



S_{oil} INFER_{ence} S_{ystem}

LIVE DEMO

José Padarian, Edward Jones, Alex McBratney

Why SINFERS?

A software engine for the systematic
prediction of...

“methodical in procedure or plan”

“marked by thoroughness and regularity”

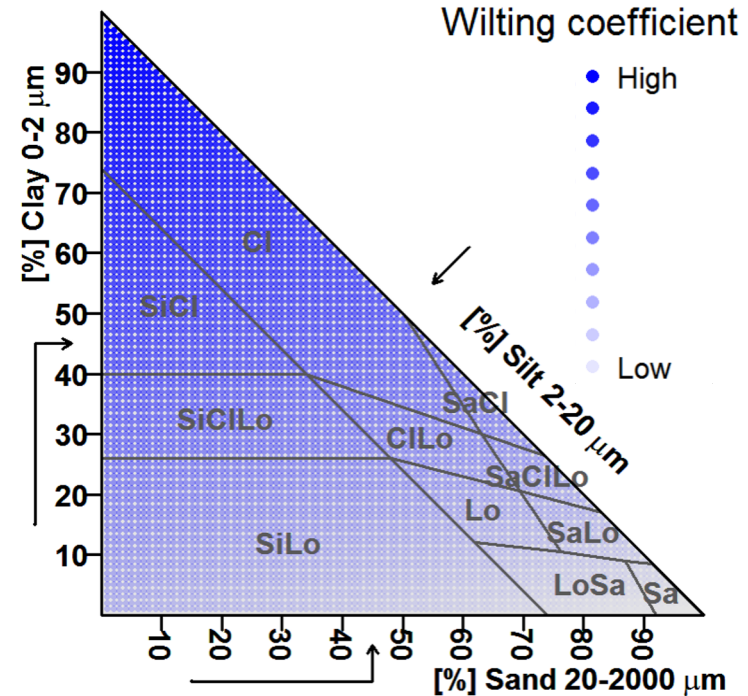


Pedotransfer Functions (PTFs)

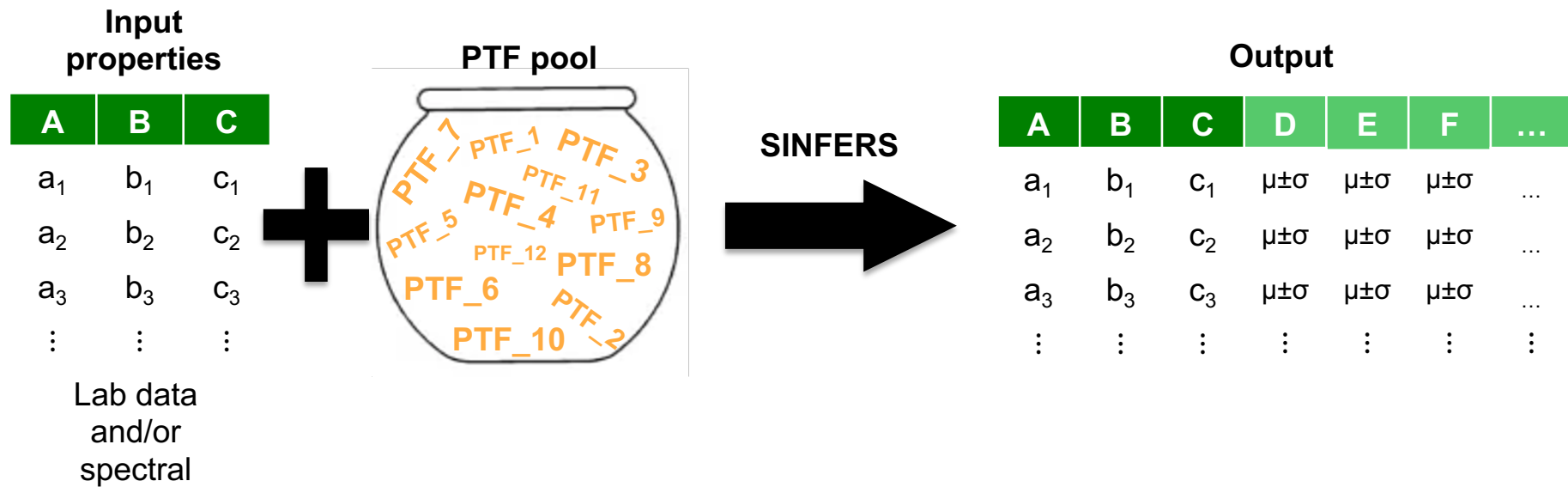
- Functions to predict soil properties from other soil properties
- Regression equations, machine learning models, spectral models



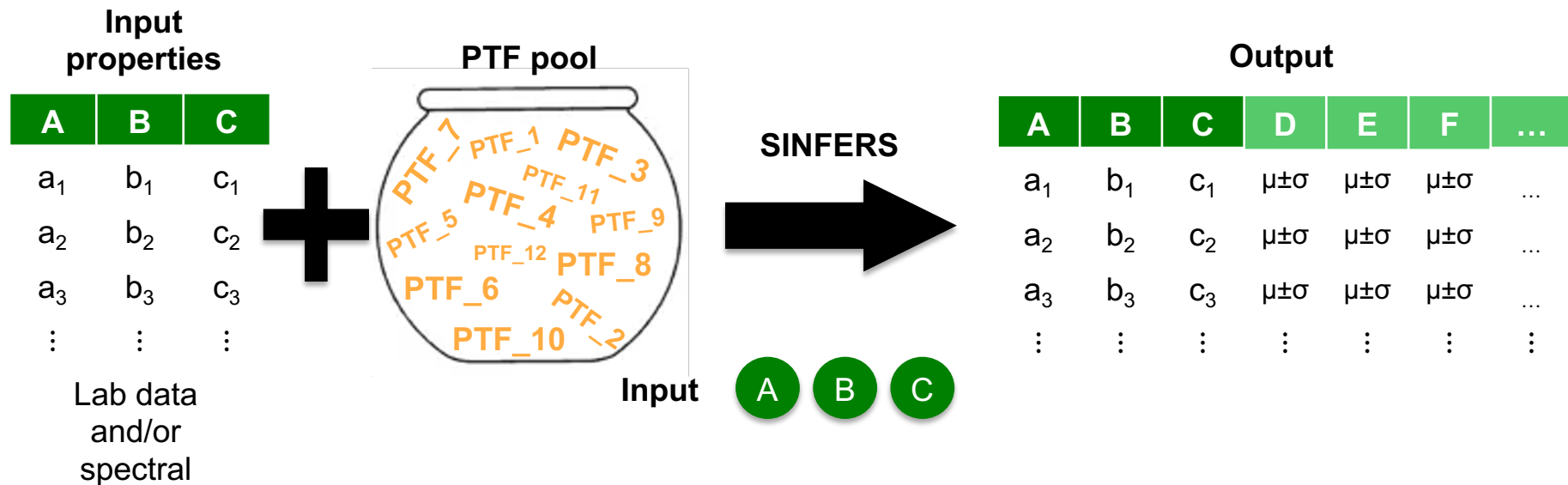
Wilting coefficient = $0.01 \text{ sand} + 0.12 \text{ silt} + 0.57 \text{ clay}$
Briggs and McLane (1907)



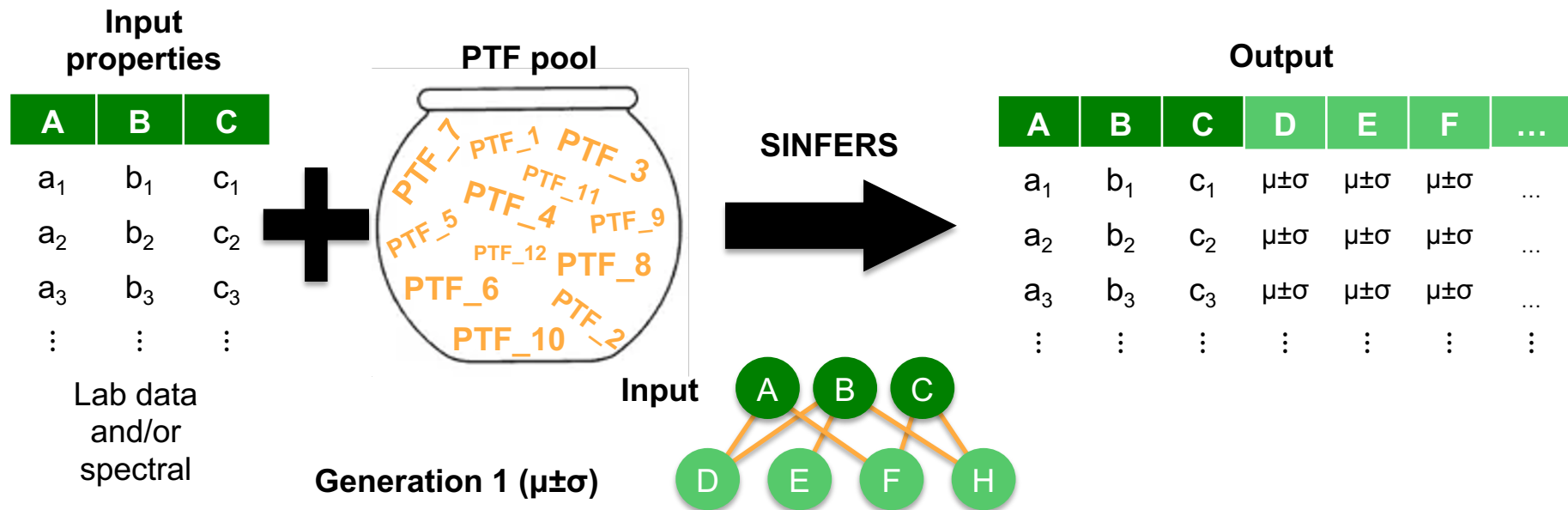
How SINFERS works



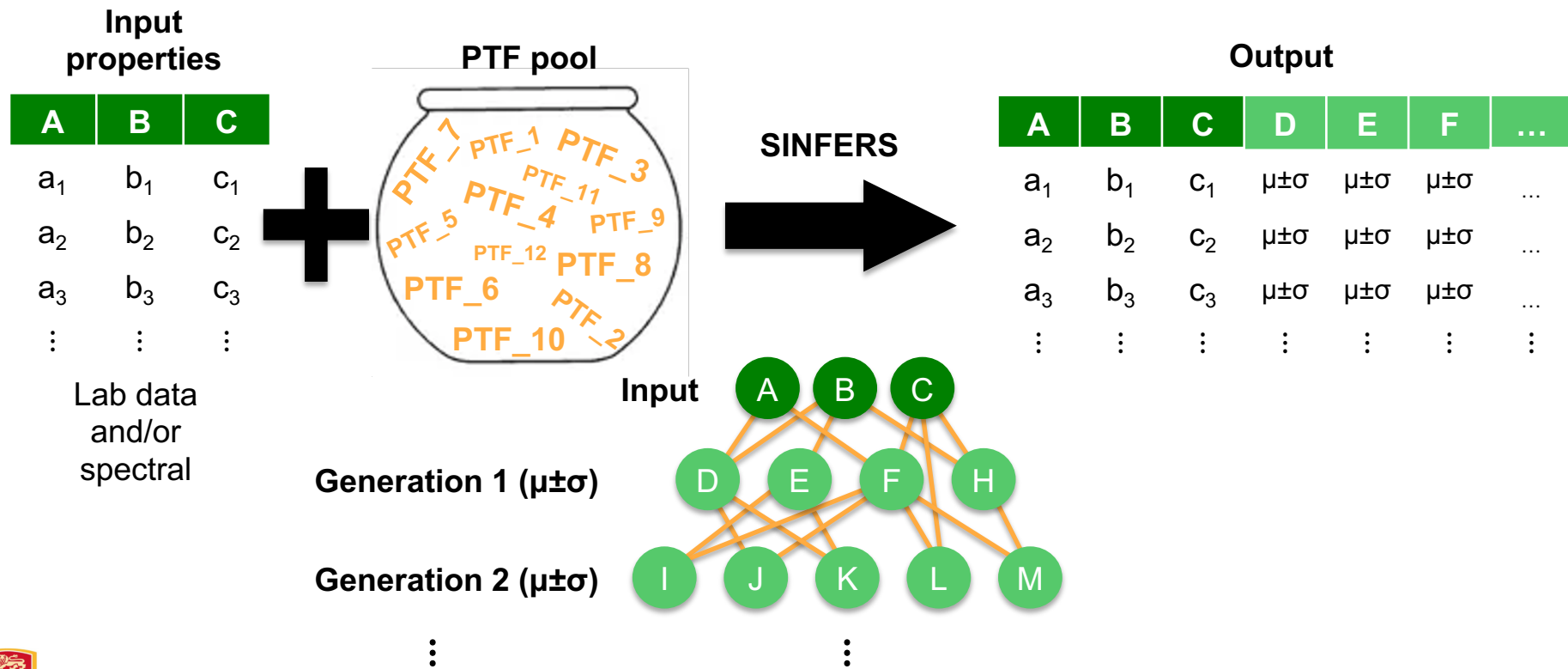
How SINFERS works



How SINFERS works



How SINFERS works



Why SINFERS?

- Knowledge organisation
 - Description, indexing, classification
- Inference
 - Prediction (trivial)



Why SINFERS?

- Knowledge organisation
 - Description, indexing, classification
- Inference
 - Prediction (trivial)

$$0.364 + 4.83 \times 10^{-5} * \text{sand} * \text{clay} - 0.003 * \text{sand}$$

$$\text{sand} = 20$$

$$\text{clay} = 50$$



Why SINFERS?

- Knowledge organisation
 - Description, indexing, classification

- Inference

$$0.364 + 4.83 \times 10^{-5} * \text{sand} * \text{clay} - 0.003 * \text{sand}$$

$$\text{sand} = 20 \pm 2.5$$

$$\text{clay} = 50 \pm 5.5$$

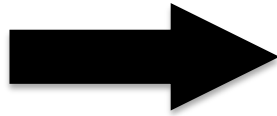
- Prediction (trivial)
- Uncertainty propagation (non-trivial)



SINFERS use cases

- Small amount of soil data → large amount of soil data

A	B	C
a_1	b_1	c_1
a_2	b_2	c_2
a_3	b_3	c_3
$\vdots$	$\vdots$	$\vdots$



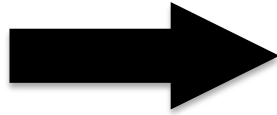
A	B	C	D	E	F	...
a_1	b_1	c_1	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	...
a_2	b_2	c_2	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	...
a_3	b_3	c_3	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$



SINFERS use cases

- Small amount of soil data → large amount of soil data
- Backfill missing data points (data imputation)

A	B	C
a_1	b_1	NA
NA	b_2	c_2
a_3	NA	c_3
$\vdots$	$\vdots$	$\vdots$



A	B	C	D	E	F	...
a_1	b_1	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	...
$\mu \pm \sigma$	b_2	c_2	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	...
a_3	$\mu \pm \sigma$	c_3	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$



SINFERS use cases

- Small amount of soil data → large amount of soil data
- Backfill missing data points (data imputation)
- Estimate properties that are difficult, expensive, or impossible to obtain



SINFERS use cases

- Small amount of soil data → large amount of soil data
- Backfill missing data points (data imputation)
- Estimate properties that are difficult, expensive, or impossible to obtain
- Estimate complex and obscure properties, e.g. integral energy of the SMC



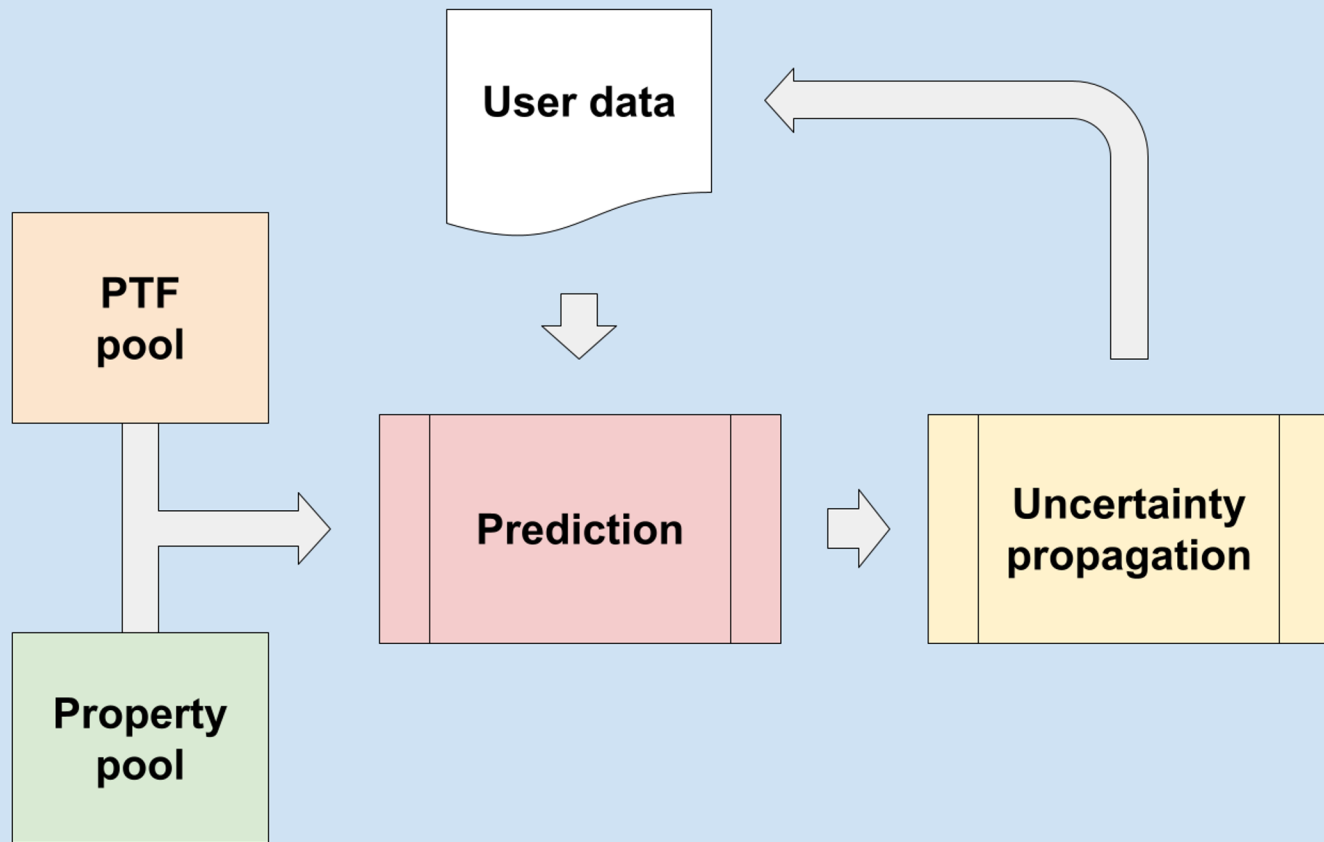
History of (SPEC-)SINFERS

- McBratney et al. (2002) 
- Trantner et al. (2009) & Morris et al. (2015) 
- Jones and McBratney (2017) 
- Padarian, Jones and McBratney (2021)   



SINFERS

SINFERS



Property pool

- **Strict** *lab_code* needed
 - Soil Chemical Methods Australasia
 - “Synonyms”

```
{
  "lab_code": {
    "scma": "6A1"
  },
  "name": "oc_wb",
  "group": "oc",
  "constraints": {
    "min": 0,
    "max": 58
  },
  "units": "g 100 g-1",
  "description": "Organic carbon - Walkley and Black",
  "metadata": {
    "reference": "http://anzsoil.org/def/au/scma/6A1"
  }
}
```

Link to GLOSOLAN



PTF pool

```
{
  "expression": "0.1766 + 0.00255 * scma_P10_HYD_C - 0.001487 * scma_P10_HYD_S",
  "sinfer_id": "theta_pwp",
  "target": [
    "scma_2E1"
  ],
  "predictors": [
    "scma_P10_HYD_C",
    "scma_P10_HYD_S"
  ],
  "uncertainty": {
    ...
  },
  "metadata": {
    "type": "<PTF_TYPE>",
    "reference": ...,
    "contact": ...,
    ...
  },
  "stats": {
    "rmse": ...,
    "r2": ...,
    ...
  }
}
```



PTF pool

- Prediction
 - Valid expression
(lab_code)
- Uncertainty
 - Fuzzy *k*-mean with extragrades
 - “Simple” uncertainty
 - Other methods
 - Some exceptions

```
{
  "expression": "0.1766 + 0.00255 * scma_P10_HYD_C - 0.001487 * scma_P10_HYD_S",
  "sinfer_id": "theta_pwp",
  "target": [
    "scma_2E1"
  ],
  "predictors": [
    "scma_P10_HYD_C",
    "scma_P10_HYD_S"
  ],
  "uncertainty": {
    ...
  },
  "metadata": {
    "type": "<PTF_TYPE>",
    "reference": ...,
    "contact": ...,
    ...
  },
  "stats": {
    "rmse": ...,
    "r2": ...,
    ...
  }
}
```



Clients



API

SINFERS

User data

PTF
pool

Property
pool

Prediction

Uncertainty
propagation



ACCURACY ASSESSMENT

Accuracy	R ²	Concordance	RPIQ	St. Bias
A	0.901 (0.82-0.99)	0.946	2.353 (0.82-1.85)	0.001
B	0.847 (0.76-0.86)	0.915	1.263 (0.54-1.71)	0.001
C	0.665 (0.58-0.74)	0.800	0.821 (0.20-1.34)	0.007
D	0.486 (0.10-0.60)	0.659	0.490 (0.02-1.10)	0.029

Multiple criteria:

R² = variance explained

Concordance: agreement at 1:1 line

RPIQ = Interquartile range/RMSE

St. Bias = Bias/Interquartile range