The benefits of using a reference sampling for mitigating the impact of legacy soil data errors on Digital Soil Mapping outputs.

Philippe Lagacherie, Maider Arregui, David Fages

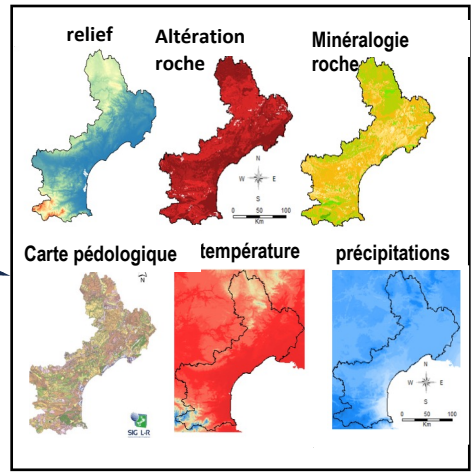
The benefits of using a reference sampling for mitigating the impact of legacy soil data errors on Digital Soil Mapping outputs.

Philippe Lagacherie[1], Maider Arregui[2] and David Fages[1]

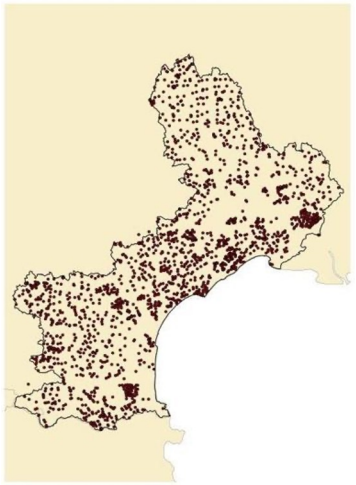
1. *LISAH, INRAE, Montpellier, France*
2. *BRL Exploitation, Nimes, France*

Source of errors in Digital Soil Mapping

Lack of/ imprecision of covariates

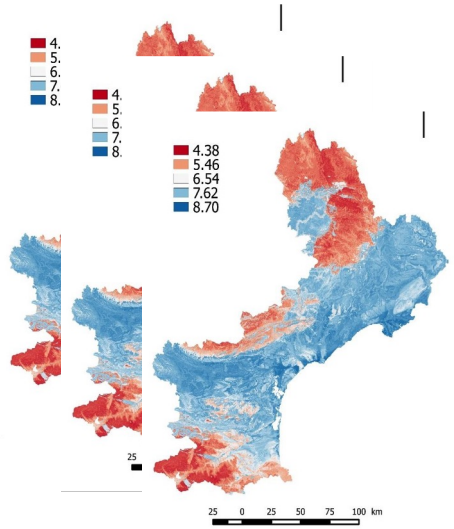


Scarce soil observations



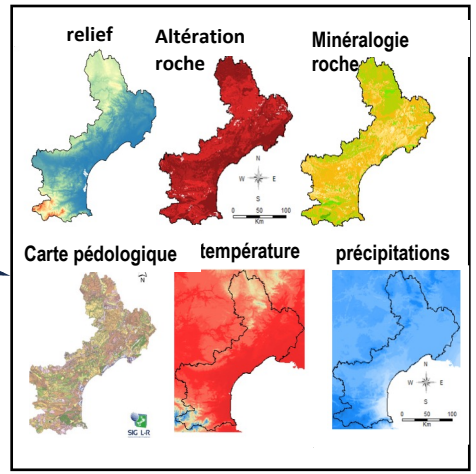
Sub-optimal algorithm

Training /interpolation

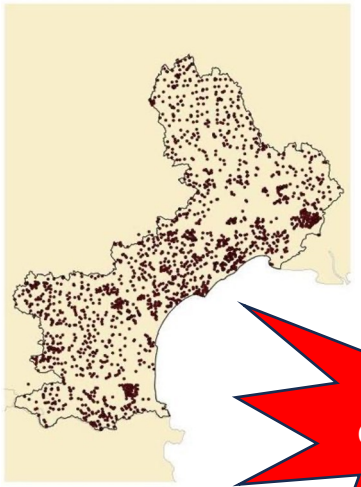


Source of errors in Digital Soil Mapping

Lack of/ imprecision of covariates



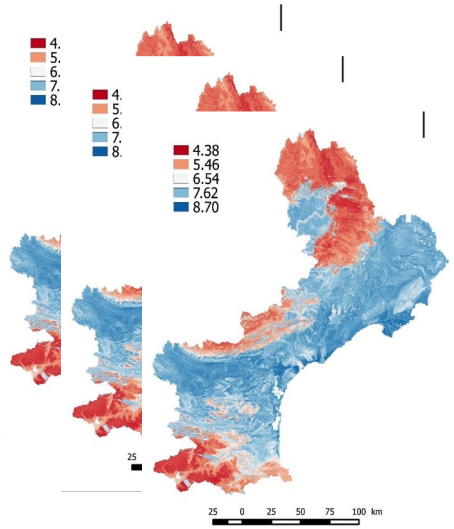
Scarce soil observations



Soil observations Errors

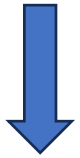
Sub-optimal algorithm

Training /interpolation



Sources of errors on legacy soil information

- Analysis protocols
- Geo-referencing (mostly in the pre-GPS era)
- Sample processing before soil analysis (sieving, grinding, decarbonatation, …)
- Soil database entry errors
- Age of observations (for time-variant soil properties)



Lagacherie, P., Arregui, M., Fages D., Evaluating the quality of soil legacy data used as input of Digital Soil Mapping models. Accepted in EJSS

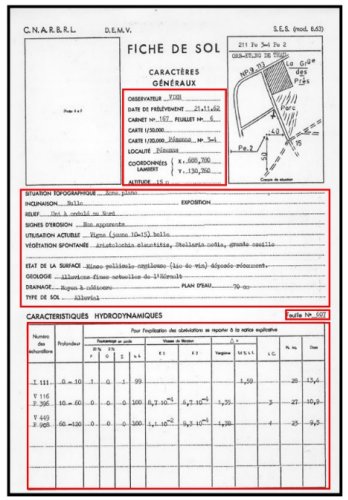
Objectives of the day

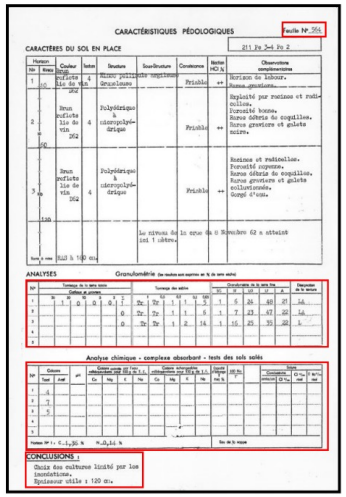
- To evaluate the impact of the legacy soil observation errors on DSM prediction performances

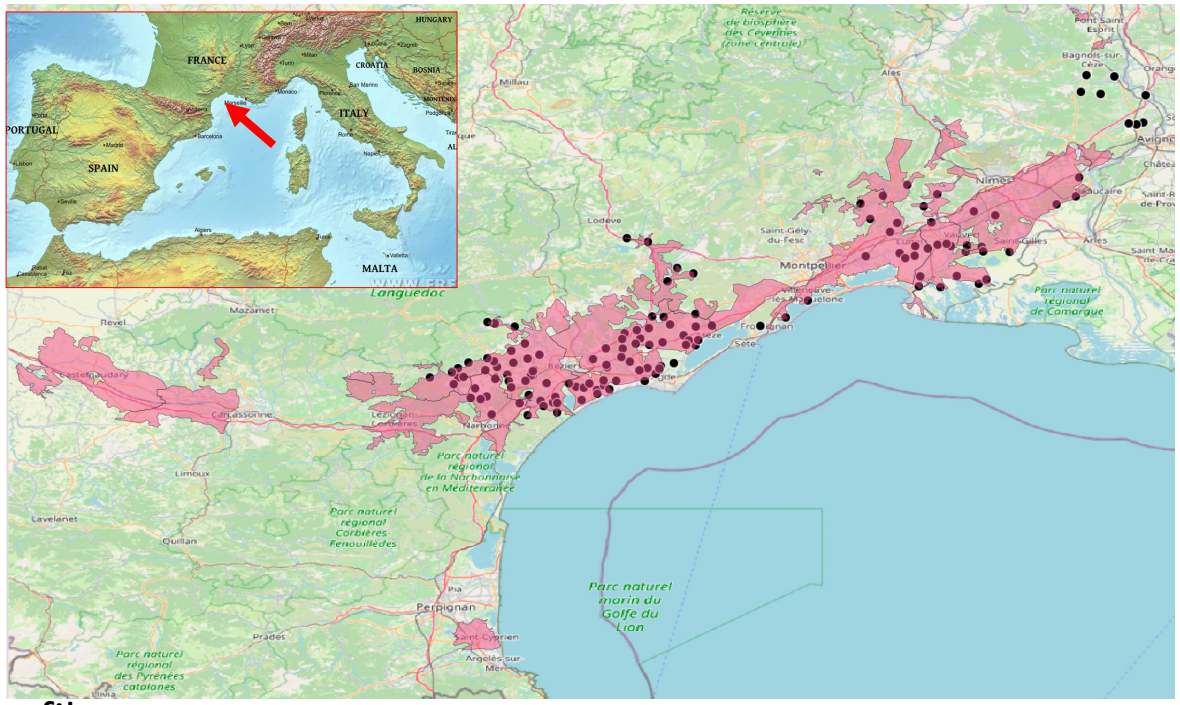
- To evaluate the benefits on DSM predictions of prior corrections of input soil data errors by using a control sampling.

The data

- Study area : BRL irrigation perimeter (6 636 km2 in the Coastal plain of Occitanie, southern France)

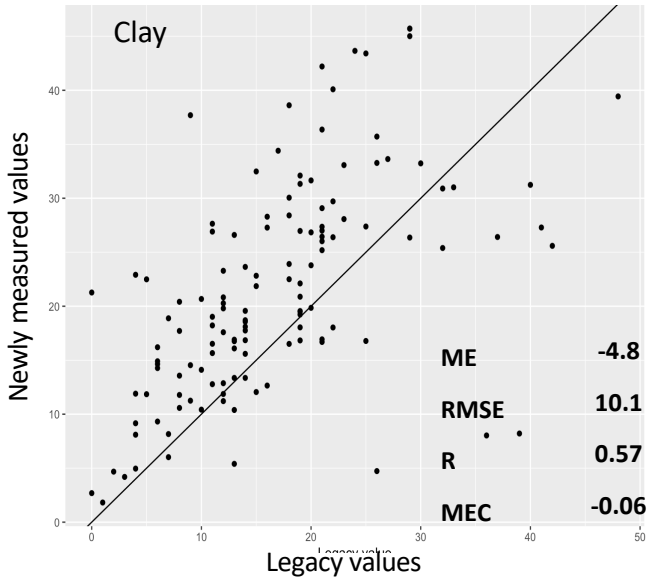


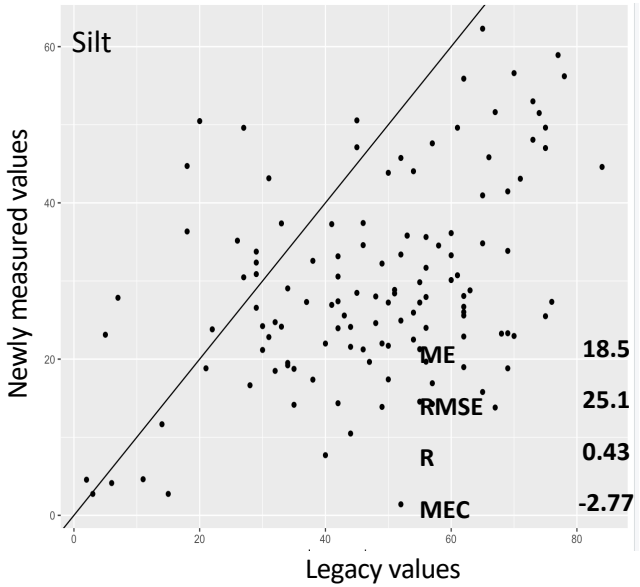


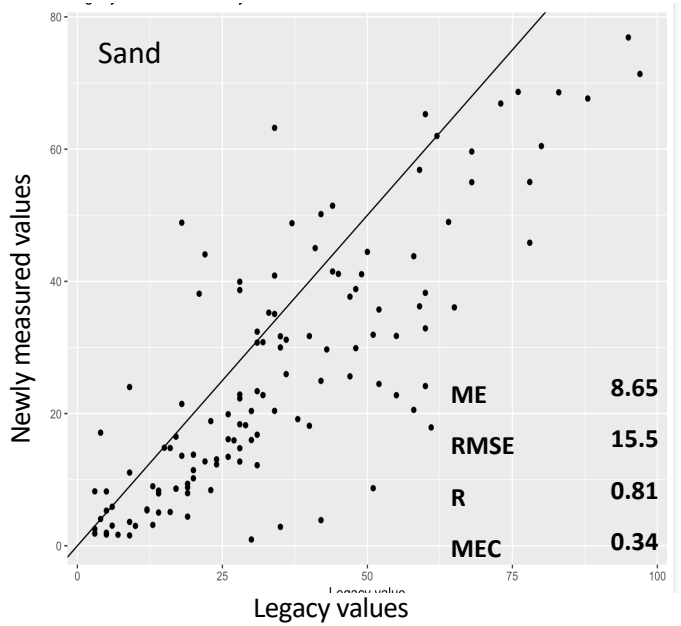


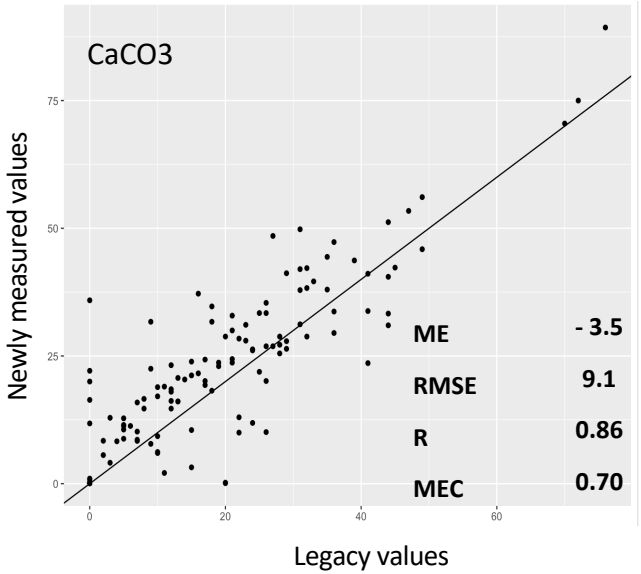
- 6872 legacy measured soil profiles (1955- 1992) digitized by automatic text recognition *(ChemChem et al, submitted)*
- Control sampling at 129 locations with a registered legacy soil profile
- 6 topsoil properties analysed by a certified soil laboratory (AUREA): Clay%, Silt%, Sand%, CaCO3%, SOC% and CEC (meq/100g)

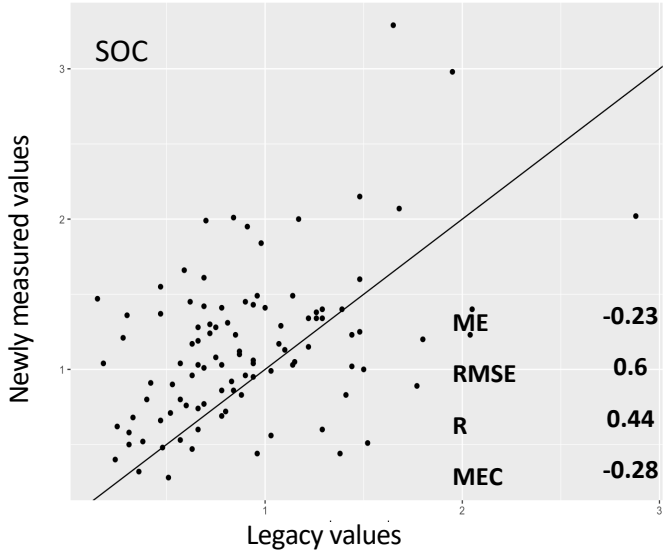
Soil Legacy data errors

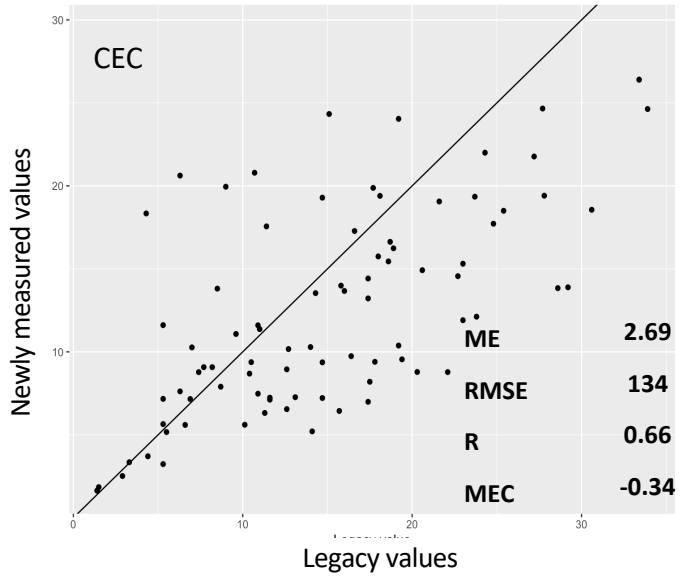








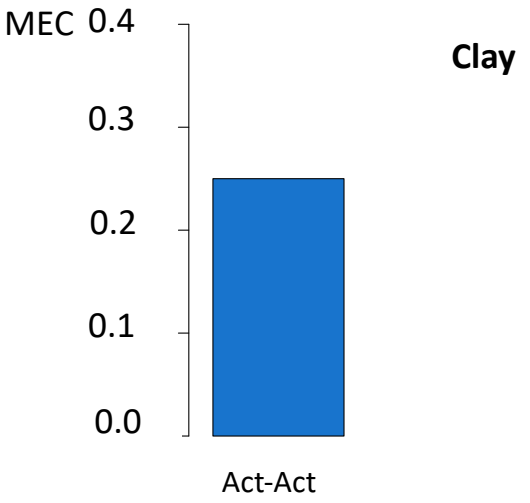


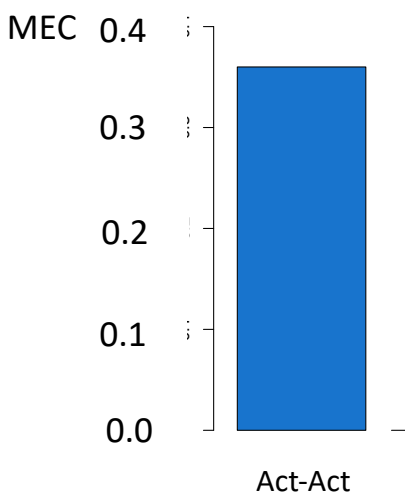


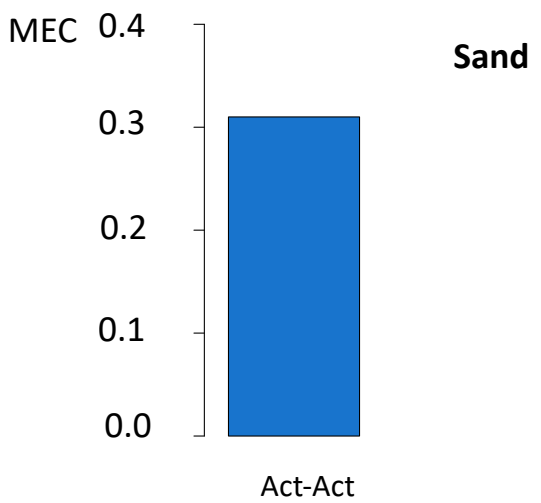
The DSM Experiment

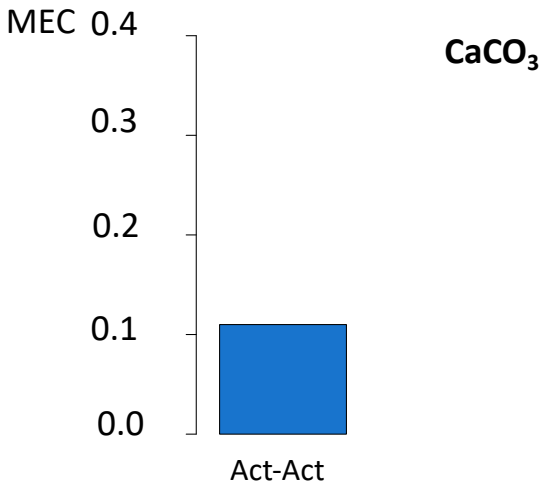
- Two alternate input datasets describing the same 129 locations: actual soil analyses (Act) and Legacy soil analyses (Leg)
- A classical DSM approach
  - Training algorithm : Random Forest
  - Soil covariates : geological maps, 1:250 000 pedological map, DEM and its derivative and a set of Remote sensing products
  - Evaluation method : 10 fold Cross Validations repeated 100 times
- 3 DSM scenario of input/evaluation data

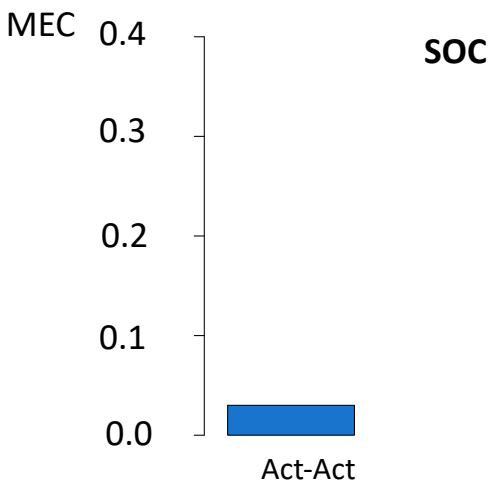
Reference scenario using actual analysis (Act-Act) for training and evaluating RF

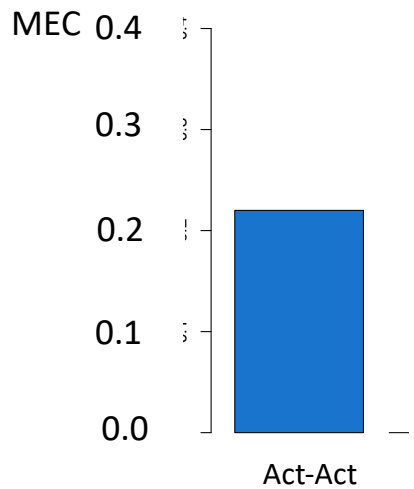






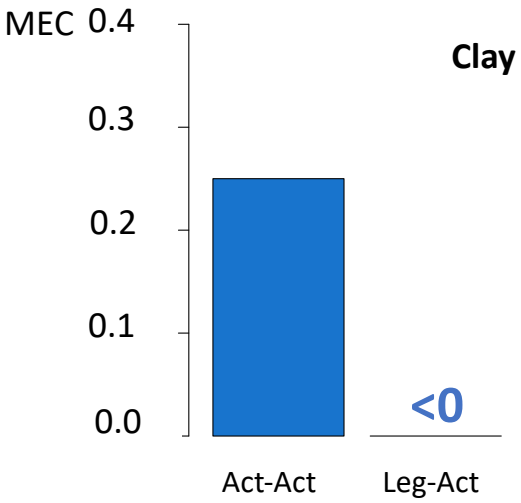


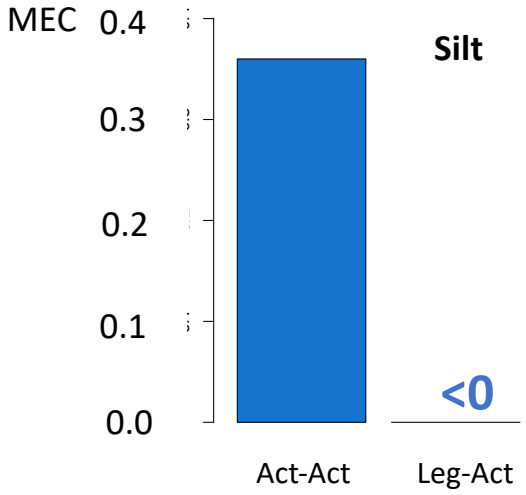


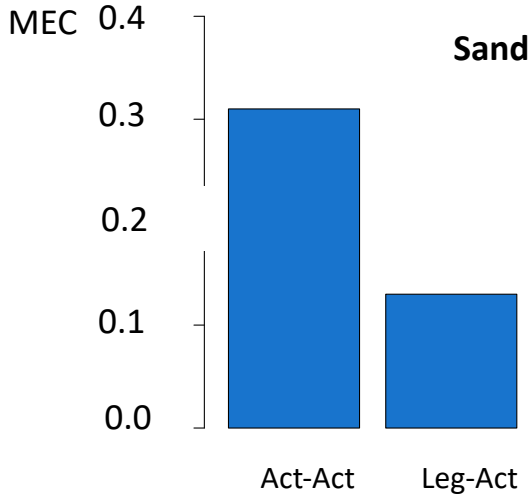


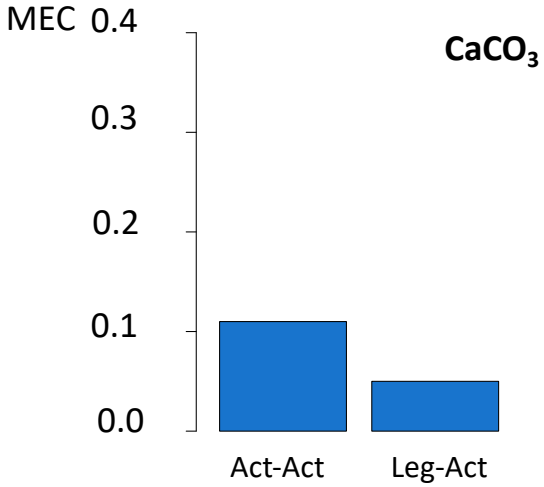
*** MEC = 1 – MSE/Variance**

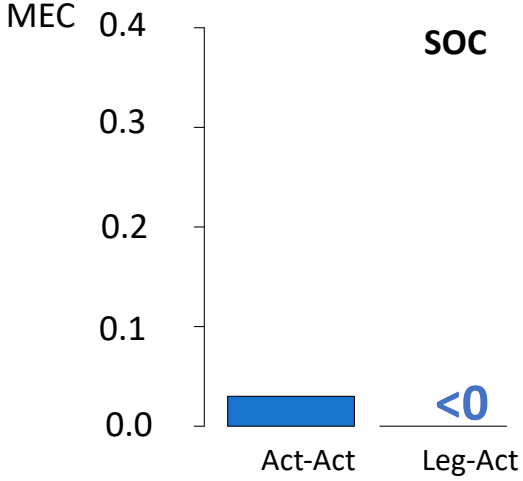
RF trained with legacy value and evaluated with actual analyses (Leg-Act)

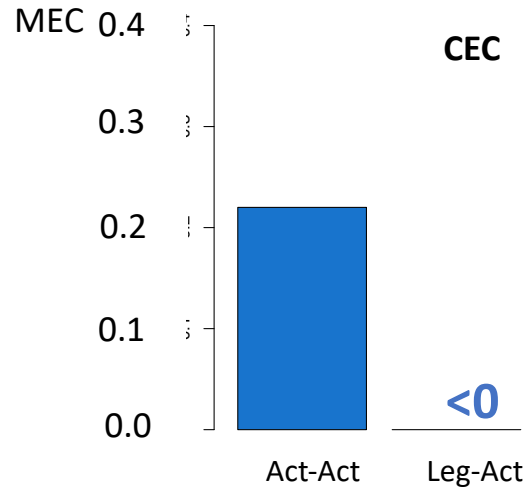










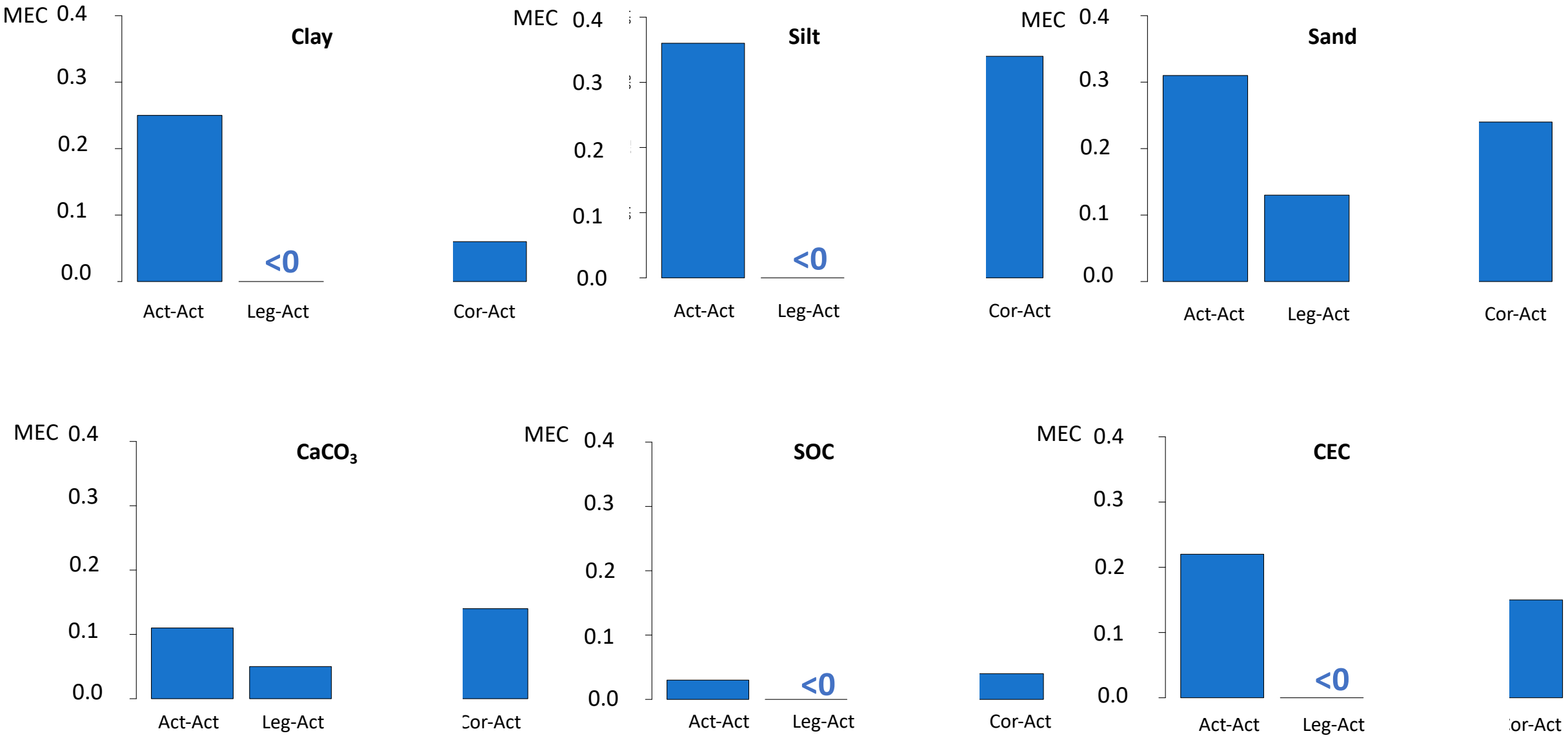


*** MEC = 1 – MSE/Variance**

Bias corrections ($Z_{cor} = aZ_{leg} + bCaCO3_{leg} + c$)

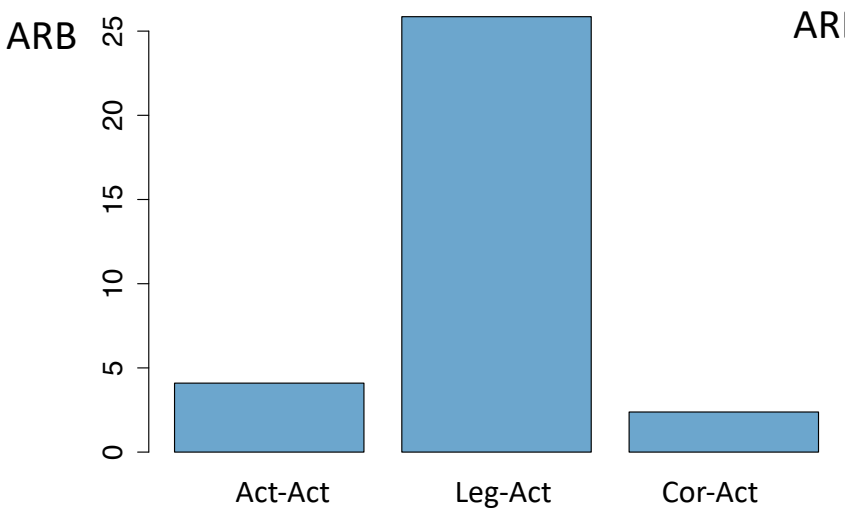
Soil property	c	a	b	R^2
Clay	11.3646	0.6064	-	0.32
Silt	17.5961	0.4246	- 0.4723	0.49
Sand	1.2228	0.7133	-	0.65
CaCO3	5.2143	0.9089	-	0.75
SOC	0.6451	0.5449	-	0.20
CEC	4.577	0.522	-	0.44

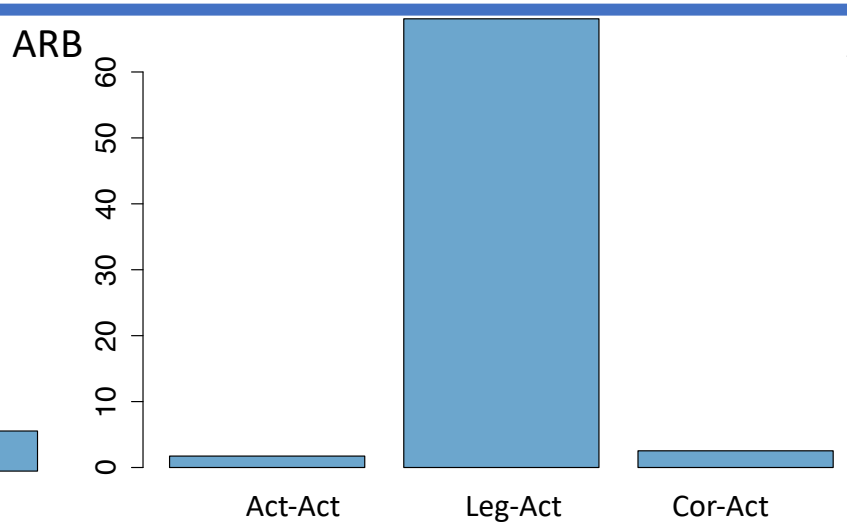
RF trained with corrected legacy values (Cor-Act)

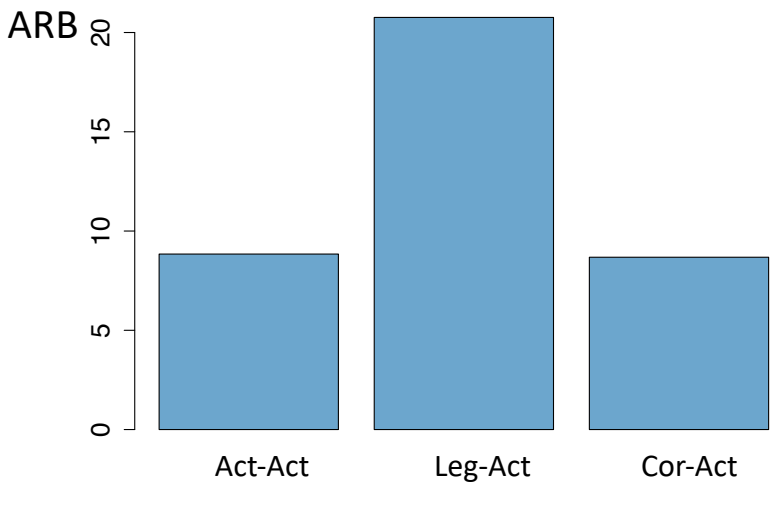


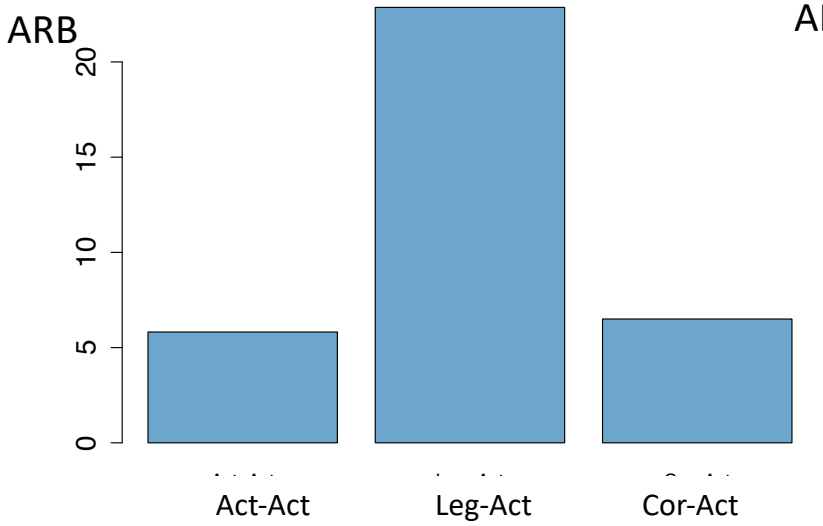
*** MEC = 1 – MSE/Variance**

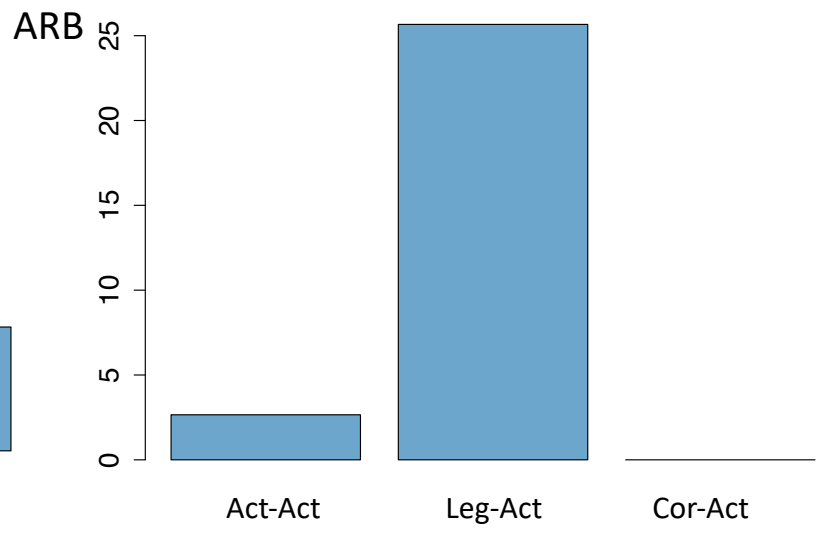
Bias removals with corrections

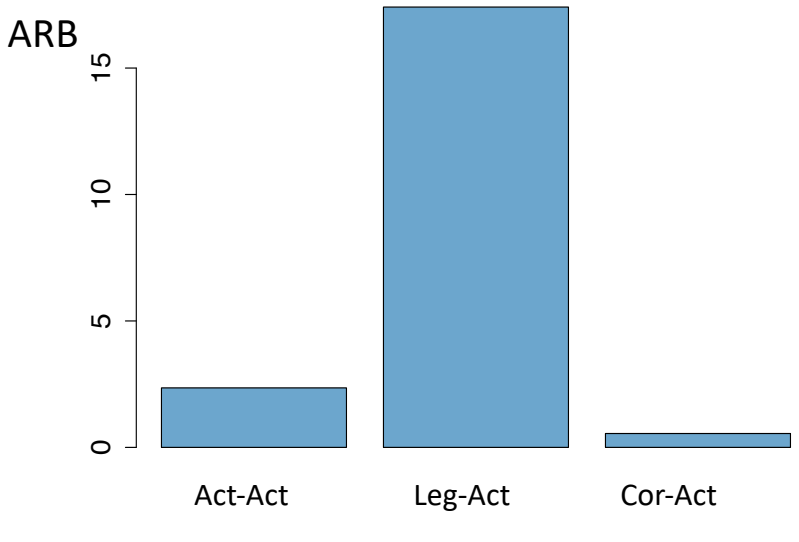












ARB : Absolute relative biases

Conclusions

- The errors affecting legacy soil measurements of soil properties can be substantial, with possible large biases
- These errors can severely affect the performance of the DSM models trained on soil legacy data
- A control sampling using recent soil analyses performed at legacy soil profile locations can partly mitigate these effects
- Low-cost and optimized control sampling approaches should be investigated in the future