

Chemometrie in der Analytischen Chemie¹

2. Optimierung mittels Response-Surface Modelling

Florian Klingler, Eckhard Reh*

* Korrespondenz: Prof. Dr. E. Reh, Technische Hochschule Bingen, Email: e.reh@th-bingen.de

Einleitung

Die Optimierung von Prozessen ist in der Analytischen Chemie von großer Bedeutung, z.B. bei der Derivatisierung von Proben vor der Chromatographie oder die Einstellung der Messbedingungen bei der Atomabsorptionsspektroskopie. Sind mehrere Einflussgrößen (Faktoren, $x_1, x_2, x_3 \dots$) in ihrer Auswirkung auf die Zielgröße (z.B. Signal, y) zu beachten sollte immer eine der beiden etablierten Optimierungs-Strategien (Simplex-, RSM-Optimierung) eingesetzt werden. Liegen zu viele Parameter vor, müssen zuvor mit Hilfe der statistischen Versuchsplanung (z.B. VPlan [5]) die 2 bis 5 wichtigsten Faktoren ermittelt werden. An dieser Stelle soll der Einsatz des Response-Surface-Modelling (RSM) vorgestellt werden, da hier auch einige negative Aspekte zu berücksichtigen sind.

Es stehen verschiedene kommerzielle Programme zur Verfügung, unter anderem Design-Expert (Stat-Ease) oder RMSOft [3], welches für die folgenden Anwendungen zum Einsatz gekommen ist.

Das Vorgehen bei RSM ist wie folgt:

- Festlegung des Versuchs-Designs, d.h. Untersuchungsbereich und Lokalisierung der Faktoren z.B. im zentral zusammengesetzten Design (siehe Abbildung 1, Abbildung 4 Insert)
- Durchführung aller Untersuchungen bei den vorgegebenen Faktor-Einstellungen
- Simulation der Mess-Ergebnisse durch ein numerisches Modell, meist Polynom 2-ten Grades z.B. $y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_1^2 + b_5x_2^2 + b_6x_3^2 + b_7x_1x_2 + b_8x_1x_3 + b_9x_2x_3$.
- Beurteilung der Modell-Größen wie Maximum, Minimum der Zielgröße und Qualität des Modells.

Dem letzten Punkt kommt eine besondere Bedeutung zu, da das verwendete Polynommodell als empirisches Modell numerisch leicht zu bestimmen ist. Es ist aber in fast allen Fällen das falsche Modell, da die physiko-chemisch zugrunde liegenden (mechanistischen) Modelle zumeist nicht bekannt oder numerisch nicht geschlossen lösbar sind.

Mit Hilfe der Varianzanalyse ist daher zu klären, ob das verwendete Polynom die Mess-Ergebnisse angemessen beschreibt. Wichtige Größen sind hier

- Residuen-Varianz $s_R =$ Abweichung der Messwerte von den theoretischen Werten des Modells zur Bewertung der Unschärfe der Modellkoeffizienten b_i (siehe Anhang a)
- F-Wert gibt an, ob das Modell akzeptiert wird:

$$F\text{-Wert} = \frac{\text{Abweichung } y\text{-Gruppenmittelwerte von Regressionsmodell}}{\text{Abweichung } y\text{-Werte von } y\text{-Gruppenmittelwert}}$$

Der so ermittelte Prüfwert wird mit dem Vergleichswert der F-Tabelle verglichen und das Modell wird akzeptiert wenn gilt: Prüfwert < Vergleichswert (siehe Anhang b, vgl. [3] Kap. 4.2.5).

Eine ausführliche Behandlung der theoretischen Grundlagen sollen nicht Ziel der Ausführungen sein, diese sind in einschlägigen Fachbüchern [1, 2, 3, 4] ausreichend dargestellt.

Beispiel 1: Messparameter AAS

Häufige Aufgabenstellung ist die Optimierung der Messbedingungen eines Analysen-Geräts. In dem Beispiel sollen bei der elektrothermischen Atomabsorption (AAS) die Veraschungsbedingungen optimiert werden.

Zielgröße ist ein maximales AAS-Signal (relativ zum Untergrund-Signal), die Einflussfaktoren sind Veraschungszeit und -temperatur. Als Versuchsdesign wurde, um Zeit und Kosten zu sparen, das partielle zentral-zusammengesetzte Design gewählt (siehe Abbildung 1). Die Messbereiche wurden zu Beginn vorgegeben zu Temperatur 250 - 1000°C bzw. Zeit 5 - 50 s, die resultierenden Bedingungen sind in Tabelle 1 angegeben.

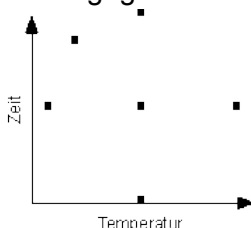


Abb. 1: Partiiell zentral zusammengesetztes Design

Analyse-Nr.	Temperatur [°C]	Zeit [s]
1	250	28
2	625	50
3	1000	28
4	625	5
5	625	28
6	359	12

Tab. 1: Veraschungs-Bedingungen

Nach Ausführung der Analysen wird mit den resultierenden Absorptionen mittels linearer Regression (vgl. [4] Kapitel 3.4.3) das Modell-Polynom bestimmt zu

$$\text{rel. Absorption} = 0,13 + 2,85 \cdot 10^{-4} [\text{1/s}] \cdot \text{Zeit} + 7,20 \cdot 10^{-3} [\text{1/}^\circ\text{C}] \cdot \text{Temp} - 3,59 \cdot 10^{-7} [\text{1/s}^2] \cdot \text{Zeit}^2 - 7,11 \cdot 10^{-5} [\text{1/}^\circ\text{C}^2] \cdot \text{Temp}^2 - 6,86 \cdot 10^{-6} [\text{1/s } ^\circ\text{C}] \cdot \text{Zeit} \cdot \text{Temp}.$$

Numerisch kann daraus das maximale AAS-Signal bei den optimalen Bedingungen 250 °C und 38,6 s bestimmt werden (Minimum bei 1000 °C, 50 s). Abbildung 2 zeigt die entsprechende Hyperfläche.

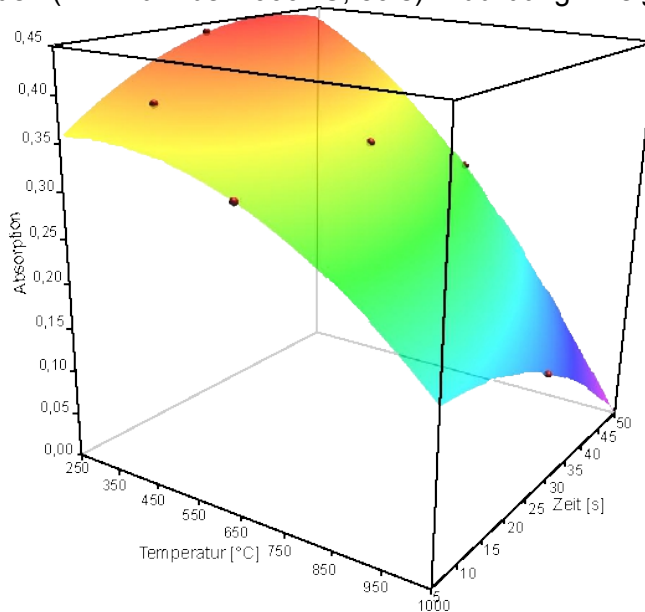


Abb. 2: RSM-Graf der AAS von Mn in Concanavalin A

Ohne zusätzliche Freiheitsgrade liegen alle Messpunkte in der Modell-Hyperfläche, eine statistische Aussage ist nicht gegeben.

Auffällig ist, dass das Maximum am Rand des gewählten Untersuchungsbereichs liegt (250°C), es kann angenommen werden, dass geringere Temperaturen günstiger sind. Diese Problematik einer ungeschickten Bereichswahl zu Beginn ist einer der RSM-Nachteile. In diesem Fall müssten die Untersuchungen mit verändertem Design wiederholt werden.

Beispiel 2: Hydrolyse Polysaccharide

Vor der Analyse von Polysacchariden ist die schonende Hydrolyse oft der erste Schritt, hier mit Hilfe von Trifluoressigsäure (TFA). Zielgröße ist die Hydrolyse-Ausbeute von Trehalose als Modells substanz analysiert mittels HILIC-HPLC-Trennung und ESI-TOF-MS-Detektion. Als Einflussgrößen wurden die Reaktionszeit bzw. die Konzentration TFA optimiert. Aus Zeitgründen wurde auch hier ein partielles zentral-zusammengesetztes Design verwendet.

Abbildung 3 zeigt die resultierende grafische Darstellung des Modells aus den Analysen.

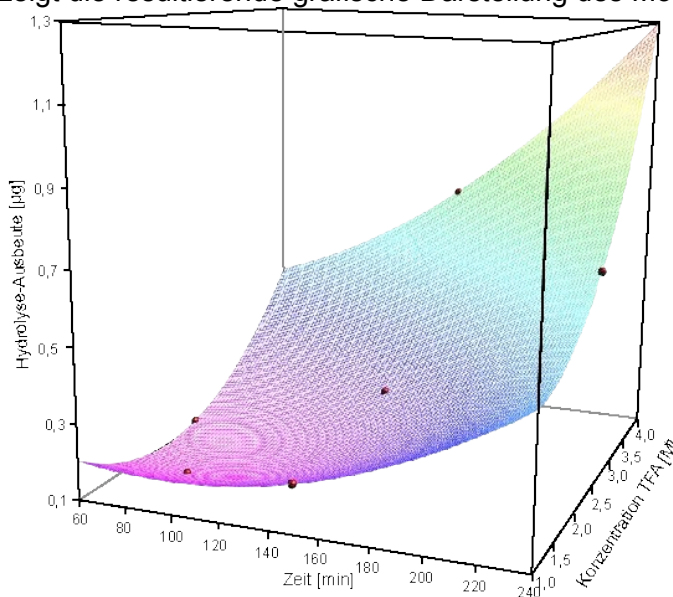


Abb. 3: RSM-Graf der Hydrolyse von Trehalose mittels TFA

Das geschätzte Modell lautet in diesem Fall:

$$\text{Ausbeute } [\mu\text{g}] = 6891,5 [\mu\text{g}] - 36,9 [\mu\text{g}/\text{min}] \cdot \text{Zeit} - 4306,6 [\mu\text{g}/\text{M}] \cdot \text{Konz.} + 0,14 [\mu\text{g}/\text{min}^2] \cdot \text{Zeit}^2 + 908,4 [\mu\text{g}/\text{M}^2] \cdot \text{Konz.}^2 + 9,97 [\mu\text{g}/\text{min M}] \cdot \text{Zeit} \cdot \text{Konz.}$$

Numerisch kann hieraus die maximale Ausbeute bei den optimalen Bedingungen 240 min und 4 M TFA bestimmt werden (Minimum bei 60 min, 2M). Aus dem Graf könnte interpretiert werden, dass in diesem Fall das empirische Polynom die Zusammenhänge angemessen beschreibt. Ohne Freiheitsgrade (zusätzliche Messpunkte, Wiederholungsmessungen) ist aber eine statistische Aussage nicht gegeben.

Beispiel 3: Derivatisierung zur GC-MS

Bei der Linkage-Analyse der Proteinglycosylierung ist ein wichtiger Schritt vor GC-MS-Analyse die Reduktion der partiell O-methylierten Monosaccharide mit Natriumbordeuterid. Als Zielgröße (y) wurde die relative Peakfläche des Glucitol (zum intern. Standard) als Maß für die Derivatisierung von 2,3,4,6-Tetra-O-Methylglucose gewählt. Als Einflussgrößen wurden die Reaktionszeit (x_1) bzw. die Temperatur (x_2) optimiert.

Hier wurde ein vollständiges zentral-zusammengesetztes Design mit 5 Wiederholungen verwendet (vgl. Abbildung 4 Insert). Abbildung 4 zeigt das resultierende grafische Modell aus den GC-MS-Daten.

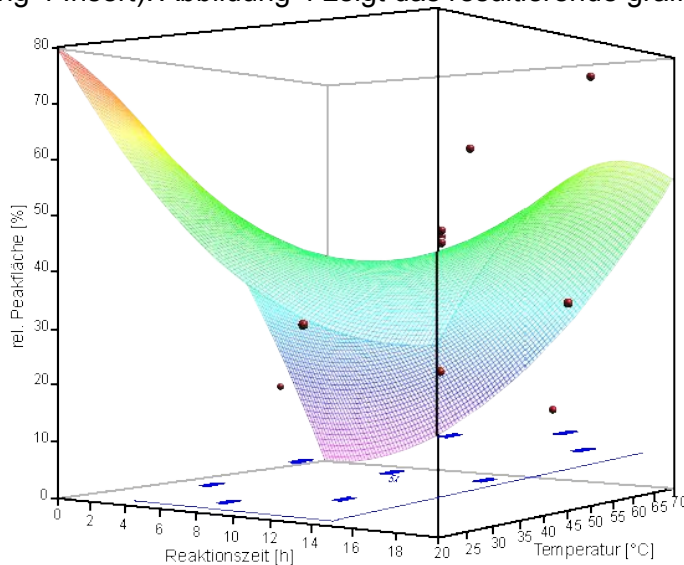


Abb. 4: RSM-Graf Reduktion 2,3,4,6-Tetra-O-Methylglucose (Insert blau: zentral zusammengesetztes Design)

Aus dem geschätzten Modell $y = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_1^2 + b_4 \cdot x_2^2 + b_5 \cdot x_1 \cdot x_2$

kann numerisch eine maximale rel. Peakfläche (73 %) bei optimalen Bedingungen 1,0 h und 20 °C abgeleitet werden (Minimum bei 1,1 h / 70°C).

Bereits die grafische Darstellung deutet auf eine Diskrepanz zwischen dem empirischen Modell und den Messwerten, dies wird durch die statistischen Parameter bestätigt:

$s_R = 15,1$ und F-Wert = 6,6 ($f_1 = 3$, $f_2 = 5$; Modell nicht akzeptiert)

Das Resümee ist daher: die RSM-Optimierung ist für diese Anwendung ungeeignet!

Als Alternative kann hier die Simplex-Optimierung herangezogen werden ([4, 6] z.B. SimSoft [5]).

Diese führt auch in kritischen Fällen schrittweise zum Optimum, erlaubt aber keine Aussage zu den Verhältnissen außerhalb der Messpunkte.

Zusammenfassung

Es lässt sich festhalten, dass die Optimierung mittels RSM eine valide Methodik in der Analytischen Chemie ist. Die Vorteile des Verfahrens liegen darin, dass durch einen relativ geringen Arbeitsaufwand (Messumfang) gute Modellierungen durchgeführt werden können, aus denen neben dem Optimum viele Aussagen abgeleitet werden können. Die grafische Darstellung der Messdaten bzw. des mathematischen Modells erleichtert dabei die Interpretation erheblich. Durch die Untersuchung weiterer Messpunkte mit Wiederholungen (Freiheitsgrade f_1 , f_2) ist eine statistisch validierte Bewertung möglich (ANOVA).

Allerdings stößt das Verfahren dann an seine Grenzen, wenn zu Beginn die Optimierungsbereiche der zu untersuchenden Faktoren ungünstig gewählt sind. Da das Verfahren erst nach Abschluss aller Messungen ein Resultat liefert ist somit gegebenenfalls eine Wiederholung der gesamten Optimierung nötig. Kritisch ist, wenn die statistische Behandlung (F-Wert) zeigt, dass das empirische Polynommodell ungeeignet ist, die Messungen zu simulieren. Hier muss dann alternativ die Simplex-Optimierung eingesetzt werden.

Abschließend bleibt festzuhalten, dass der Einsatz von Optimierungsverfahren, wie RSM- und Simplex-Optimierung unabdingbar sind für die Untersuchung von mehreren Einflussfaktoren (eventuell mit Wechselwirkungen). Nur damit ist eine Optimierung von Analysen-Verfahren sicher und effizient möglich.

Literatur

- [1] Box, George EP; Wilson, Kenneth B. On the experimental attainment of optimum conditions. In: Breakthroughs in statistics. Springer, New York, NY, S. 270-310, 1992
- [2] Box, George EP; Hunter, W. H.; Hunter, J. Stuart. Statistics for experiment., 1978
- [3] Brereton, R.G., Chemometrics, Data Analysis for Laboratory and Chemical Plant, Wiley, 2006
- [4] Reh, Eckhard. Chemometrie: Grundlagen der Statistik, Numerischen Mathematik und Software Anwendungen in der Chemie. Walter de Gruyter GmbH & Co KG, 2017
- [5] www.chemometrie.info
- [6] Deming, Stanley N.; Parker JR, Lloyd R.; Bonner Denton M. A review of simplex optimization in analytical chemistry. CRC Critical Reviews in Analytical Chemistry, 7/3, 187-202, 1978

Anhang

a) Residuen-Varianz

$$s_R^2 = \frac{1}{n-q} \sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \hat{y}_j)^2$$

mit n: Gesamtzahl der Untersuchungen; n_j : Zahl der Werte bei Rubrik j,
k: Zahl unterschiedliche Rubriken (z.B. 2 Temperatur-Einstellungen), q: Zahl Modellkoeffizienten b_i
 y_{ij} : einzelne Untersuchungsgröße, \hat{y}_j : aus Modell berechnete Zielgröße bei der Rubrik j,
 $f = n - q$: Freiheitsgrad

b) F-Wert

$$F - \text{Wert} = \frac{1/(k-q) \sum_{j=1}^k n_j (\hat{y}_j - \bar{y}_j)^2}{1/(n-k) \sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2}$$

mit \bar{y}_j : über die j-te Rubrik gemittelte Zielgröße,
 $f_1 = k - q$: Freiheitsgrad 1, $f_2 = n - k$: Freiheitsgrad 2