Procter&Gamble

Beating complexity in automated method qualification via Tailored Split-*(Split)*-Plot design with JMP Pro

JMP Europe Discovery Summit

March 14th -17th, 2016

Amsterdam

Authors:

Ph.D. Zhiwu Liang

Pablo Moreno Peláez

Davy Van den Bosch

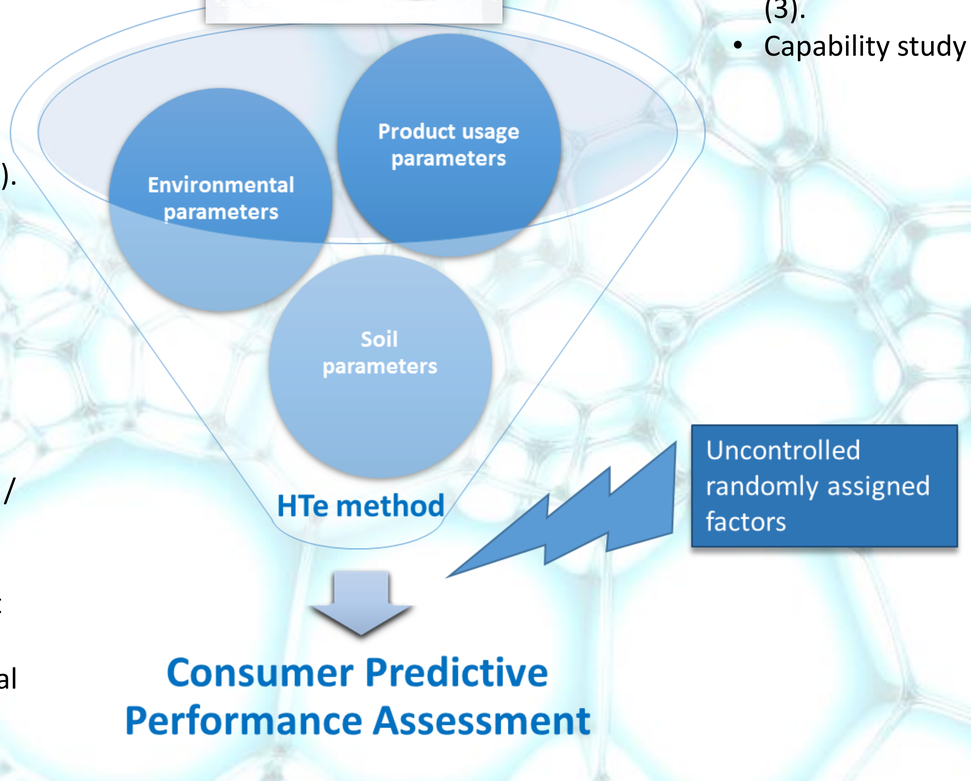
Context

High-Throughput Test Method

- Automated method mimicking consumer behavior & predicting consumer acceptance.
- Throughput ~600 samples/week vs. max. 20 samples/week for a human operator.
- Controlled variables:
  - Environmental parameters (2);
  - Product usage parameters (4);
  - Soil parameters (2).
- Uncontrolled randomly assigned parameters (3).

Method used for

- Technology development: from early development stage technologies to market qualification.
- Product optimization: fine-tune product performance based on geographical conditions / consumer habits / cost optimization
- Competitive benchmarking: across all geographies where we're present in the market (WE, EMEA, NA, LA, AAIK)
- Claim support: competitive advantage numerical exploitation



Validation of automated high-throughput test method

- Typical validation consists of machine/instrument variation + **operator-operator variability** + **day-to-day variation** + method variability → due to automation no human bias.
- Number of factors in the method = 8 (many) + random factors (3).
- Capability study rather than minimum validation.

What do we want to get out of this study?

- Validation allows to right-size & exploit the potential of our method.
- Internal QA validation level confirmed:
  - Screening;
  - Technical;
  - Consumer-predictive.

The Design Challenge

Complexity in design

- Many factors in system need to be investigated
- Some factors has too many levels to be designed balance.
- Some factors needed to be handled as “Discrete numeric” to be able to get to the level of detailed answer that we were interested in.
- Balancing across all levels of continuous factors was a key.
- There is at least one uncontrolled randomly assigned factor “sink” that we also need to account for in the final model, but cannot balance in the design due to the robot setup.

Factors

Add Factor | Remove | Add N Factors 1

Name	Role	Changes	Values
Product_F2	Continuous	Easy	0.1, 0.2
Soil_F1	Categorical	Easy	Type E, Type Z
UCR_F1	Continuous	Hard	-1, 1
Product_F1	Discrete Numeric	Easy	1, 2, 3, 4, 5
UCR_F2	Categorical	Easy	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17
Environmental_F1	Discrete Numeric	Easy	1, 7.5, 15
Environmental_F2	Discrete Numeric	Easy	17, 26

Model

Main Effects | Interactions | RSM | Cross | Powers | Remove Term

Name	Estimability
Intercept	Necessary
Product_F2	Necessary
Soil_F1	Necessary
Product_F1	Necessary
Product_F1*Product_F1	If Possible
Product_F1*Product_F1*Product_F1	If Possible
Product_F1*Product_F1*Product_F1*Product_F1	If Possible
Environmental_F1	Necessary
Environmental_F1*Environmental_F1	If Possible
Environmental_F2	Necessary
Product_F2*Product_F2	Necessary
Product_F2*Soil_F1	Necessary
Product_F2*Product_F1	Necessary
Soil_F1*Product_F1	Necessary
Product_F2*Environmental_F1	Necessary
Soil_F1*Environmental_F1	Necessary
Product_F1*Environmental_F1	Necessary
Product_F2*Environmental_F2	Necessary
Soil_F1*Environmental_F2	Necessary
Product_F1*Environmental_F2	Necessary
Environmental_F1*Environmental_F2	Necessary

… And even more complexity ahead

- Complex robotic setup with certain restrictions that need to be reflected in design.
- Terms used for balancing the design were not the same ones as needed and used for the later models.
- Different models for Type E and Type Z were identified.
- One design calculation in JMP took about 20 minutes on a 64-bit i7 machine (!).
- Capability study → accurate model prediction needed.

How to execute?

- We need to run 408 experiments but only can do 102 test runs in a day.
- We have 17 tubes, each of which can hold a sample amount that can be used in a maximum of 6 runs → 102 = 17 * 6
- Tube balancing via Subplots (Split-Split-Plot design) did not give actionable balanced designs.
- To achieve balance design for 5 levels factor we had to add higher order Model Terms that we were not interested in to get a balanced design.
- Based on pilot data we calculated that we need to run a 4-day study to achieve the power we wanted! This is the first split plot.

Design Generation

Number of Whole Plots: 4

Number of Runs:

- Minimum 17
- Default 24
- User Specified 408

Make Design

File Edit Tables Rows Cols DOE Analyze PG Scripts Graph Tools Add-Ins View Window Help

WASP DoE for post...
Design: Custom Design
Criterion: I Optimal
Model
DOE Dialog

Columns (10/0)
Whole Plots
Product_F2
Soil_F1
UCR_F1
Product_F1
UCR_F2
Environmental_F1
Environmental_F2
Sink
Y

Rows
All rows 408
Selected 0
Excluded 0
Hidden 0
Labelled 0

	Whole Plots	Product_F2	Soil_F1	UCR_F1	Product_F1	UCR_F2	Environmental_F1	Environmental_F2	Sink	Y
1	1	0.2	Type E	0.161208822	1	9	1	17	1	45.3
2	1	0.1	Type Z	0.161208822	5	11	1	17	2	10
3	1	0.1	Type E	0.161208822	1	16	15	17	3	32
4	1	0.1	Type Z	0.161208822	4	9	15	26	4	22.6
5	1	0.2	Type Z	0.161208822	5	11	1	26	2	11
6	1	0.1	Type Z	0.161208822	5	2	15	26	2	20.4
7	1	0.1	Type Z	0.161208822	3	13	7.5	26	4	31.8
8	1	0.15	Type Z	0.161208822	2	16	7.5	17	3	57.8
9	1	0.15	Type E	0.161208822	3	4	7.5	26	1	17.5
10	1	0.1	Type E	0.161208822	2	2	15	26	2	11.2
11	1	0.2	Type Z	0.161208822	5	7	15	26	2	5.5
12	1	0.2	Type Z	0.161208822	3	15	15	17	1	26.5
13	1	0.2	Type Z	0.161208822	2	8	7.5	26	4	33.2
14	1	0.1	Type Z	0.161208822	3	8	15	17	2	22.8
15	1	0.15	Type Z	0.161208822	1	14	1	26	3	19.6
16	1	0.2	Type E	0.161208822	1	2	1	26	1	12
17	1	0.15	Type Z	0.161208822	2	5	1	26	2	12.6
18	1	0.2	Type Z	0.161208822	1	6	15	26	4	17.7
19	1	0.2	Type Z	0.161208822	5	3	7.5	26	1	21.4
20	1	0.15	Type Z	0.161208822	2	17	7.5	17	2	21.3
21	1	0.15	Type E	0.161208822	2	14	7.5	17	3	6.1
22	1	0.15	Type E	0.161208822	2	3	7.5	17	4	15.2
23	1	0.15	Type Z	0.161208822	2	10	7.5	17	3	22.1
24	1	0.2	Type Z	0.161208822	2	15	15	26	1	20.7
25	1	0.15	Type E	0.161208822	4	7	1	17	2	31
26	1	0.2	Type E	0.161208822	1	1	15	17	4	10.3
27	1	0.2	Type E	0.161208822	4	14	15	26	3	23.3
28	1	0.2	Type Z	0.161208822	1	8	1	17	4	11.3
29	1	0.15	Type E	0.161208822	4	8	7.5	17	1	8.5
30	1	0.15	Type E	0.161208822	4	1	1	26	4	55.9
31	1	0.1	Type Z	0.161208822	1	3	1	17	1	21
32	1	0.1	Type Z	0.161208822	5	17	1	17	2	21.2
33	1	0.15	Type Z	0.161208822	4	17	7.5	26	3	30.2
34	1	0.15	Type E	0.161208822	2	12	7.5	17	1	14
35	1	0.1	Type Z	0.161208822	1	12	15	26	2	10.9
36	1	0.2	Type Z	0.161208822	1	10	15	17	1	18.7
37	1	0.2	Type E	0.161208822	5	10	15	26	2	18.9
38	1	0.2	Type Z	0.161208822	3	14	15	17	3	19.5
39	1	0.2	Type E	0.161208822	2	11	1	17	1	55.1

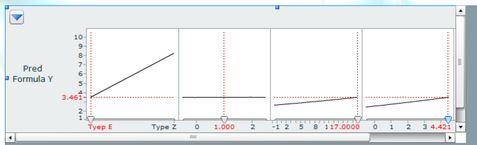
evaluations done

Results

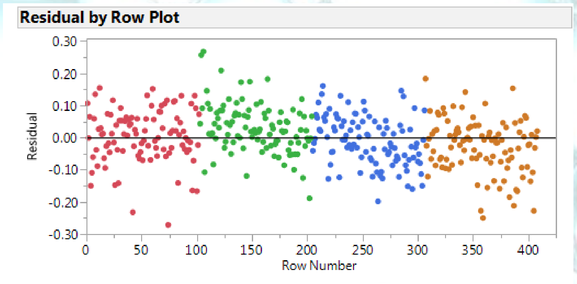
Prediction Model

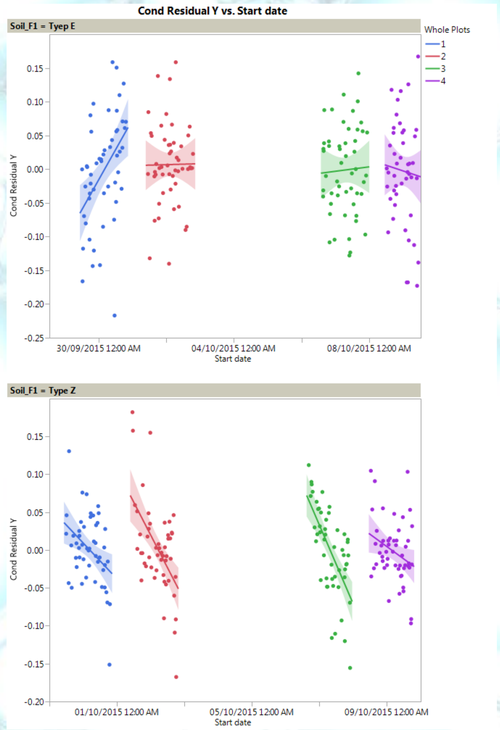
- For each soil type we were able to calculate a separate very accurate prediction model, since soil types were quite different and will always be used separately in the robot.
- New insights were gained: one soil shows changing behavior over time, which is now closely investigated.
- All objectives of the study could be met.

Model for different soils



Response trend with time





From the "customer" point-of-view

- Full space coverage validation for court case defense.
- Identification of areas to further improve the method and/or the equipment to keep competitive advantage by delighting our consumers.
- Power analysis & sample size simulator to distinguish Minimum Meaningful Difference in specific conditions or for specific claims.
- Followed approach is also useable for other DOE on this equipment.

- ✓ How many tests do I need to run to see a significant difference?
- ✓ What variability can I expect?
- ✓ What test conditions have the highest chance of showing sign. differences?

Inputs

- Products
- Soils
- Environmental conditions

"Technical Test Recommendation Tool"

Model

Outputs

- ✓ Sample Size
- ✓ Power analysis
- ✓ Expected error

Touching lives, improving life. P&G™