

Abbildung 1 Titelbild Blog #60 Stefan Moser

DoE Blog #060 – Wenn die Streuung uns die Sicht versperrt

In Blog #58 habe ich eine Liste mit über 17 möglichen Ursachen gesammelt, warum erwartete Faktoreffekte in DoEs manchmal ausbleiben – selbst wenn man „alles richtig gemacht hat.“

Heute steigen wir tiefer in Punkt b.) dieser Serie ein:

Große Streuung in Messergebnissen – also dann, wenn externe Einflüsse oder Messfehler aus Umgebung, Messsystem oder Bedienung dafür sorgen, dass unsere Daten mehr rauschen als sprechen.

„Schon wieder unklare Ergebnisse ... kann nicht sein, das Material ist doch stabil!“

Ein wenig abstrahiert – aber mit einer ganz ähnlichen Kernaussage – begann mein Austausch mit einem Entwicklerteam aus dem Bereich Spezialverpackungen.

Der Kunde testete eine mehrlagige Hochleistungsverbundfolie für pharmazeutische Anwendungen, kaschiert aus fünf funktionalen Schichten, jeweils mit spezifischen Aufgaben: Barriere, Festigkeit, Siegelbarkeit, UV-Schutz etc.

Das Ergebnis:

- **Theorie:** Top-Materialwerte, sauber spezifiziert.
- **Praxis:** Starke Schwankungen bei Zugprüfungen – Kriechverhalten, Dehnung, Schmelzpunkt: alles inkonsistent, ohne erkennbare Trends.

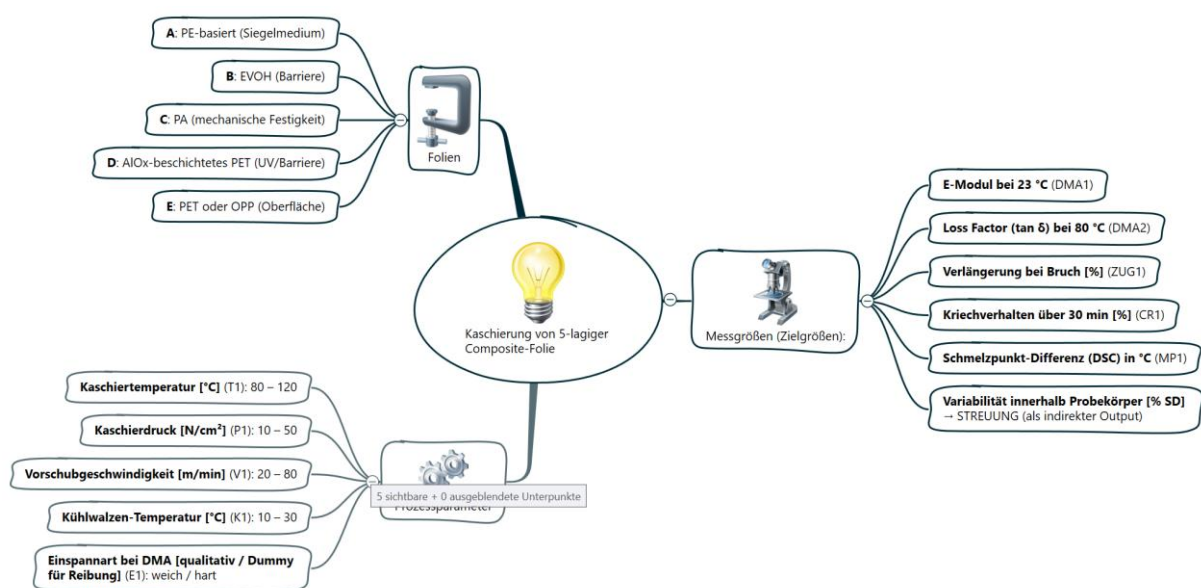
Das Team hatte – auf Basis der gängigen Praxis – eine Design of Experiments (DoE)-Studie aufgesetzt, um den Ursachen auf den Grund zu gehen.

Konkret kam ein D-Optimales Design zum Einsatz – ein flexibles Versuchsplanverfahren, das sich besonders bei nicht vollständigen Faktorkombinationen und qualitativen Einflussgrößen (wie hier: „Einspannart“) eignet.

Ziel war es, mit möglichst wenigen, aber clever gewählten Experimenten, Zusammenhänge zwischen diesen Faktoren und den gemessenen Materialeigenschaften sichtbar zu machen.

Doch keiner der Faktoren erschien wie erwartet signifikant.

Und an genau diesem Punkt wurde das Projekt spannend...



DoE als letzter Strohhalm – oder Beginn echter Ursachenforschung?

Keine Firma setzt „einfach so“ auf DoE – zumindest nicht, wenn die Methode intern (noch) nicht etabliert ist. In vielen Fällen kommt Design of Experiments erst dann ins Spiel, wenn die klassischen Werkzeuge ausgeschöpft sind und die Lösung weiterhin auf sich warten lässt.

So auch hier: Das Team war bereits mit dem Rücken zur Wand – die Ergebnisse waren widersprüchlich, der Projektdruck hoch, die Erwartung an die DoE entsprechend groß.

Meine Rolle beginnt in solchen Situationen oft nicht am Reißbrett, sondern in einem Gespräch. Ich frage, was dem Team besonders auffällt, welche Hypothesen schon im Raum stehen, und wo die Unsicherheiten liegen.

Für diesen Einstieg nutze ich gerne strukturierende Methoden wie Ishikawa-Diagramme, Mindmaps oder auch Ursachen-Wirkungs-Cluster. Warum?

Weil sie helfen, komplexe Systeme übersichtlich abzubilden, ohne vorschnell in die Tiefe zu springen. Ich beginne dabei immer „vom Groben ins Detail“ – sozusagen vom Blick aufs Ganze hin zum kritischen Zoom.

In diesem Fall startete ich – ganz pragmatisch – mit der Zusammensetzung der Folie selbst. Denn was von außen als „ein Produkt“ erscheint, ist in Wahrheit ein hochkomplexer Materialverbund aus mehreren Einzelschichten, die sich gegenseitig beeinflussen können.

Die Folie, der Prozess – und was wir wirklich wissen müssen

Aufbau der 5-lagigen Composite-Folie:

- A – PE-basiert: Dient als Siegelmedium, also der thermisch aktivierbaren Schicht.
- B – EVOH: Die klassische Barriere-Schicht gegen Gase.
- C – PA (Polyamid): Verantwortlich für die mechanische Stabilität.
- D – PET mit AlOx-Beschichtung: Kombiniert UV- und Sauerstoffbarriere – wichtig für lichtempfindliche Wirkstoffe.
- E – PET oder OPP: Die Außenschicht, je nach Anforderung (z. B. Bedruckbarkeit, Reibung, Kratzfestigkeit).

Jede dieser Schichten bringt nicht nur ihre Funktion, sondern auch ihre Eigenheiten mit: Schrumpfverhalten, Adhäsion, Wärmeausdehnung, Additivreste ... und natürlich: Spannungseinträge. Diese Aspekte wurden in der Diskussion besonders wichtig.

Die Prozessfaktoren – was wurde überhaupt variiert?

Um mögliche Einflüsse auf die Materialeigenschaften zu identifizieren, wurden fünf Prozessparameter in der Versuchsplanung berücksichtigt – vier quantitativ, einer qualitativ:

- T1 – Kaschier Temperatur [°C]: 80 – 120
- P1 – Kaschierdruck [N/cm²]: 10 – 50
- V1 – Vorschubgeschwindigkeit [m/min]: 20 – 80
- K1 – Kühlwalzentemperatur [°C]: 10 – 30
- E1 – Einspannart (hart/weich): *Qualitativer Dummyfaktor* für Reibungseinfluss beim DMA-Messaufbau

Letzterer Faktor wurde bewusst aufgenommen – nicht als technischer Prozessparameter, sondern als Dummy, um potenzielle Einflüsse durch das Einspannen und damit die Reibung an der Probe zu erfassen.

Die Zielgrößen – worauf wurde eigentlich geschaut?

Zur Beurteilung der kaschierten Folie wurden folgende Zielgrößen definiert – mit Fokus auf mechanische und thermische Stabilität:

- DMA1 – E-Modul bei 23 °C
→ Steifigkeit, mechanische Belastbarkeit im Ausgangszustand
- DMA2 – Loss Factor (tan δ) bei 80 °C
→ Dämpfungsverhalten im Übergangsbereich – empfindlich auf molekulare Mobilität
- ZUG1 – Verlängerung bei Bruch [%]
→ Dehnung, ein Maß für Flexibilität bzw. Materialversprödung

- CR1 – Kriechverhalten über 30 min [%]
→ Zeigt, ob das Material unter konstanter Belastung „nachgibt“
- MP1 – Schmelzpunkt-Differenz (DSC)
→ Indikator für Homogenität, Kaschiereffekte, mögliche Störungen im Aufbau

Aufbau der Versuchsplanung – oder: Wenn Statistik keine Klarheit bringt

Aufgrund des qualitativen Faktors „Einspannart“ (weich/hart) entschied sich die Gruppe vor meiner Unterstützung für ein D-optimales Design. Warum? Klassische Designs wie Central Composite (CCC) oder Box-Behnken stoßen an ihre Grenzen, sobald qualitative oder kategoriale Faktoren ins Spiel kommen – D-optimale Designs sind hier deutlich flexibler. Sie erlauben die gezielte Modellierung von Wechselwirkungen und nicht-linearen Effekten, auch bei gemischten Faktortypen.

Das Team entschied sich für:

- 25 Grundversuche + 2 Centerpoints + 1 Wiederholung dieses Versuchsplans zur Beurteilung der Reproduzierbarkeit im ganzen Versuchsraum

Soweit, so gut.

Doch dann kam der Moment, den viele Teams nicht erleben möchten – obwohl genau dieser Moment einer der ehrlichsten in einem DoE-Prozess ist:

☞ **Das Modell erklärte nichts.**

🧐 **Kein Effekt war signifikant – was tun?**

Trotz systematischer Planung und bewusst gewählter Faktorbereiche ließen sich keine klaren Ursache-Wirkung-Beziehungen erkennen.

- **Ein Teil des Teams reagierte mit Frustration:** „Das war’s – DoE bringt uns auch nicht weiter. Wir verschwenden nur Zeit mit Statistik.“
- **Ein anderer Teil war reflektierter:** „Jetzt sehen wir es wenigstens systematisch. Was wir vorher diffus gespürt haben, ist jetzt schwarz auf weiß – aber ohne Klarheit.“

Und ich?: Ich wollte es genauer wissen. Also stellte ich eine scheinbar einfache Frage:

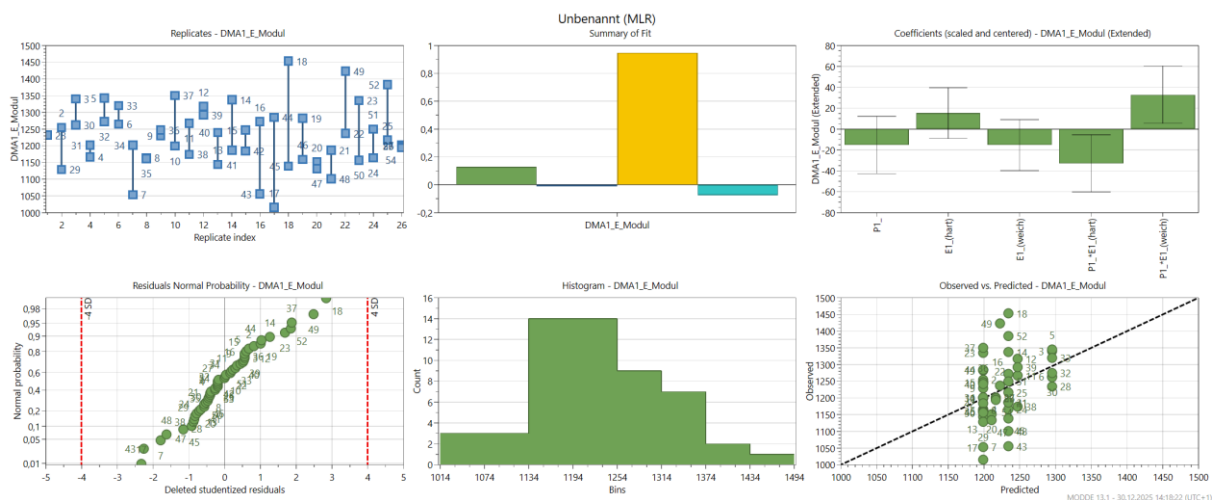
- „Wusstet ihr eigentlich vorher schon, dass die Ergebnisse bei ähnlichen Einstellungen oft schwanken?“ - Kurze Pause. Dann Stirnrunzeln.

Die Frage schien zu direkt, also versuchte ich es etwas weicher:

- „Gab es auch vor der DoE bereits Situationen, in denen bei gleichen Parametern sehr unterschiedliche Ergebnisse rausgekommen sind?“ Diese Variante kam an. Die Antwort lautete: „Ja. Genau deswegen haben wir überhaupt erst die DoE gemacht.“

Warum das ein entscheidender Punkt ist

- Das war der Wendepunkt – denn damit wurde klar:
Die DoE wurde zwar präzise geplant, aber auf einem unsicheren Fundament gebaut. Wenn die Reproduzierbarkeit schon vorab schlecht war, dann ist jede Analyse ein Blick in den Nebel.
- Denn: Ein DoE kann nicht erkennen, welches Ergebnis durch echte Effekte und welches durch versteckte Störungen verzerrt wurde – wenn diese Informationen nicht im Modell enthalten sind.
- Das Modell kann bestenfalls mitteln, glätten oder rauschen – aber es kann keine Wahrheit zaubern, wo keine Konstanz vorhanden ist. Selbst mit Wiederholungen (Replikaten) lässt sich die Unsicherheit zwar abschätzen, aber nicht „entwirren“.



In dieser DoE-Studie standen die mechanischen Eigenschaften einer mehrschichtig kaschierten Folienstruktur im Fokus. Der Aufbau war anspruchsvoll – fünf funktionale Schichten, sorgfältig aufeinander abgestimmt und in einem sensiblen Kaschierprozess verarbeitet. Doch trotz dieses durchdachten Designs zeigten sämtliche Messgrößen ein ähnlich widersprüchliches Bild: inkonsistent, ohne erkennbare Trends. Um der Sache methodisch auf den Grund zu gehen, entschied ich mich zunächst, eine repräsentative Zielgröße auszuwählen – als Einstieg in die vertiefte Analyse.

Als repräsentative Zielgröße diente das

☞ Elastizitätsmodul bei 23 °C (DMA1_E_Modul) – eine klassische Kennzahl zur Bewertung der Steifigkeit und Festigkeit.

Was zeigt der erweiterte Overview-Plot in MODDE?

Nach dem Setup der DoE und der ersten Irritation über die unklaren Ergebnisse ging es ans Eingemachte: Der erweiterte Overview-Plot aus MODDE visualisiert alle wichtigen Aspekte der Modellgüte – und liefert wertvolle Hinweise auf die wahre Ursache des Problems.

1. Replikate-Plot (oben links)

- Der erste Hinweis auf das Kernproblem: Starke Schwankungen bei identischen Versuchsbedingungen.
- Mehrere Versuche – z. B. 7 vs. 43 oder 18 vs. 47 – unterscheiden sich um teils mehr als ± 100 MPa, obwohl alle Eingabewerte laut Plan identisch waren.

🤔 Vermutung / Tipp:

- So etwas deutet in der Praxis fast immer auf nicht geregelte Störgrößen hin – z. B. Einspannschwankungen, Handlingeffekte oder unerkannte Toleranzen im Messprozess.
- Ein erstes Warnsignal, das man ernst nehmen sollte – auch wenn es „nur ein paar Ausreißer“ zu sein scheinen.

2. Summary of Fit (oben Mitte)

- $R^2 \approx 0,09$ – das Modell erklärt weniger als 10 % der beobachteten Gesamtvariation
- $Q^2 < 0$ – die Vorhersagekraft ist faktisch nicht vorhanden.

🤔 Vermutung / Tipp:

- Ein so niedriges R^2 kombiniert mit negativem Q^2 ist typisch für Situationen, in denen Streueffekte dominieren, aber nicht auf modellierte Faktoren zurückführbar sind.
- Da hilft auch leider keine „Modelloptimierung“ – sondern nur Ursachenforschung.

3. Koeffizienten-Diagramm (oben rechts)

- Der einzige sichtbare Effekt: Einspannart (hart vs. weich).
- Alle anderen Faktoren bleiben unterhalb der Signifikanzgrenze.
- Auffällig breite Konfidenzintervalle → hohe Unsicherheit.

🤔 Vermutung / Tipp:

- Hier könnte sich ein typischer Effekt zeigen: Ein einzelner qualitativer Faktor (wie z. B. Einspannung) wirkt so dominant, dass die anderen überdeckt werden.
- Oder anders gesagt: Solange dieser Einfluss nicht stabilisiert wird, kommen wir an die anderen nicht Erfassten gar nicht ran.

4. Normalverteilung der Residuen (unten links)

- Die Residuen – also die Vorhersagefehler – folgen grob einer Normalverteilung.
- Einzelne moderate Ausreißer (z. B. Versuch 7, 43) sind zu erkennen, ein Entfernen der Residuen (Exkludieren) bringt hier keinen Mehrwert.

🤔 Vermutung / Tipp:

- Die Verteilung wirkt „zufällig“, nicht systematisch verzogen – das spricht gegen Modellierungsfehler und eher für externe Streuquellen.
- Gut erkennbar: Die Statistik ist formal okay, das System ist das Problem.

5. Histogramm (unten Mitte)

- Die Zielgrößen verteilen sich breit, unscharf und annähernd normalverteilt.
- Kein Peak, keine Clusterbildung, keine erkennbare Unterstruktur.

🤔 Vermutung / Tipp:

- Das spricht für systematisch überlagerte Effekte – möglicherweise ein Mix aus mechanischen, thermischen oder elektrostatischen Einflüssen, die nicht als DoE-Faktoren erfasst wurden.
- Hier ist oft die reale Welt „rauschiger“ als das beste Modell erlaubt.

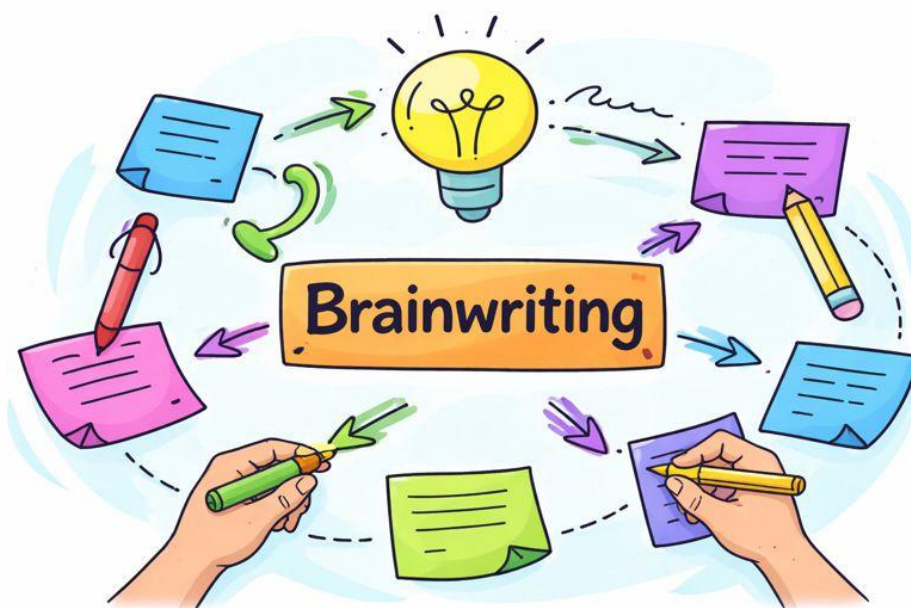
6. Observed vs. Predicted (unten rechts)

- Die Datenpunkte liegen weit verstreut rund um die Diagonale.
- Selbst eine gute Modellierung kommt hier nicht auf Vorhersageniveau.
- Keine prädiktive Qualität des Modells erkennbar.

🤔 Vermutung / Tipp:

- Wenn selbst Replikate und Centerpoints weit auseinanderliegen, lohnt es sich nicht, am Modell zu schrauben.
- Erst die Reproduzierbarkeit sichern – dann modellieren.

Die Suche nach den Störgrößen



Natürlich gingen wir nicht einfach zur Tagesordnung über. Ich nutzte einen meiner liebsten Perspektivwechsel – Brainwriting.

Diese Methode zählt zu den eher stillen, aber unglaublich effektiven Kreativitätstechniken – perfekt, wenn es darum geht, Fehlerquellen aufzuspüren, über die sonst selten offen gesprochen wird.

Ich stellte der Gruppe daher eine bewusst provokante Aufgabe:

„Versetzen Sie sich in die Rolle eines Saboteurs. Wo im Prozess könnten Sie ganz gezielt eingreifen, damit die Zielgrößen schlechter werden – ohne, dass es auf den ersten Blick auffällt?“

Der ein oder andere schaute erst verdutzt, dann zum Teamleiter – aber der war eingeweiht. Und so entstand eine kurze, fast andächtige Schreibphase. ...Jeder für sich. ... Keine Diskussion, ... kein Einfluss von außen.

Die Stifte kratzten über die Zettel – und das erste Lachen ließ nicht lange auf sich warten.

Doch damit nicht genug.

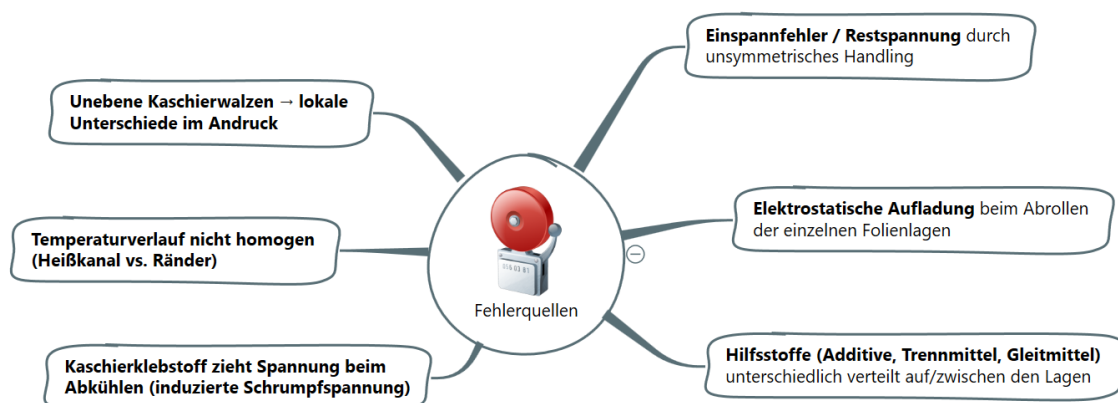
Im nächsten Schritt wechselten die Zettel reihum – jeder bekam einen fremden Gedanken zur Weiterbearbeitung.

Die Aufgabe war jetzt:

„Spinne die Idee weiter. Ergänze, widersprich, bring eine neue Perspektive rein.“

Das hatte Wirkung.

Was vorher vielleicht wie eine Einzelmeinung wirkte, bekam plötzlich Tiefe, Varianten, Zusammenhänge.



Erkenntnisse und Anmerkungen – wie z.B. „Elektrostatische Aufladung beim Abrollen“ – wurde ergänzt um „hängt stark von der Luftfeuchte ab“, „beim Wechsel von PET zu PA besonders auffällig“ oder „wird oft nicht entladen, wenn die Folie die Maschine durchläuft“.

Innerhalb weniger Minuten entstand so ein kollaborativer Hypothesenraum, gefüllt mit realistischen, praxisnahen Störgrößen – viele davon bisher nur Vermutungen, jetzt systematisch gesammelt und bewertet.

Das Ergebnis war aufschlussreich – und für manche im Raum auch ein kleiner Aha-Moment:

Nicht der DoE ist schuld. Die Realität ist komplexer, als das Modell zulässt.

Wenn das Rauschen lauter ist als das Signal

Was in diesem Projekt sichtbar wurde, war nicht etwa ein Mangel an Planung, Struktur oder Methodik. Im Gegenteil: Das Team hatte sauber gearbeitet, die Versuchsbedingungen klar definiert, Wiederholungen integriert und sogar einen qualitativen Faktor über ein geeignetes D-Optimales Design eingebunden.

Und trotzdem: Kein klares Bild. Kein roter Faden. Kein signifikanter Zusammenhang.

Spätestens jetzt war klar: Nicht das Modell war das Problem – sondern das, was nicht modelliert wurde.

Und genau hier lag der Hund begraben:

Störgrößen, Umwelteinflüsse, unerwünschte Nebeneffekte – Dinge, die nicht geplant waren, aber dennoch passieren.

Ein Beispiel? Elektrostatische Aufladung, minimale Unterschiede im Einspannen, Mikroklima im Raum, lokal unterschiedliche Klebstoffverteilung, oder unerkannte Spannungen im Laminat.

Einzelnen vielleicht harmlos – zusammen jedoch genug, um den Versuchsraum zu übertönen.

Erst nachdem das Team – basierend auf den Hypothesen aus dem Brainwriting – gezielt auf mögliche Störgrößen einging, entstanden reproduzierbare Ergebnisse.

Kleine Maßnahmen mit großer Wirkung wie: Entladungsvorrichtung installiert. Spannungsfreie Einspannung etabliert. Klimabedingungen stabilisiert.

Und siehe da: Plötzlich zeigte sich eine klare Struktur. Die Modelle wurden erklärbar. Vorhersagen ergaben Sinn. Die Faktoren entfalteten Wirkung – und das Team verstand plötzlich, warum „es vorher nicht funktioniert hat“.

Learning des Blogs:

Jede DoE bildet einen Ausschnitt der Realität ab – und alles, was wir nicht explizit variieren, bleibt im Dunkeln. Störgrößen, die nicht erfasst oder als Faktoren integriert sind, bleiben unsichtbar – ihre Wirkung jedoch nicht.

Die Statistik kann nur sortieren, was auch im Modell drin ist. Und wenn das Rauschen zu laut ist, geht das Signal unter – selbst wenn es da ist.

Nicht mehr Experimente sind die Lösung – sondern bessere Bedingungen.

🎯 Ausblick – Blog #061:

Mit dem heutigen Beitrag haben wir Punkt 3 der „17+ Ursachen für fehlende Signifikanz“ beleuchtet: (b) Externe Störgrößen und Messrauschen.

Im nächsten Beitrag geht's weiter mit: (c) Schwache oder grenzwertige Signifikanz – wenn nur ein Teilbereich des Faktors wirkt, der Rest aber „leer läuft“.

Sollten Sie weitere Gründe vermissen oder kennen so senden Sie mir die gerne zu, ich nehme Sie gerne in diese Blog Staffel mit auf!!

Die Darstellung bleibt (*oder sollten*) bewusst kompakt, verständlich und orientiert sich – wie immer – an konstruierten, aber realitätsnahen Beispielen aus meiner Praxis. Denn: Nicht jeder fehlende Effekt ist gleich ein methodischer Fehler. Viel häufiger liegt die Ursache in einem kleinen, vielleicht unscheinbaren – aber entscheidenden – Detail, das im ersten Moment leicht übersehen wird.

Und ja, ich gebe zu: Ich bin thematisch wieder etwas eskaliert. Aber ohne Absicht – in der Hoffnung, dass der zusätzliche Tiefgang Ihnen hilft, ähnliche Situationen künftig schneller zu erkennen und besser einzuordnen.

👉 Vielleicht war ja ein Aha-Moment für Sie dabei? Dann freue ich mich über Feedback oder einen Kommentar – wo hat es bei Ihnen Klick gemacht?

Teilen Sie den Beitrag gerne – Wissen wird nicht kleiner, wenn man es teilt.

Und ein kleines Highlight zum Schluss

Nach intensiver Vorbereitung habe ich alle bisherigen [Hashtag#DoE](#)-Beiträge in unsere neue LinkedIn-Gruppe übertragen. So haben alle DoE-Interessierten nun einen zentralen Ort, an dem auch ältere Inhalte leicht auffindbar sind, neue Themenvorschläge

eingebraucht werden können und der fachliche Austausch rund um Design of Experiments weiter wächst.

Im Abstand von etwa 1–2 Wochen erscheinen dort neue Blogbeiträge aus meiner DoE-Reihe. Ich freue mich besonders auf den Austausch in der Gruppe, eure Erfahrungen, Fragen und lebendige Diskussionen.

Ich lade euch herzlich ein, der neuen Gruppe beizutreten:

<https://lnkd.in/d8t4gt74>



Abbildung 04 Logo Stefan Moser Process Optimization

Mehr aus Ihren Prozessen rausholen?

Ob DoE-Grundlagen oder Spezialthemen wie Troubleshooting, Screening, Optimierung, Mischungsdesigns oder Robustheit – ich unterstütze Sie mit praxisnahen DoE-Trainings, gezielter Beratung und methodischer Begleitung. Auch bei MVDA, DFSS und QFD bin ich an Ihrer Seite – vom ersten Workshop bis zur robusten Umsetzung.

✉ Schreiben Sie mir: info@stefan-moser.com

🌐 Mehr unter: www.stefan-moser.com

Linkedin : Meine DoE Gruppe: <https://lnkd.in/d8t4gt74>

Bleiben Sie experimentierfreudig und neugierig!

Ihr DoE-Trainer Stefan Moser