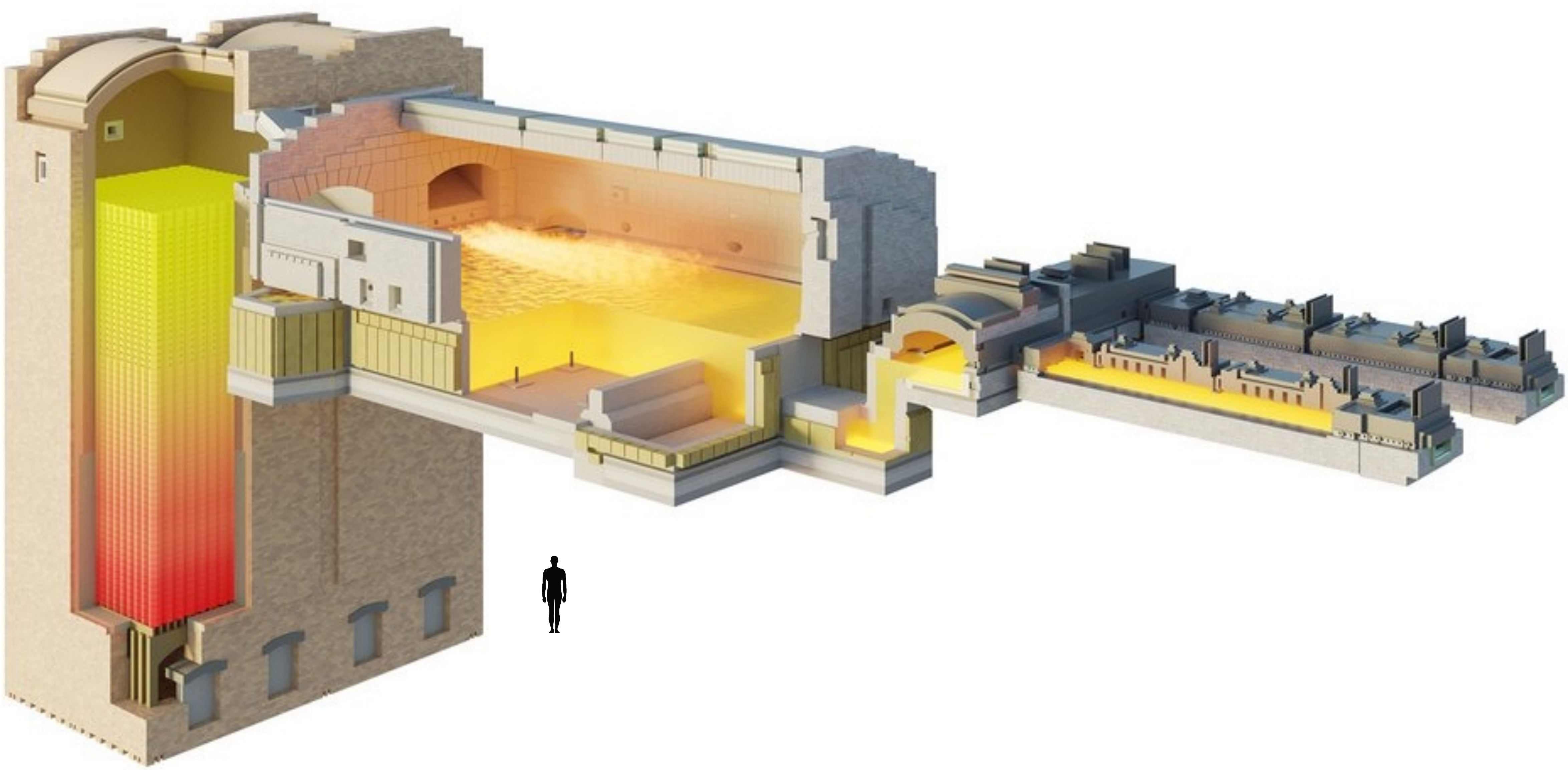
Predicting Effluents from Glass Melting Process for Sustainable Zero-Waste

Modeling of continuous process with sparse batch sampling of filter dust

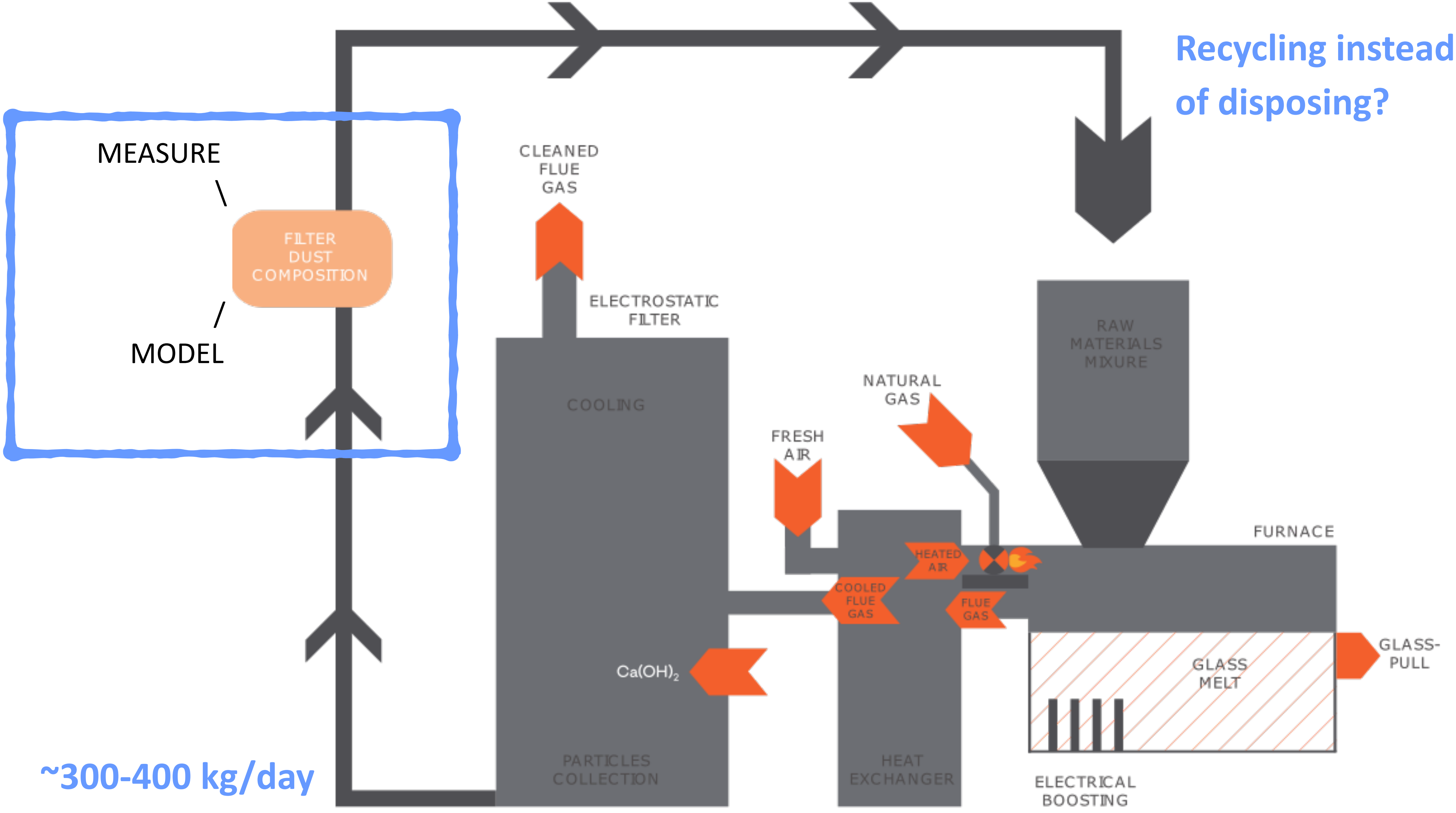
Dr Drejc Kopac (drejc.kopac@culmium.com)

Matej Emin (matej.emin@culmium.com)

Dr Philip Ramsey (philip.ramsey@predictum.com)



Recycling instead of disposing?



~300-400 kg/day

Overview

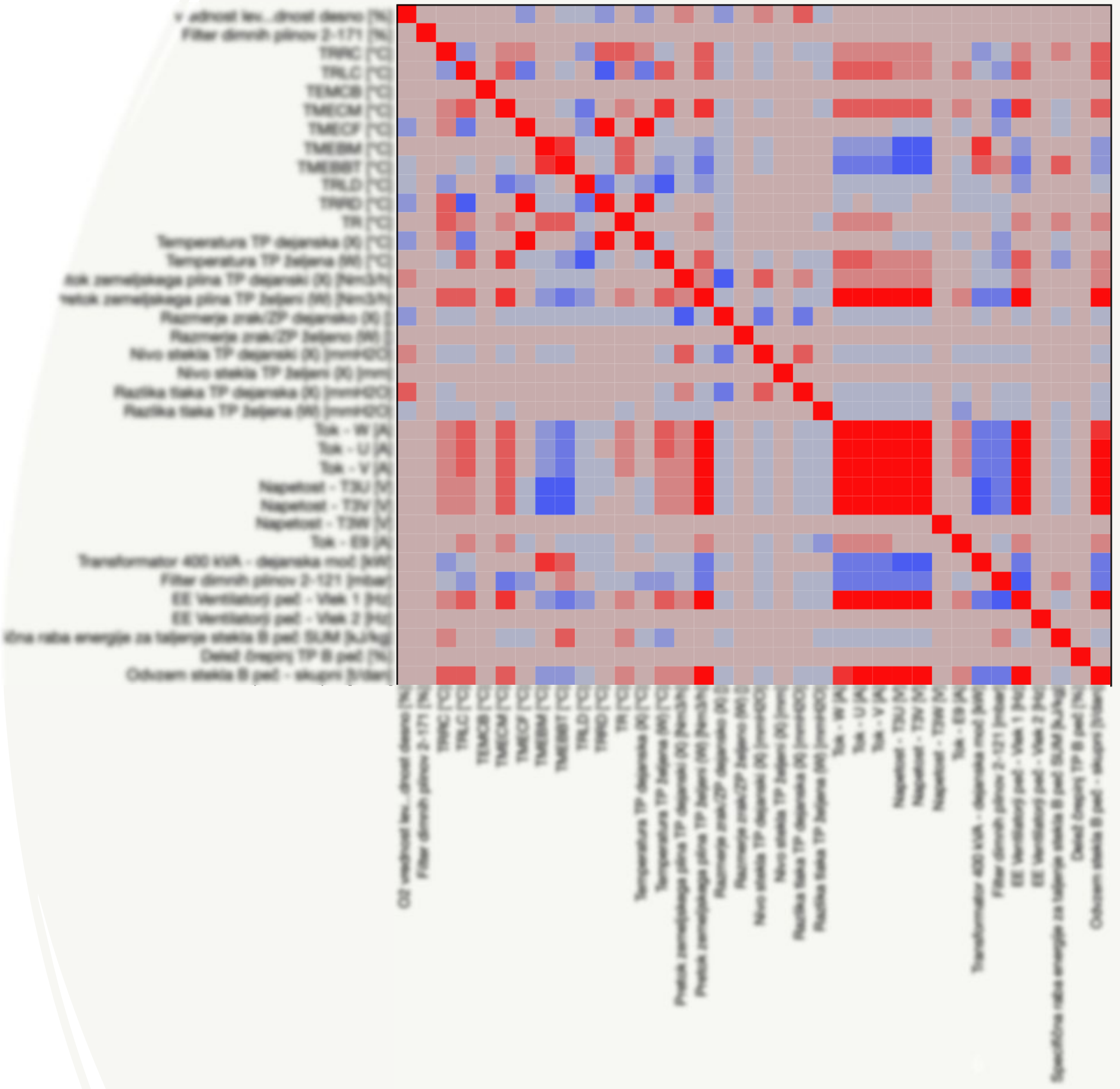
- Extra white flint glass - SeO_2 content vital
- SeO_2 is highly volatile (60-80% loss) and is captured by the filter
- Filter dust is a hazardous waste, disposal is expensive
- Enabling circular production: if SeO_2 concentration is known, the dust can be recycled in the furnace
- **Chemical analyses are expensive and time consuming. Possible solution: predictive modeling**

- The glass manufacturing process is continuous with 1-minute sensor readings
- 40+ on-line sensors measure various temperatures, gas consumption, heater power, glass pull rate, etc.
- **Challenge: predict SeO_2 content in a dust sample based on sparse sampling on different time scale and many parameters**

CULMIUM

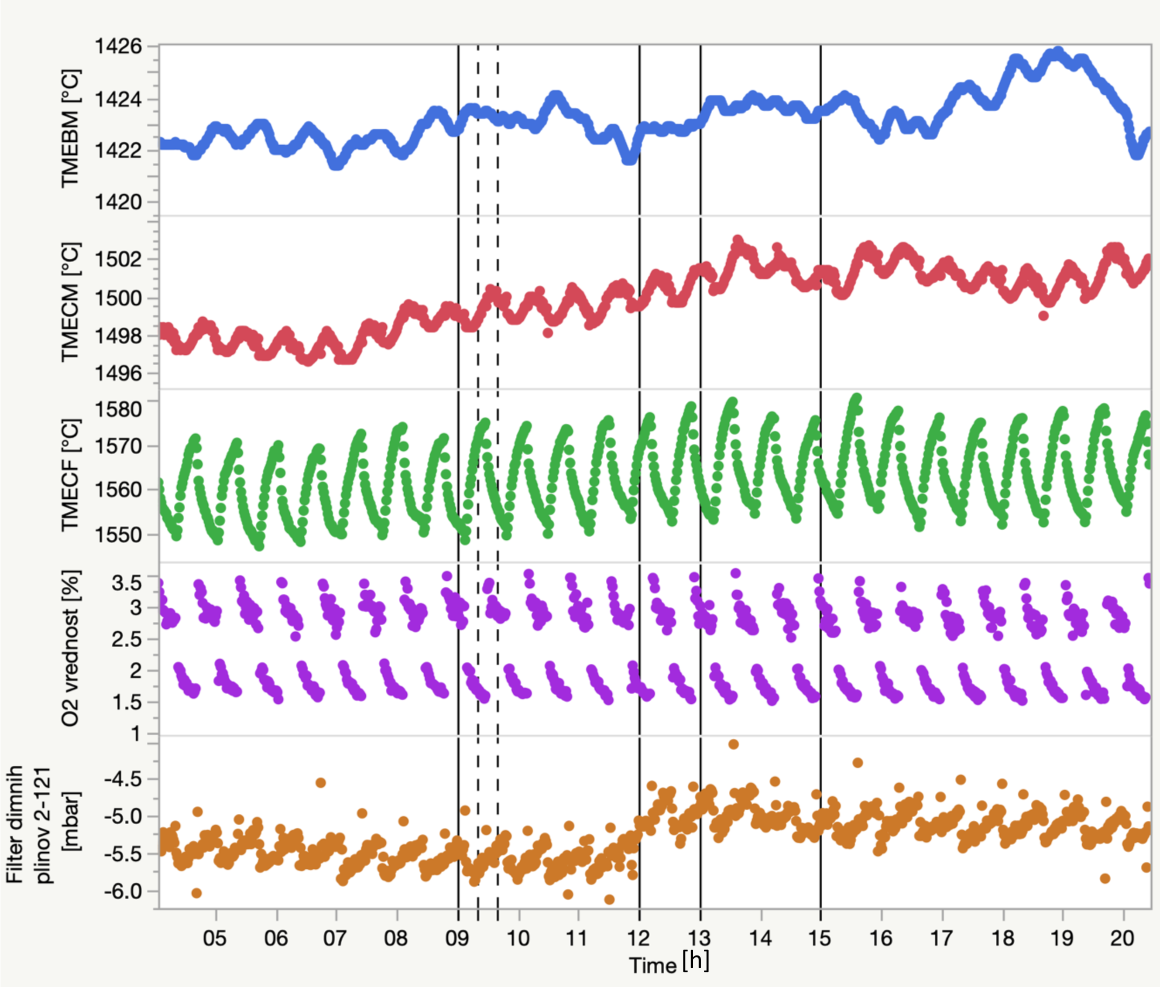
Initial data collection and screening

- Historical data overview (over 1 year)
- Identification of influential parameters with data screening and SME brainstorming
- **Distributions and multivariate analysis** platforms -> determine correlations, cross-correlations, and non-influential parameters (missing, unique, etc.)



Initial data collection and screening

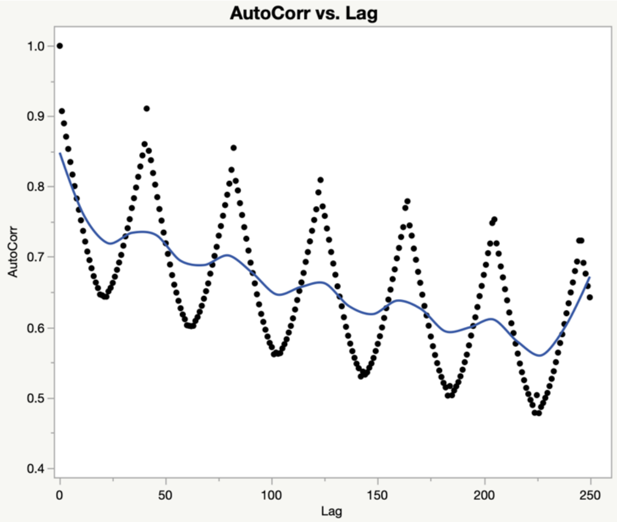
- X-Y graphs to understand behavior
- 20-minute interval between switching burner side (left/right)
- Most of the short-term variability occurs during this 20-minute interval

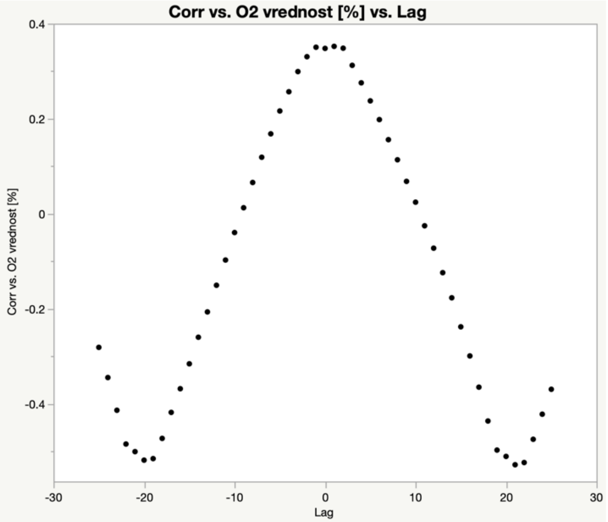


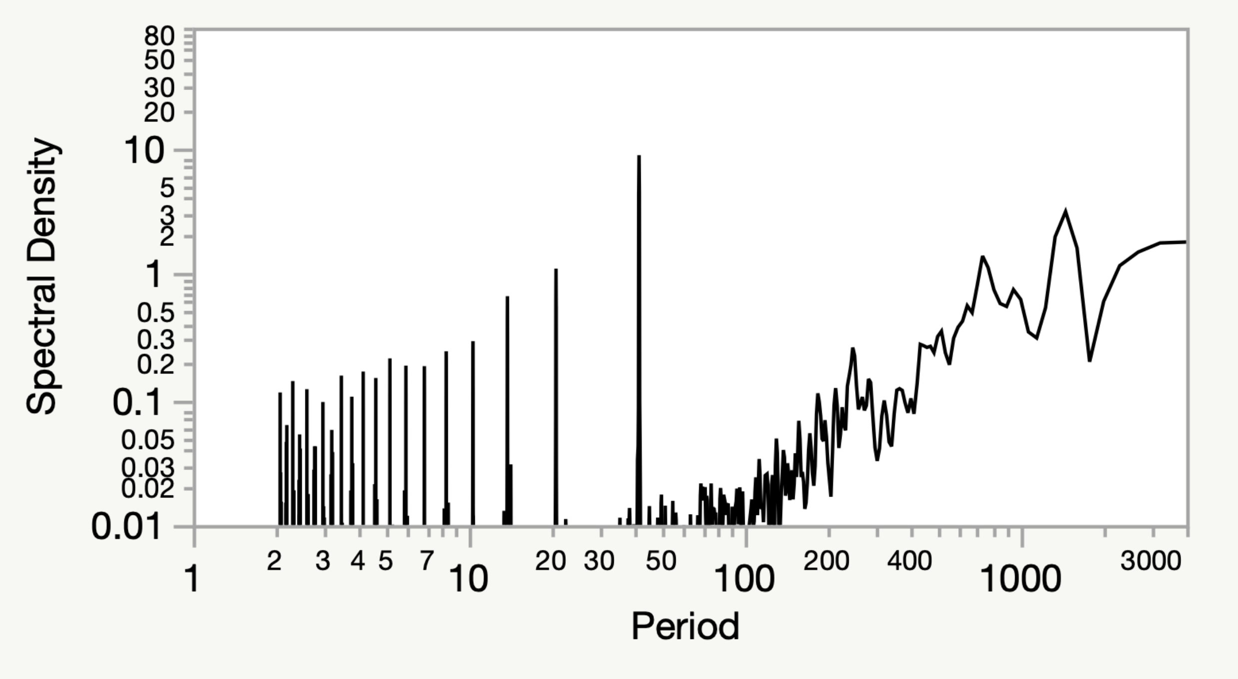
CULMIUM

Auto-correlation and delay using “Time series”

- SeO_2 evaporation is rapid once the raw material touches the hot melt
- Effluent transition time from melt to filter: ~15 s
- Dust extraction time from filter: ~ 1 min

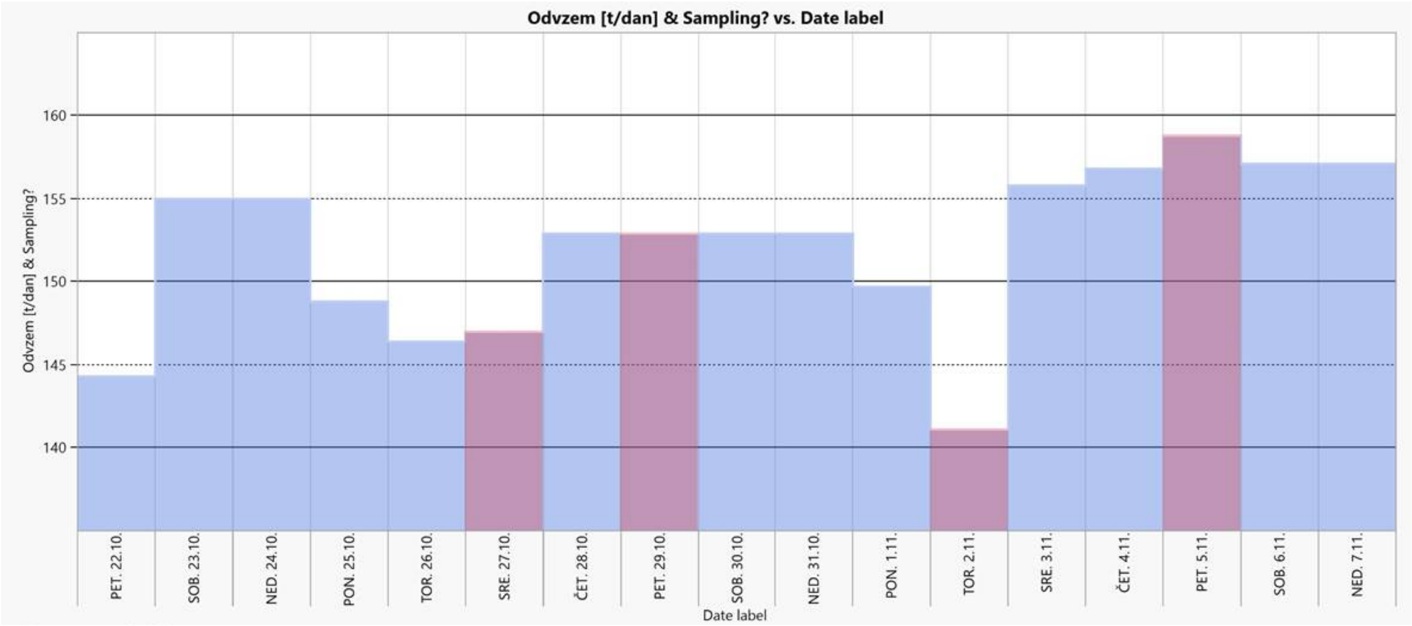






Dust collection plan

- Based on subject-matter expertise and previous tests, one of the most influential parameters for the **SeO_2 content** should be the **glass pull rate**
- Based on the glass manufacturer **monthly pull rate plan**, the optimal **dates for dust collection** were chosen
- Targeting **min, max, and middle pull rate, to optimally cover the parameter space**



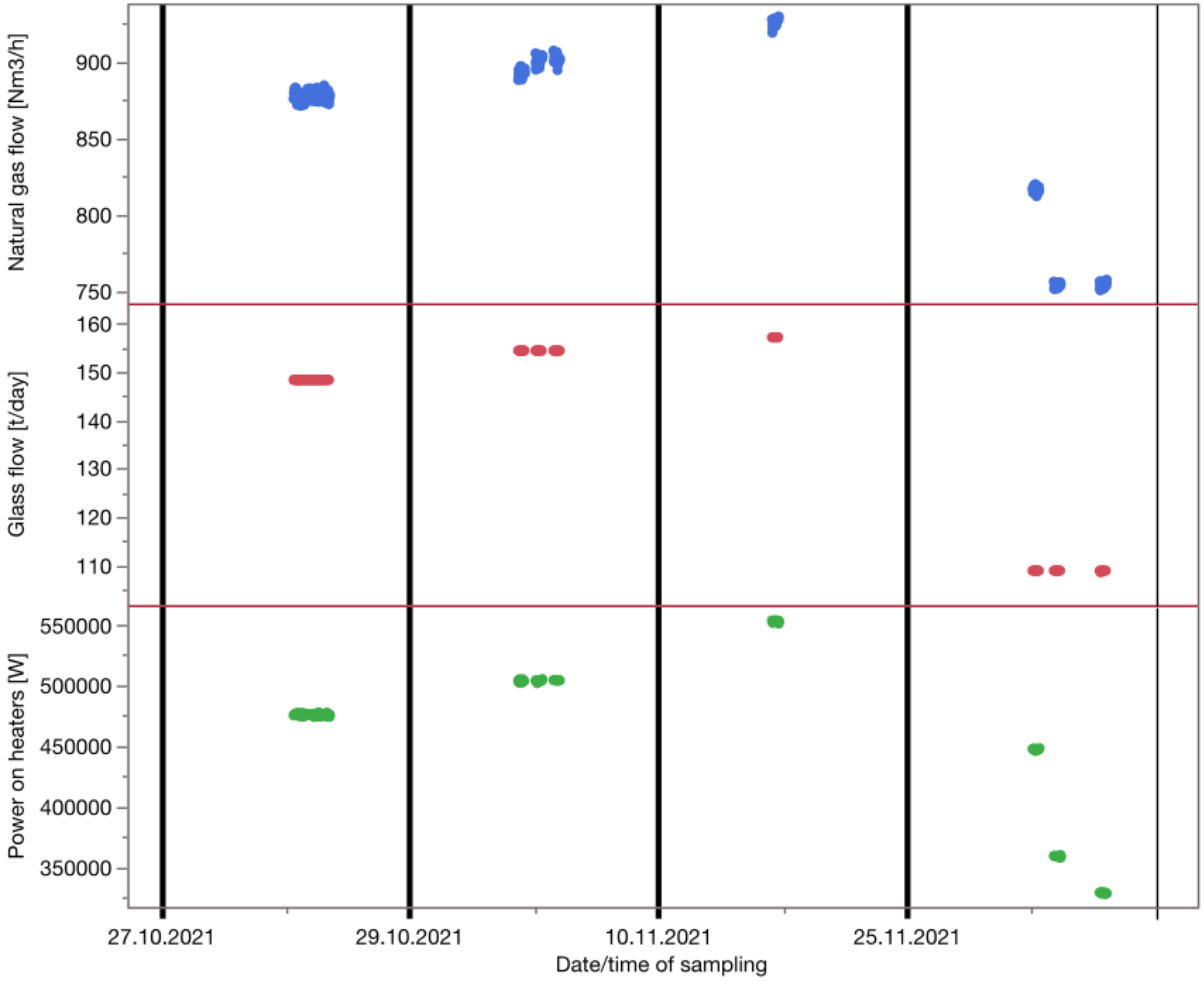
Dust sampling

- Dust is collected and homogenized
- Sent to external institute for ICP-OES analysis -> known SeO_2 content per batch

CULMIUM

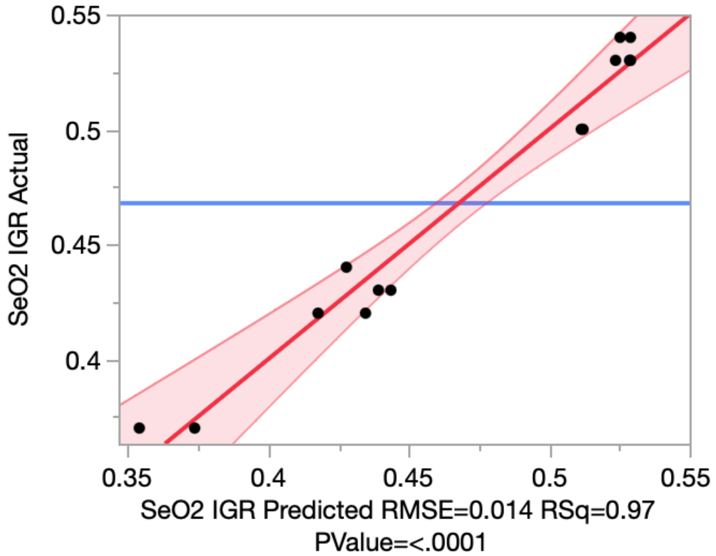
Simple model

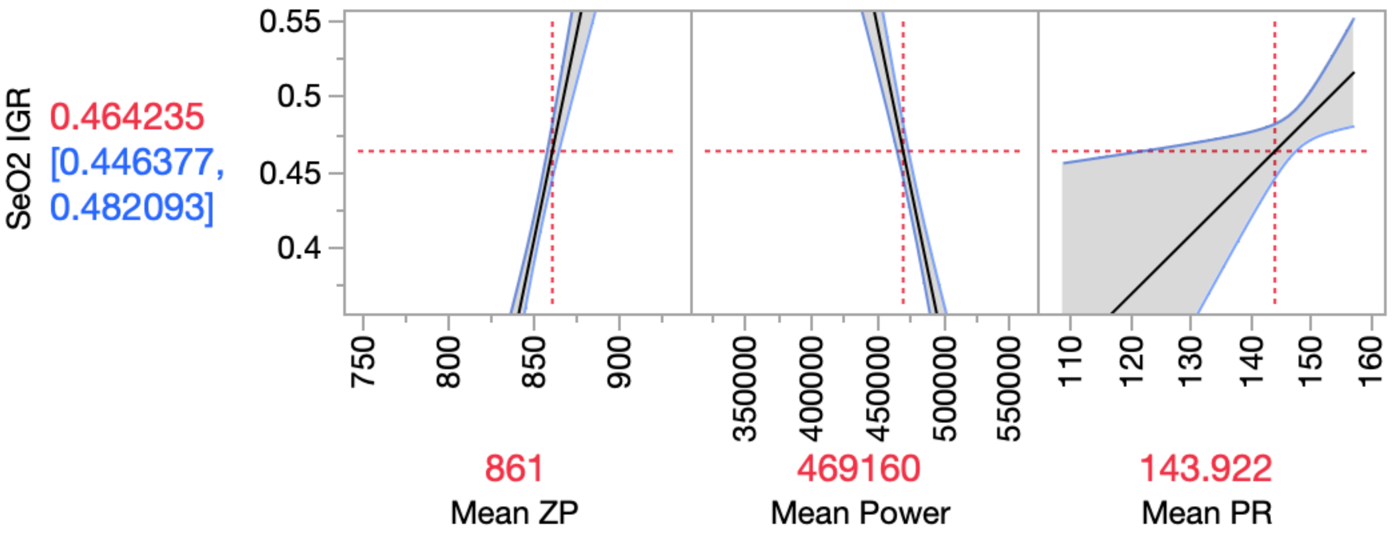
- ICP-OES determined SeO2 content:
  - **9 data points** from 20-minute Culmium sampling (time well known, homogenized dust)
  - **5 data points** from various big-bags sampling (date+time only approximate)
- Using only 3 parameters which are approximately constant during dust sampling (**taking median values**)
  - Natural gas flow used for heating [Nm3/h]
  - Glass pull rate from the glass melting furnace [t/day]
  - Power on electrical heaters [W]
- Using a simple modeling technique (std least squares) with 2nd order interactions



Simple model – std least squares, 3 parameters

Actual by Predicted Plot





Effect Summary

Source	Logworth		PValue
Mean Power	5.114		0.00001
Mean ZP	5.023		0.00001
Mean Power*Mean PR	3.300		0.00050
Mean ZP*Mean PR	3.145		0.00072
Mean PR	1.567		0.02708 ^

Remove Add Edit Undo FDR ('^' denotes effects with containing effects above them)

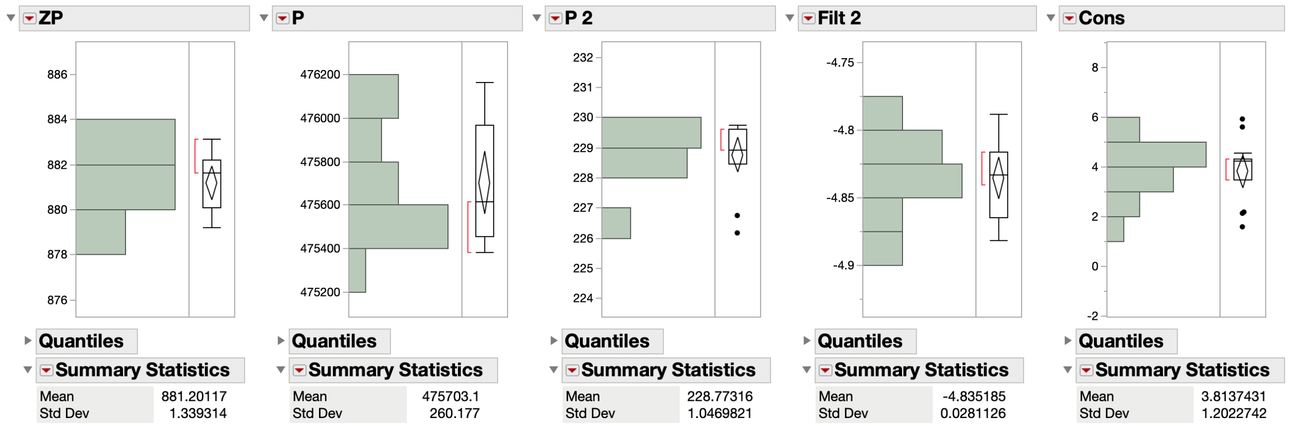
Simple model caveats

- **Simple model does not predict new measurements very well (physically unrealistic predictions)**
- Various other parameters could affect the SeO2 level in the filter dust (e.g., furnace temperatures)
- Prediction profiler indicated large uncertainty at extreme values
- More process parameters need to be considered in the model, but **only 9 measurements** are available

- Could the info on process variation bring additional information to the existing dataset?
- Additional parameters might be influential, but dataset is too scarce, needs to be expanded to be used

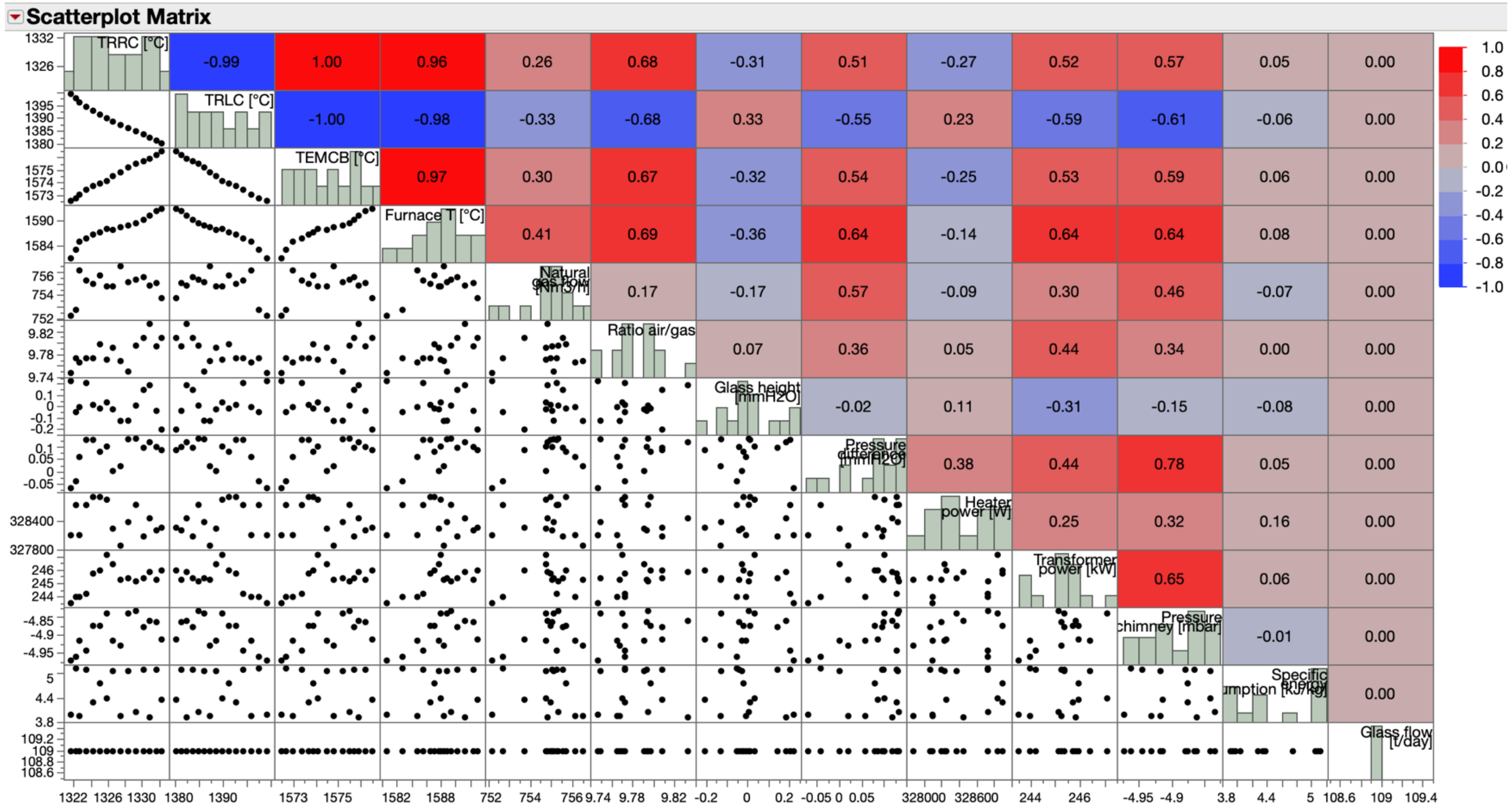
Adding information to the dataset

- SeO2 determination has a measurement uncertainty (+/- 200 ppm @ 95% CI)
- For each 20 minute sampling interval –> 18 individual sensor readings available (excluding 2 minutes during the burner switch), with the underlying distributions:



- Given the distribution means (median) and dispersions, we can randomly pick values from the parent populations, thus generating a synthetic dataset using a **Monte Carlo (MC) approach**
- Together with process experts we consider **13 input parameters,** which are normally distributed during dust sampling periods and have no outliers, no strange artefacts, and no short-term temporal drifts
- **MC simulations must consider parameter correlations via covariance matrices**

MC simulations – covariance matrix



- For each dust sampling, the corresponding covariance matrix is generated from data during dust collection interval
- We exported the covariance matrix and used python (numpy multivariate_normal method) to generate multidimensional normal distributions, taking covariances into account

Synthetic dataset modeling

- Using covariance matrices, 500 synthetic values for each chosen sample with known SeO2 content are generated
- 9 trustworthy samples + 1 additional measurement from big-bag sampling with very low SeO2 content (to extend the parameters space towards low SeO2 values)
- Using a **simple neural network (3 TanH functions, 1 hidden layer)** to model the synthetic dataset

Model Launch

Model NTanH(3)

Training

SeO2 IGR

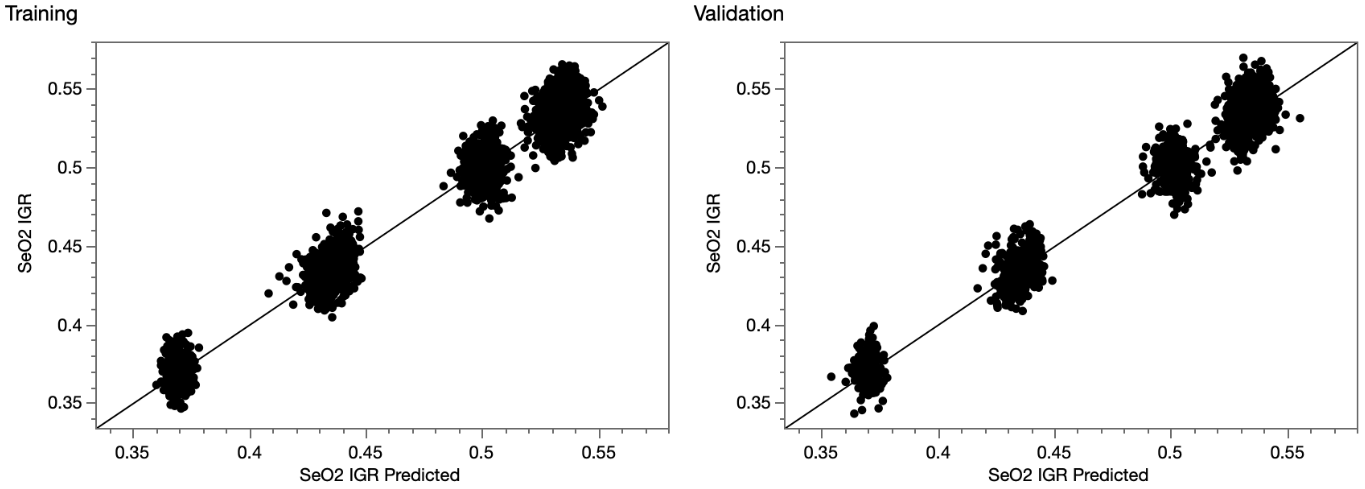
Measures	Value
RSquare	0.9615807
RASE	0.0105279
Mean Abs Dev	0.0084748
-LogLikelihood	-10257.36
SSE	0.369418
Sum Freq	3333

Validation

SeO2 IGR

Measures	Value
RSquare	0.9598819
RASE	0.0109285
Mean Abs Dev	0.0087097
-LogLikelihood	-5084.641
SSE	0.1990947
Sum Freq	1667

Actual by Predicted Plot



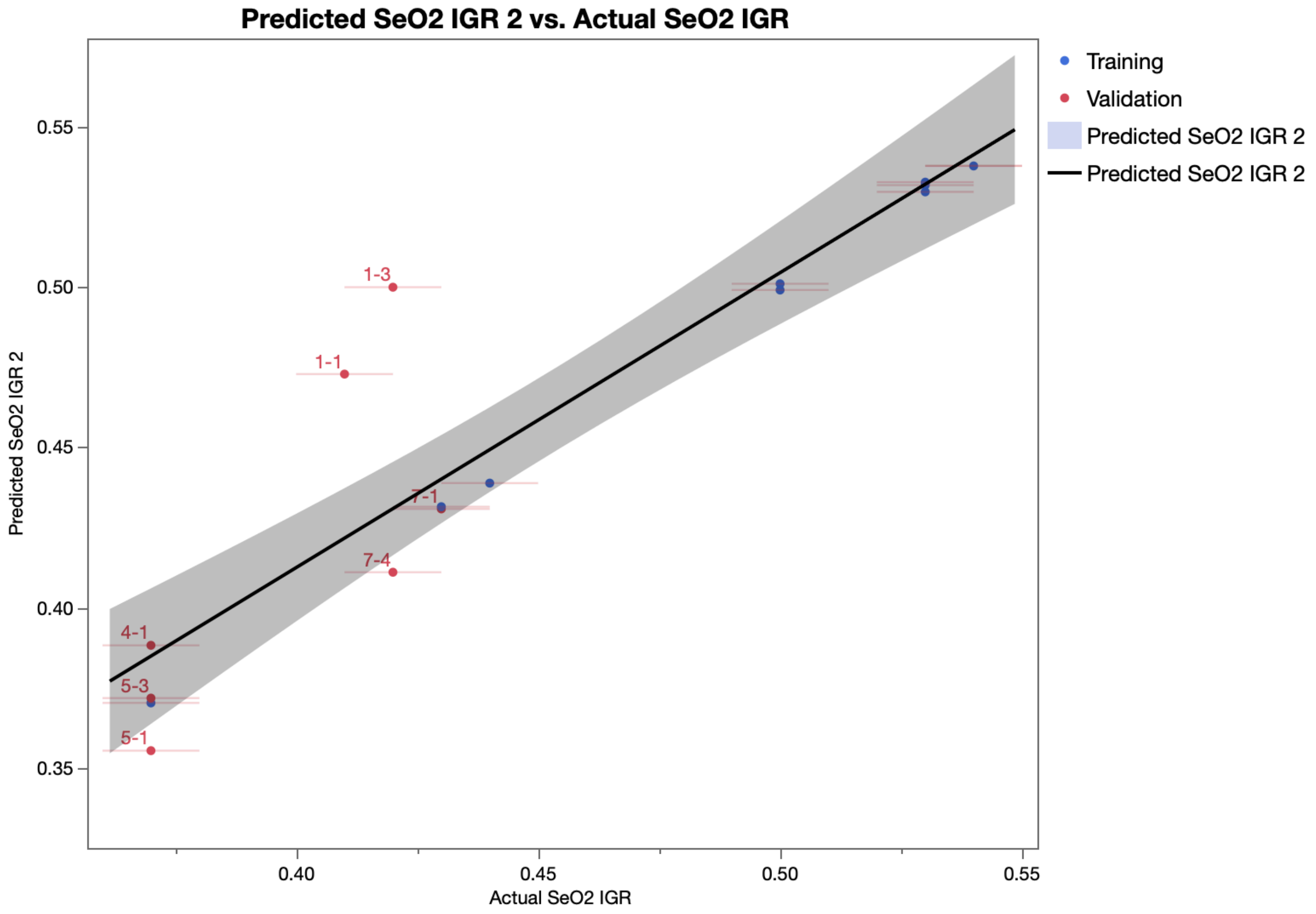
Synthetic dataset modeling

For prediction validation, additional measurements of SeO2 in dust samples from big-bag sampling were used

22/0	Tag	Validation	Meritev	Vreca - vzorec	Mean(Date+Time)	Median(PR)	Actual SeO2 IGR	Predicted SeO2 IGR	Predicted SeO2 IGR (no covariance)
	Training Validation	1 0	1 1-1 1-3 2 3 12 others	1-1 1-3 4-1 5-1 5-3 2 others	23 05 2022 27 10 2021	157 109	0.54 0.37	0.54 0.37	0.78 0.095
1	Training	0	1		27 10 2021 12:50	148.31712	0.54	0.539655148	0.4460315177
2	Training	0	2		27 10 2021 15:34	148.31712	0.53	0.5308072358	0.7615582288
3	Training	0	3		27 10 2021 15:54	148.31712	0.53	0.5300398664	0.7751513624
4	Training	0	4		29 10 2021 13:58	154.33919	0.5	0.5002547579	0.5693055114
5	Training	0	5		29 10 2021 14:18	154.33919	0.5	0.499798521	0.5807976746
6	Training	0	6		10 11 2021 11:04	157.0896	0.53	0.5314731767	0.476936639
7	Training	0	7		10 11 2021 11:25	157.0896	0.54	0.5407084003	0.4195303983
8	Training	0	8		25 11 2021 18:43	108.97919	0.44	0.4395317694	0.4957211272
9	Training	0	9		25 11 2021 19:03	108.97919	0.43	0.4298490467	0.5060467358
10	Validat…	1	1-3	1-3	12 05 2022 20:16	144.24191	0.42	0.4999983473	0.5725768166
11	Training	0	4-4		15 05 2022 03:11	144.06912	0.37	0.3685427824	0.5009021572
12	Validat…	1	5-3	5-3	15 05 2022 22:46	144.06912	0.37	0.3753195089	0.2380037006
13	Validat…	1	6-3	6-3	•		0.38	•	•
14	Validat…	1	7-4	7-4	22 05 2022 21:13	146.8224	0.42	0.3841568466	0.1707579696
15	Validat…	1	1-1	1-1	13 05 2022 04:27	144.24191	0.41	0.4874607197	0.539133005
16	Validat…	1	4-1	4-1	15 05 2022 13:25	144.06912	0.37	0.4135389822	0.0952703526
17	Validat…	1	5-1	5-1	16 05 2022 04:55	144.06912	0.37	0.406577133	0.2341834666
18	Validat…	1	6-1	6-1	•		0.37	•	•
19	Validat…	1	7-1	7-1	23 05 2022 08:27	149.990	0.43	0.4359921694	0.4018419192

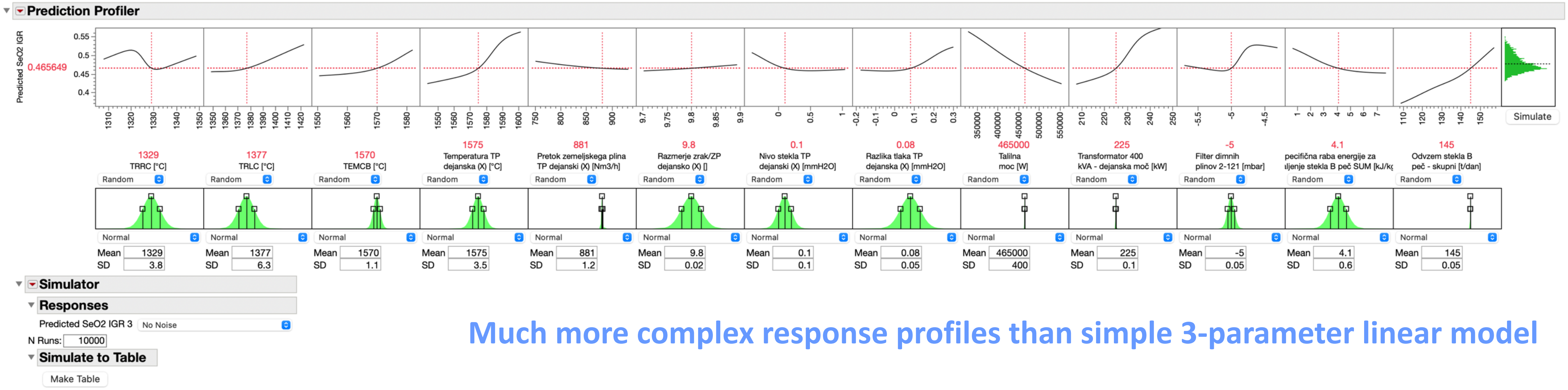
Neural model

- Validation with new sampling
- Only 1 big-bag (nr. 1) is significantly overpredicted
- **Actual by Predicted $R^2 = 0.86$**



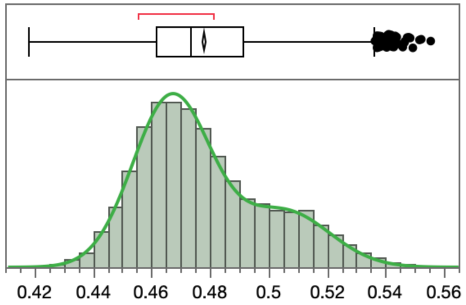
CULMIUM

Model profiler and simulator



Much more complex response profiles than simple 3-parameter linear model

Predicted SeO2 IGR 3



Compare Distributions

Show	Distribution	AICc	BIC	-2*LogLikelihood
☑	Normal 2 Mixture	-49075.95	-49039.9	-49085.95

Simulator for prediction scatter

Quantiles

100.0%	maximum	0.55547561
99.5%		0.5377955
97.5%		0.52695358
90.0%		0.51139431
75.0%	quartile	0.49133397
50.0%	median	0.473391
25.0%	quartile	0.46138594
10.0%		0.452298
2.5%		0.44292856
0.5%		0.43453752
0.0%	minimum	0.41797497

Fitted Normal 2 Mixture Distribution

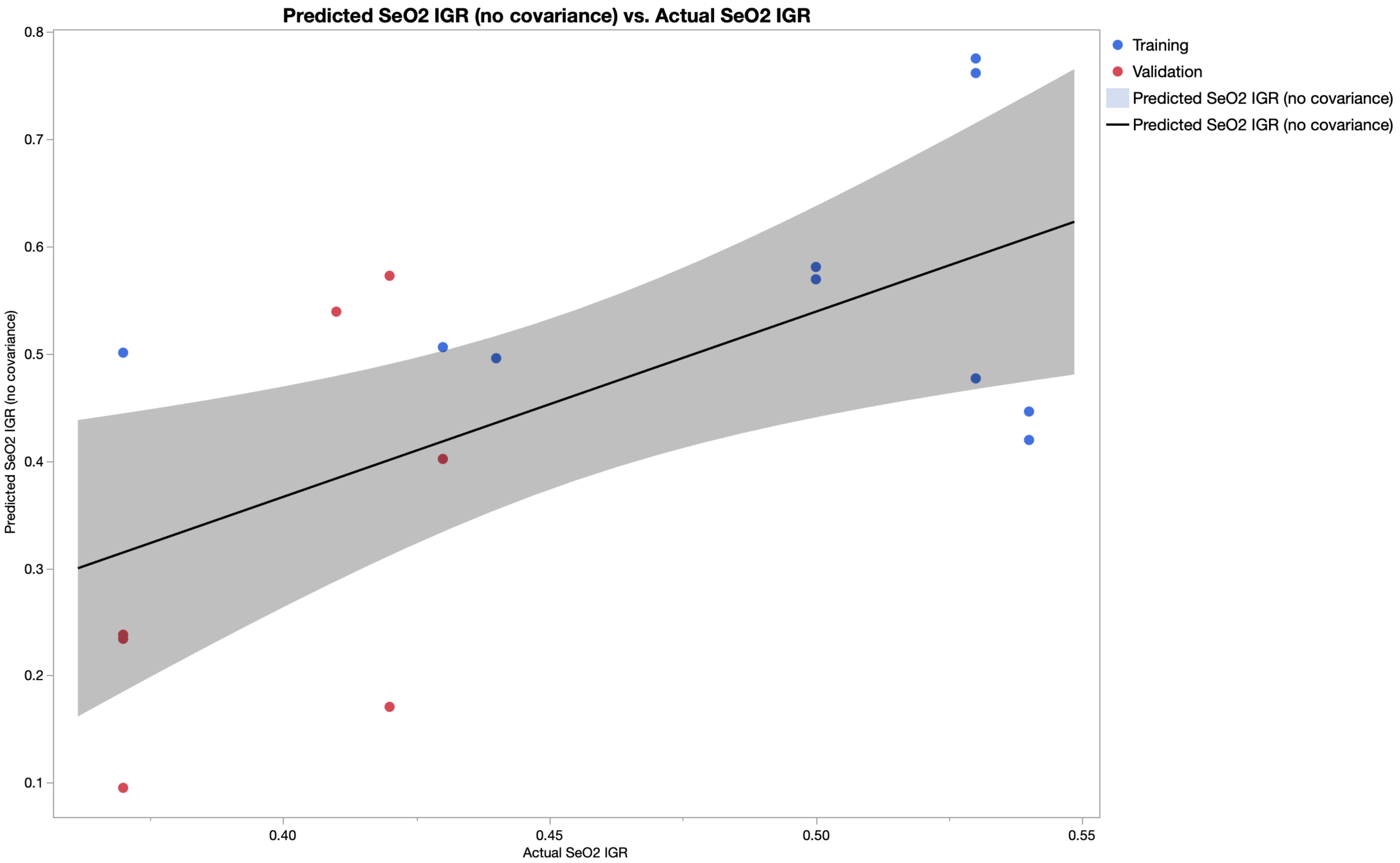
Parameter		Estimate	Lower 95%	Upper 95%
Location	μ1	0.4669502	0.4666426	0.4672577
Location	μ2	0.5054775	0.5048953	0.5060597
Dispersion	σ1	0.0133514	0.0130475	0.0136624
Dispersion	σ2	0.0156087	0.015038	0.0162011
Probability	π1	0.7237826	0.7068018	0.7401416
Probability	π2	0.2762174	0.2660351	0.2866372

Measures

-2*LogLikelihood	-49085.95
AICc	-49075.95
BIC	-49039.9

No correlation/ covariance

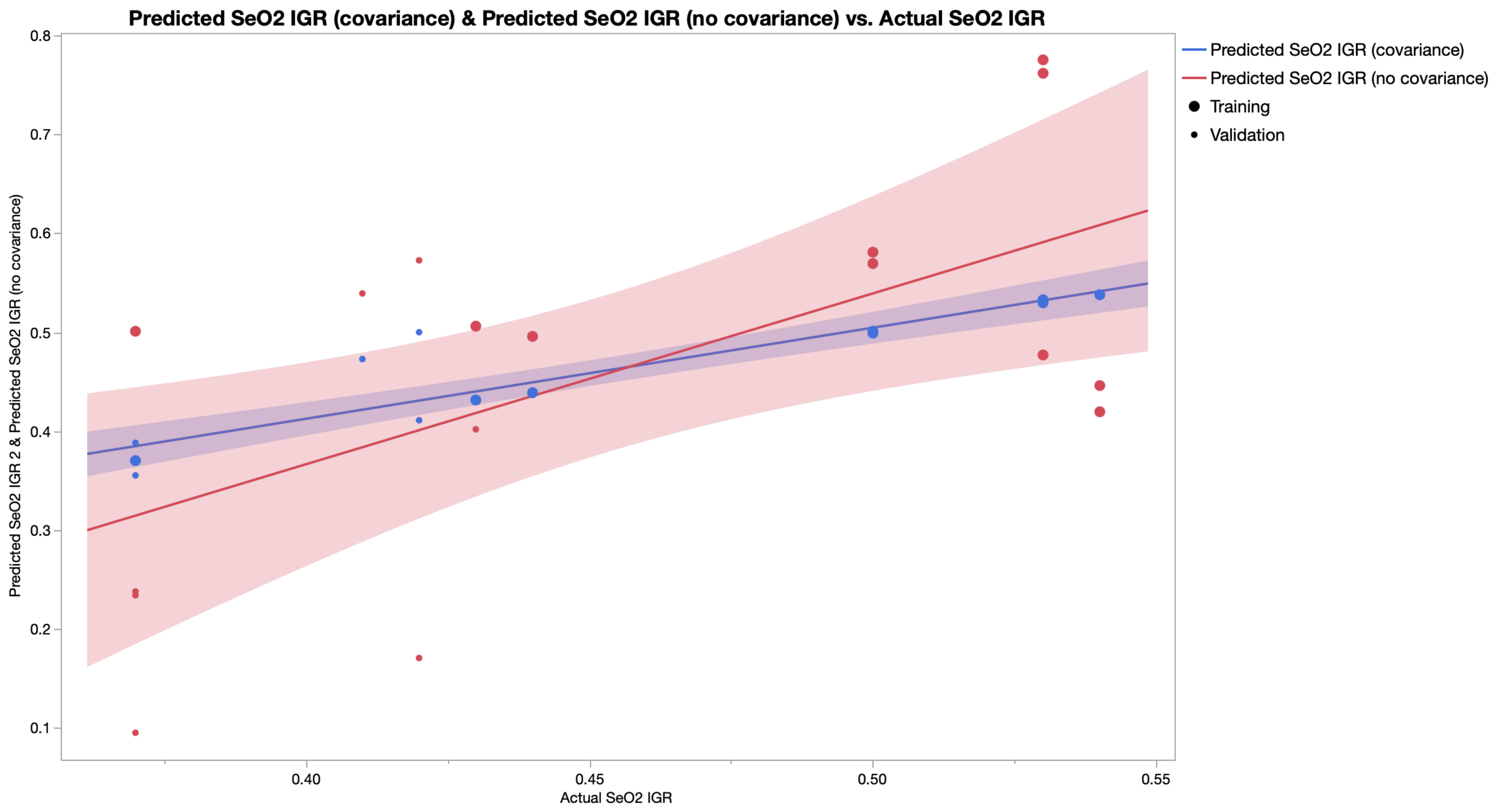
- **What if we do not consider correlation/covariance of params with MC simulations?**
- Model is inappropriate, with non-physical predictions
- The span of predicted SeO2 values is unrealistic
- **Actual by Predicted $R^2 = 0.37$**



CULMIUM

No correlation/ covariance

- **Comparison of “Actual by predicted” plots**
- Taking covariances into account shows much more realistic predictions for SeO_2



CULMIUM

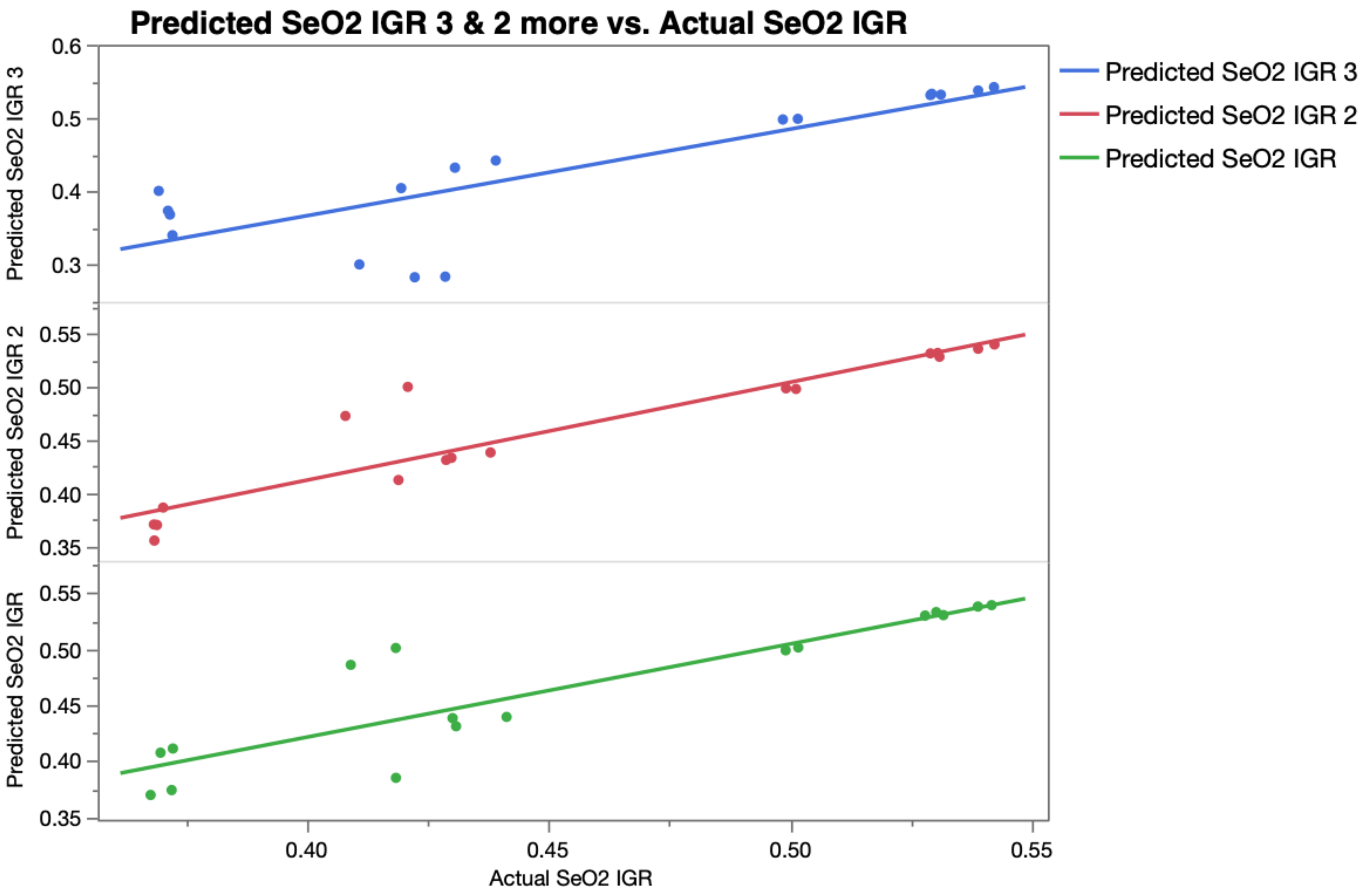
Conclusion

- Despite sparse dust sampling and small dataset, we derived a useful predictive model for SeO_2 concentration in filter dust
- Monte Carlo approach required, considering the complexity of the process
- Predictive power of the model was tested new dust sampling and showed realistic, practically useful predictions
- Approach led to a 100% return of the waste filter dust in a pilot attempt, thereby replacing 60% of an expensive primary raw material for glass decolorization, and completely removing the cost of filter dust disposal

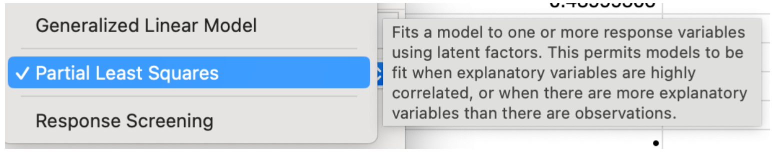
Appendix

Neural model

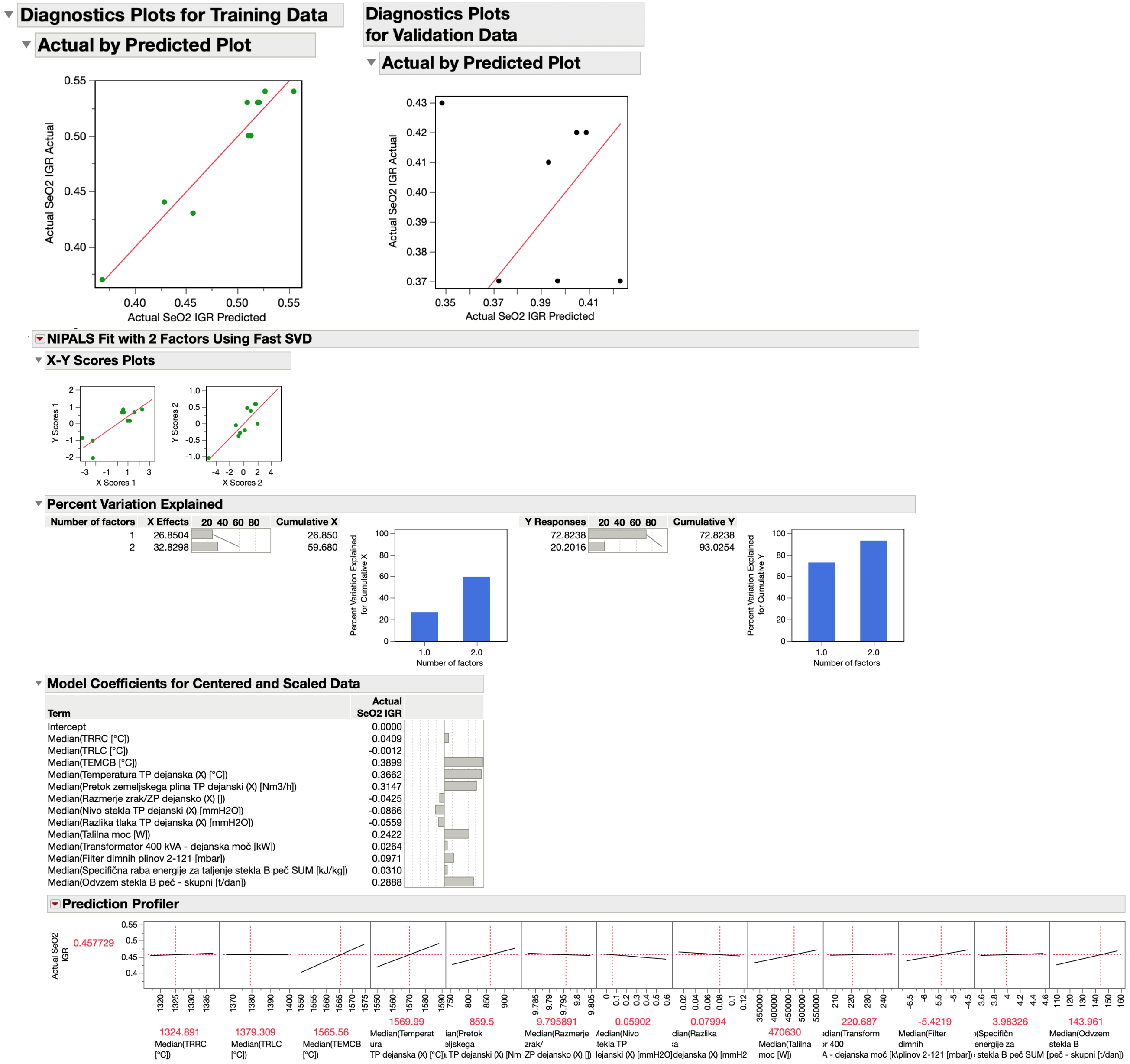
- Prediction is consistent even when taking different random seeds
- Big-bag nr. 1 is slightly over- or underpredicted in most cases



Partial Least Squares (vs neural network)



- Testing Partial Least Squares on original (non-simulated) dataset (10 data points for training, 13 params):
- **Partial Least Squares model shows worse predictive power**
- Actual by Predicted plot for validation data is “poor”
- **Profiler shows somehow similar behaviour and influence of parameters, but only in linear regime**



Number of factors	X Effects	Cumulative X	Y Responses	Cumulative Y
1	26.8504	26.850	72.8238	72.8238
2	32.8298	59.680	20.2016	93.0254

Term	Actual SeO2 IGR
Intercept	0.0000
Median(TRRC [°C])	0.0409
Median(TRLC [°C])	-0.0012
Median(TEMCB [°C])	0.3899
Median(Temperatura TP dejanska (X) [°C])	0.3662
Median(Pretok zemeljskega plina TP dejanski (X) [Nm3/h])	0.3147
Median(Razmerje zrak/ZP dejansko (X) [])	-0.0425
Median(Nivo stekla TP dejanski (X) [mmH2O])	-0.0866
Median(Razlika tlaka TP dejanska (X) [mmH2O])	-0.0559
Median(Talilna moc [W])	0.2422
Median(Transformator 400 kVA - dejanska moč [kW])	0.0264
Median(Filter dimnih plinov 2-121 [mbar])	0.0971
Median(Specifična raba energije za taljenje stekla B peč SUM [kJ/kg])	0.0310
Median(Odvzem stekla B peč - skupni [t/dan])	0.2888