Wind Experiment

Thomas Wollerdorfer, Daniel Panagger, Melissa Schmidt
Report to presentation in statistical projects on 27.06.2018

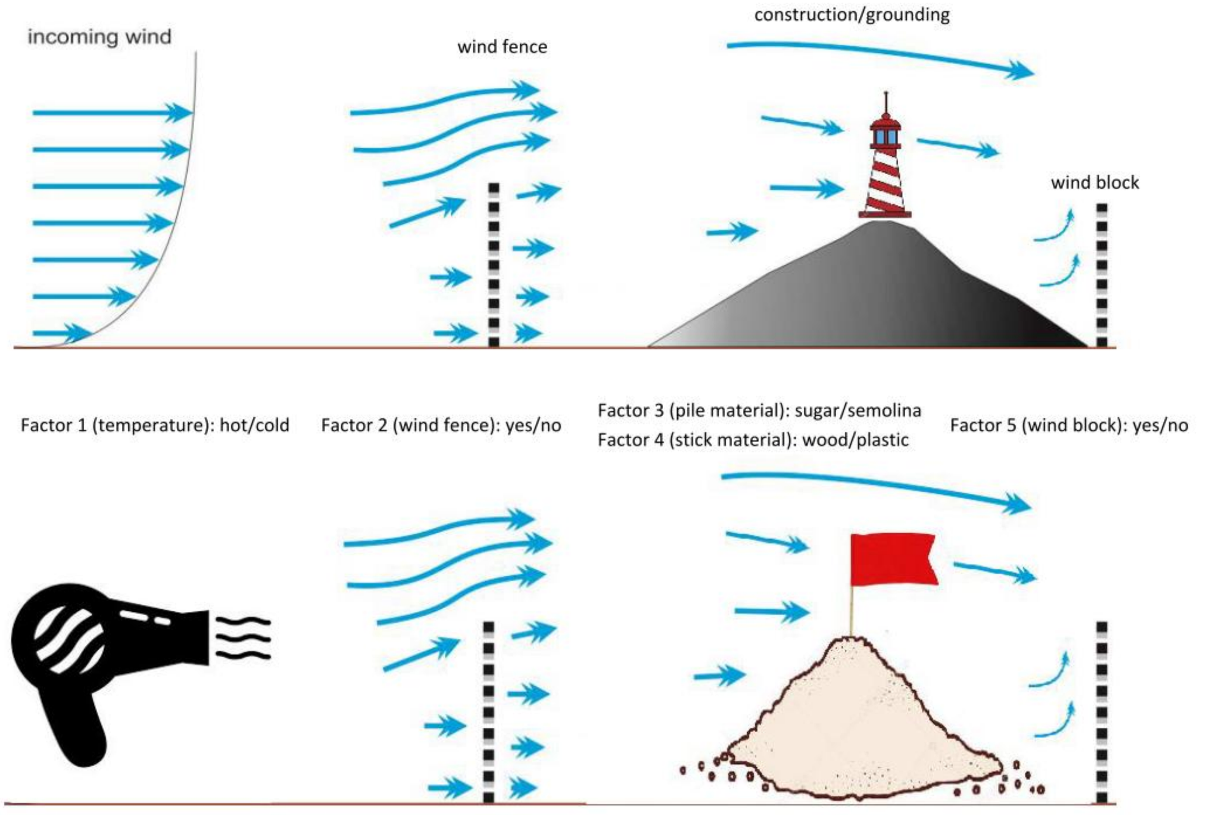
Initial idea

The basic idea of the experiment is to evaluate the lifetime of certain constructions in a wind exposed environment.

In the real world, constructors have to decide to choose or avoid different types of locations. The quality of a location depends on the natural setting, like sediment conditions and additional disturbance factors like a wind block on the slipstream side. The climate and especially the temperature could also have influence on the stability of the buildings.

The constructors can decide about the construction material itself and they have the possibility to build a wind fence for additional protection. Both factors are costly and therefore the possible effect should be significant.

Such an experiment would be very difficult to accomplish. The financial afford is tremendous and it would take years or even decades to get results.



Therefore we modelled a minimized experiment, which could be easy handled and replicated. For the simulation of the factor cold and hot wind we used a hairdryer. A pile of sugar or semolina simulated the hill. The sticks of wood and plastic are supposed to symbolize the buildings. A small fence and a wind block were constructed and could easily put in position or be removed.

The experimental setup should have as few external influence factors as possible. So this setup is done indoor, the hairdryer is fixed, the position of the pile, the fence and the barrier should always be precisely the same.

Design evaluation:

Factor	Level -1	Level 1
Pile material	Sugar	Semolina
Fence	No	Yes
Construction type	Wooden	Plastic
Block	No	Yes
Temperature	Cold	Hot

The experiment has five factors with two levels each. This would lead to a full factorial with two to the power of five runs (32 runs). It would allow us to estimate all main effects and all interactions.

Even in the minimalized model, 32 runs are too expensive and time consuming.

If we assume that the highest order of interactions is two, we could use a fractional factorial with 16 runs. (Morris, 2011, Chapter 13.2)

Furthermore, we have the factor pile material, which is quite messy to change. Therefore, the pile material is a hard to change factor, which allows us no complete randomization.

The final design we used was the split plot design. It allows us the mixture of hard and easy to change factors. Therefore, no complete randomization is necessary and it is easier and quicker to conduct. The disadvantage is a trade off against statistical power for the hard to change factor. A split plot design has at least two experimental units. In this experiment, the whole plot units are the blocks and the split plot units are the runs. The whole plots need at least one replication otherwise an estimation of the hard to change factor would not be possible, i.e. we had to prepare four blocks instead of two.

The notation of the model consists of an alpha for the whole plot factor coefficient and a beta for the sub plot factors coefficient. There are two error terms in the model, one that corresponds to the whole-plot unit and one that corresponds to the split plot unit. Both error terms are normal distributed and independent. The presence of two error terms causes the parameters to have different variances. (Morris, 2011, Chapter 10.2.)

$$\square_{\square\square\square} = \dot{\square}_{\square} + \dot{\square}_{\square} + (\square\dot{\square})_{\square\square} + \square_{\square\square(\square)} + \square_{\square\square(\square)\square}$$

The Design

The Design has 4 blocks and 21 runs. Factor 1 is the hard to change factor. Coincidentally, this model matrix looks like we had to change the pile material only once. But even if the levels are the same, the factor settings were reset between groups. The first two blocks, therefore the runs from 1-10, are the one with semolina, the second two blocks, run 11-21, are the one with sugar. In the table there are all possibilities of factors visible.

Group	Run	Factor 1 a:Pile	Factor 2 B:Fence	Factor 3 C:Construction	Factor 4 D:Block	Factor 5 E:Temperature	Response 1 Duration Seconds
1	1	Semolina	No	Plastic	Yes	Cold	4.0
1	2	Semolina	No	Wooden	Yes	Hot	3.7
1	3	Semolina	No	Plastic	No	Hot	3.7
1	4	Semolina	Yes	Wooden	No	Cold	6.9
1	5	Semolina	Yes	Plastic	Yes	Hot	5.8
2	6	Semolina	No	Wooden	No	Cold	3.2
2	7	Semolina	Yes	Plastic	No	Cold	7.0
2	8	Semolina	No	Plastic	Yes	Hot	3.8
2	9	Semolina	Yes	Wooden	Yes	Cold	6.5
2	10	Semolina	Yes	Wooden	No	Hot	6.2
3	11	Sugar	No	Plastic	No	Hot	4.5
3	12	Sugar	Yes	Plastic	Yes	Cold	23.5
3	13	Sugar	Yes	Wooden	No	Cold	15.6
3	14	Sugar	Yes	Wooden	Yes	Hot	23.0
3	15	Sugar	No	Wooden	Yes	Cold	5.0
4	16	Sugar	Yes	Plastic	No	Hot	12.6
4	17	Sugar	No	Wooden	No	Hot	4.8
4	18	Sugar	Yes	Wooden	Yes	Cold	17.3
4	19	Sugar	No	Plastic	No	Cold	5.7
4	20	Sugar	No	Plastic	Yes	Hot	6.0
4	21	Sugar	No	Wooden	No	Cold	4.8

The second table is the design matrix. All factors have two levels, -1 and 1. In the last column the duration of the stick until it falls, is listed.

Group	Run	Factor 1 a:Pile	Factor 2 B:Fence	Factor 3 C:Construction	Factor 4 D:Block	Factor 5 E:Temperature	Response 1 Duration Seconds
1	1	{ 1 }	{ -1 }	{ 1 }	{ 1 }	{ -1 }	4.0
1	2	{ 1 }	{ -1 }	{ -1 }	{ 1 }	{ 1 }	3.7
1	3	{ 1 }	{ -1 }	{ 1 }	{ -1 }	{ 1 }	3.7
1	4	{ 1 }	{ 1 }	{ -1 }	{ -1 }	{ -1 }	6.9
1	5	{ 1 }	{ 1 }	{ 1 }	{ 1 }	{ 1 }	5.8
2	6	{ 1 }	{ -1 }	{ -1 }	{ -1 }	{ -1 }	3.2
2	7	{ 1 }	{ 1 }	{ 1 }	{ -1 }	{ -1 }	7.0
2	8	{ 1 }	{ -1 }	{ 1 }	{ 1 }	{ 1 }	3.8
2	9	{ 1 }	{ 1 }	{ -1 }	{ 1 }	{ -1 }	6.5
2	10	{ 1 }	{ 1 }	{ -1 }	{ -1 }	{ 1 }	6.2
3	11	{ -1 }	{ -1 }	{ 1 }	{ -1 }	{ 1 }	4.5
3	12	{ -1 }	{ 1 }	{ 1 }	{ 1 }	{ -1 }	23.5
3	13	{ -1 }	{ 1 }	{ -1 }	{ -1 }	{ -1 }	15.6
3	14	{ -1 }	{ 1 }	{ -1 }	{ 1 }	{ 1 }	23.0
3	15	{ -1 }	{ -1 }	{ -1 }	{ 1 }	{ -1 }	5.0
4	16	{ -1 }	{ 1 }	{ 1 }	{ -1 }	{ 1 }	12.6
4	17	{ -1 }	{ -1 }	{ -1 }	{ -1 }	{ 1 }	4.8
4	18	{ -1 }	{ 1 }	{ -1 }	{ 1 }	{ -1 }	17.3
4	19	{ -1 }	{ -1 }	{ 1 }	{ -1 }	{ -1 }	5.7
4	20	{ -1 }	{ -1 }	{ 1 }	{ 1 }	{ 1 }	6.0
4	21	{ -1 }	{ -1 }	{ -1 }	{ -1 }	{ -1 }	4.8

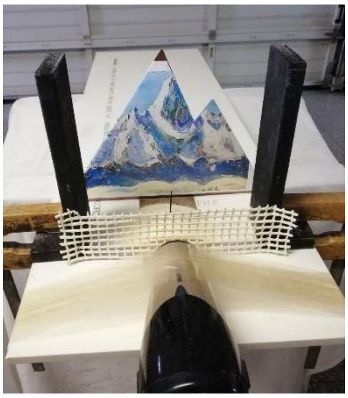
Some examples:

Just pile

The pile material is semolina (+1), there is no fence (-1) or block (-1). The stick is made from wood (-1) and the temperature of the hairdryer is cold (-1).

Block, but no fence

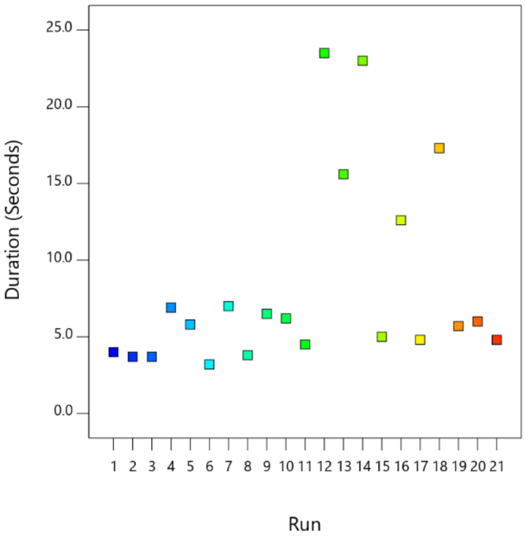
Again there is semolina (+1), no fence (-1), a cold temperature (-1) of the wind, the stick is wooden (-1) but now there is a block (+1) behind the pile



Block and fence

The last thing which changes in those three examples is the fence in the front of the pile.

Summary of the runs



The first 10 runs took a duration of 3- 8 seconds to blow the stick away. From run 11 to 21 there are lots of differences of the duration. The maximum duration is 23.5 at run 12 and the minimum duration is 3.2 at run 6. The graph shows that the duration of sugar is more spread than semolina.

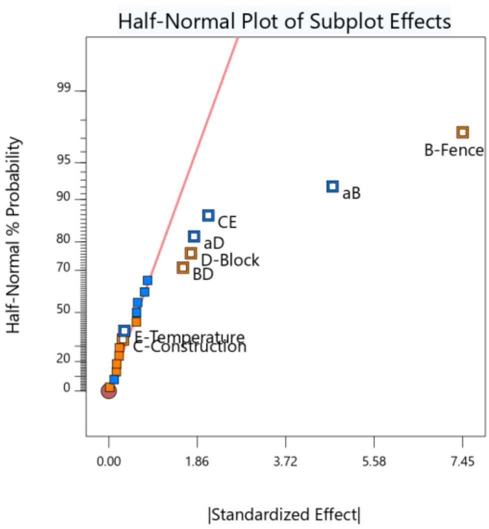
Model selection

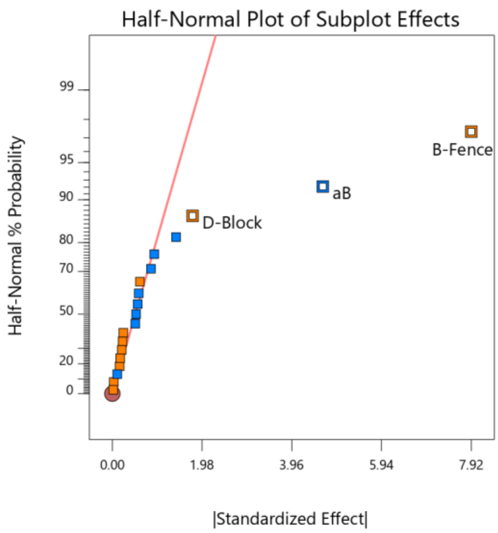
The model is done with a backward selection and a significance level of α = 0.05. This leads to a model with 10 parameters. 5 main effects and 4 interactions. One of the main effects is the hard to change factor. The confirmation is through a half normal plot. The intercept is 8.71 and the whole plot term and 2 other sub plot main effects are significant. The subplot term C (construction) and the subplot term E (temperature) are not significant, but their interaction is so. So as we conduct a model with hierarchy we need those insignificant effects to put the interaction into the model. (Morris, 2011, Chapter 11.5.3)

Source	Coefficient Estimate	Standard Error	p-value	95 % Confidence Interval	
Intercept	8.71	0.4268		6.87	10.54
Whole-plot Terms:			0.0177		
a-Pile	-3.27	0.4268	0.0177	-5.10	-1.43
Subplot Terms:			< 0.0001		
B-Fence	3.72	0.3390	< 0.0001	2.99	4.46
C-Construction	0.1489	0.3328	0.6650	-0.57	0.87
D-Block	0.8563	0.3374	0.0307	0.13	1.59
E-Temperature	-0.1636	0.3328	0.6346	-0.88	0.56
aB	-2.33	0.3374	< 0.0001	-3.06	-1.60
aD	-0.8976	0.3390	0.0258	-1.63	-0.16
BD	0.7678	0.3271	0.0433	0.06	1.47
CE	-1.03	0.3271	0.0116	-1.74	-0.32

Interpretation and plots

With a backward selection the program “Design Expert” suggests a model with the 5 main effects and 4 interactions. The highest influence of the easy to change factors, has the fence with a coefficient of 3.72. The combination of the whole plot effect and the fence has a high negative influence on the duration. In the plot it is visible that the effects of temperature and construction do not have much influence, but their interaction do have much influence. The $R^2$is 0.9685 and the AIC of the model is 94.43.



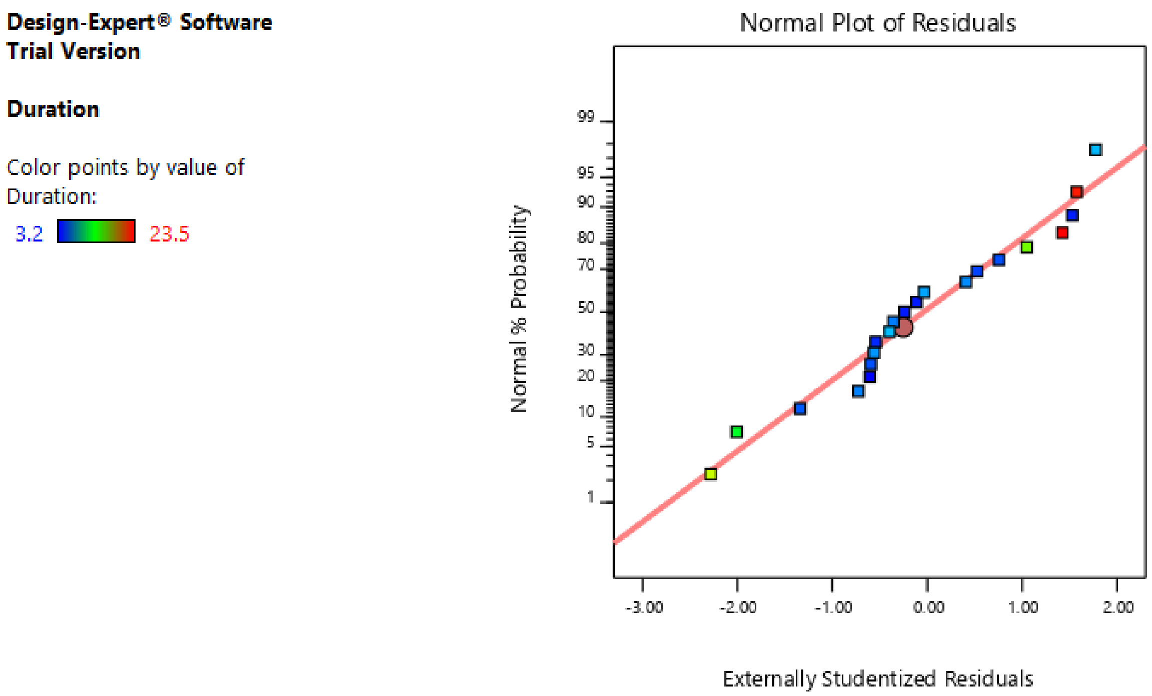


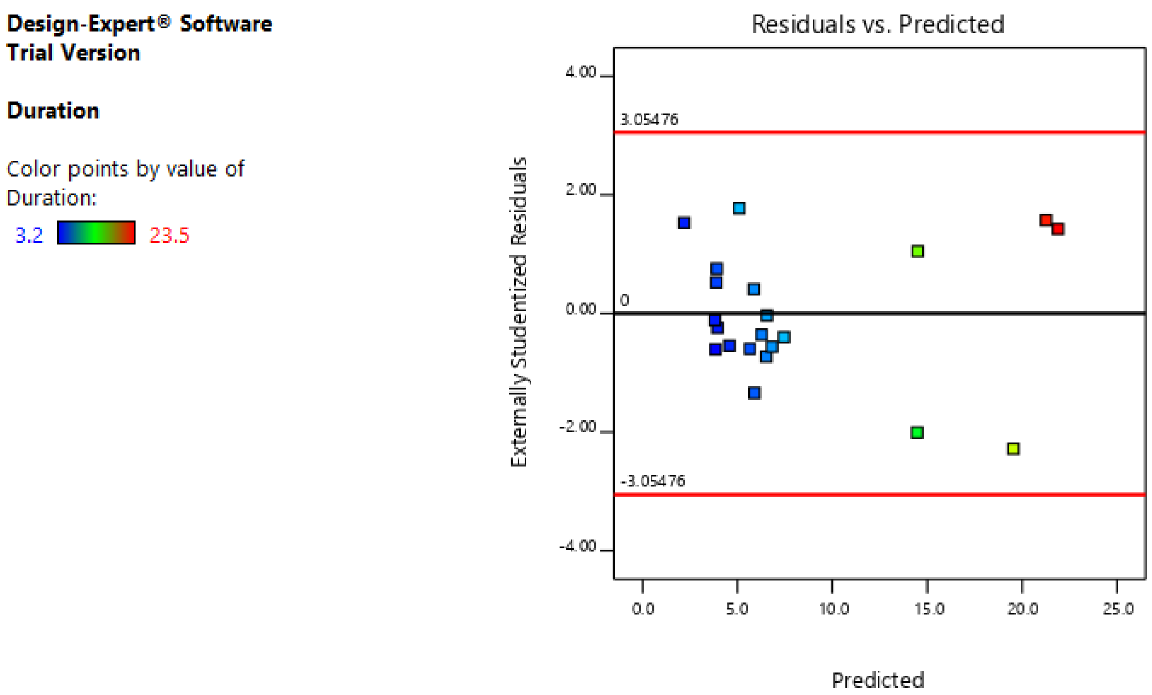
Another possible model is, to only get those effect in the model which are significant without hierarchy. Then there are only the fence, the block and the interaction between the pile and the fence in the model. The R^2 goes down to 0.9037 but the AIC is better, with a result of 99.46.

For further analysis we took the first model, because of the number of effects in the model and the included interactions.

Residual analysis

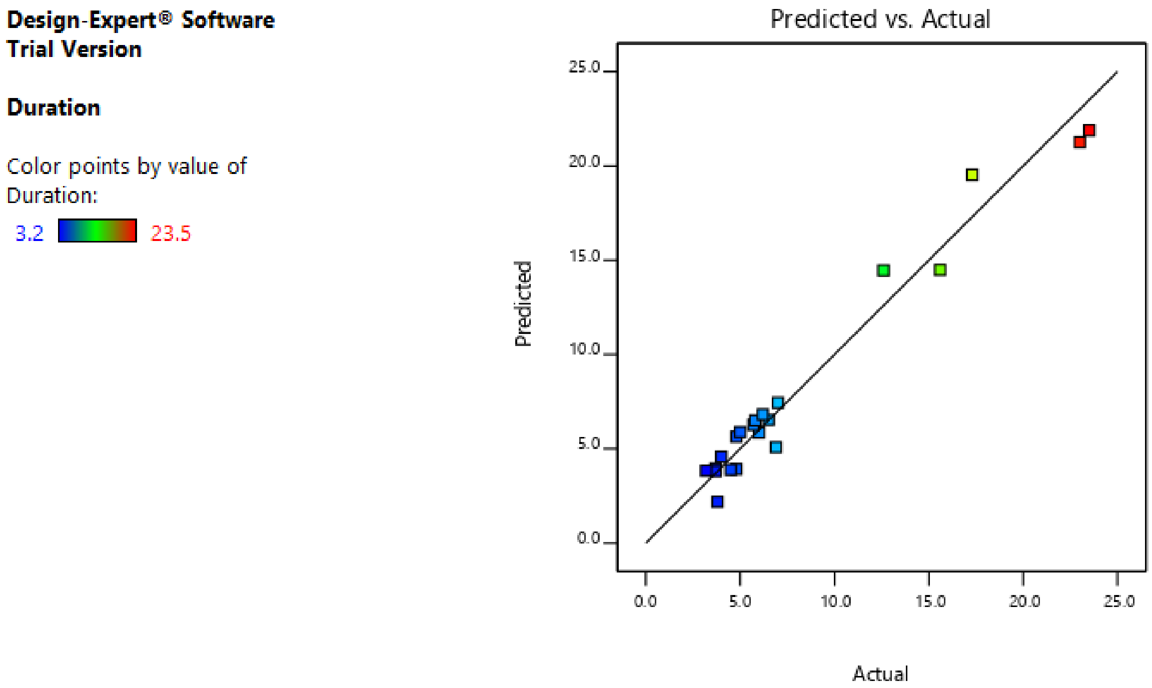
To verify whether the model produces homoskedastic errors, a residual analysis was conducted. The two following plots show that the model is well designed and that there are no real abnormalities regarding the residuals. Nevertheless one should mention that especially observations with high values of duration also lead to bigger residuals. This is not really surprising and altogether one can say that the model is well designed and a linear model is suitable to model the duration.





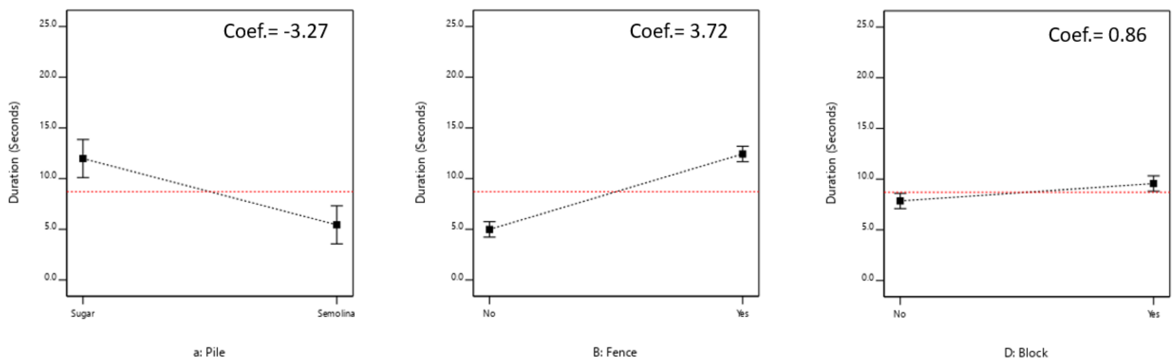
Predicted vs. Actual

The plot shows the real data plotted against the predicted data. The blue dots are mainly the measurements with semolina. The highest values of the real data are overfitted with that model, but it is also visible that the other high points are above the predicted line and the model seems to fit pretty good the real data.



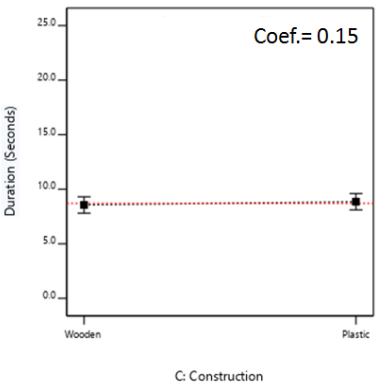
Significant main effects

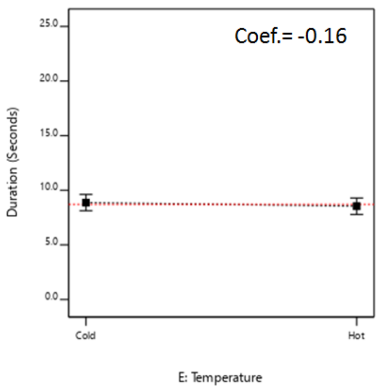
There are three significant main effects. The whole plot factor and two subplot factors. The whole plot factor pile has a negative effect on semolina and a huge confidence interval compared to the other ones. The duration of blowing the stick away lasts significantly longer, if there is a fence in front of the pile. Another small but significant effect is, if there is a block behind the pile. The coefficient is 0.86 and means, that if there is a block behind the pile, the duration will last longer than with no block behind.



Insignificant main effects

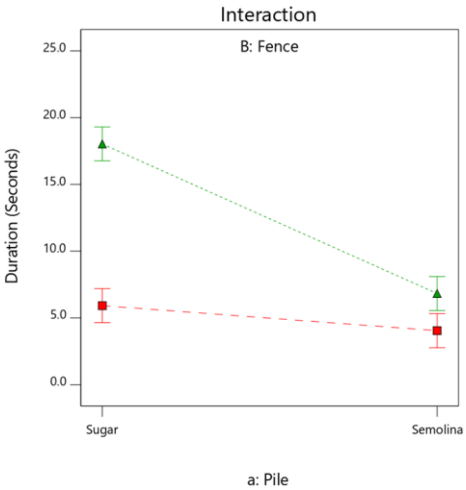
The model also contains two insignificant main effects, the construction and the temperature. Both coefficients are very low with 0.15 and -0.16. So we conclude that there is no difference for the duration, if you use a wooden or plastic sticks and if there is hot or cold wind.





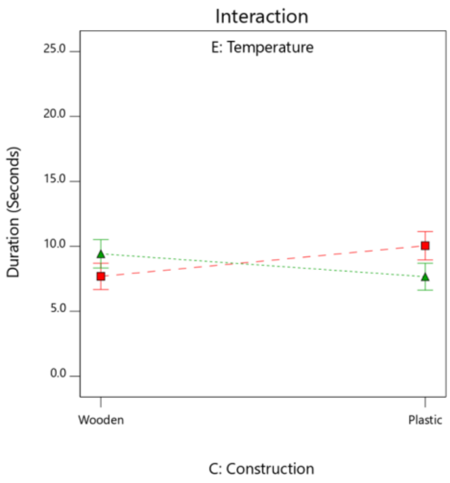
Interactions

We would like to look a little bit closer at two interactions in the model. First the interaction between the pile and the usage of a fence.



The red squares represent the setting “no fence” whereas the green triangles represent the setting “fence”. The plot shows that the usage of a fence leads to an increase in seconds for both types of piles. However, a fence is much more effective in combination with a pile made of sugar. For sugar piles the duration nearly triples when using a fence.

The second interaction we would like to look closer, is the combination between the construction and the temperature. As we could see earlier, both main effects are not significant. Nevertheless, the interaction between those two effects is significant.



The red squares represent the setting “cold wind” whereas the green squares represent the setting “hot wind”. This means that hot wind has a negative effect on the duration for plastic sticks and a positive effect for wooden sticks. This result may seem a little bit odd, but there is a kind of a contrariwise relationship between the construction and the temperature of the hairdryer, although the size of the effect is quite small.

Maximize duration

To maximize the duration, the following levels of the different main effects have to be used:

- Pile: Sugar
- Fence: Yes
- Construction: Plastic
- Block: Yes
- Temperature: Hot

For this setting the estimated duration is 21.9 seconds. The 95% confidence interval has a lower bound of 19.4 and an upper bound of 24.4 seconds. One run in our design exactly had this combination of treatments and led to a time of 23.5 seconds, which falls in our estimated confidence interval. With this combination the duration in our experiment was maximized.

References

Max D. Morris. *Design of Experiment - An Introduction Based on Linear Models.* 2011