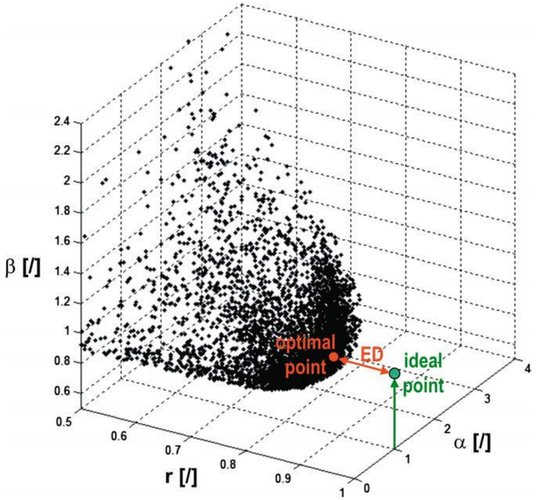
$$ED = \sqrt{(R-1)^2 + (\gamma-1)^2 + (\varepsilon-1)^2}$$

- KGE: $KGE = 1 - ED$

$$\boxed{KGE = 1 - \sqrt{(R-1)^2 + (\gamma-1)^2 + (\varepsilon-1)^2}}$$

R = Correlation
$\varepsilon = \mu(\theta_e) / \mu(\theta_i)$ = bias ratio
$\gamma = (\sigma(\theta_e)/\mu(\theta_e)) / (\sigma(\theta_i)/\mu(\theta_i))$ = variability ratio

KGE range from $-\infty$ to 1, where 1 corresponding to a perfect fit



2.2. Methodology

Temporal Stability Analysis

- In the network, there has a location that its soil moisture measurement is invariant during time (time stable location)
- Soil moisture from a point-based time stable location can represent field-average soil moisture within network (representative location)

Temporal Stability Analysis

Relative Difference (RD_{ij})

$$\frac{\theta_{ij} - \overline{\theta_j}}{\overline{\theta_j}}$$

where θ_{ij} is the soil moisture observed at location i (i = 1,…,N) at the time j (j = 1,…,M); the mean of each sampling j:

$$\overline{\theta_j} = \frac{1}{N}\sum_{i=1}^{N}\theta_{ij}$$

Mean of RD_{ij} (MRD_i)

$$MRD_i = \frac{1}{M}\sum_{j=1}^{M} RD_{ij}$$

Standard Deviation of RD_{ij} ($SDRD_i$)

$$SDRD_i = \sqrt{\frac{1}{M-1}\sum_{j=1}^{M}\left(RD_{ij} - MRD_i\right)^2}$$

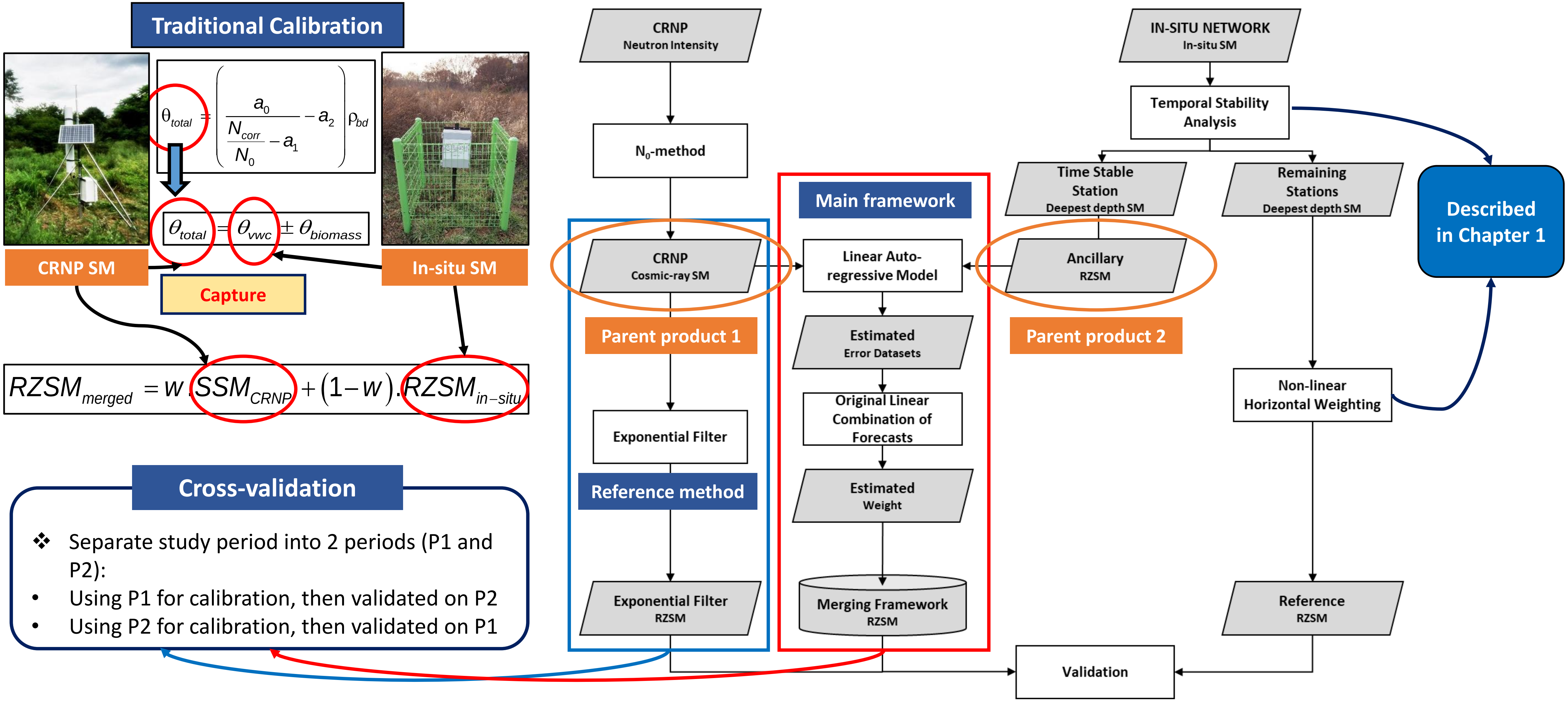
The representative location have ITS value close to zero

Index of Temporal Stability (ITS)

$$ITS = RMSE_i = \sqrt{MRD_i^2 + SDRD_i^2}$$

2.2. Methodology

Flowchart of the merging framework



2.2. Methodology

Detailed processes of merging framework

Main framework

❖ **Linear combination of forecasts:**

$$\theta_{merged} = w\theta_1 + (1-w)\theta_2$$

- The weight (w):

$$w = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$$

σ_1, σ_2, and ρ are SDs of **error datasets** and the correlation of those **2 error datasets**, respectively

❖ **Linear autoregressive model (AR):**

$$E_t = \sum_{i=1}^{p} \phi_i . E_{t-i} + \varepsilon_t$$

E_t = estimated error at time t between a parent soil moisture and the reference
E_{t-i} = historical error value at the lagged i hrs,
Φi = corresponding coefficient for E_{t-I}
ε_t = white noise.
p = order of AR model. Forecast based on past p-hindcasts

Here Selected AR(12)

Exponential Filter

- **The recursive formulation of the original exponential filter:**

$$SWI_2(t_i) = SWI_2(t_{i-1}) + K_i\left[SWI_1(t_i) - SWI_2(t_{i-1})\right]$$

- **Gain K_i:**

$$K_i = \frac{K_{i-1}}{K_{i-1} + \exp(-(t_i - t_{i-1})/T)}$$

K_i range from 0 to 1
K_1 and $SWI_2(1)$ were set to 1 and $SWI_1(1)$, respectively
T = characteristic time length (obtained by calibration)

Metrics for performance evaluation

Pearson Correlation (R)

Root-Mean-Square Error (RMSE)

Bias

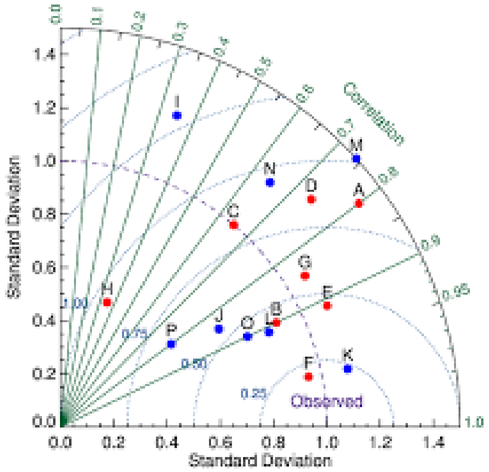
Kling-Gupta Efficiency (KGE)

Taylor Diagram

The relationship of 3 metrics:
Centered RMSD (E')
Correlation (R)
Normalized SD (SDV, ratio of estimated and reference SD)

$$E'^2 = SDV^2 + 1 - 2.SDV.R$$

$$SDV = \frac{\sigma(\theta_e)}{\sigma(\theta_r)}$$



2.3. General Results

Validation results

Tonzi Ranch (Grassland)					
Weighting method	Wetness conditions	R	RMSE	Bias	KGE
Nonlinear	Driest (Q1)	0.91	0.099	0.071	0.43
	Mod. Dry (Q2)	0.91	0.118	0.090	0.31
	Mod. Wet (Q3)	0.91	0.067	0.039	0.63
	Wettest (Q4)	**0.92**	**0.039**	**-0.016**	**0.74**

Kendall (Shrubland)					
Weighting method	Wetness conditions	R	RMSE	Bias	KGE
Nonlinear	Driest (Q1)	0.79	0.033	0.011	0.75
	Mod. Dry (Q2)	0.79	0.049	0.034	0.63
	Mod. Wet (Q3)	0.79	0.036	0.016	0.73
	Wettest (Q4)	**0.79**	**0.032**	**0.010**	**0.75**

SM-FC (Mixed Forest)					
Weighting method	Wetness conditions	R	RMSE	Bias	KGE
Nonlinear	Driest (Q1)	0.77	0.099	0.090	0.54
	Mod. Dry (Q2)	0.77	0.079	0.070	0.61
	Mod. Wet (Q3)	0.76	0.032	0.009	0.75
	Wettest (Q4)	**0.76**	**0.031**	**0.007**	**0.75**

Results

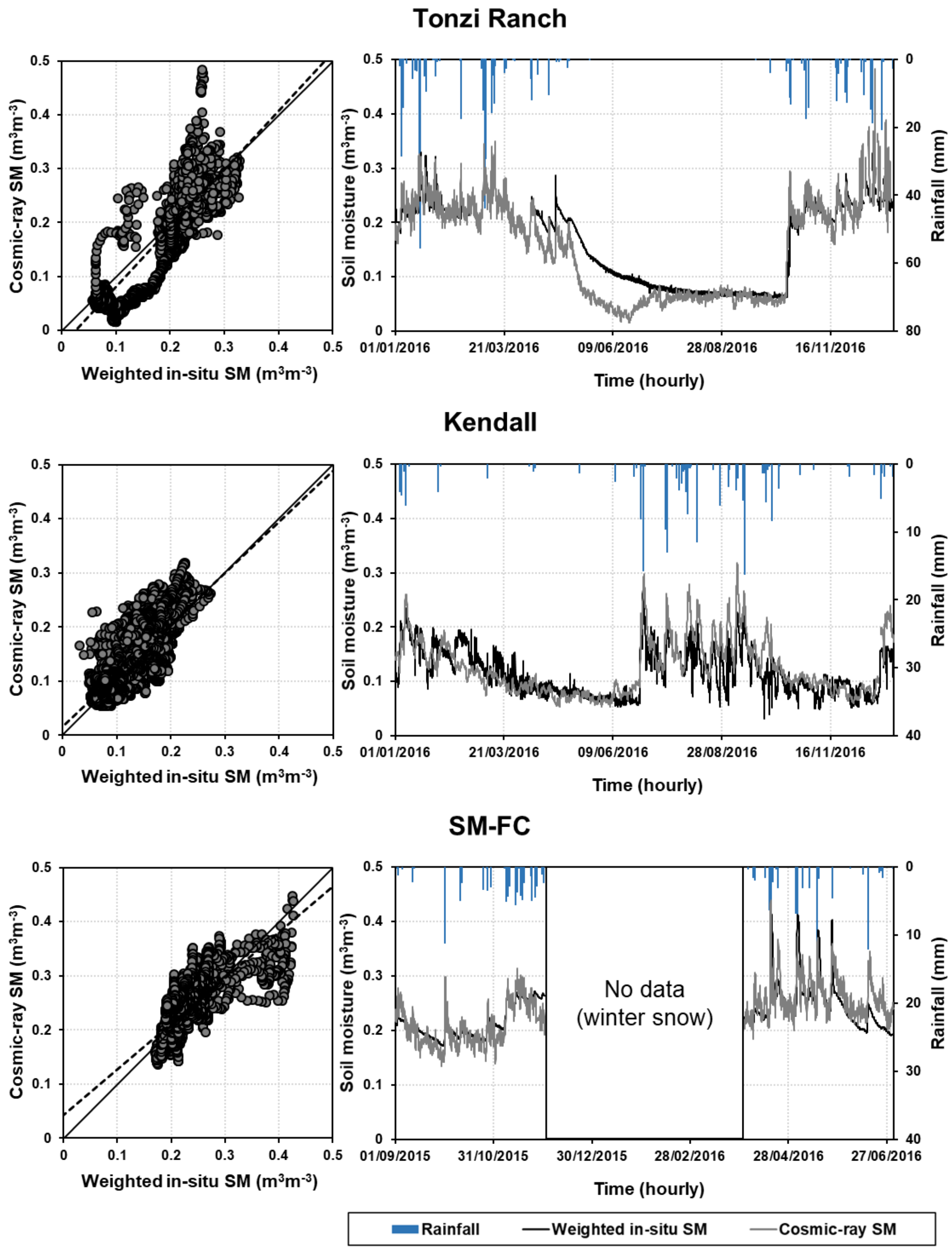
- Calibration considering **wettest conditions** mostly generate the **best cosmic-ray soil moisture products**.
- Relatively dry soil conditions generate the worst performance due to the **change in aboveground hydrogen pools**, especially the litter layer, which **is less dominant under drier conditions.**

2.3. General Results

Surface Soil Moisture (SSM) estimation from CRNP

Sites	Tonzi Ranch (Grassland)		Kendall (Shrubland)		SM-FC (Mixed Forest)	
SM product	Cosmic ray SM	Weighted SM	Cosmic-ray SM	Weighted SM	Cosmic-ray SM	Weighted SM
Mean	0.149	0.156	0.125	0.115	0.237	0.230
CV	0.58	0.47	0.40	0.36	0.19	0.18
Calibration conditions	Wettest (Q4)		Wettest (Q4)		Wettest (Q4)	
N_0	1358		3929		1356	
R	0.92		0.79		0.76	
RMSE	0.039		0.032		0.031	
Bias	-0.016		0.010		0.007	
KGE	0.74		0.75		0.75	

- **Highest correlation (R value)** in **grassland**, followed by shrubland and mixed forest. It is due to low aboveground and belowground biomass accumulation grasses, revealing that cosmic-ray neutron signal is not much affected by other hydrogen sources signal variation (mostly vegetation).
- **Highest variation (CV value)** for **both neutron intensity and soil moisture** in **grassland**, followed by shrubland and mixed forest. It is due to low canopy density of grasses cannot intercept much rainfall water, leading to the high temporal variation directly linked to rainfall variation, compared to forest, where rainfall water is mostly intercepted by canopy.
- The **KGE values** are almost **similar** over 3 vegetation covers (around 0.75) => similar efficiency of calibration method over vegetated areas



Reference: Nguyen, H. H., Kim, H., & Choi, M. (2017). Evaluation of the soil water content using cosmic-ray neutron probe in a heterogeneous monsoon climate-dominated region. *Advances in Water Resources*, *108*, 125-138.

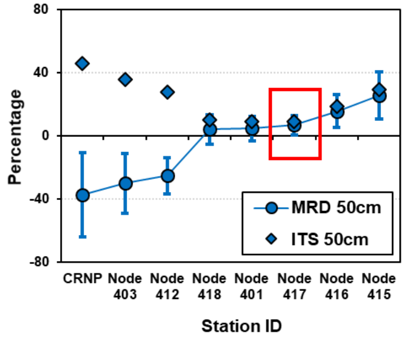
2.3. General Results

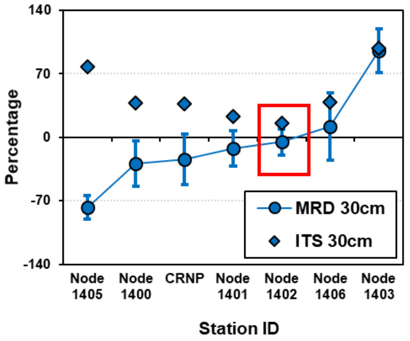
Selection of representative point RZSM using TSA

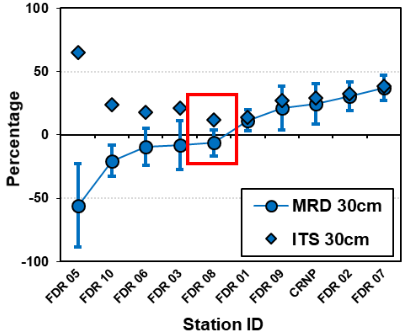
Temporal Stability Analysis applied for the selected RZSM

Rank of small ITS	FDR stations		
	Tonzi Ranch	Kendall	SM-FC
1	Node 417	Node 1402	FDR 08
2	Node 401	Node 1401	FDR 01
3	Node 418	**CRNP**	FDR 06
4	Node 416	Node 1400	FDR 03
5	Node 412	Node 1406	FDR 10
6	Node 415	Node 1405	FDR 09
7	Node 403	Node 1403	**CRNP**
8	**CRNP**		FDR 02
9			FDR 07
10			FDR 05

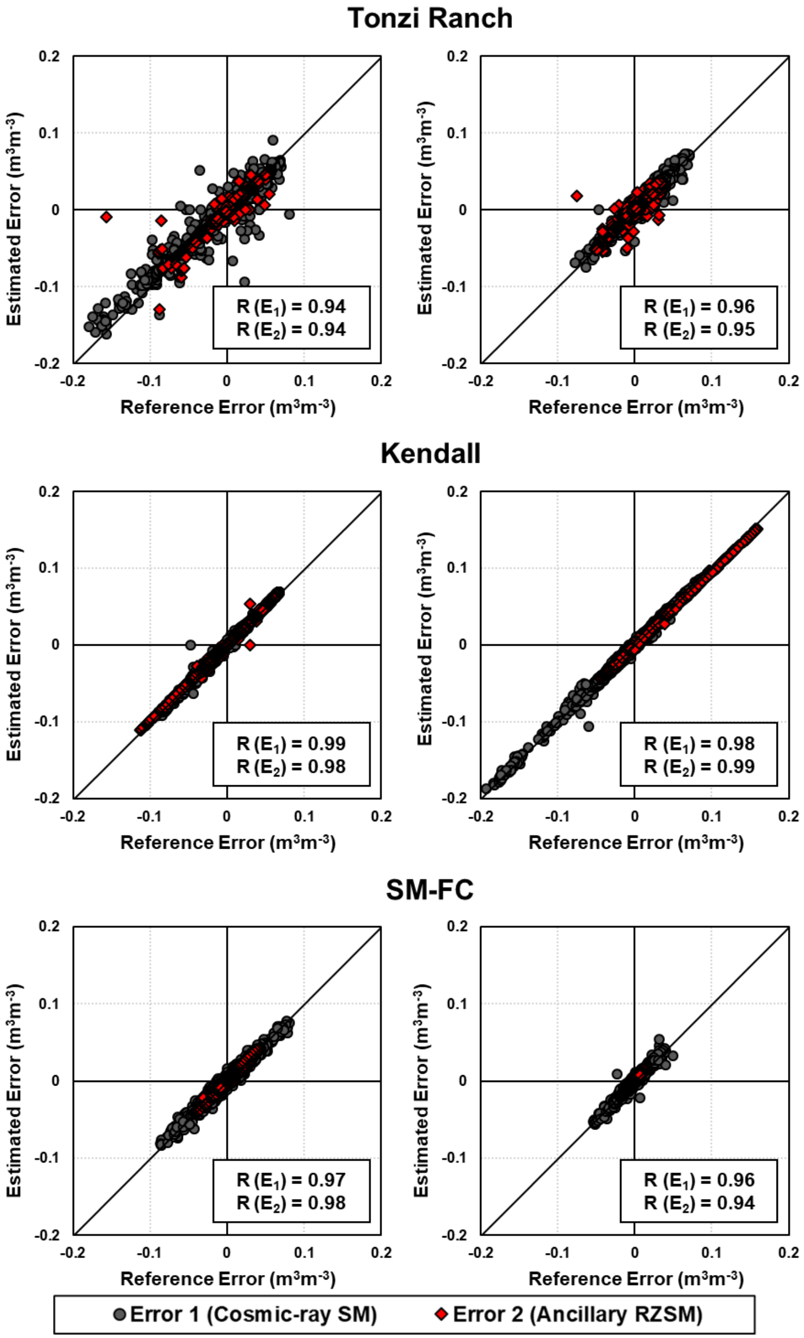
- For 3 sites, the **CRNP have higher ITS values** -> not stable at RZ layer -> **cannot represent field-scale RZSM**







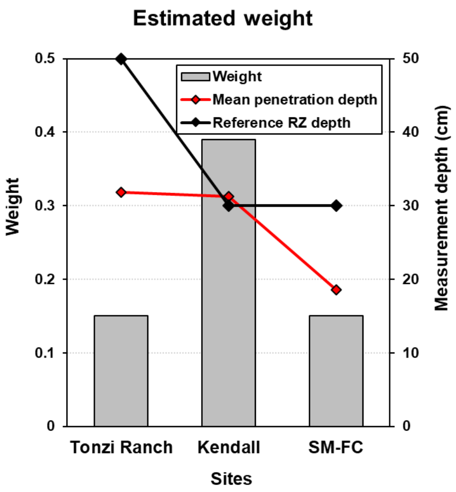
Linear Autoregressive Model AR(12) for Error forecasts



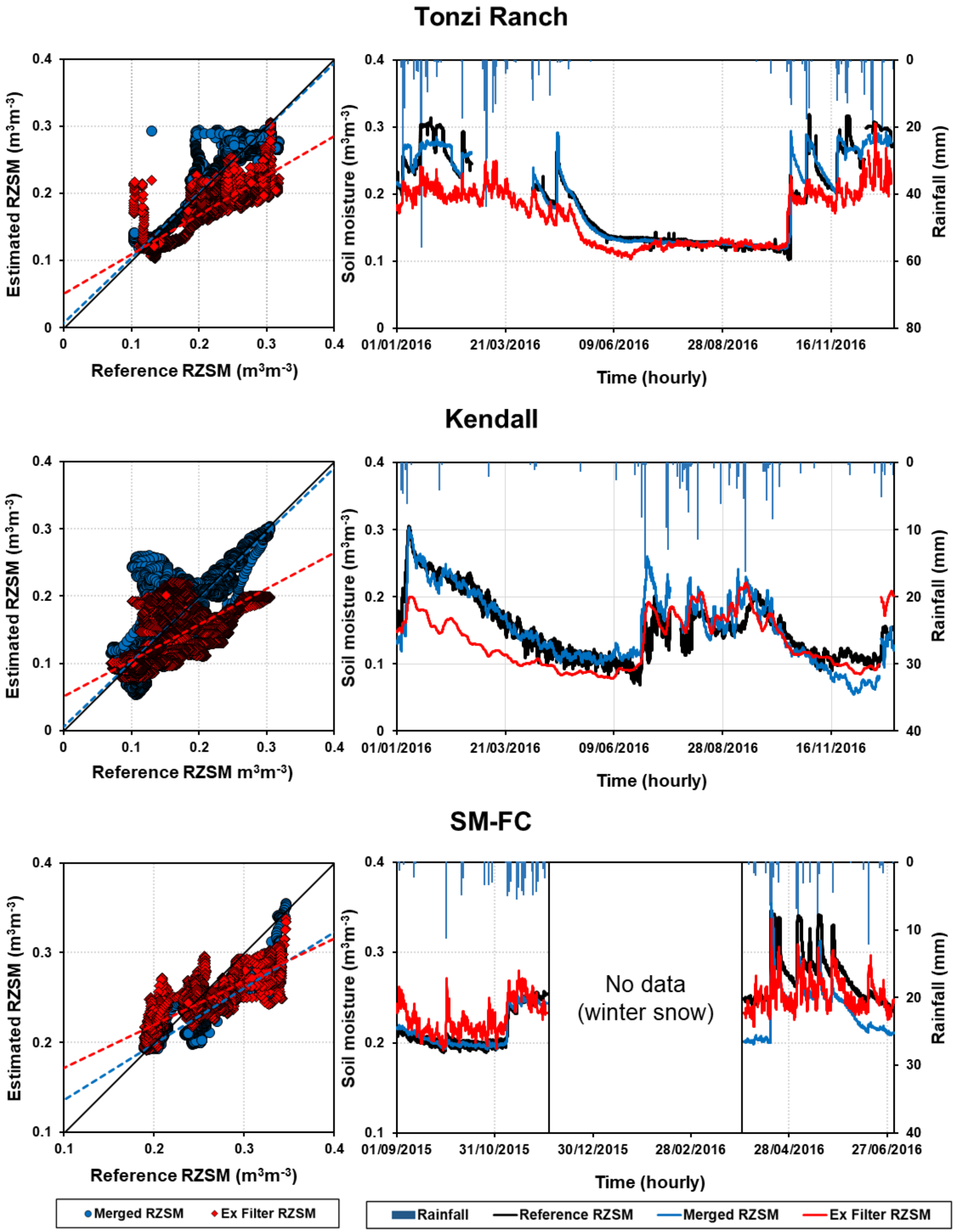
2.3. General Results

RZSM Estimation

		Tonzi Ranch (grassland)		Kendall (shrubland)		SM-FC (mixed forest)	
		P1	P2	P1	P2	P1	P2
Reference RZSM	CV (RZSM)	0.336		0.307		0.161	
Merging framework	Weight (w)	0.11	0.19	0.37	0.41	0.24	0.06
	Mean w	0.15		0.39		0.15	
	R (RZSM)	0.97		0.85		0.88	
	CV (RZSM)	0.331		0.344		0.123	
Exponential Filter	T_{opt} (hrs)	1	4.6	100	100	2	1.8
	Mean T_{opt}	2.8		100		1.9	
	R (RZSM)	0.92		0.64		0.82	
	CV (RZSM)	0.242		0.294		0.095	



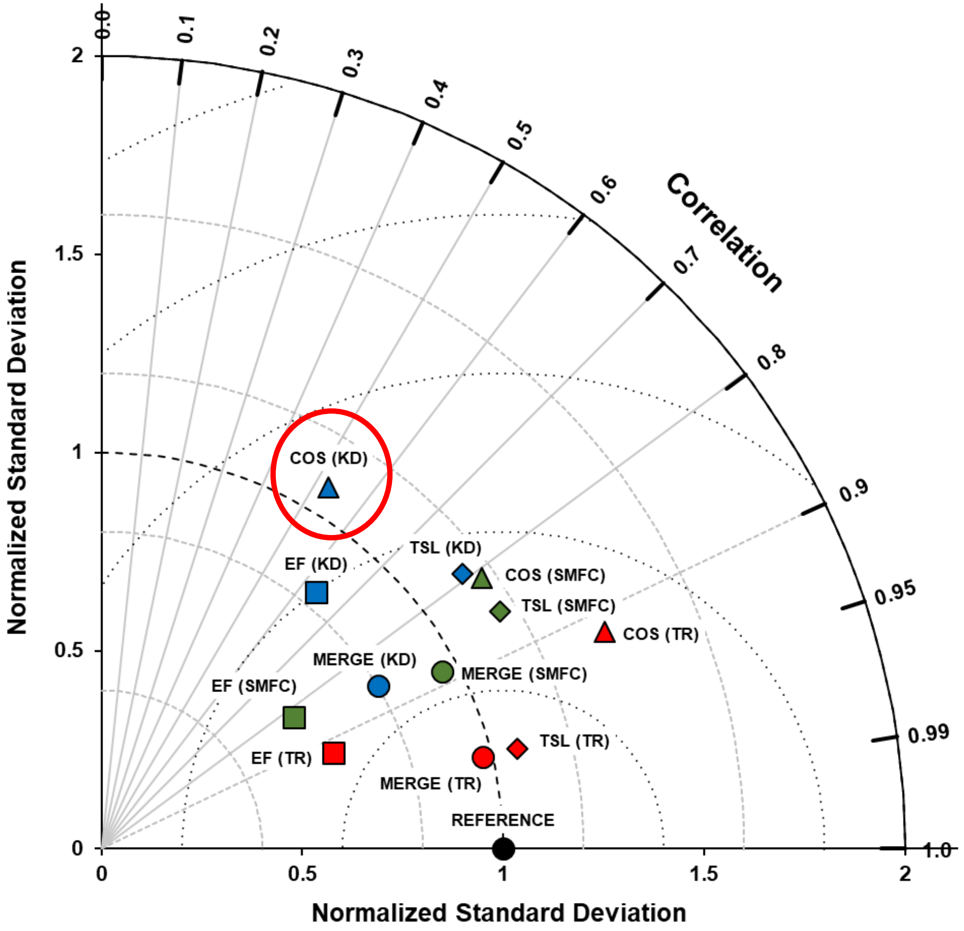
- **Highest variation (CV)** for **Reference RZSM** in **grassland**, followed by shrubland and mixed forest. It is due to (1) the **highest CV in grassland SSM** (mentioned in Chapter 1) and (2) the **grasses have higher root water uptake rate** compared to forest tree -> consume more water at root zone layer
- Merged RZSM have higher variation than Ex. Filtered RZSM but close to Reference RZSM variation. Ex. Filter fail to estimate T in Kendall site -> **limitation of Ex. Filter**
- **The weight (w)** depends on **soil wetness conditions. Dry conditions** -> **deeper CRNP penetration depth** -> **more CRNP contribution** -> **more weight added to CRNP-SSM** and vice versa



2.3. General Results

Performance Evaluation

Sites	Products	R	RMSE	Bias	KGE
Tonzi Ranch (TR) Grassland	Cosmic-ray SM (COS)	0.92	0.062	-0.040	0.23
	Ancillary RZSM (TSL)	0.97	0.033	0.030	0.82
	Merged RZSM (MERGE)	0.97	0.015	-0.004	0.97
	Ex Filtered RZSM (EF)	0.92	0.041	-0.024	0.68
Kendall (KD) Shrubland	Cosmic-ray SM (COS)	0.53	0.055	-0.027	0.41
	Ancillary RZSM (TSL)	0.79	0.034	0.007	0.76
	Merged RZSM (MERGE)	0.86	0.024	0.000	0.76
	Ex Filtered RZSM (EF)	0.64	0.042	-0.019	0.62
SM-FC (SMFC) Mixed Forest	Cosmic-ray SM (COS)	0.81	0.027	-0.006	0.73
	Ancillary RZSM (TSL)	0.86	0.067	-0.063	0.36
	Merged RZSM (MERGE)	0.88	0.018	0.000	0.88
	Ex Filtered RZSM (EF)	0.82	0.024	-0.003	0.55



- For all three vegetation covers, the **merged RZSM outperformed 2 parent products (Cosmic-ray SM and Time stable RZSM) and Exponentially Filtered RZSM -> robust for application in most vegetation covers**
- **Lower performance** of **Cosmic-ray SM** and **Time stable location (TSL)** for RZSM -> **standalone use** of **CRNP and point-based TSL cannot fully represent field-scale RZSM**.
- **Lower performance** of **Exponential Filter** -> limitation because **1 parameter (T_{opt}) cannot fully interpret the physical processes controlling infiltrated water**

- **The quality of merged RZSM** depends on the **quality of each parent product against reference RZSM**
- **More weights** were considered to added to a **better product (higher R and SDV-Normalized SD close to 1)**

Reference: Nguyen, H. H., Jeong, J., & Choi, M. (2019). Extension of cosmic-ray neutron probe measurement depth for improving field scale root-zone soil moisture estimation by coupling with representative in-situ sensors. *Journal of Hydrology*, *571*, 679-696.

I. INTEGRATION OF NOVEL FIELD-SCALE SOIL MOISTURE OBSERVATION SYSTEMS

3. Integration of CRNP and SAR Sentinel-1 for SM estimation over vegetation covers

3.1. Study areas and Datasets

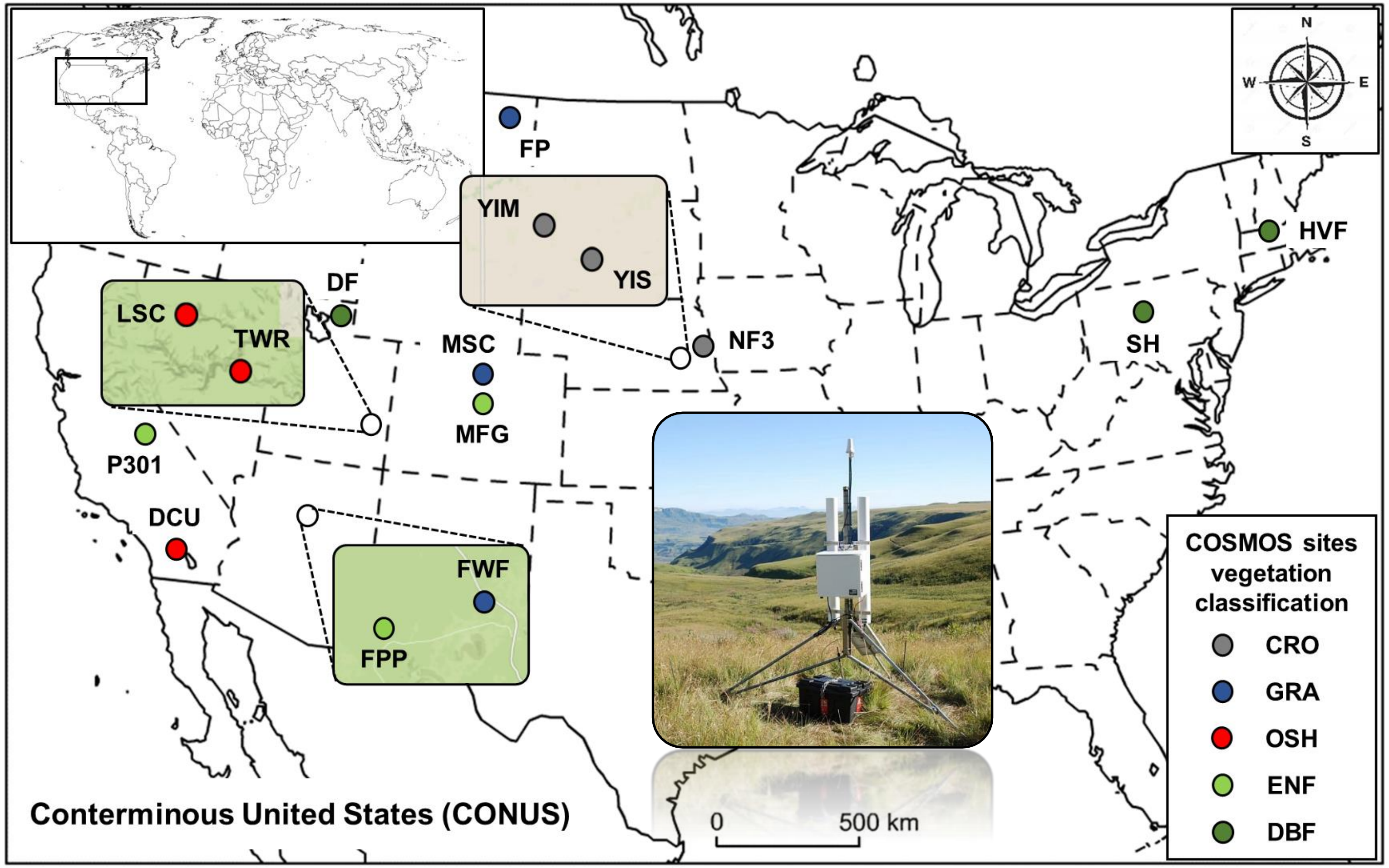
Field-scale ground-based soil moisture dataset

Cosmic-ray Soil Moisture

- **COSMOS Network (USA)**
- **15 sites** over **5 vegetation covers**
- **Level 3 Soil moisture data**
- Study period: **1 year (2017)**

Vegetation Cover				
Cropland	Grassland	Open Shrubland	Evergreen Needleleaf Forest	Deciduous Broadleaf Forest
CRO	GRA	OSH	ENF	DBF

Veg. cover	COSMOS Sites	
CRO	Neb Field 3	NF3
	York Irrigated Maize	YIM
	York Irrigated Soybean	YIS
GRA	Marshall Colorado	MSC
	Fort Peck	FP
	Flag Wildfire	FWF
OSH	Desert Chaparral UCI	DCU
	Lower Salt Creek	LSC
	Tower Ruin	TWR
ENF	P301	P301
	Flag Ponderosa Pine	FPP
	Manitou Forest Ground	MFG
DBF	Daniel Forest	DF
	Havard Forest	HVF
	Shale Hills	SH

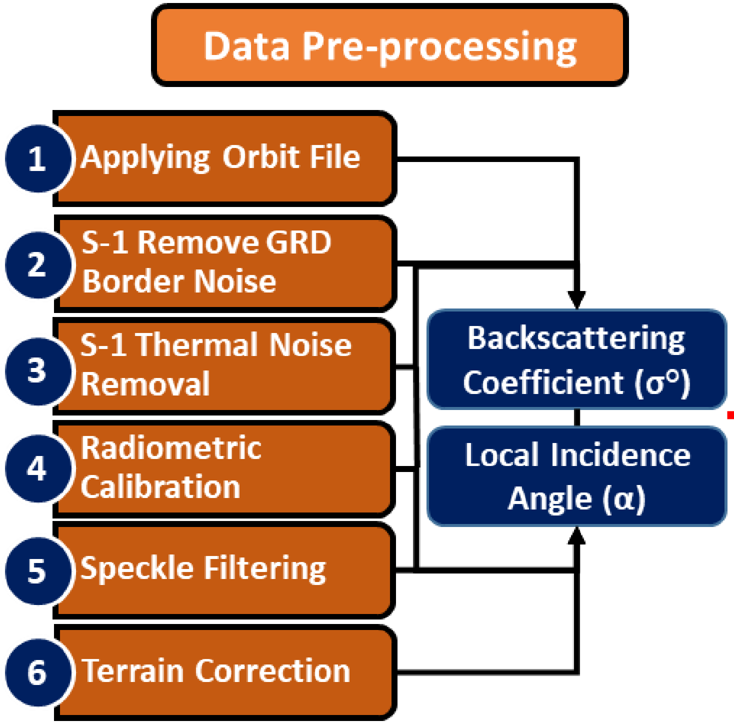


3.1. Study areas and Datasets

SAR dataset

SAR Sentinel-1
Backscattering Coefficient (σ°)

SENTINEL-1	
Sensor Type	Active Microwave (SAR)
Band	C-band (5.4 GHz)
Duration	2014 - present
Temporal Res.	**6 – 12 days**
Spatial Res.	20 m
Selected Specification	**VV + VH polarization, Level1 GRD, IW mode**
Target Spatial Res.	**500 m**
Target Period	**1 year (2017)**



LIA Normalization

- Regression Slope (β_r):

$$\beta_r = (-0.01725)S + (0.00553)\mu(\sigma^o) + (0.02546)$$

- LIA Normalization to reference LIA of 40° for backscattering coefficient:

$$\sigma^o(40,t) = \sigma^o(\alpha,t) - \beta_r(\alpha - 40^o)$$

3.2. Methodology

Copulas and Dependence Modelling

Copulas

❖ **Sklar's Theorem**

- If random variables $x_1,\ldots,x_n$ follow arbitrary marginal distributions $F_1(x_1),\ldots,F_n(x_n)$, respectively, there then exists a copula, C, that combines these marginal distribution functions to give a joint distribution function, $F(x_1, \ldots, x_n)$:

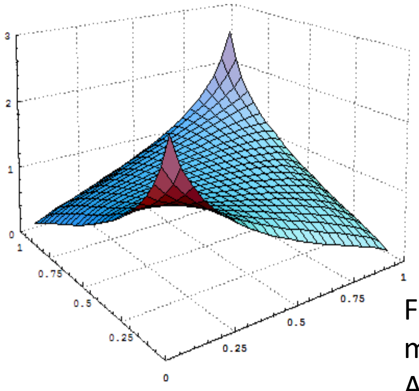
$$F(x_1,\ldots,x_n) = C\{F_1(x_1),\ldots,F_n(x_n)\} = C(u_1,\ldots,u_n), \quad x_1,\ldots,x_n \in R$$

- A **copula** is defined as a **joint distribution function** on the n-dimensional unit cube with all marginal distributions are uniform on [0, 1]. Formally: $C:[0, 1]^n \to [0, 1]$

❖ **Advantages:**

Separate **(1) Distribution Analysis** & **(2) Dependence Analysis**

- Each variable can be modelled by **its arbitrary marginal distribution** => strengthen the nature of variable data
- Each **dependence structure** can be modelled by different **copula**



Conditional Copulas

The **conditional distribution** **($C_{u2|u1}$)** of variable U_2 at u_2 given variable $U_1 = u_1$ and a copula C_{12} that joint two variables U_1 and U_2:

$$C_{u_2|u_1} = \frac{\partial C_{12}(u_1,u_2)}{\partial u_1}$$

1 Marginal Distribution Selection for each Variable

Trying to fit all candidate distributions and parameters → **Determining optimal parameters for all candidate distributions** (using Maximum Likelihoods) → **Selecting best-fit marginal distribution** (using Goodness-of-Fit Tests)

2 Dependence Structure-based Copula Selection

Bivariate Archimedean Copulas

For ϕ is a generators with inverse $(\phi)^{-1}$ completely monotonic on $[0,\infty)[0,\infty)$, then the bivariate Archimedean copula can be defined as:

$$C(u,v) = \varphi^{-1}(\varphi(u)+\varphi(v))$$

Name of Copula	*Bivariate Copula* $C_\theta(u,v)$
Ali-Mikhail-Haq[11]	$\frac{uv}{1-\theta(1-u)(1-v)}$
Clayton[12]	$[\max\{u^{-\theta}+v^{-\theta}-1;0\}]^{-1/\theta}$
Frank	$-\frac{1}{\theta}\log\left[1+\frac{(\exp(-\theta u)-1)(\exp(-\theta v)-1)}{\exp(-\theta)-1}\right]$
Gumbel	$\exp\left[-((-\log(u))^\theta+(-\log(v))^\theta)^{1/\theta}\right]$
Independence	uv
Joe	$1-[(1-u)^\theta+(1-v)^\theta-(1-u)^\theta(1-v)^\theta]^{1/\theta}$

name	*generator* $\psi_\theta(t)$	*generator inverse* $\psi_\theta^{-1}(t)$
Ali-Mikhail-Haq[11]	$\log\left[\frac{1-\theta(1-t)}{t}\right]$	$\frac{1-\theta}{\exp(t)-\theta}$
Clayton[12]	$\frac{1}{\theta}(t^{-\theta}-1)$	$(1+\theta t)^{-1/\theta}$
Frank	$-\log\left(\frac{\exp(-\theta t)-1}{\exp(-\theta)-1}\right)$	$-\frac{1}{\theta}\log(1+\exp(-t)(\exp(-\theta)-1))$
Gumbel	$(-\log(t))^\theta$	$\exp\left(-t^{1/\theta}\right)$
Independence	$-\log(t)$	$\exp(-t)$
Joe	$-\log(1-(1-t)^\theta)$	$1-(1-\exp(-t))^{1/\theta}$

Best-fit Copula Selection

Trying to fit all candidate copulas and their parameter range → **Determining optimal parameter θ for all candidate copulas** (using Maximum Likelihoods) → **Selecting best-fit copula among candidate copulas** (using Goodness-of-Fit Tests (Anderson-Darling/ Kolmogorov-Smirnov/ Cramer-von Mises)

3.2. Methodology

Multivariate dependence modelling using Vine copula

Vine Copula

- Allow modelling high-dimensional multivariate dependence structures (for more than 2 variables)
- Enable the flexibility in choosing copulas

D-Vine Copula Quantile Regression (DVQR)

❖ **Fitting ACs to each bivariate copula:**

$$C_{VVVH}(u_{VV}, u_{VH}) = \varphi^{-1}[\varphi(u_{VV}) + \varphi(u_{VH})]$$
$$C_{VHM_v}(u_{VH}, u_{M_v}) = \varphi^{-1}[\varphi(u_{VH}) + \varphi(u_{M_v})]$$
$$C_{VVM_v|VH}(u_{VH}, u_{M_v}) = \varphi^{-1}[\varphi(H_{VV|VH}) + \varphi(H_{M_v|VH})]$$

❖ **Computing conditional copula:**

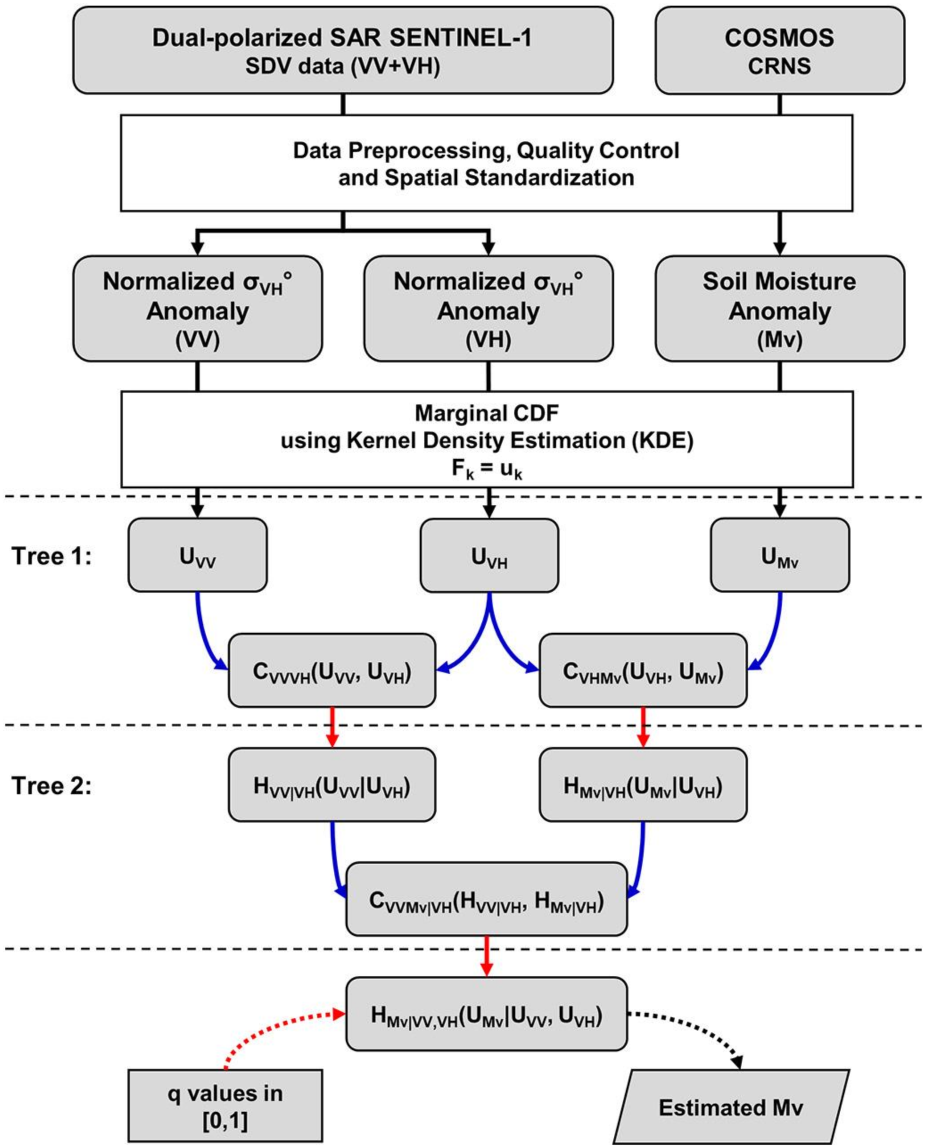
$$H_{VV|VH}(u_{VV}|u_{VH}) = \frac{\partial C_{VVVH}(u_{VV}, u_{VH})}{\partial u_{VH}}$$

$$H_{M_v|VH}(u_{M_v}|u_{VH}) = \frac{\partial C_{VHM_v}(u_{VH}, u_{M_v})}{\partial u_{VH}}$$

$$H_{M_v|VV,VH}(u_{M_v}|u_{VV}, u_{VH}) = \frac{\partial C_{VVM_v|VH}(H_{VV|VH}, H_{M_v|VH})}{\partial H_{VV|VH}}$$

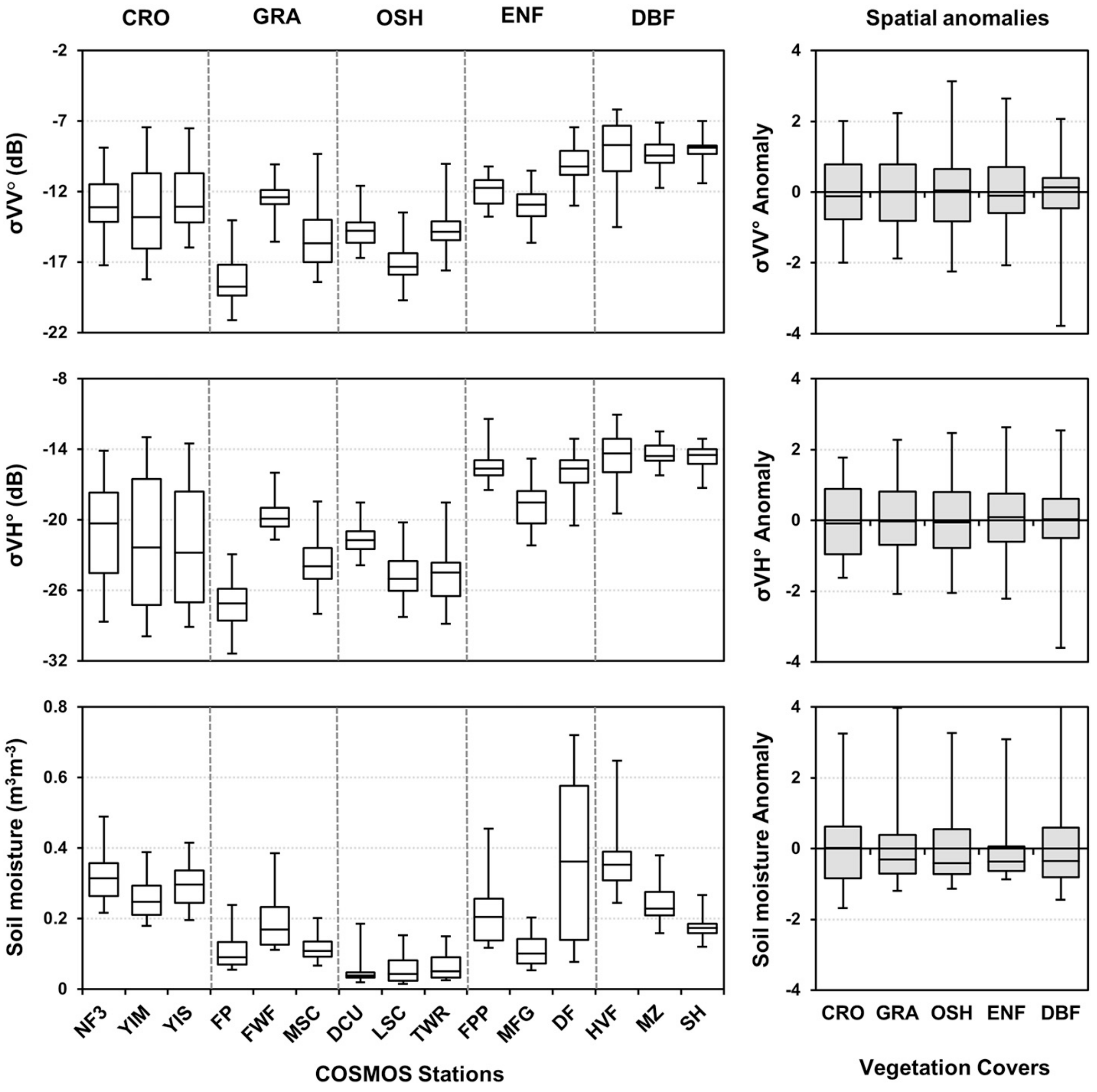
❖ **Backward simulation of soil moisture (Mv):**

$$x_{M_v} = F_{M_v}^{-1}(u_{M_v}) = F_{M_v}^{-1}\left(H_{M_v|VH}^{-1}\left(H_{M_v|VV,VH}^{-1}(q|u_{VV}, u_{VH})\right)\right)$$

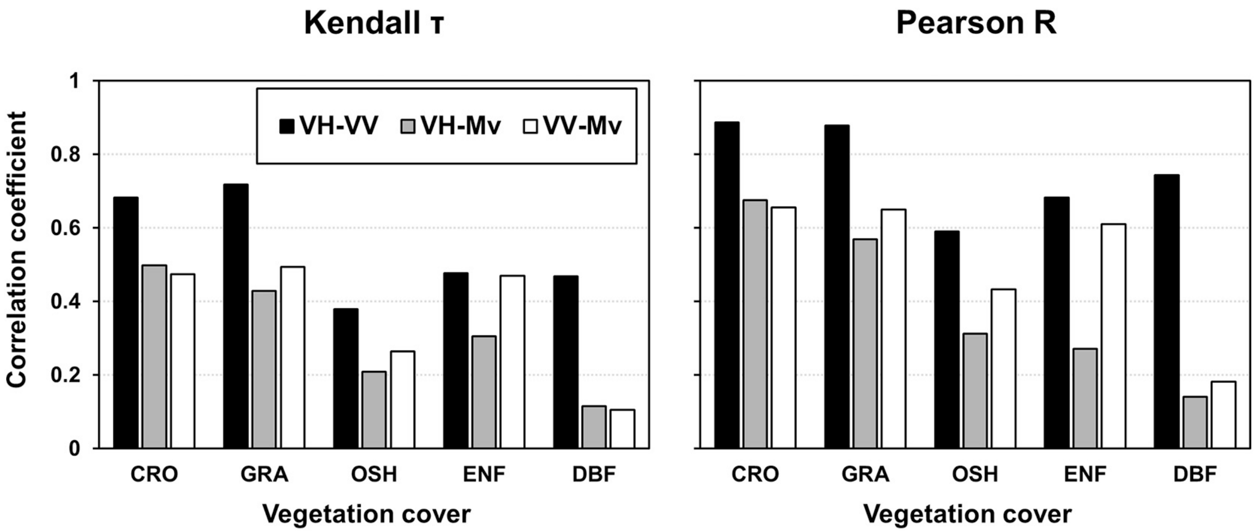


3.3. General results

Overview of variables



Interdependences of variables

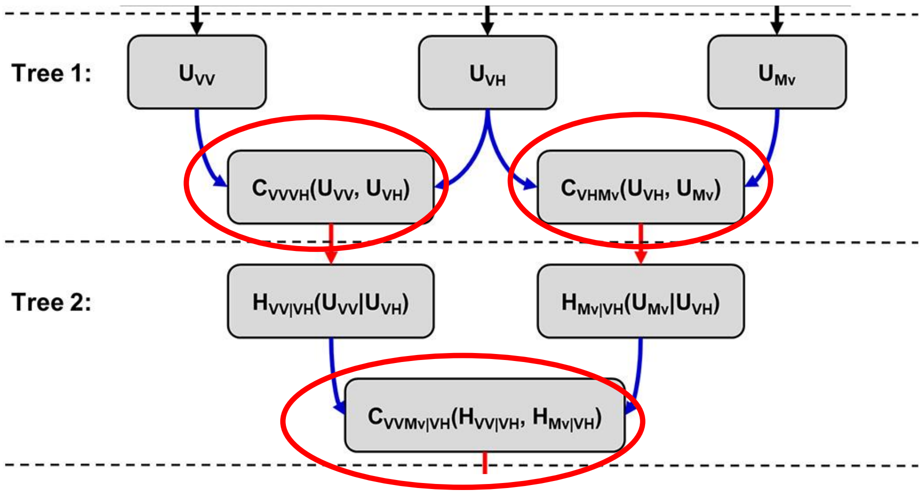


- the highest correlation values of three dependence pairs were generally obtained in the two herbaceous areas (CRO and GRA) compared to the woody regions (OSH, ENF and DBF)
- correlation values decreased with increasing vegetation density

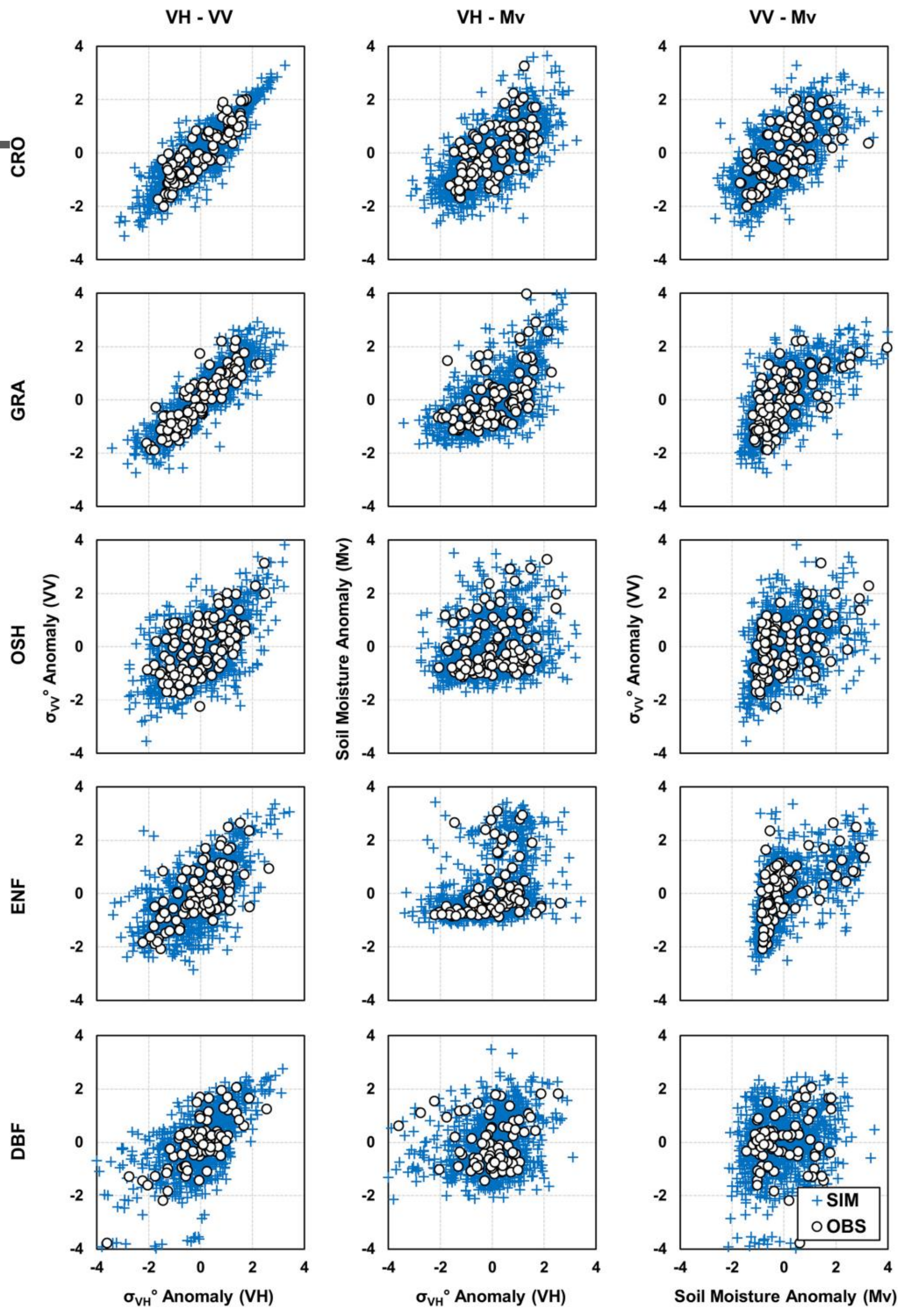
-> **attenuation effects of dense and complex vegetation on radar backscatter signal (C-band SAR signal cannot penetrate the densely forested canopies)**

3.3. General results

Dependence structures modelling



Vegetation covers	D-Vine Copula		
	Type	AC	θ
CRO	C_{VVVH}	G	3.43
	C_{VHMV}	F	6.12
	$C_{VVMv\|VH}$	F	1.00
GRA	C_{VVVH}	F	13.06
	C_{VHMV}	G	1.77
	$C_{VVMv\|VH}$	F	2.76
OSH	C_{VVVH}	G	1.71
	C_{VHMV}	F	2.08
	$C_{VVMv\|VH}$	A	0.93
ENF	C_{VVVH}	G	1.89
	C_{VHMV}	F	3.18
	$C_{VVMv\|VH}$	G	1.60
DBF	C_{VVVH}	G	1.87
	C_{VHMV}	J	1.37
	$C_{VVMv\|VH}$	C	0.22

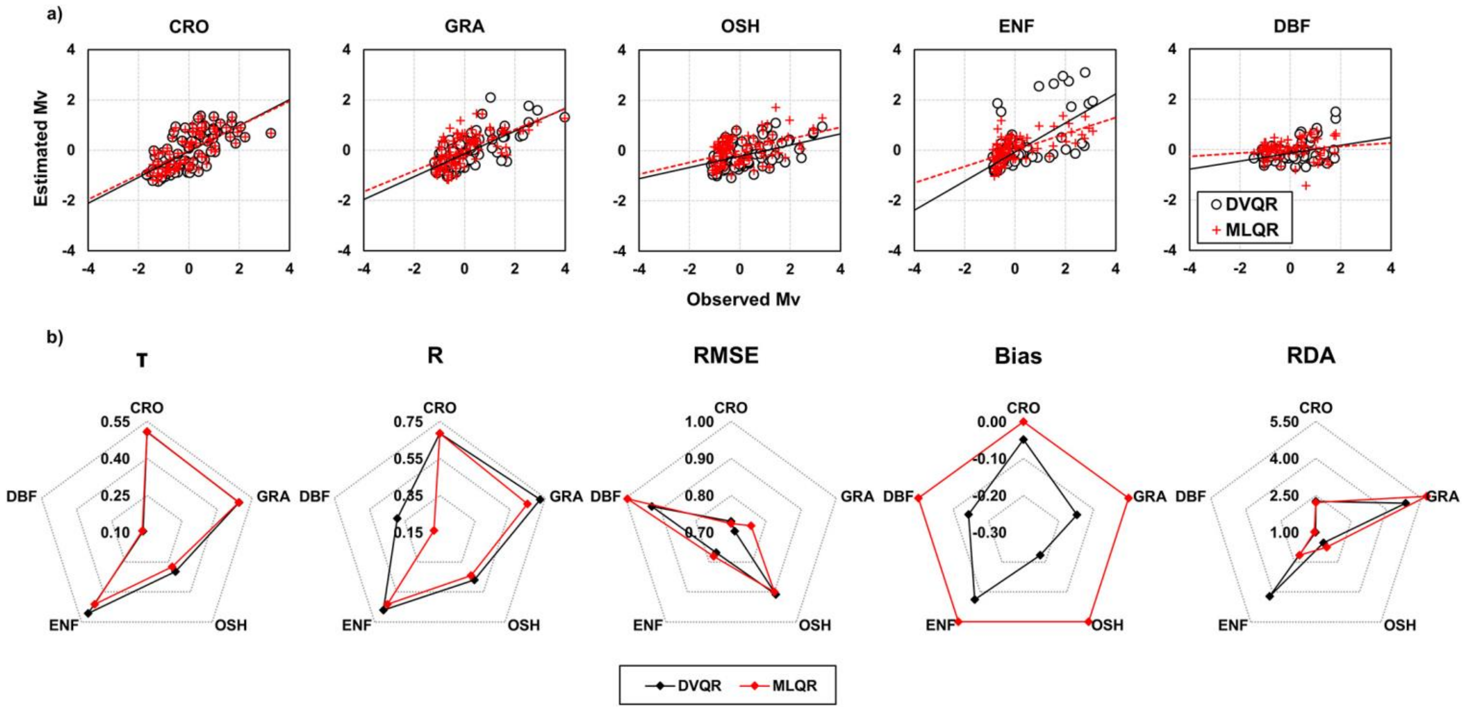


- Interdependence structures among variables can be well captured and simulated by the D-Vine Copula
- The interdependence structures over most of the vegetation covers are nonlinear and asymmetric

3.3. General results

SAR Sentinel-1 SSM retrieval over different vegetation covers

Intercomparison of D-Vine Copula Quantile Regression (DVQR) and Multi-Linear Quantile Regression (MLQR)



- **Relative average deviation amplitude:**

$$RDA = \frac{1}{N}\sum_{i=1}^{N}\left|\left(\frac{1}{2}\left(q_i^u + q_i^l\right) - Mv_i\right)\Big/ Mv_i\right|$$

where N is the sample size, q_i^u and q_i^l are upper (q = 95%) and lower (q = 5%) prediction bounds at the i^{th} observation, respectively; Mv_i is the i^{th} observed CRSM anomaly.

- Higher τ and R, lower RMSE, Bias and RDA can be regarded as a superior product

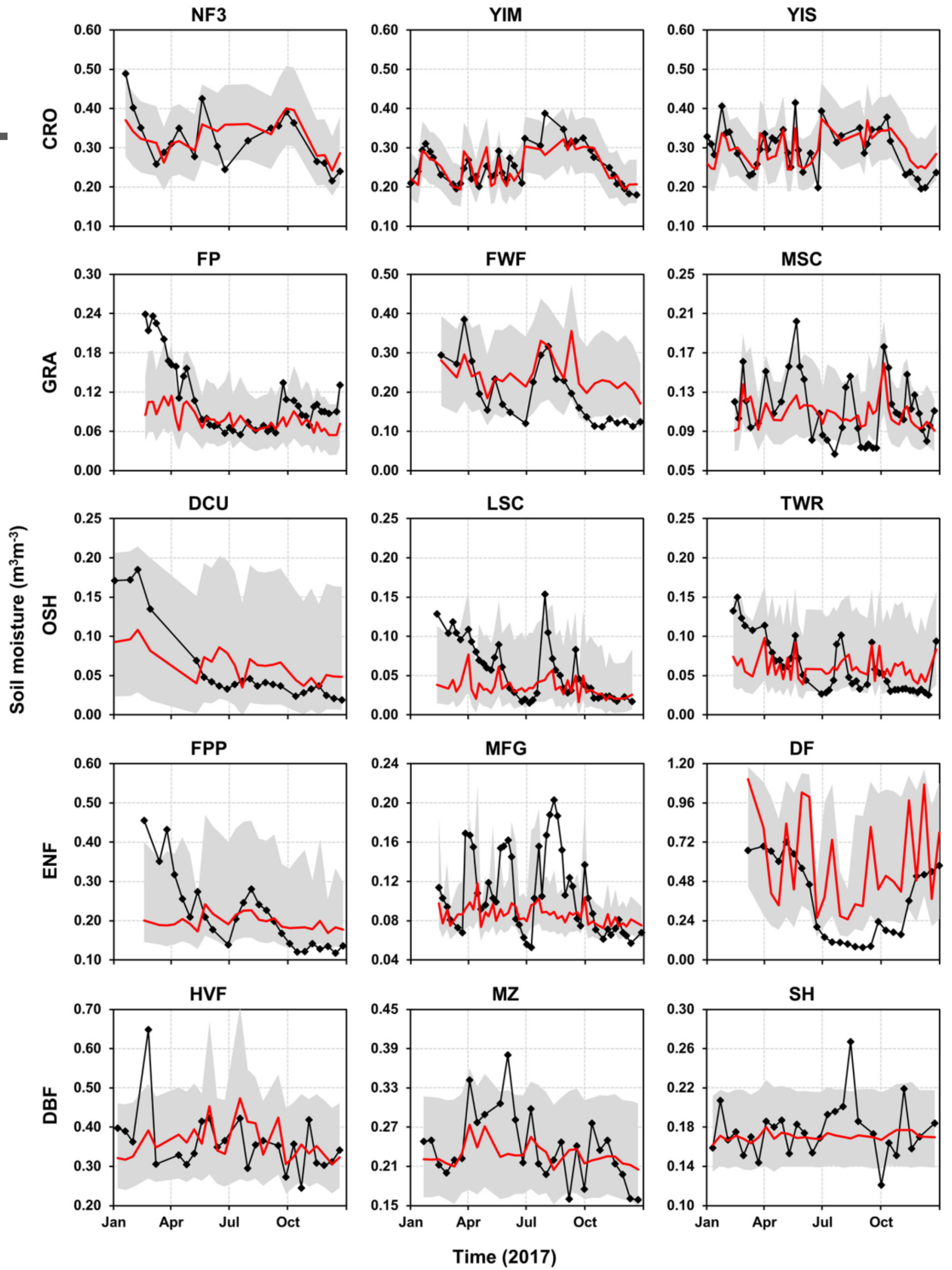
- The **superior performances obtained with the DVQR** compared with the MLQR considering all evaluation metrics in most VCs -> **robustness of the DVQR for capturing highly nonlinear dependence structures among variables**
- Over VCs, **superior performance was generally obtained at low-canopy herbaceous regions**, especially grasslands and croplands, according to the high correlations between each pairs of variables.

3.3. General results

SAR Sentinel-1 SSM retrieval at each COSMOS site

Vegetation covers	Sites	Deterministic evaluation				Probabilistic evaluation
		τ	R	RMSE (m^3m^{-3})	Bias (m^3m^{-3})	RDA
CRO	NF3	0.56	0.69	0.050	0.004	0.133
	YIM	0.66	0.82	0.028	-0.005	0.084
	YIS	0.53	0.69	0.041	-0.003	0.124
GRA	FP	0.36	0.66	0.054	-0.031	0.272
	FWF	0.57	0.68	0.074	0.048	0.488
	MSC	0.22	0.50	0.028	-0.006	0.227
OSH	DCU	0.46	0.73	0.039	0.004	1.447
	LSC	0.41	0.52	0.038	-0.021	0.702
	TWR	0.32	0.44	0.030	-0.002	0.773
ENF	FPP	0.31	0.22	0.092	-0.021	0.495
	MFG	0.33	0.40	0.043	-0.021	0.270
	DF	0.34	0.46	0.354	0.227	1.803
DBF	HVF	0.10	0.27	0.076	0.001	0.158
	MZ	0.42	0.55	0.047	-0.011	0.168
	SH	0.07	0.18	0.027	-0.006	0.109

Reference: Nguyen, H. H., Cho, S., Jeong, J., & Choi, M. (2021). A D-vine copula quantile regression approach for soil moisture retrieval from dual polarimetric SAR Sentinel-1 over vegetated terrains. *Remote Sensing of Environment*, *255*, 112283.



I. INTEGRATION OF NOVEL FIELD-SCALE SOIL MOISTURE OBSERVATION SYSTEMS

4. Conclusions

4. Conclusions

- Calibration using wettest conditions can generate reasonable cosmic-ray soil moisture products.
- The merging framework combined CRNP and in-situ soil moisture measurements by providing the weight considering the CRNP penetration depth. When the CRNP penetration depth is shallower than RZSM depth, less weight is given to CRNP and more weight is provided to representative point measurement.
- The merging framework outperformed 2 original products and exponential filter, indicating that independent uses of both CRNP and in-situ sensor cannot fully represent field-scale RZSM.
- In root-zone layer, vegetation type and root water uptake is the major factors controlling the temporal variation of merged RZSM, which can be partly revealed through the merging framework procedure
- Radar backscatter is more sensitive to soil moisture over low-canopy herbaceous areas rather than dense canopy woodland areas, leading to superior performances of SAR-based soil moisture retrieval at low-canopy regions

- Vegetation is the major factor controlling spatio-temporal variation of soil moisture -> needs to be considered carefully before integrating different soil moisture observation systems

II. POTENTIAL FOR AUTOMATED HYDROMETEOROLOGICAL MONITORING USING AI

1. The needs for data assimilation and Artificial Intelligence

1.1. The need for data assimilation

Pros and Cons of precipitation observation systems

Rain gauges

❖ **Advantages:**
- High temporal resolution (1 min.)
- Most accurate observation

❖ **Disadvantages:**
- Point-scale measurement (low spatial coverage)
- Errors due to external sources (winds, high rainfall intensity…)

Weather radars

❖ **Advantages:**
- High temporal resolution (5-15 min.)
- Good spatial coverage

❖ **Disadvantages:**
- High errors
- Require rain gauges data for calibration and correction

Satellite remote sensing

❖ **Advantages:**
- Large scale observation (high spatial coverage)

❖ **Disadvantages:**
- Lower temporal resolution (30 min.-1 day)
- Coarse spatial resolution
- Require rain gauge network and radar data for downscaling and correction

1.2. The need for Artificial Intelligence

Automated monitoring system for precipitation data

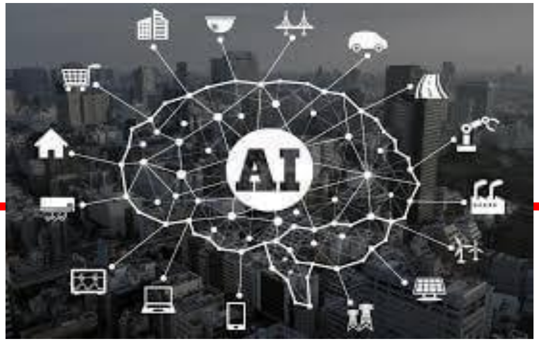
Precipitation data from different observation systems

Automated Data Assimilation

Data updated and stored in a Database

Applications

- Rain gauges
- Weather radars
- Satellites



- Embedded Devices & Hardwares
- Decision-making
- Forecasting
- Natural Disasters Monitoring

Automation

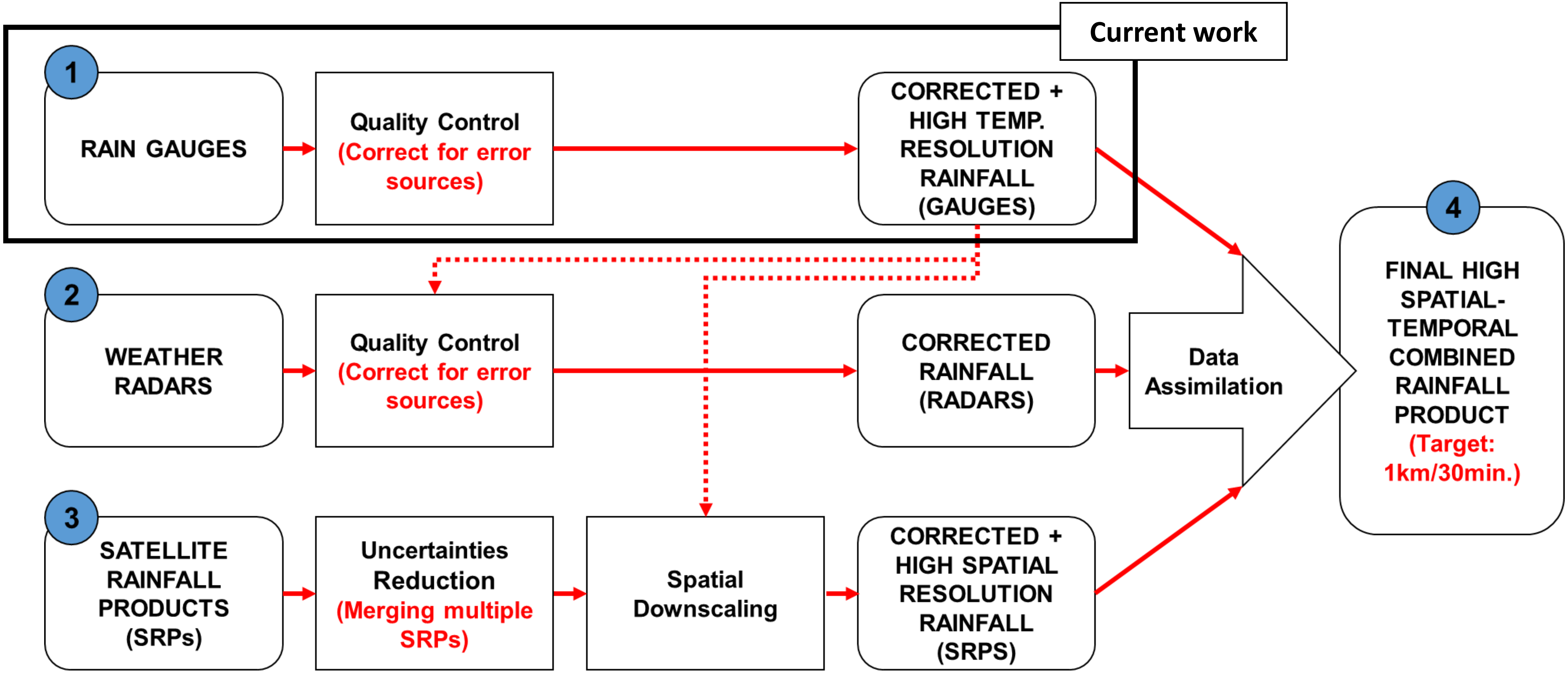
Real-time/ Near Real-time Processing

II. POTENTIAL FOR AUTOMATED HYDROMETEOROLOGICAL MONITORING USING AI

2. What we are working on

2.1. Roadmap

Roadmap for data assimilation of precipitation observation systems



2.2. Current work

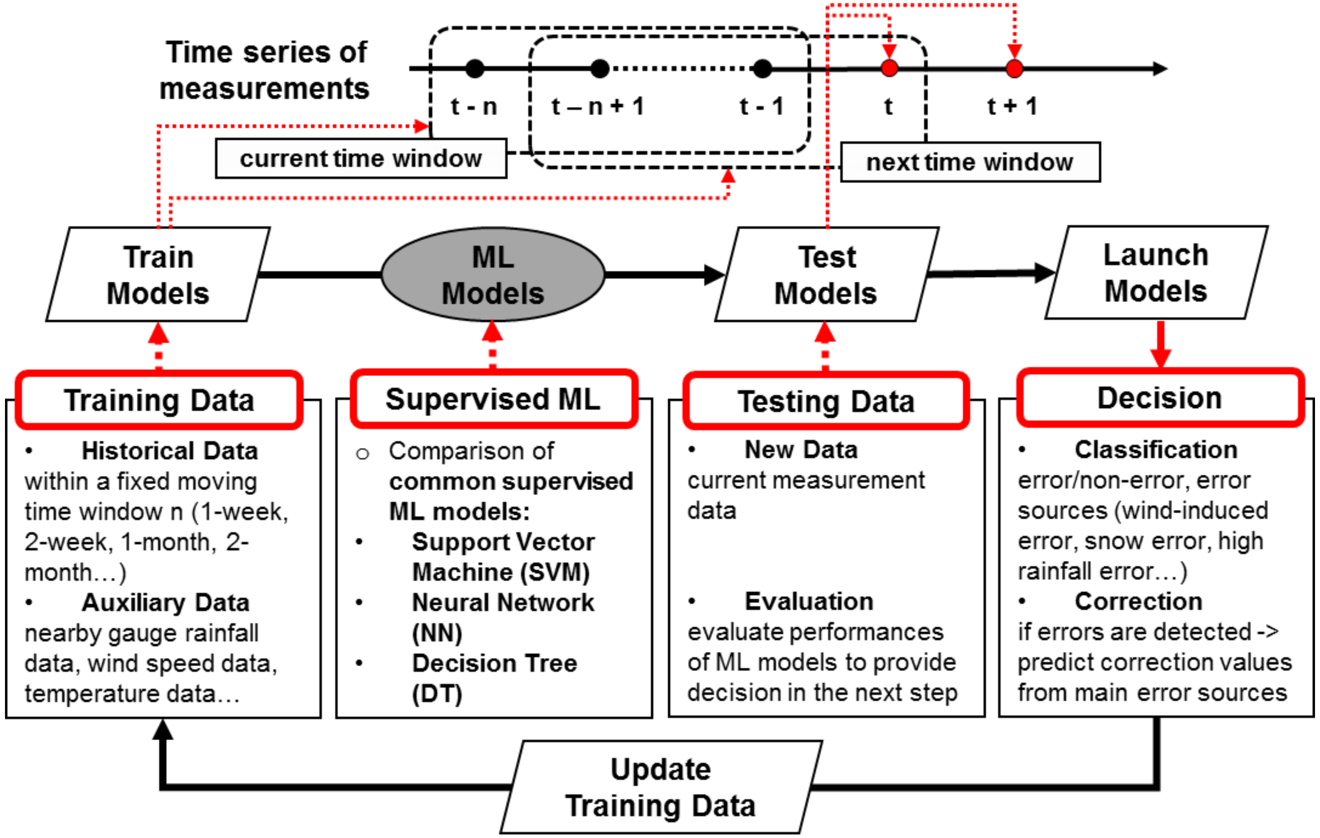
Development of a smart rain gauge

Smart Rain Gauge

Real-time Automated Quality Control

Automated Classification of error sources (malfunction error, wind-induced error, high rainfall intensity error, snow error…) and **Correction Errors** -> require **supervised Machine Learning (ML) models for both Classification and Regression**

Real-time Processing -> require **reasonable accuracy** and **easily adaptive to changes with low computational cost (fast processing)**



- The datasets within a fixed moving time window n (t-n, …, t-1) can be used as training datasets for decision at measurement time t
- After measurement at time t was corrected, the corrected value can be continuously used as training datasets (update training data) for decision at measurement time t+1
- This procedure will be iteratively processed whenever a new measurement comes.

THANK YOU FOR YOUR ATTENTION!