LIÈGE université
Gembloux
Agro-Bio Tech

SOPHIE @ Paris2022
Ring test, results, outlook & questions

Benjamin Guillaume – Alexandre Pomes Bordedebat - Aurore Degré

Challenge for soil physics labs

- Soil physics laboratories aim to quantify the hydrophysical properties of soils (i.e., retention and conductivity)
  - Important role in a wide range of societal issues
  - Data need to be reliable

- These properties are mainly structure dependent
  - The challenge of soil physics is to work on **undisturbed samples**

- There is no guarantee that two laboratories would give the same result on the same soil sample

- **SOPHIE demonstrates the need for interlab comparisons**

We identified 3 issues with increasing levels of complexity

- To ensure the **reproducibility** of a given protocol, over time, **within a laboratory**
- To ensure **consistency** between analyses performed using the same protocol in **different laboratories**
- To ensure consistency **(harmonization)** between similar hydro physical characterizations performed with **different protocols** in different laboratories

Leading to 3 questions

- **Are the measurements on a same sample stable in a same lab ?**
- **Are the measurements on a same sample stable in different labs ?**
- **Are the samples affected by transfers between labs ?**

It became obvious that we needed

- **A reference sample**
  - After a quick benchmarking, we identified a good candidate provided by UGent
  - mix of glass beads and cement

Wet end of the WRC - 1st ring test (ever)

- **14 labs** involved
- **84** reference **samples** (6 per lab)
  - 1 example from UGent + 5 manufactured by each lab
- **3 rounds** of measurements
  - Saturation
    - Saturation time: 48h (in box with water: water level incrementing at regular time intervals with 2 cm steps)
    - Water used: demineralized water
    - Presence of a bottom cloth: yes
    - Presence of a lid: yes
  - Mass measurement at 4 points of the retention curve
    - Equilibration time :
      - 10 hPa : 5 days --> mass measurement
      - 50 hPa : 7 days --> mass measurement
      - 100 hPa : 10 days --> mass measurement
      - 300 hPa : 15 days --> mass measurement
  - Drying :
    - 72h at 60°C
    - mass measurement

Wet end of the WRC - 1st ring test (ever)

From each lab : 6 samples :

Samples 1 and 2 :

Round 1 to 2 : **Keep**

Round 2 to 3 : **Keep**

Are the measurements on a same sample stable in a given lab ?

Samples 3 and 4 :

Round 1 to 2 : **Send** to + 1
Receive from -1

Round 2 to 3 : **Send** to + 1
Receive from -1

Are same samples giving the same data in different labs ?

Samples 5 and 6 :

Round 1 to 2 : **Send** to +1
Receive from -1

Round 2 to 3 : **Send back** to -1
Receive back from +1

Are the samples affected by transfers between labs ?

State of affairs

- The ring test is almost complete
  1st round : 14/14 labs
  2nd round : 12/14 labs
  3rd round : 12/14 labs

- All our analyses are based on these data

WR results – Data and outliers

- For some WRC, water content increased with tension
  - Physical nonsense -> deleted data points

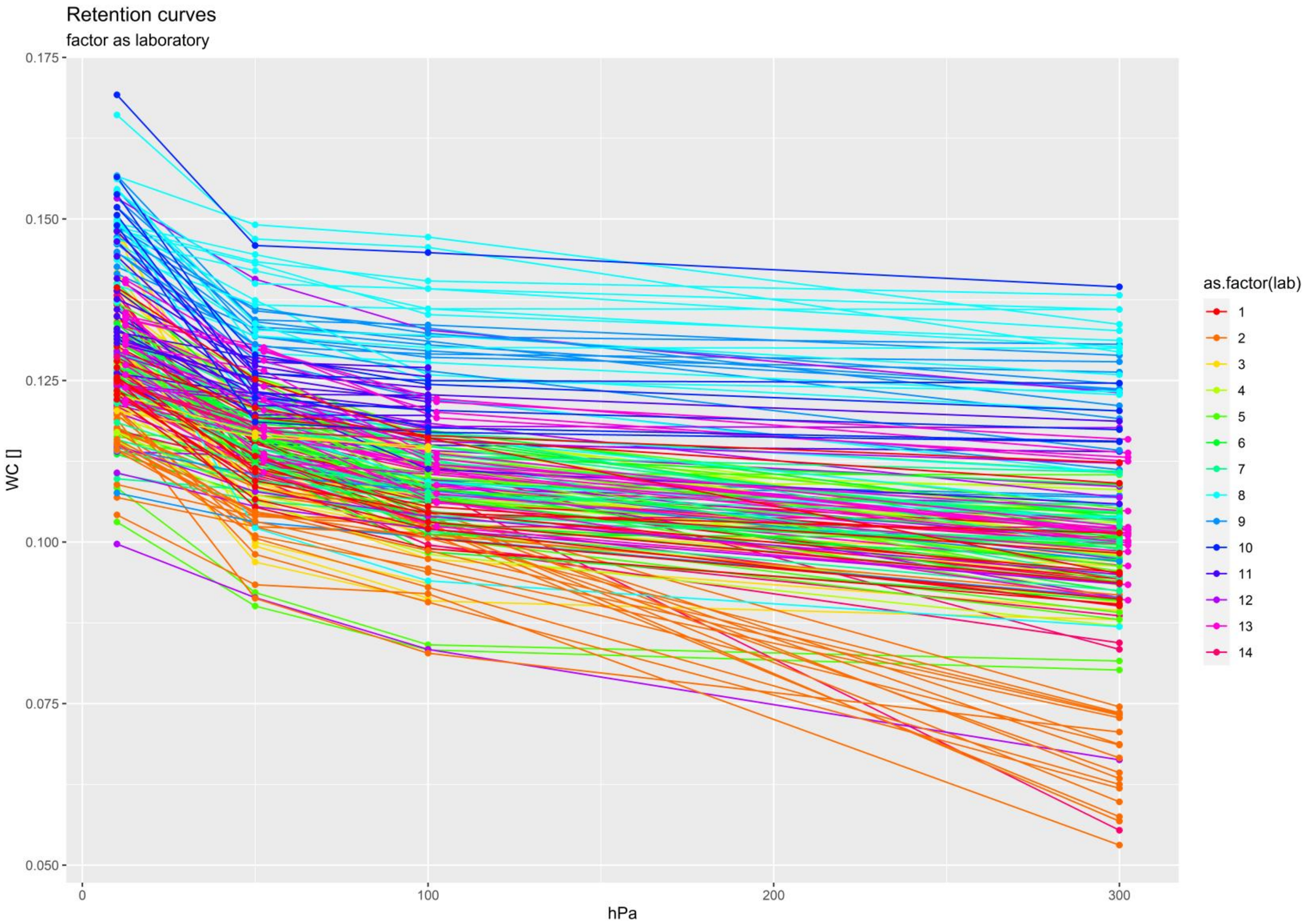
- Final dataset – 235 curves:

	Quadruplet	Triplet (10-50-300hPa)	Triplet (10-50-100hPa)	Doublet (10-50hPa)	Doublet (10-100hPa)	Total
Round 1	69	1	13	1	0	84
Round 2	52	0	18	5	0	75
Round 3	58	1	15	0	2	76

Results are not obvious ...

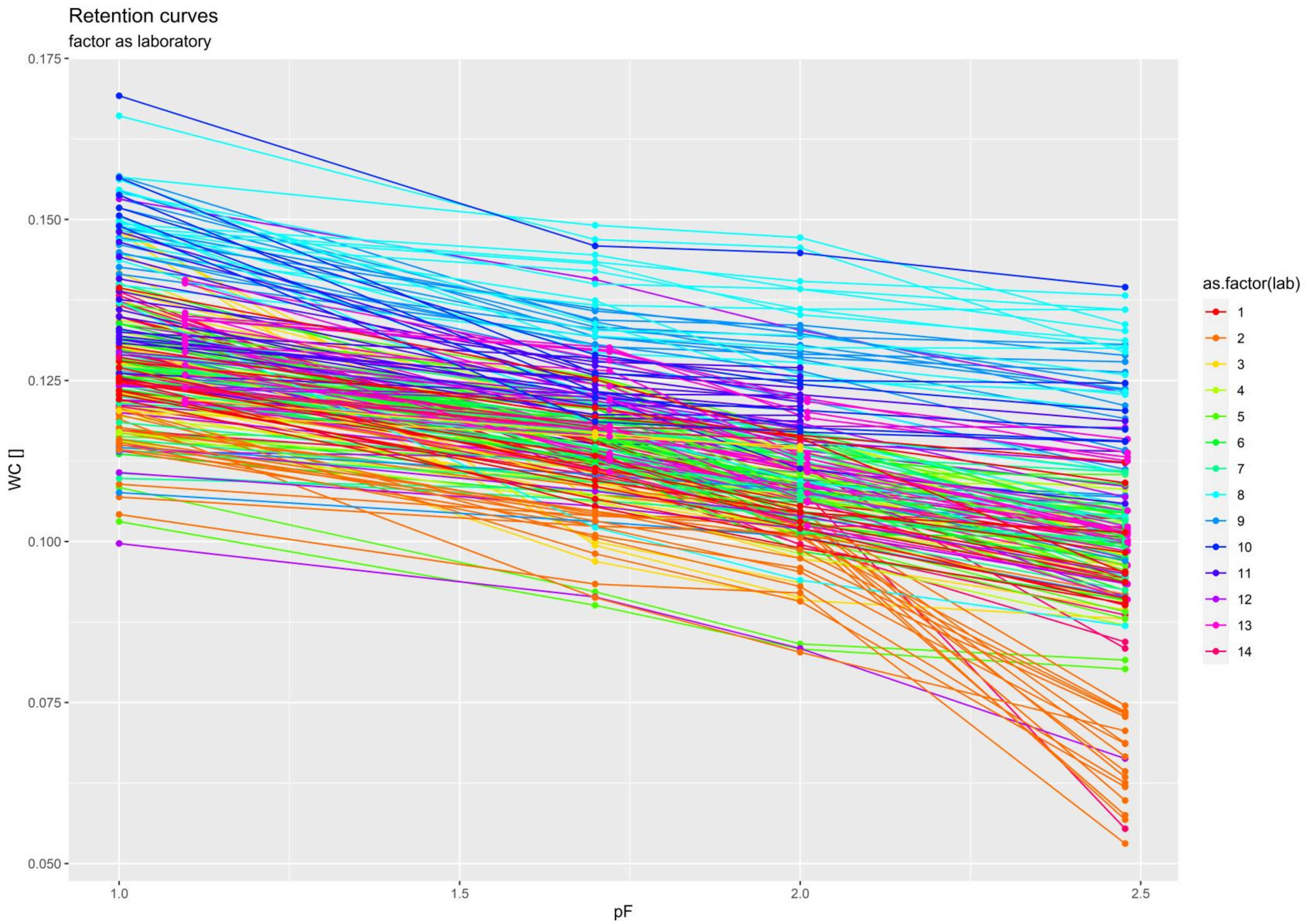
- One curve is 4, 3 or 2 points

- WRCs have a negative logarithmic trend



Results ... Overcome dependence

- Linear trend
- Lack of data independence
  - Repeated measurements
- We will use **linear model** to express these data (with independent intercept and slope)



Linear mixed model... Investigating sources of variability

- If observations were not affected by the reference **samples**, **transport** and were stable in each /or between different **laboratories,** this data set would be expressed by a single observation. **This is obviously not the case !**

- We must consider **random effects** to explain the variability of these observations :
  - **Laboratories** *(14 levels → 14 parameters)*
  - **Samples** *(84 levels → 84 parameters)*
  - **Transport** *(2 levels → 2 parameters)*

- **Parameters related to each level of random effects** are estimated and, **according to their variability**, it is decided whether **they exert an influence**, or not, on the output of the model.

Results ... Linear Mixed Random Effect Model

- Building up increasingly complex models
  - *Linear model :*

$$y_{ijk} = a + bx_j + \epsilon_{ijk}$$

  - *Linear mixed effect model with random effects on the* ***intercept*** *:*

$$y_{ijk} = \alpha + \beta_{L(ik)} + \gamma_{T(ik)} + \delta_{S(i)} + bx_j + \epsilon_{ijk}$$

Lab **Transport** **Sample**

  - *Linear mixed effect model with random effects on the* ***slope*** *:*

$$y_{ijk} = a + \left(\zeta + \eta_{L(ik)} + \theta_{T(ik)} + \iota_{S(i)}\right) * x_j + \epsilon_{ijk}$$

Lab **Transport** **Sample**

  - *Linear mixed effect model with random effects on the* ***intercept*** *and the* ***slope*** *:*

$$y_{ijk} = \alpha + \beta_{L(ik)} + \gamma_{T(ik)} + \delta_{S(i)} + \left(\zeta + \eta_{L(ik)} + \theta_{T(ik)} + \iota_{S(i)}\right) * x_j + \epsilon_{ijk}$$

Lab **Transport** **Sample** **Lab** **Transport** **Sample**

- **Parameters related to each level of random factors** are estimated and, **according to their variability**, it is decided whether **they exert an influence**, or not, on the output of

y : The wc []
x : The tension [pF]
a : The intercept
b : The slope
e : The error []
i : The i^{th} sample
j : The j^{th} tension
k : The k^{th} round
Random effect on interception
α : The general mean intercept
β_L : Laboratory
L : [1,...,14]
Υ_T : Transport
T : [0,1]
δ_S : Sample
S : [0,...,84]
Random effect on slope
ζ : The general mean slope
η_L : Laboratory
L : [1,...,14]
θ_T : Transport
T : [0,1]
ι_S : Sample
S : [0,...,84]

Results ... But how to estimate the random effects (parameters) from the observations?

- Bayesian statistics
  - *Information based on pre-existing knowledge can be incorporated*
  - *Complex models with many variance components can be fit*

Posterior **Likelihood** **Prior**

$$P(Parameters|Data) \propto P(Data|Parameters)\ P(Parameters)$$

Posterior — *Output : Probability **distributions** of the parameters knowing the data*

Likelihood — *Input : Probability **distributions** to get the data for a given parameter value*

Prior — *Input : A priori probablity **distributions** of the parameters based on previous knowledge*

A quick example of Bayesian statistics: Let's play a game

The game : A card game with 1's and 0's. If you get a 1 you win 1 euro, if you get a 0 you lose 1 euro.

Objective : Estimate the probability of losing (or winning) → Estimate the posterior (parameter knowing the data)

Your experiment :

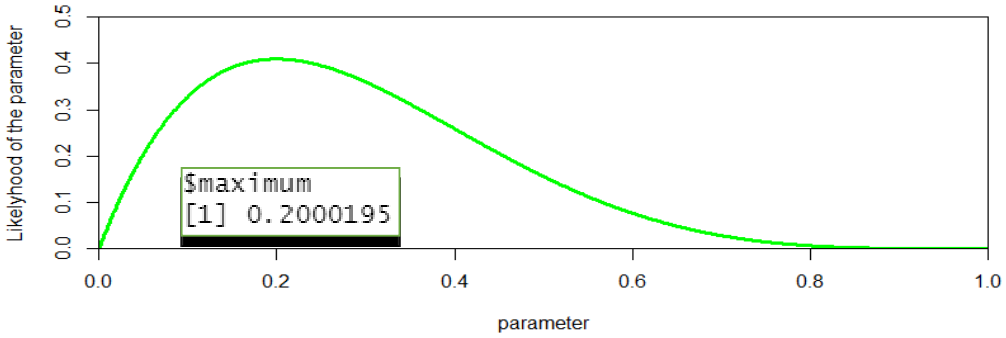
```
> data
[1] 1 1 1 0 1
```

Posterior $\propto$ **Likelihood** x **Prior**

Likelihood

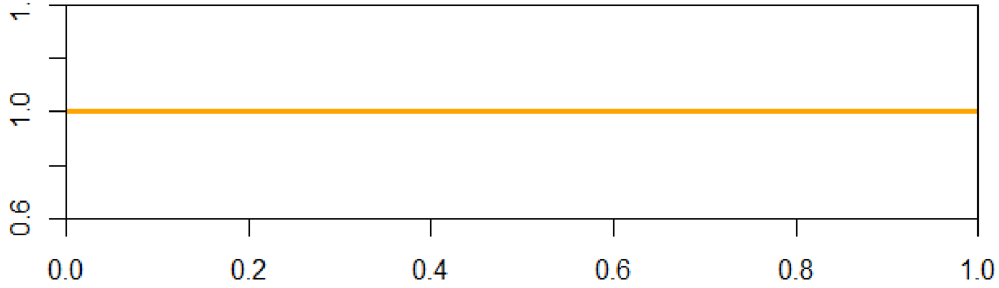
```
lik.fun <- function(parameter) {
  ll <- dbinom(x=1, size = 5, prob=parameter)
  return(ll)
}
test_param <- seq(from = 0, to = 1 , by = 0.001)
likelyhood <- lik.fun(test_param)

optimize(lik.fun, c(0,1), maximum=TRUE)
```



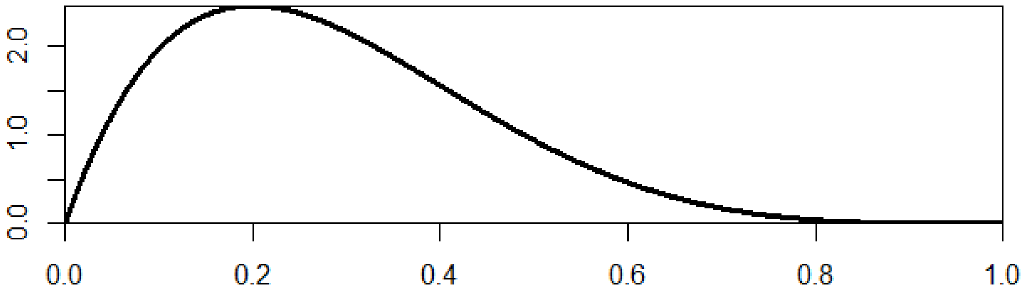
Prior

```
prior1 <- dbeta(p,1,1) # uniformative
```



Posterior

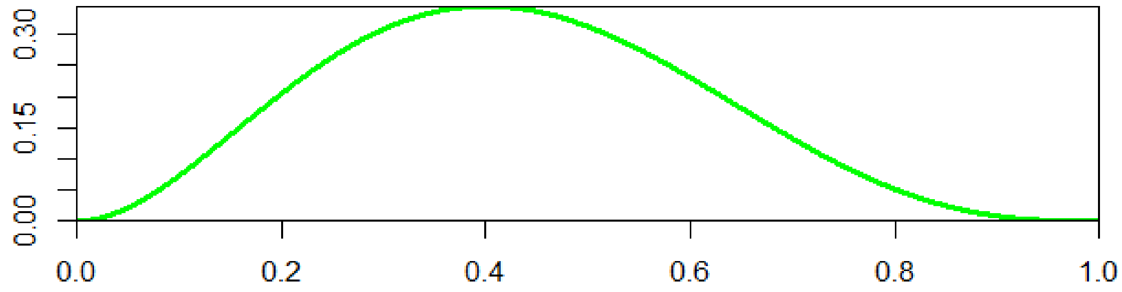
```
numerator1 <- function(p) dbinom(x,size,p)*dbeta(p,a1,b1)
denominator1 <- integrate(numerator1,0,1)$value
posterior1 <- numerator1(p)/denominator1
```



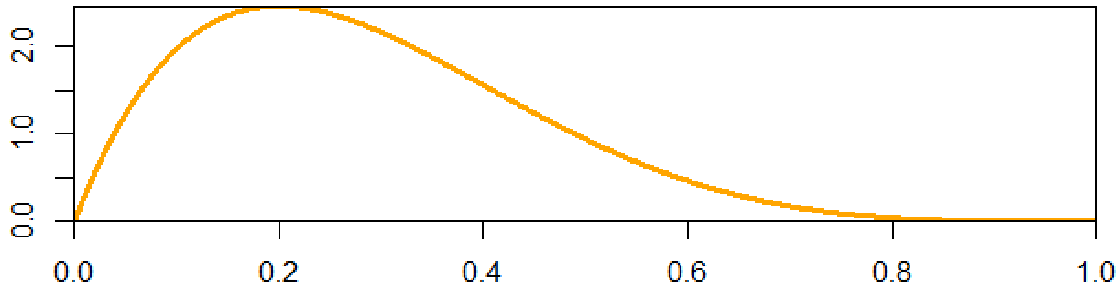
A quick example : Second draw

```
> data
[1] 1 1 0 1 0
```

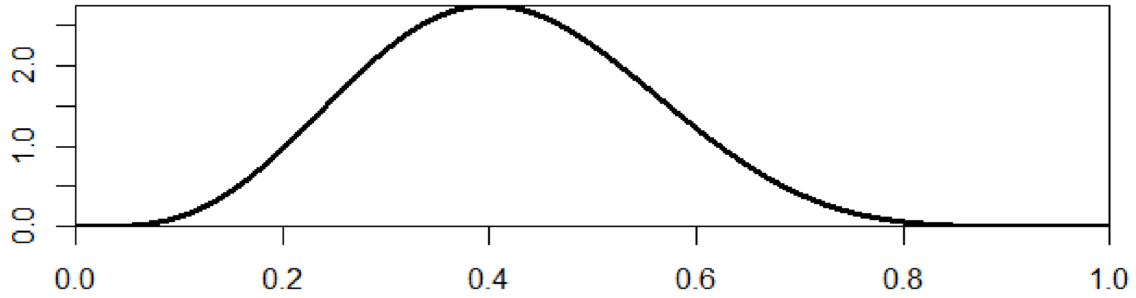
Likelihood 2



Prior 2 = **Posterior 1**

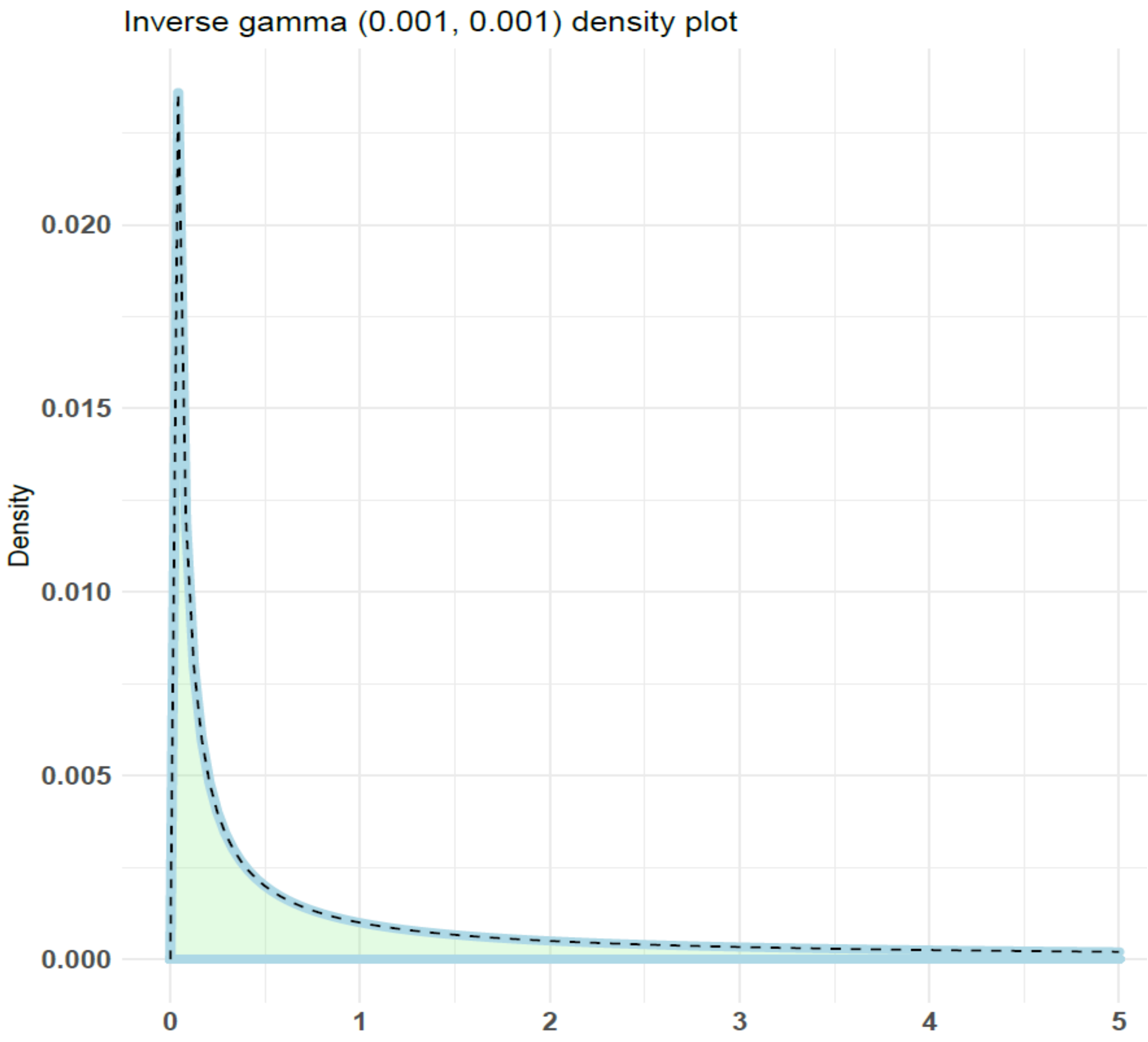


Posterior 2

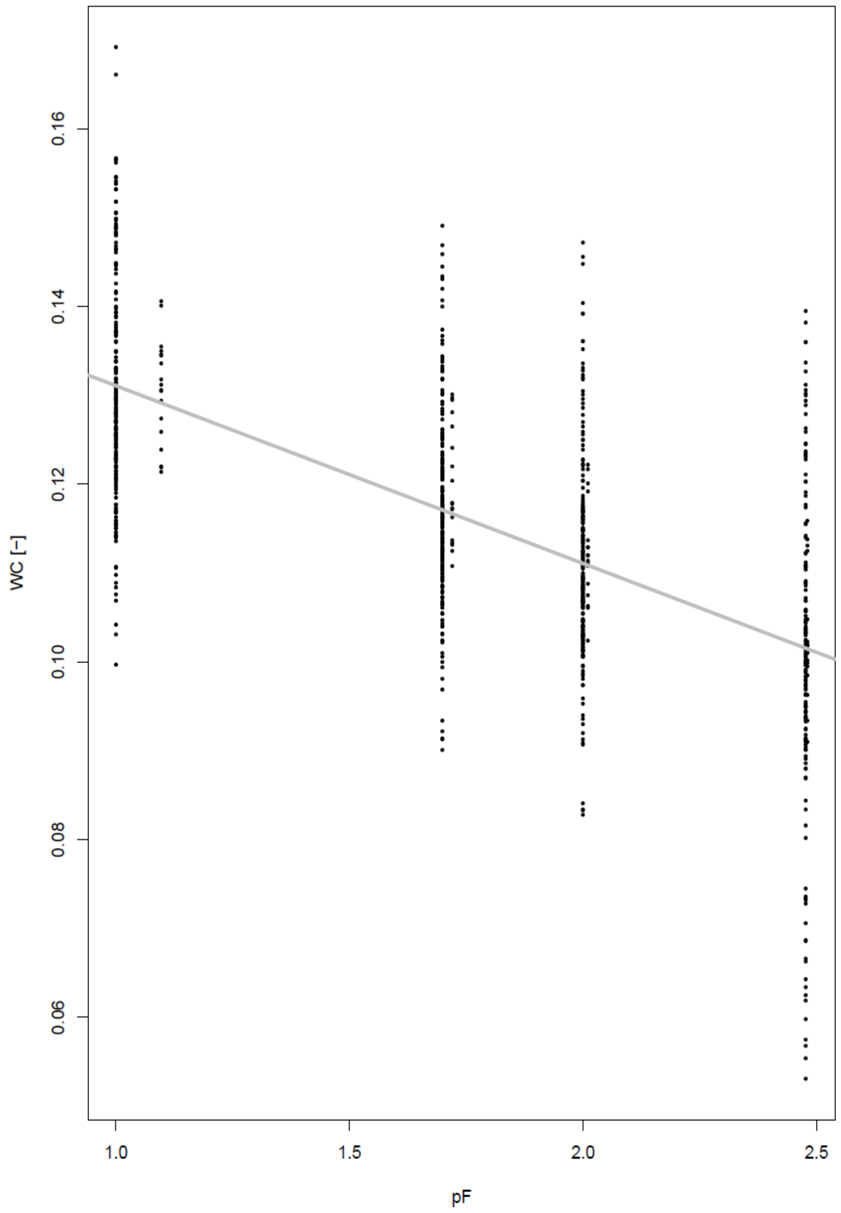


Choice of the priors

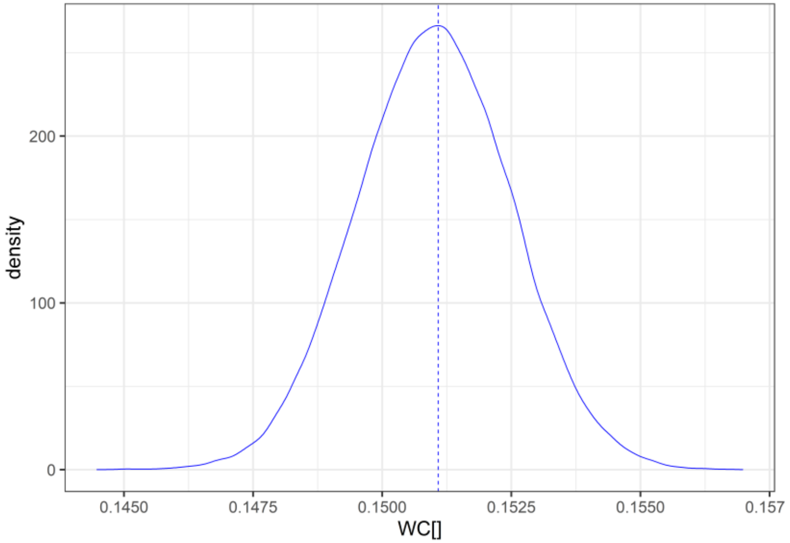
- Subjective belief - transparency
- $(\beta_L, \delta_S, \Upsilon_T, \eta_L, \iota_S, \theta_S) \sim \text{Normal}(0,\tau)$
  - A priori, parameters from each random effects are normally distributed around 0 with an unknown variance.
  - i.e., Laboratory 1 may overestimate retention ($\beta_1>0$) while laboratory 2 may underestimate it ($\beta_2<0$). But in general, we expect β_L to be normally distributed around a mean value = 0.
- $\tau\ (\beta_L, \delta_S, \Upsilon_T, \eta_L, \iota_S, \theta_S) \sim$ Inverse gamma(0.001, 0.001)
  - « A just proper default uninformative prior » for variance parameters *(Spiegelhalter et al., 2003)*



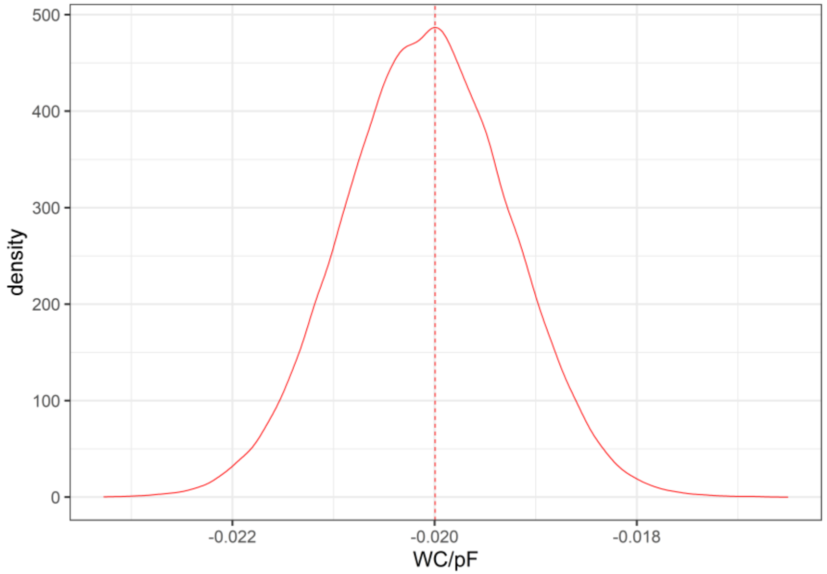
Linear model without random effects: $y_{ijk} = a + bx_j + \epsilon_{ijk}$



Posterior density of a
the mode is 0.1511 [-]

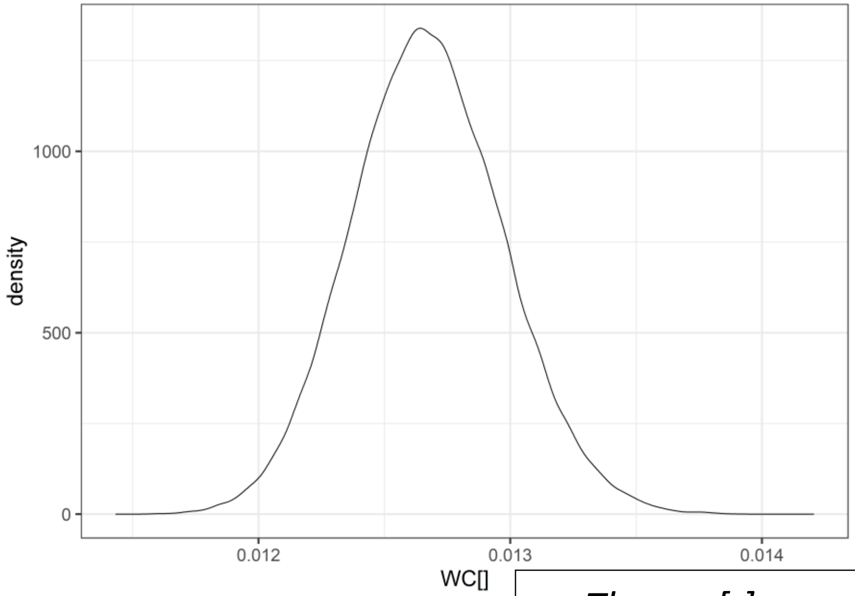


b : the mode is - 0.01999 [-/pF]



Standard deviation of the error : σ_{ε}

- Represents the variability not considered in the model

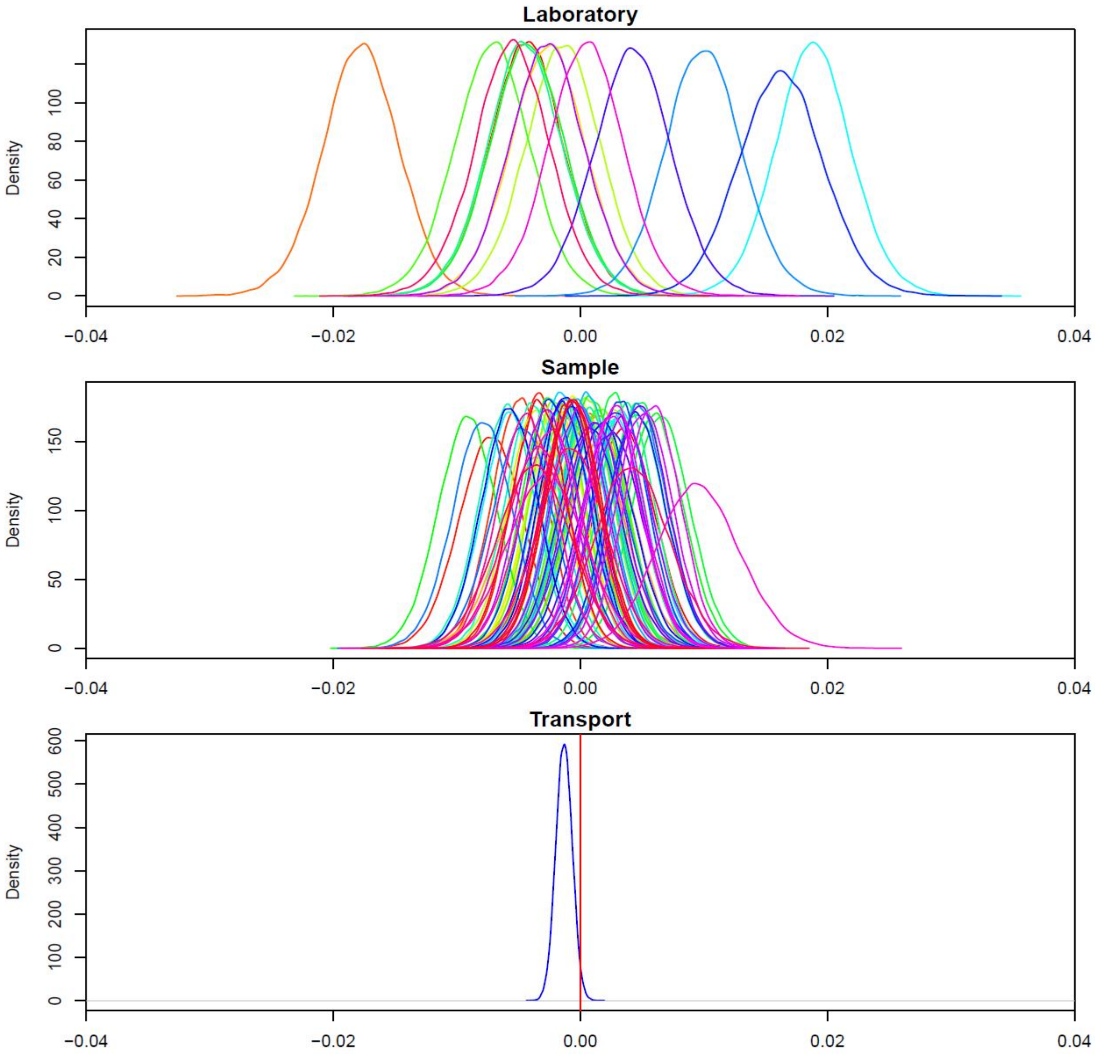


y : The wc [-]
x : The tension [pF]
a : The intercept [-]
b : The slope [-/pF]
e : The error [-]
i : The i^{th} sample
j : The j^{th} tension
k : The k^{th} round

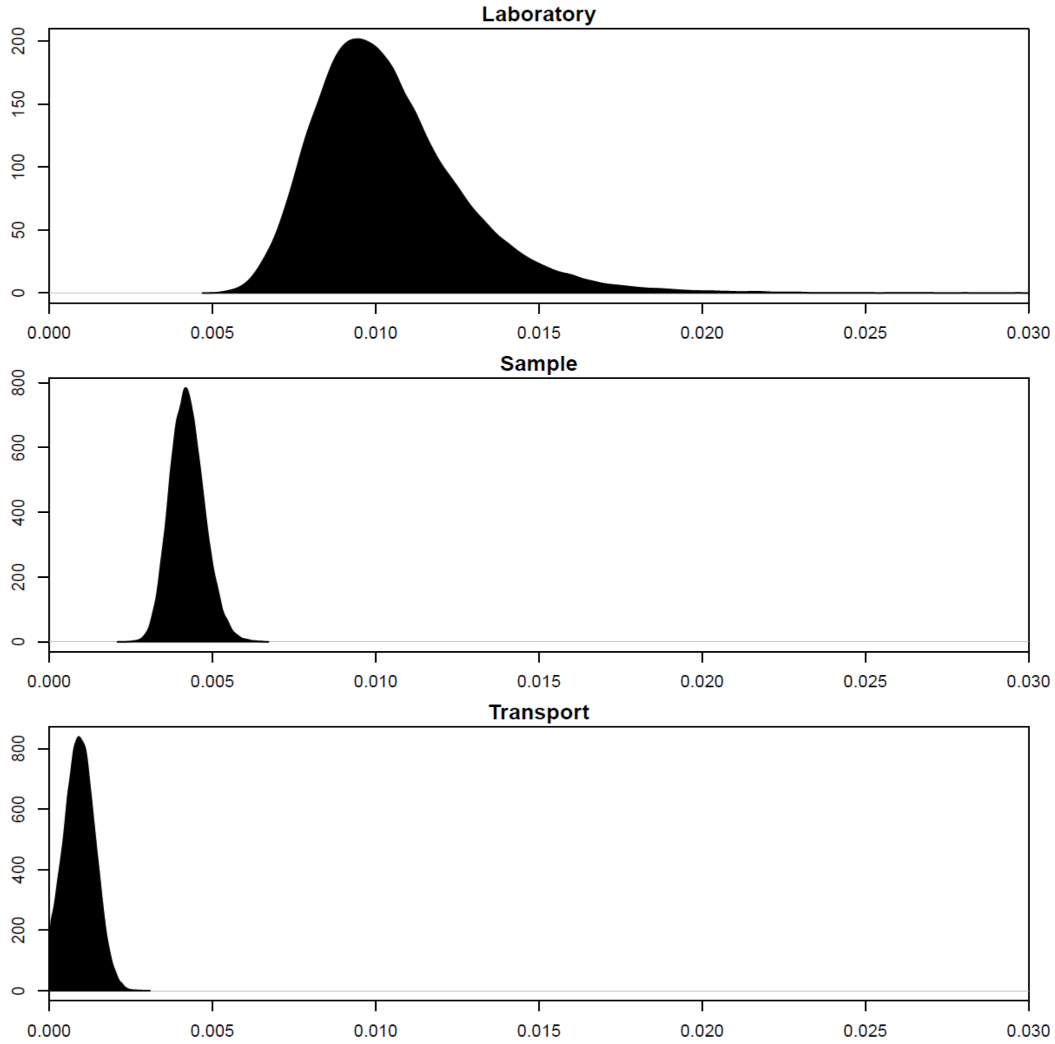
Linear model with « **Lab** », « **Transport** » and « **Sample** » as random effect on the **intercept** :

$$y_{ijk} = \alpha + \beta_{L(ik)} + \gamma_{T(ik)} + \delta_{S(i)} + bx_j + \epsilon_{ijk}$$

- We can observe the relative influence of each random effect on the model output (and thus on the data) thanks to the variability of the parameters (linked to each level of random effects).



Posterior distributions of the parameters estimates for the random effects (wc [])

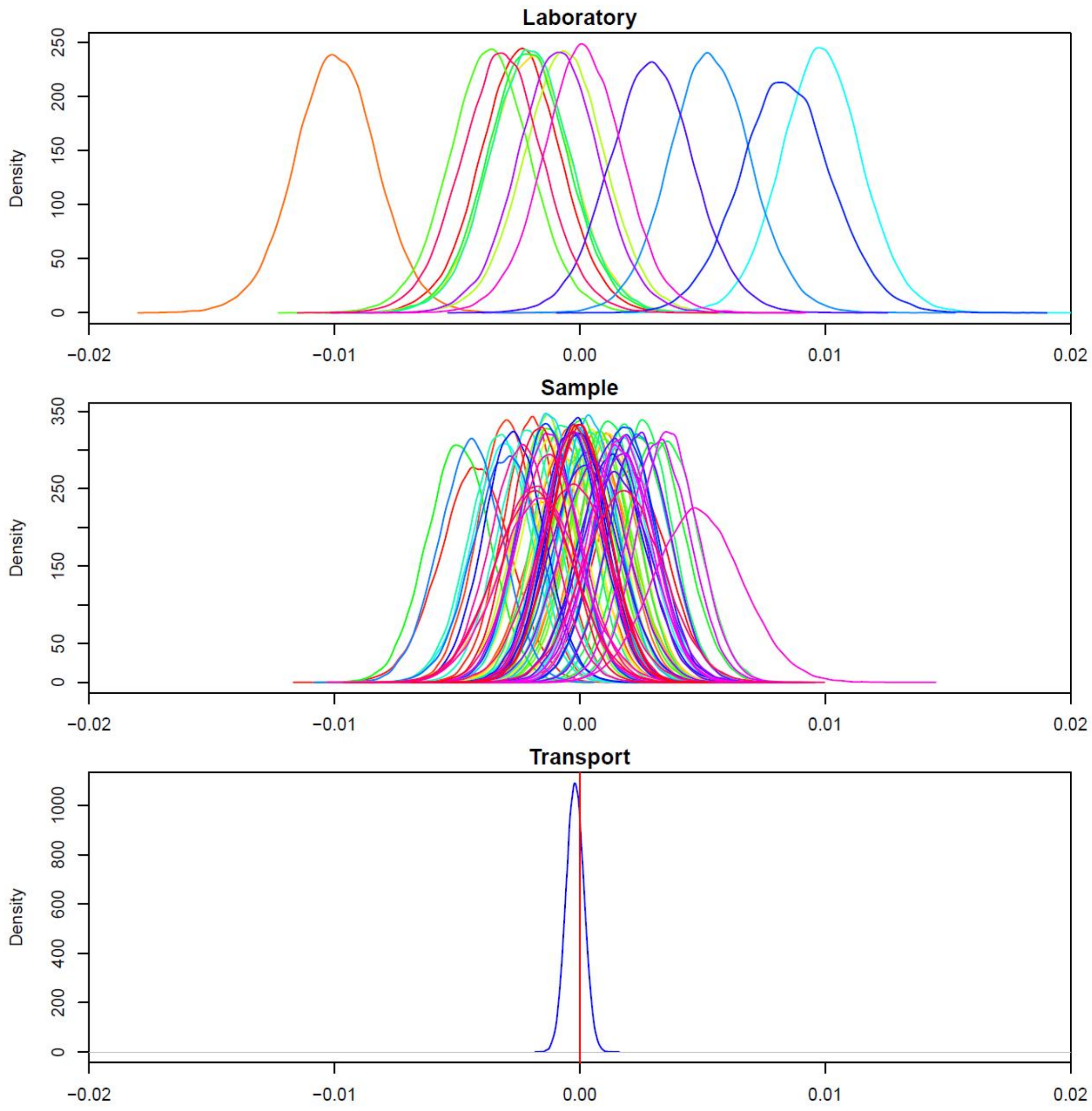


Posterior distributions of the ***standard deviation*** *of the parameters estimates for the random effects (wc [])*

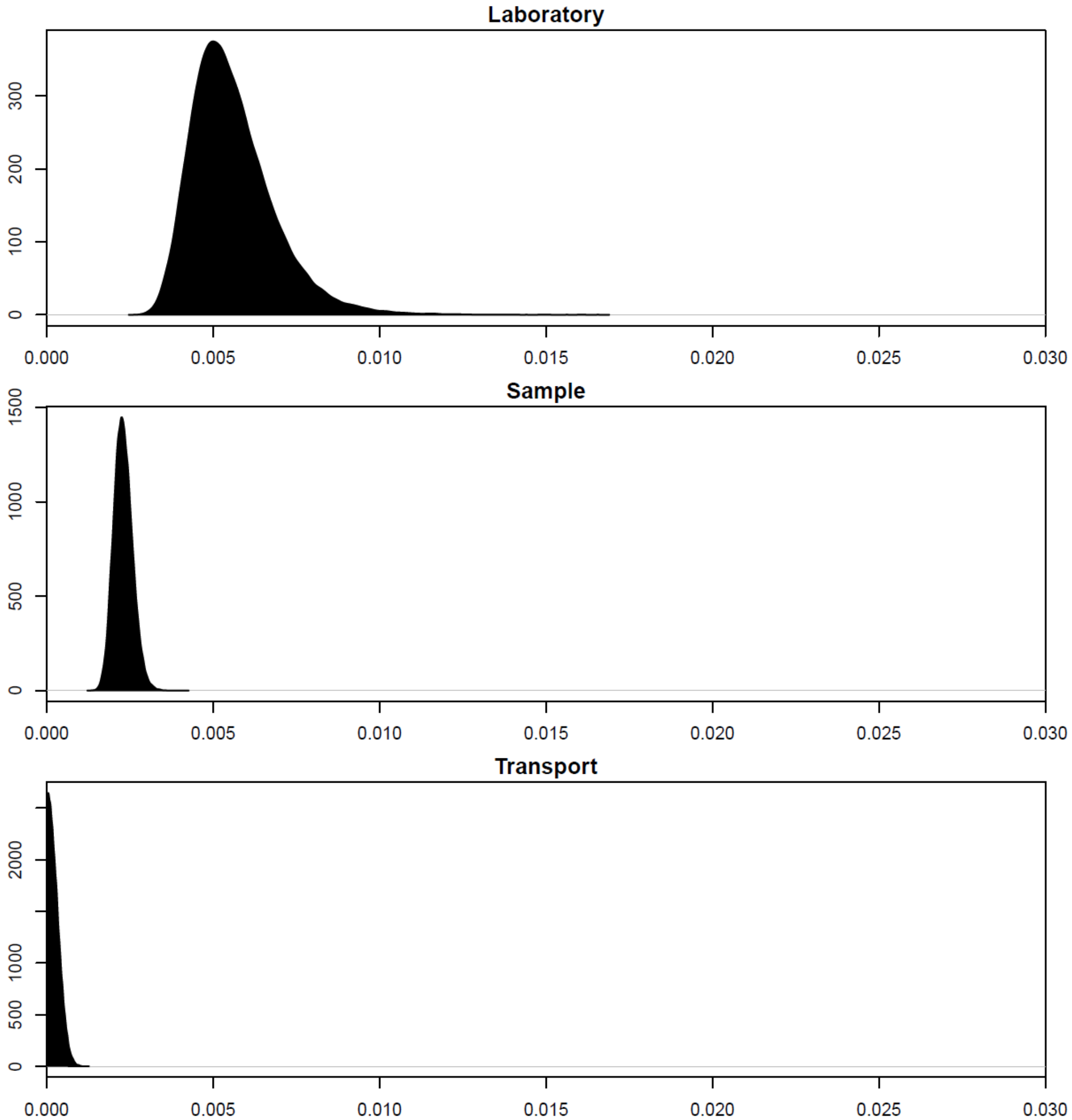
y : The wc []
x : The tension [pF]
b : The slope
e : The error []
i : The i^{th} sample
j : The j^{th} tension
k : The k^{th} round
Random effect on interception
α : The general mea intercept
β_L : Laboratory
Υ_T : Transport
δ_S : Sample

Linear model with « **Lab** », « **Transport** » and « **Sample** » random effect on the **slope** :

$$y_{ijk} = a + \left(\zeta + \eta_{L(ik)} + \theta_{T(ik)} + \iota_{S(i)}\right) * x_j + \epsilon_{ijk}$$



Posterior distributions of the parameters estimates for the random effects (wc/pF)



Posterior distributions of the ***standard deviation*** *of the parameters estimates for the random effects (wc/pF)*

y : The wc []
x : The tension [pF]
a : The intercept
e : The error []
i : The i^{th} sample
j : The j^{th} tension
k : The k^{th} round

Random effect on slope

ζ : The general mean slope
η_L : Laboratory
θ_T : Transport
ι_S : Sample

Random effects on the **intercept**

*From the posterior distribution **of standard deviation** estimates of random effects*

- Laboratory > Sample > Transport

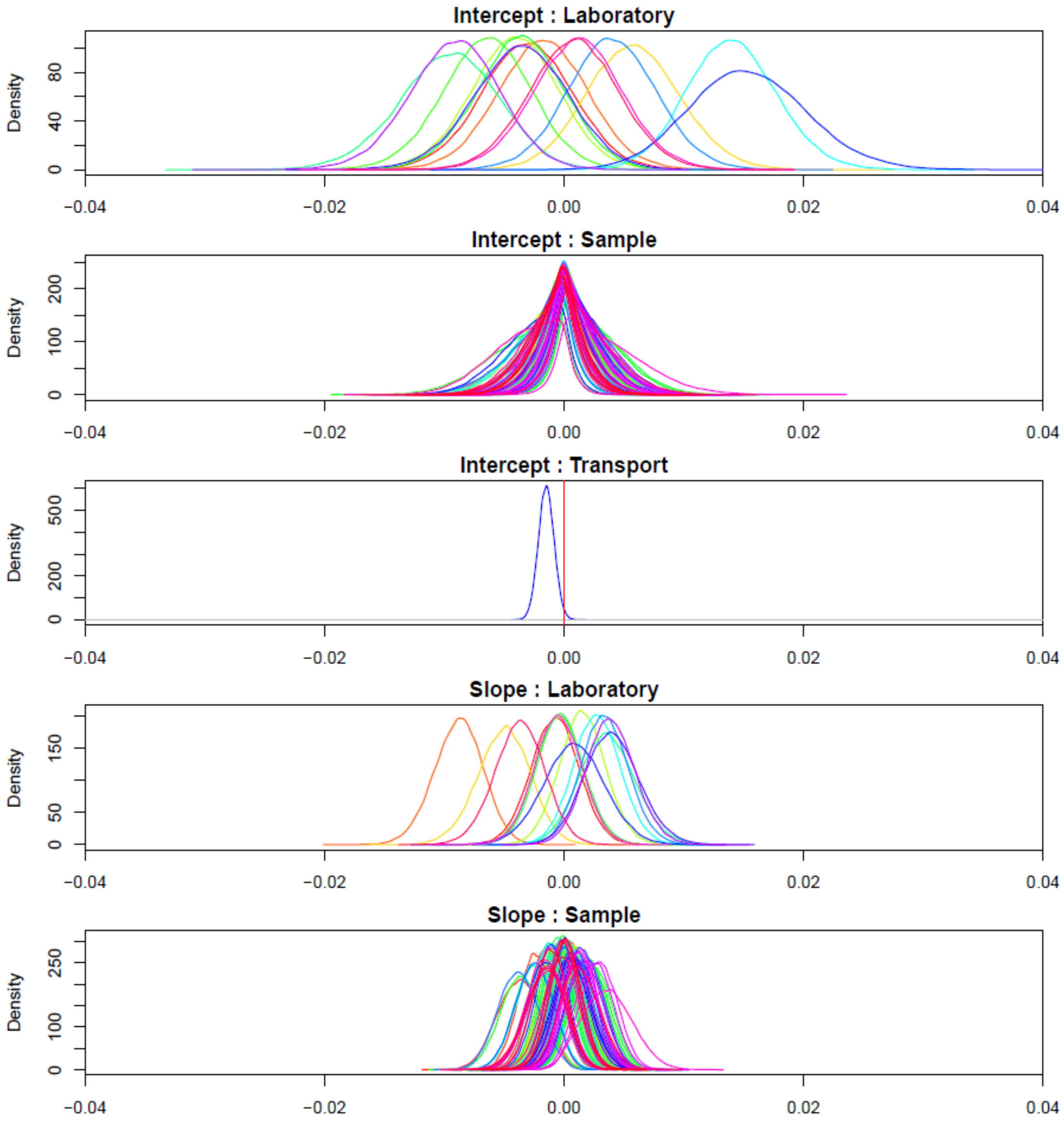
Random effects on the **slope**

*From the posterior distribution **of standard deviation** estimates of random effects*

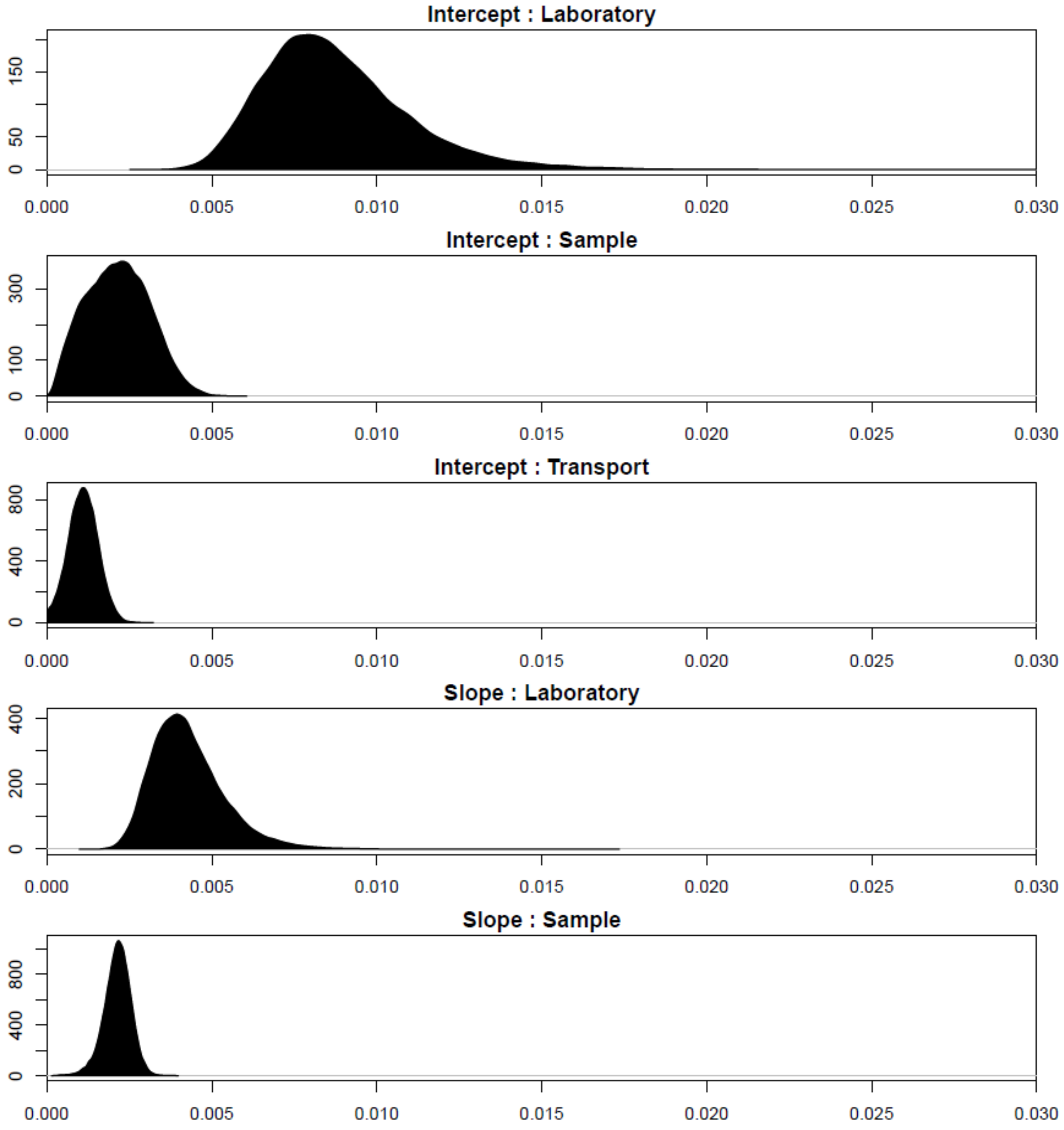
- Laboratory > Sample > Transport
- **Transport** effect is negligible -> **fixed** slope

Linear model with « **Lab** », « **Transport** » and « **Sample** » varying **interception** and with « **Lab** » and « **Sample** » varying **slope** :

$$y_{ijk} = \alpha + \beta_{L(ik)} + \gamma_{T(ik)} + \delta_{S(i)} + \left(\zeta + \eta_{L(ik)} + \iota_{S(i)}\right) * x_j + \epsilon_{ijk}$$



Posterior distribution estimates for random effects (wc and wc/pF)



Posterior distribution ***of standard deviation*** *estimates for random effects (wc and wc/pF)*

y : The wc []
x : The tension [pF]
a : The intercept
b : The slope
e : The error []
i : The i^th^ sample
j : The j^th^ tension
k : The k^th^ round

Random effect on interception

α : The general mean intercept
$β_L$: Laboratory
$ϒ_T$: Transport
$δ_S$: Sample

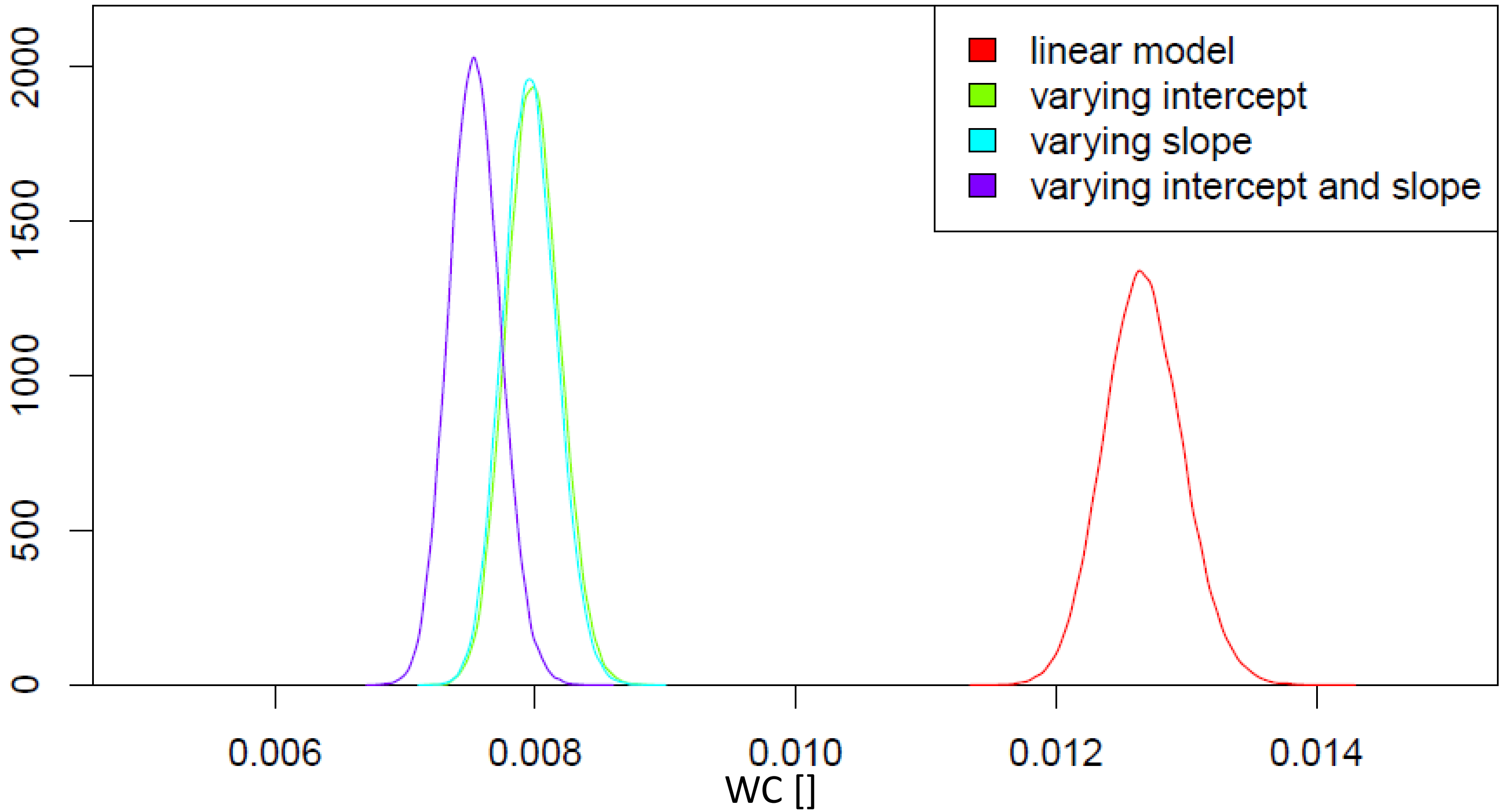
Random effect on slope

ζ : The general mean slope
$η_L$: Laboratory
$θ_T$: Transport
$ι_S$: Sample

Evolution of the SD of the error of the model

- Represents the variability not considered in the model

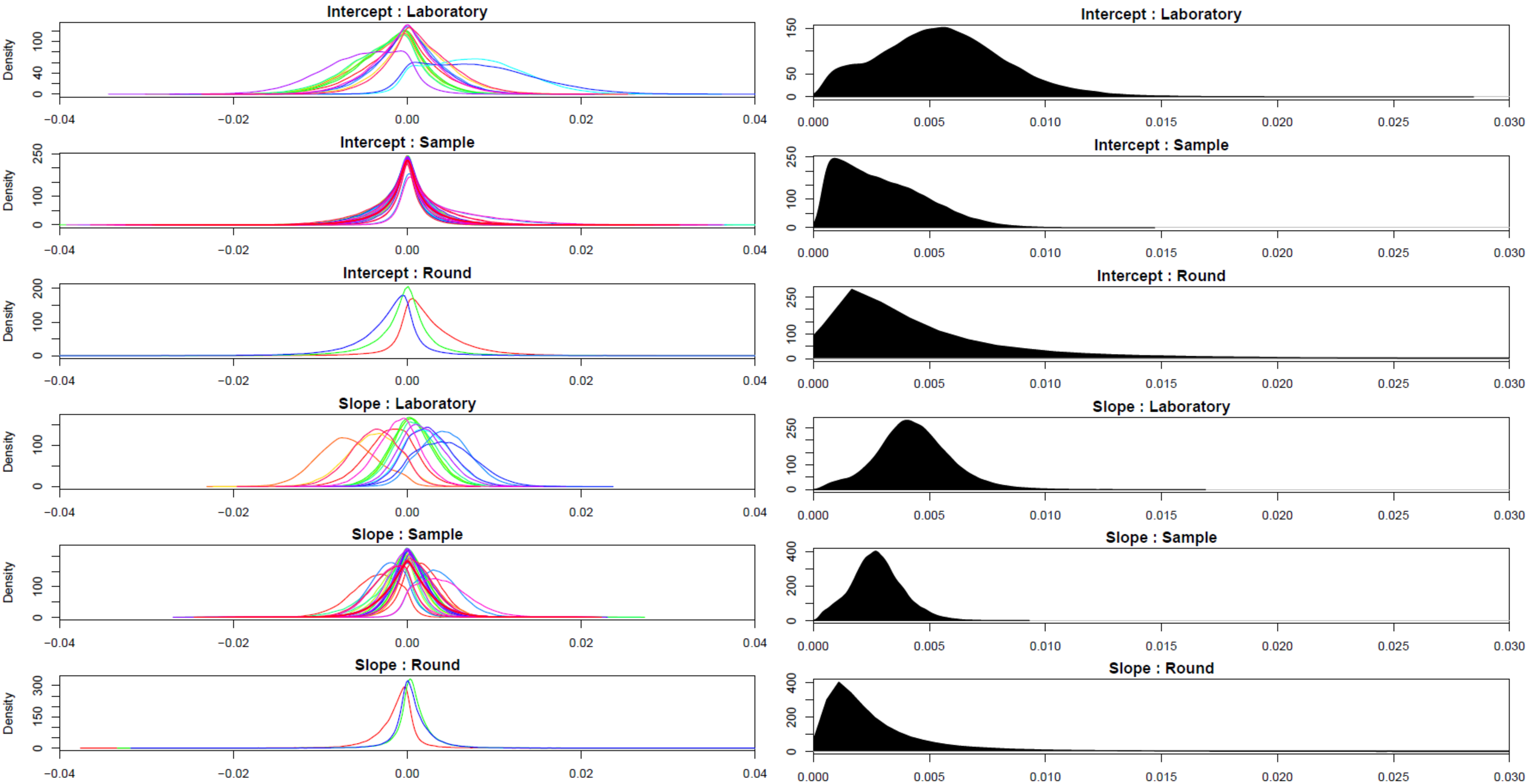
- Two main reasons :



- Linear model does not perfectly fit the measured data

- Variation of the measured retention of a same sample across rounds in a given laboratory
  - *Sample instability or measurement instability*

Round effect on staying samples



Conclusion of the analysis

Are the measurements on a same sample stable in a given lab ?

- **No, but also due to changes of samples themselves**

Are same samples giving the same data in different labs ?

- **No and the labs seem to account for most of the explained variability**

Are the samples affected by transfers between labs

- **Probably, but not that much**

Reference sample issues – Bulk densities

(from the prepared masses after curing, assuming rings of 100 cm^3)

- Bulk densities are different between crafting labs
- Bulk densities from the examples from UGent are lower than the replicas from the other labs
- SD are variable depending the crafting lab

Newman and Keuls' groups of populations of bulk densities (g/cm^3)

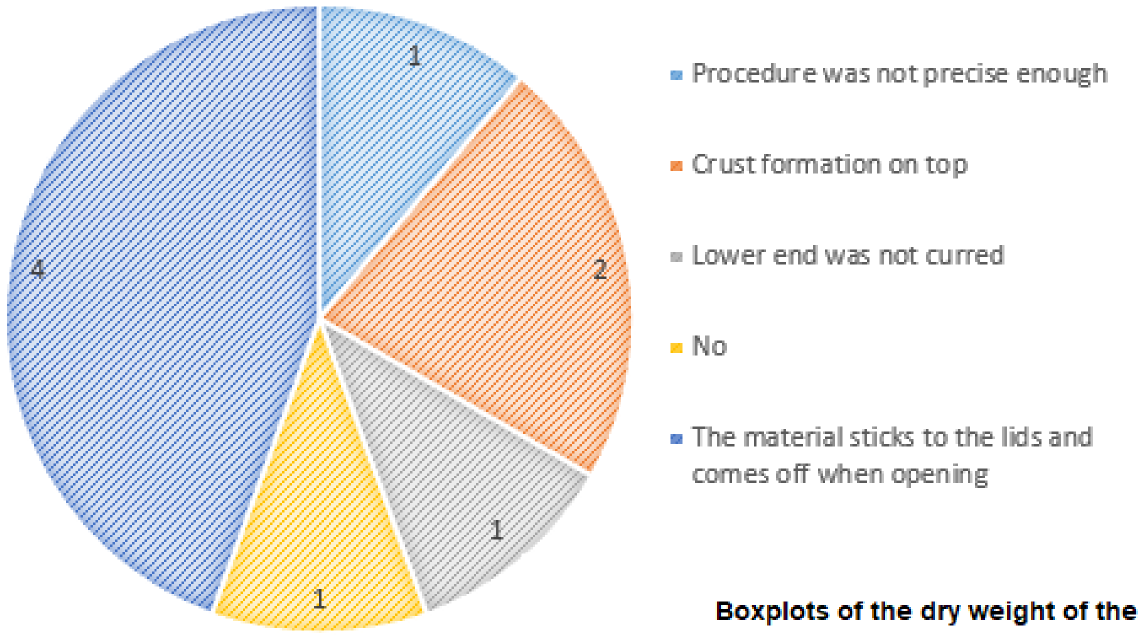
Lab number	Mean	SD	Pop. size	NK Group
1	1.8035	0.0094	5	a
2	1.7781	0.0141	5	b
8	1.7639	0.0494	5	b c
3	1.7551	0.0049	5	b c d
11	1.7540	0.0090	5	b c d
7	1.7528	0.0046	5	b c d
10	1.7425	0.0062	5	c d
12	1.7314	0.0168	5	d
4	1.6948	0.0198	5	e
13	1.6657	0.0291	5	f
5	1.6579	0.0133	5	f
6	1.6574	0.0177	5	f
9	1.6489	0.0056	5	f
14	1.6462	0.0136	5	f
15	1.6359	0.0113	14	f

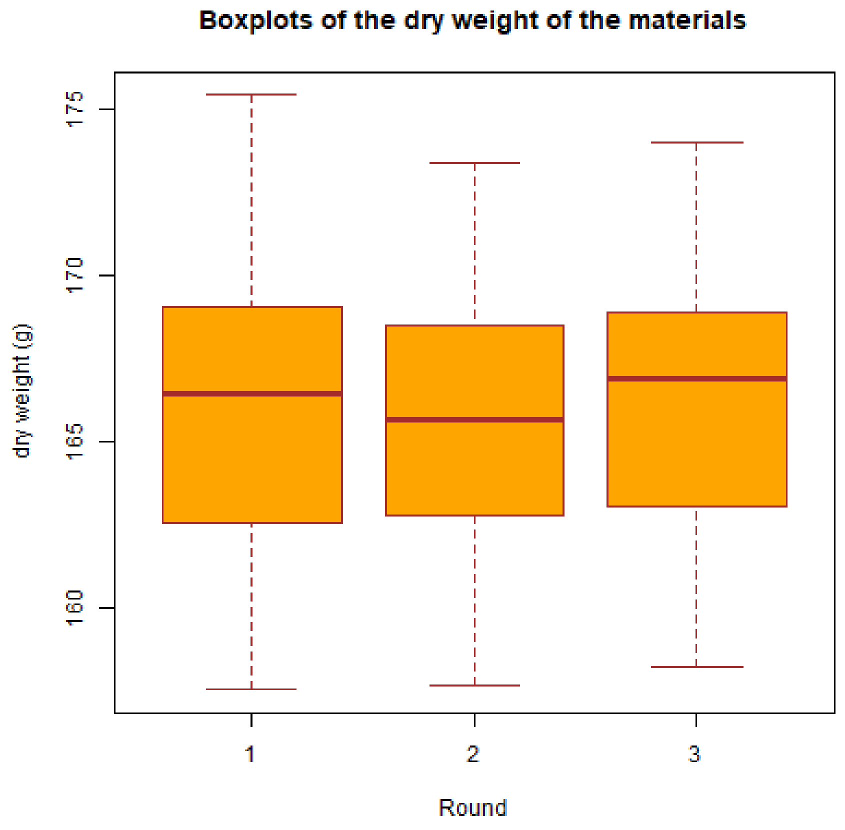
Reference samples issues

- Sample preparation

- Biofilm formation
  - For some WRC, water content increased with tension
  - Dry masses have increased despite the reported loss of material between rounds
  - Orange spots on the samples

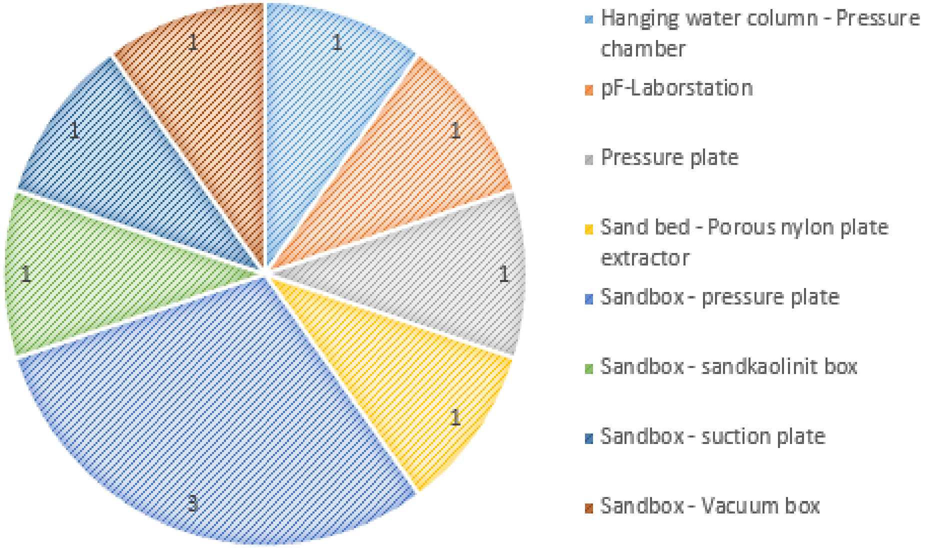
- Successive wetting and drying cycles can damage the mortar
- WC variations are too small for the range of pressures studied



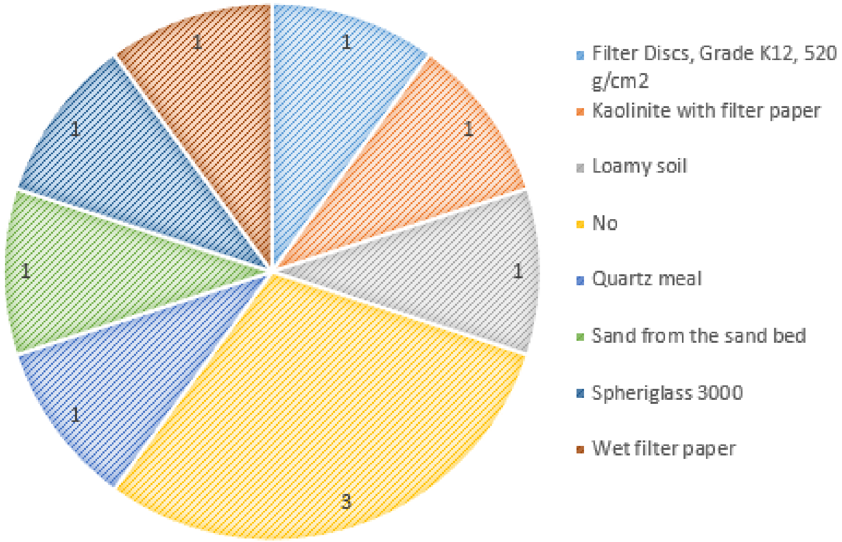


Procedure differences between laboratories

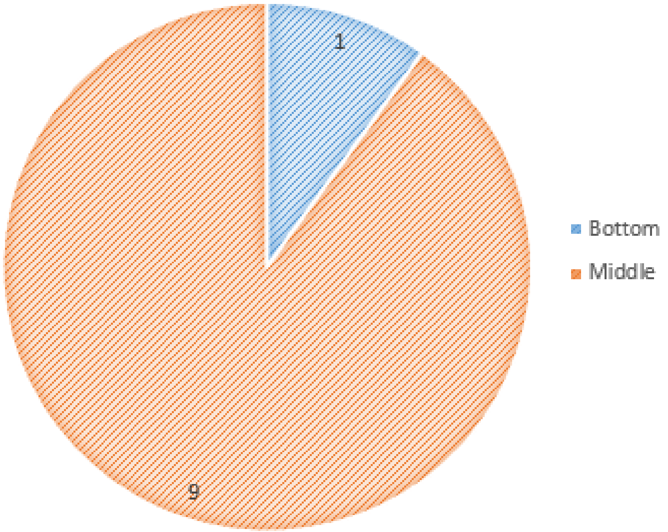
- Apparatus used



- Contact material

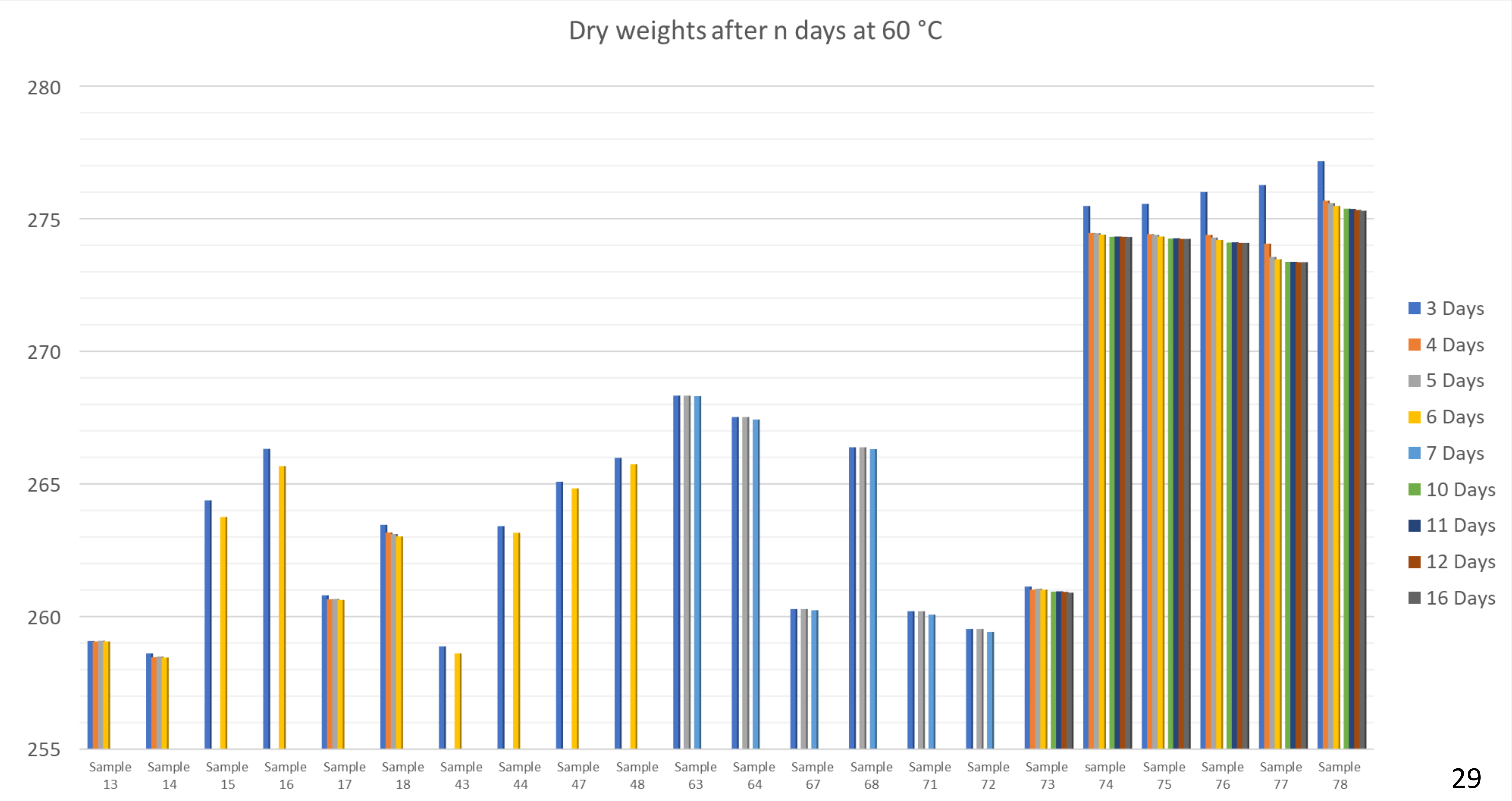


- Bottom or middle of the sample for pressure regulation



- Cleaning of the plates
- Dry weight measurement procedure (dessicator, …)
- Samples storage in the labs between rounds
- Caps to prevent evaporation
- Pressure regulation issues
- Temperature control in the laboratory (the environment)

Dry weight measurement



General conclusions of the first ring test

- The **reference samples** did not have standard retention properties
  - Manufacturing perspective
  - Unstable

- **Differences between laboratories** account for most of the explained variability (more than samples)
  - Non-harmonized SOPs (from the saturation to the dry weight measurement)

- **Differences within a same laboratory exist**
  - Reference samples unstability
  - Procedures reproducibility

Open discussion

- Remarks / Questions?
- How can we improve the analysis?
- How should we communicate these results?
- What could/should be done now?
  - Reference material (New propositions)
  - Harmonization (SOPs, GLOSOLAN, …)
  - Next ring test (Yes but maybe too long)

LIÈGE université
Gembloux
Agro-Bio Tech