Statistische Maße für rechnerische Robustheitsbewertungen CAE gestützter Berechnungsmodelle

Johannes Will[1*], Christian Bucher[1,2]

[1] DYNARDO – Dynamic Software and Engineering GmbH, Weimar

[2] Institut für Strukturmechanik, Bauhaus-Universität Weimar

Zusammenfassung

In der virtuellen Produktentwicklung werden heute zur Untersuchung des Einflusses von Eingangsstreuungen auf wichtige Ergebnisgrößen in zunehmenden Maß stochastische Berechnungen durchgeführt. Die Bewertung einer Ingenieuraufgabenstellung erfährt durch eine stochastische Betrachtungs- und Berechnungsweise allerdings unter Umständen zusätzliche Unsicherheiten. Zu den heute in deterministischen Berechnungen betrachteten Ursachen von Abweichungen der Berechnungsergebnisse gegenüber beobachteten (i.d.R. gemittelten) Strukturantworten kommen Unsicherheiten der Schätzung statistischer Maße hinzu. Deshalb ist es notwendig, die Vertrauenswürdigkeit der geschätzten statistischen Maße abzusichern. Zentraler Punkt der Vertrauenswürdigkeit statistischer Maße ist eine ausreichende Anzahl von Stichproben (Durchrechnungen) zur sicheren Bestimmung der statistischen Kenngröße. In der Folge wird diskutiert, wie mittels Konfidenzschätzungen die Vertrauenswürdigkeit von Korrelationskoeffizienten beurteilt werden kann. Es wird gezeigt, dass Latin Hypercube Samplings bezüglich der Konfidenz der Schätzung wesentlich weniger Stichproben benötigen als Monte Carlo Samplings. Weiterhin wird diskutiert, wie mittels Bestimmtheitsmaßen geschätzt werden kann, wie viel der Variation einer Ergebnisgröße durch die gefundenen Korrelationen zu Eingangsstreuungen erklärt werden kann.

Keywords: Robustheitsbewertungen, Korrelationsanalyse, Bestimmtheitsmaße, Konfidenzschätzung, Latin Hypercube Sampling

[*] Kontakt: Dr.-Ing. Johannes Will, DYNARDO – Dynamic Software and Engineering GmbH, Luthergasse 1d, D-99423 Weimar, E-Mail: johannes.will@dynardo.de

1 Einführung

In der virtuellen Produktentwicklung werden heute zur Untersuchung des Einflusses von Eingangsstreuungen auf wichtige Ergebnisgrößen in zunehmenden Maß stochastische Berechnungen durchgeführt. Die Bewertung einer Ingenieuraufgabenstellung erfährt durch eine stochastische Betrachtungs- und Berechnungsweise allerdings unter Umständen zusätzliche Unsicherheiten. Zu den heute in deterministischen Berechnungen betrachteten Ursachen von Abweichungen der Berechnungsergebnisse gegenüber beobachteten (i.d.R. gemittelten) Strukturantworten, die zum Beispiel aus der Modellierung oder aus Approximationsfehlern der numerischen Berechnungsverfahren herrühren können, kommen Unsicherheiten der Schätzung statistischer Maße hinzu. Deshalb ist es notwendig, die Vertrauenswürdigkeit der geschätzten statistischen Maße abzusichern.

Das einfachste statistische Maß ist dabei der Mittelwert einer Ergebnisgröße, der sozusagen die Brücke zur herkömmlichen deterministischen Betrachtungsweise ist. Sollen allerdings höhere statistische Momente, wie Varianz und Standardabweichung oder Korrelations- und Variationskoeffizienten zur Beurteilung der Robustheit von Systemantworten herangezogen werden, stellt sich die Frage wie vertrauenswürdig diese Schätzungen statistischer Maße von Systemantworten sind und in welcher Reihenfolge welche Maße zur Beurteilung herangezogen werden sollten.

Zentraler Punkt der Vertrauenswürdigkeit statistischer Maße ist eine ausreichende Anzahl von Stichproben (Durchrechnungen) zur sicheren Bestimmung der statistischen Kenngröße. Bei rechnerischen Robustheitsbewertungen werden in erster Linie die Sensitivitäten wichtiger Systemantworten gegenüber den Eingangsstreuungen untersucht und hierfür Mittelwerte, Variations- sowie Korrelationskoeffizienten und Bestimmtheitsmaße herangezogen. Da zum Beispiel bei der Bestimmung vertrauenswürdiger Korrelationskoeffizienten und Bestimmtheitsmaße die notwendige Anzahl von Stichproben stark vom normalerweise unbekannten Charakter der Korrelation abhängen, ist diese Fragestellung nicht a priori zu beantworten. Eine zu konservative (also zu hohe) Abschätzung der Stichprobengröße wäre wiederum aus Rechenzeitgründen häufig unmöglich. Im Normalfall wird man deshalb mit möglichst wenigen Durchrechnungen die statistischen Maße schätzen wollen und muss gleichzeitig sichern, sich nicht außerhalb der Grenzen statistisch vertrauenswürdiger und damit weitgehend zufälliger Maße zu bewegen.

Nach unseren Erfahrungen ist es zweckmäßig, die Anzahl der Durchrechnungen an den Korrelationskoeffizienten, die für die spezielle Aufgabenstellung noch signifikant sind auszurichten. Es ist dann empfehlenswert mittels

Konfidenzschätzung die statistische Vertrauenswürdigkeit zu sichern sowie mit Bestimmtheitsmaßen die Signifikanz der identifizierten Korrelationen bezüglich der Variation der Antwortgrößen zu untersuchen.

2 Wichtige Maße mathematischer Statistik für Robustheitsbewertungen in Ingenieuranwendungen

Mathematische Statistik beschäftigt sich damit, aufgrund von Kenntnissen über Eigenschaften einer Teilmenge (Stichprobe), die einer möglichen Gesamtmenge entnommen sind, etwas über die entsprechenden Eigenschaften der Gesamtmenge (Grundgesamtheit) auszusagen. Dabei handelt es sich um Merkmale, die zufallsbedingt sind und quantitativer (z.B. Durchmesser einer Welle) oder qualitativer (funktioniert oder funktioniert nicht) Natur sein können.

Die Grundgesamtheit wird mathematisch durch eine Zufallsgröße x_i modelliert. Im quantitativen Fall ist in der Regel x_i das Merkmal selbst, im qualitativen Fall ist es häufig über x_i definiert (z.B. funktioniert, wenn $x_i \geq 0$, funktioniert nicht, wenn $x_i \leq 0$).

Unter einer Stichprobenentnahme vom Umfang N versteht man eine zufällige Auswahl von N Ereignissen aus der Grundgesamtheit. Im Regelfall wird in der mathematischen Statistik davon ausgegangen, dass die Auswahl der Stichproben zufällig und unabhängig voneinander geschieht. In virtuell erzeugten Stichproben würde das einem Monte Carlo Sampling entsprechen.

Wichtige Eigenschaften von Zufallsgrößen können aus konkreten Stichprobenumfängen ($x_i^{(k)}$) wie folgt geschätzt werden:

arithmetischer Mittelwert $$\bar{x}_i = \frac{1}{N}\sum_{i=1}^{N} x_i^{(k)} \qquad \text{(Gl. 2-1)}$$

arithmetische Varianz $$\sigma_{x_i}^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i^{(k)} - \bar{x}_i\right)^2 \qquad \text{(Gl. 2-2)}$$

wobei σ der Standardabweichung (mittlere quadratische Abweichung) entspricht

Variationsbreite $$R = x_{i-\max}^{(k)} - x_{i-\min}^{(k)} \qquad \text{(Gl. 2-3)}$$

Variationskoeffizient $$COV_{x_i} = \frac{\sigma_{x_i}}{\bar{x}_i} \qquad \text{(Gl. 2-4)}$$

Zentrale Eigenschaft einer zufälligen Größe ist deren Verteilung bzw. Verteilungsfunktion. Um bei einem quantitativen Merkmal eine erste Vorstellung von der Verteilung zu erhalten, konstruiert man so genannte Histogramme, Stufenbilder relativer Häufigkeit der Stichproben in einer gleichmäßigen Intervalleinteilung des Wertebereichs der Zufallsgröße. Zur Prüfung welche Verteilungshypothese für die Beschreibung der Zufallsgröße geeignet ist, können Signifikanztests für verschiedene Verteilungshypothesen durchgeführt werden und die Verteilung mit dem besten Fit ausgewählt und in das Histogramm gezeichnet werden. Aus den Histogrammen oder aus den Verteilungsfunktionen können dann Quantilwerte beziehungsweise die Werte der zufälligen Antwortgröße mit den dazugehörigen Wahrscheinlichkeiten geschätzt werden. Da bei rechnerischen Robustheitsbewertungen in der Regel nur relativ wenige Stichproben berechnet werden, sollten nur Wahrscheinlichkeiten relativ häufiger Ereignisse (> 1%) ermittelt und bewertet werden. Für eine sichere Bestimmung kleinerer Wahrscheinlichkeiten werden Methoden der Zuverlässigkeitsanalyse empfohlen.

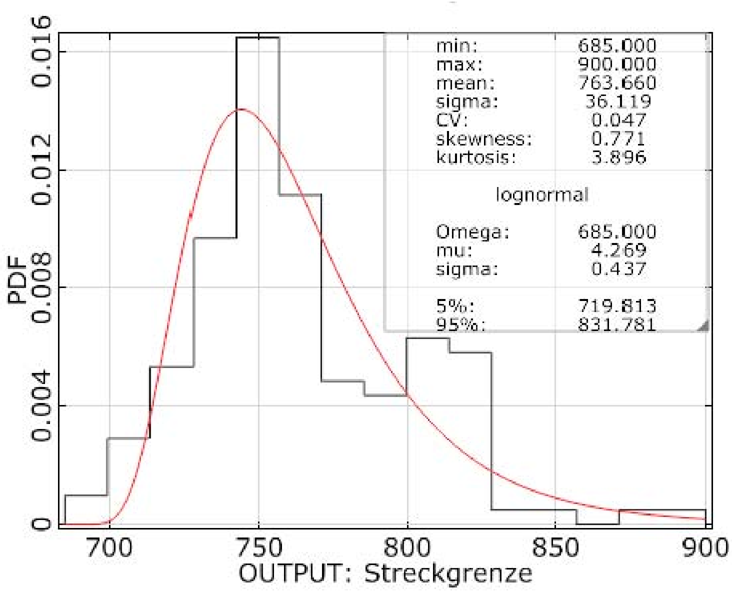

Abbildung 1 Histogramm mit Variationsbreite, Mittelwert, Standardabweichung, Variationskoeffizient (CV) sowie gefitteter Verteilungsfunktion

Neben statistischen Maßen der einzelnen zufälligen Größen (Eingangs- und Ergebnisgrößen) sind statistische Maße über deren Zusammenhang, hier insbesondere zwischen den Eingangs- und Ergebnisgrößen für Ingenieuraufgabenstellungen der Sensitivitätsanalyse und Robustheitsbewertung besonders wichtig. Hiefür werden in der mathematischen Statistik Regressions- und Korrelationsanalysen durchgeführt. Während sich die Regressionsanalyse mit der funktionalen Art des Zusammenhangs beschäftigt, ermittelt die Korrelationsanalyse den quantitativen Grad des Zusammenhangs.

3 Korrelationsanalyse

Der Grad des linearen Zusammenhangs zweier zufälliger Ereignisse wird durch den Korrelationskoeffizienten ρ_{ij} angegeben.

$$\rho_{ij} = \frac{S_{ij}}{\sqrt{S_{ii} S_{jj}}} \qquad \text{(Gl. 3-1)}$$

3.1 Lineare Korrelationskoeffizienten

Der gebräuchlichsten Korrelationsschätzung liegt eine lineare Regressionshypothese zugrunde. Der lineare Korrelationskoeffizient (auch Pearson Product Moment Correlation genannt):

$$\begin{aligned} S_{ij} &= \sum_{k=1}^{N} \left(x_i^{(k)} - \bar{x}_i\right)\left(x_j^{(k)} - \bar{x}_j\right) \\ S_{ii} &= \sigma_{x_i}^2 (N-1) \\ S_{jj} &= \sigma_{x_j}^2 (N-1) \end{aligned} \qquad \text{(Gl. 3-2)}$$

$$\rho_{ij} = \frac{1}{N-1} \frac{\sum_{k=1}^{N} \left(x_i^{(k)} - \bar{x}_i\right)\left(x_j^{(k)} - \bar{x}_j\right)}{\sigma_{x_i} \sigma_{x_j}} \qquad \text{(Gl. 3-3)}$$

bestimmt dabei die Richtung und die Übereinstimmung einer Ausgleichsgeraden mit der Stichprobe und kann dabei Werte zwischen +1 und -1 annehmen. Ist der Korrelationskoeffizient +1 existiert ein positiver linearer Zusammenhang, beide Werte werden mit miteinander größer oder kleiner. Ist der Korrelationskoeffizient -1, existiert ein negativer linearer Zusammenhang, wenn ein Wert größer wird, wird der andere kleiner.
In der Literatur werden dabei Korrelationen mit Korrelationskoeffizienten > 0.80 bzw. 0.70 als starke Korrelationen, sowie Korrelationen mit Korrelationskoeffizienten < 0.50 als schwache Korrelationen bezeichnet. Ob ein Korrelationskoeffizient für einen Zusammenhang signifikant ist, hängt allerdings immer auch davon ab, welche weiteren Korrelationen existieren. Bei kleinen Korrelationskoeffizienten (< 0.30 bzw. 0.20) spricht man davon, dass kein nennenswerter linearer Zusammenhang festgestellt werden kann. Dann ist allerdings zu prüfen, ob die zwei zufälligen Variablen voneinander unabhängig sind, also wirklich kein Zusammenhang besteht oder ob nichtlineare Zusammenhänge identifiziert werden können.

3.2 Konfidenzschätzungen linearer Korrelationskoeffizienten

Alle zuvor aufgeführten statistischen Maße sind im Sinne der mathematischen Statistik Punktschätzungen. Sie sind dadurch gekennzeichnet, dass diese Schätzwerte aus einer anderen Stichprobe der Grundgesamtheit andere Werte annehmen. Deshalb können diese Maße weitgehend wertlos sein, wenn über die Sicherheit der Schätzung nichts bekannt ist. Es ist daher notwendig, die Genauigkeit und Sicherheit der Schätzung zu untersuchen. Derartige Angaben liefern Schätzungen von Konfidenzintervallen. Weil die Korrelationskoeffizienten i.d.R. die Basis der Sensitivitäts- und Robustheitsbewertungen darstellen, werden in der Folge Konfidenzschätzungen der linearen Korrelationskoeffizienten diskutiert.

Dabei sucht man ein Intervall ($I\rho$) um den Schätzwert (ρ_{ij}) so anzugeben, dass dieser mit einer Wahrscheinlichkeit ($1-\alpha$) den unbekannten Parameter überdeckt. Dabei heißt $\gamma = 1-\alpha$ Konfidenzniveau oder auch statistische Sicherheit und α ist die tolerierbare Irrtumswahrscheinlichkeit, die Wahrscheinlichkeit, dass der geschätzte Korrelationskoeffizient außerhalb des Konfidenzintervalls liegt.

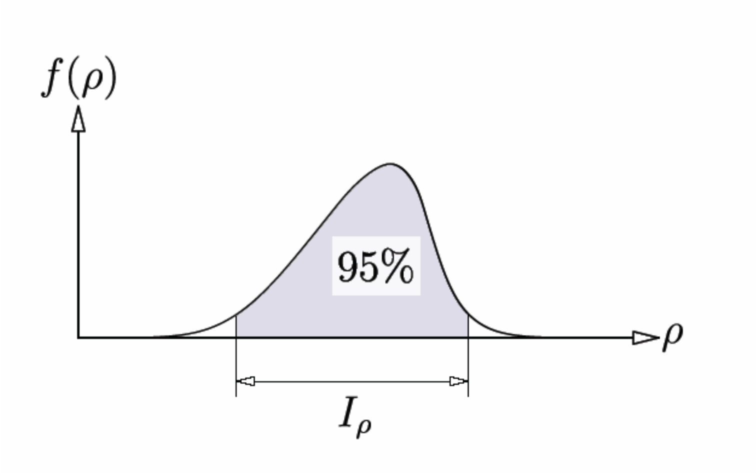

Abbildung 2 Konfidenzintervall für Korrelationskoeffizienten mit einer Irrtumswahrscheinlichkeit von 5%

Zur Schätzung der Konfidenzintervalle werden wegen der Nichtnormalverteilung der Schätzer der linearen Korrelationskoeffizienten die Korrelationskoeffizienten mittels der Fischer-z-Transformation in einen normalverteilten Raum projiziert. Das Konfidenzintervall ist dann gegeben durch:

$$\left[\tanh\left(z_{ij} - \frac{z_c}{\sqrt{N-3}}\right), \tanh\left(z_{ij} + \frac{z_c}{\sqrt{N-3}}\right)\right] \qquad \text{(Gl. 3-4)}$$

Die Irrtumswahrscheinlichkeit soll für alle zu schätzenden Korrelationskoeffizienten $M = m\frac{(m-1)}{2}$ zwischen m Eingangs- und Ergebnisgrößen der gesamten Korrelationsmatrix gelten. Deshalb wird der kritische Wert z_c mittels Bonferroni korrigierten Werten des Konfidenzniveaus $\alpha' = \frac{\alpha}{M}$ ermittelt. Damit ergibt sich für eine Sicherstellung eines geeigneten Konfidenzintervalls eine Abhängigkeit zwischen der notwendigen Anzahl von Stichproben (N) und der Anzahl der Eingangs- sowie der Ergebnisvariablen (m).

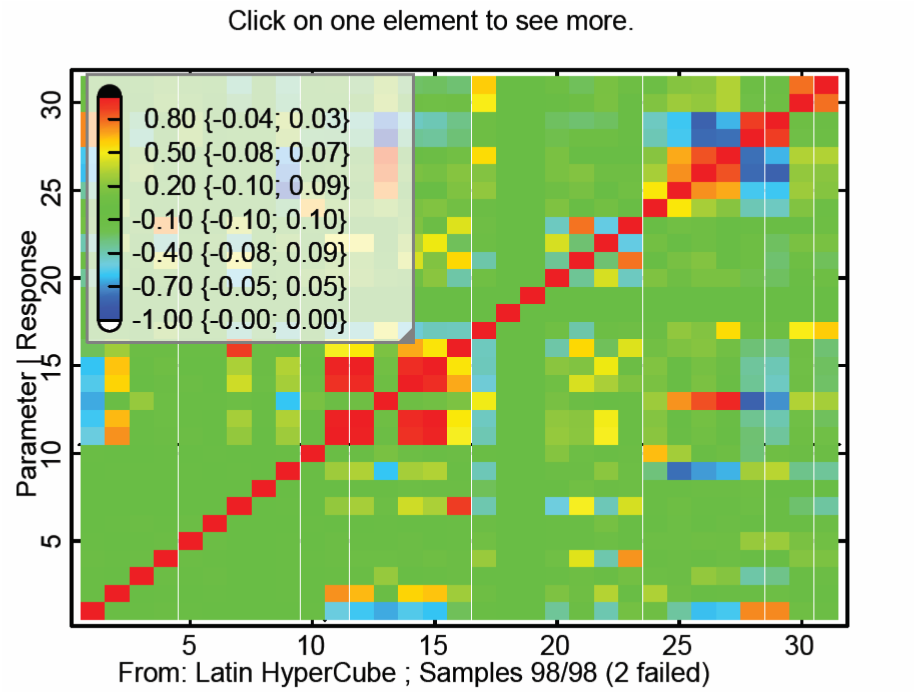

Abbildung 3 Matrix linearer Korrelationen mit Konfidenzintervallen

3.3 Einfluss der Samplingmethode auf das Konfidenzintervall linearer Korrelationskoeffizienten

Zur Schätzung vertrauenswürdiger Korrelationskoeffizienten ist eine geeignete Anzahl von Stichproben notwendig. In der Literatur reichen dabei die Empfehlungen von mindestens Anzahl der Einganggrößen $n+1$ bis n^2 für Monte Carlo Samplings. Im Gegensatz zu „beobachtender“ Statistik, wo in der Regel von Monte Carlo verteilten Stichprobensets ausgegangen werden muss, kann bei virtuellen, zu berechnenden Stichproben durch varianzminimierende Samplingverfahren die statistische Sicherheit der Schätzung von Korrelationskoeffizienten deutlich erhöht werden. Deshalb soll in der Folge der Einfluss von Latin Hypercube Sampling auf die Konfidenzintervalle untersucht werden. In den folgenden Vergleichen werden die Konfidenzintervalle der Korrelationskoeffizienten für eine Irrtumswahrscheinlichkeit von 5%

(Konfidenzlevel 95%) von Monte Carlo Samplings und Latin Hypercube Samplings ausgewertet. Es werden Schätzungen für die Konfidenzintervalle aus 1000 Wiederholungen der Samplings angegeben.

N	Korrelationskoeffizient ρ				
	0.0	0.3	0.5	0.7	0.9
10	1.261	1.231	1.054	0.757	0.299
30	0.712	0.682	0.557	0.381	0.149
100	0.409	0.374	0.306	0.199	0.079
300	0.230	0.209	0.170	0.116	0.045
1000	0.124	0.115	0.093	0.062	0.023

Tabelle 1Konfidenzintervalle zur Schätzung eines Korrelationskoeffizienten (Konfidenzlevel 95%) für Monte Carlo Sampling

N	Korrelationskoeffizient ρ				
	0.0	0.3	0.5	0.7	0.9
10	0.420	0.382	0.260	0.158	0.035
30	0.197	0.194	0.139	0.073	0.018
100	0.111	0.101	0.071	0.042	0.009
300	0.065	0.057	0.042	0.024	0.006
1000	0.038	0.033	0.025	0.014	0.003

Tabelle 2 Konfidenzintervalle zur Schätzung eines Korrelationskoeffizienten (Konfidenzlevel 95%) für Latin Hypercube Sampling

Tabelle 1 und 2 zeigen die Abhängigkeit der Konfidenzintervalle der Schätzung eines Korrelationskoeffizienten von der Anzahl der Stichproben. Für die Schätzung der Korrelationskoeffizienten ist eine signifikante Verbesserung der Schätzung bei Verwendung eines Latin Hypercube Sampling deutlich zu erkennen. Zum Beispiel wird bei der Schätzung eines Korrelationskoeffizienten von 0.5 bei Verwendung eines Monte Carlo Sampling mit 1000 Stichproben ein Konfidenzintervall von 0.093 sowie bei Latin Hypercube Sampling schon mit 100 Stichproben ein Konfidenzintervall von 0.071 erreicht. Mit 1000 Stichproben kann beim Latin Hypercube Sampling ein Konfidenzintervall von 0.025 erreicht werden, was eine signifikant bessere Schätzung als mit Monte Carlo Sampling darstellt. Zusammenfassend kann festgestellt werden, dass Latin Hypercube Sampling eine um den Faktor ~12 geringere notwendige Anzahl von Stützstellen benötigt, um vergleichbare Konfidenzintervalle wie das Monte Carlo Sampling aufzuweisen.

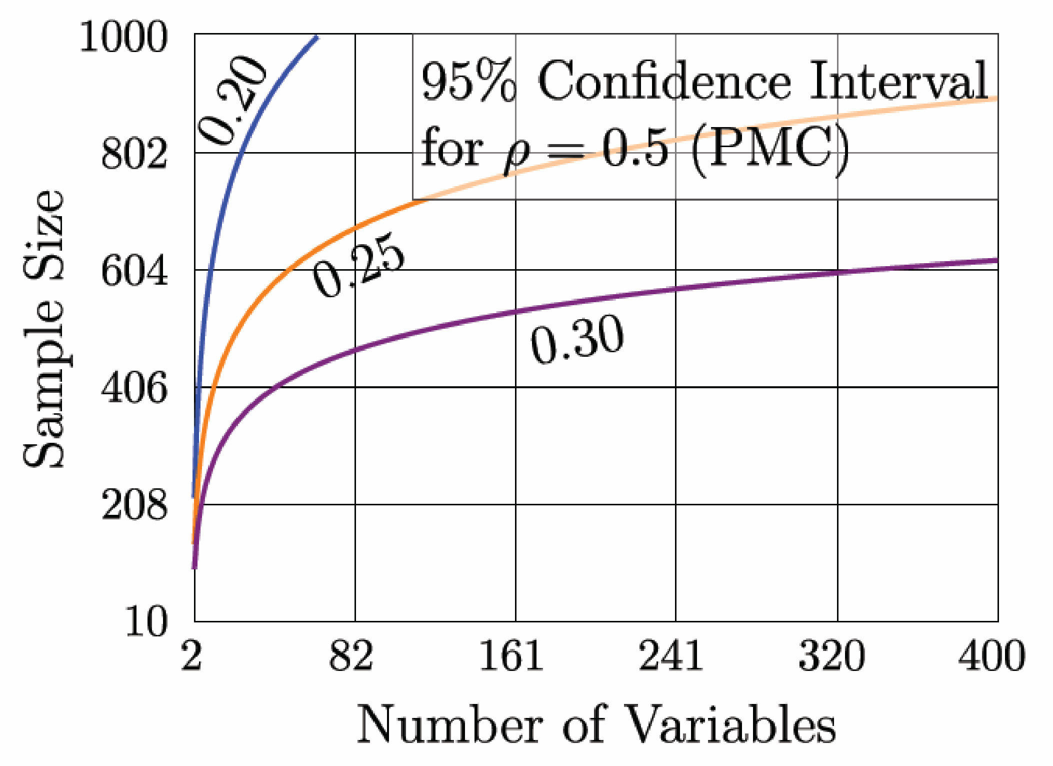

Abbildung 4 Konfidenzintervalle (Konfidenzlevel 95%) für M-Korrelationskoeffizienten 0.5 Monte Carlo Sampling

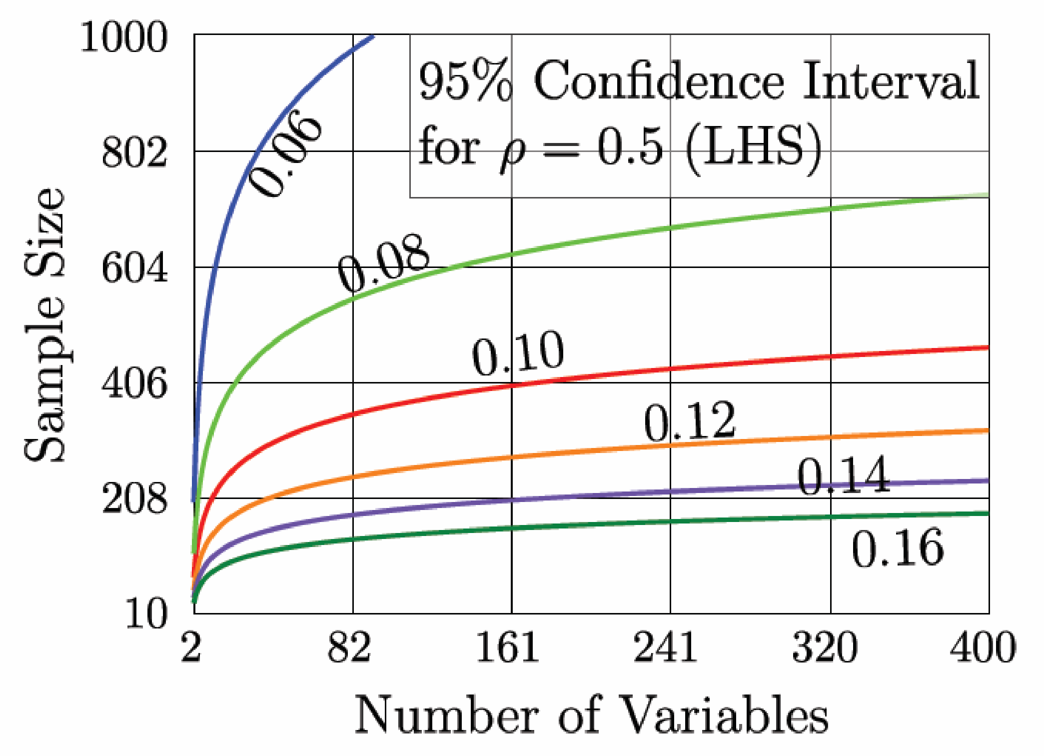

Abbildung 5 Konfidenzintervalle (Konfidenzlevel 95%) für M-Korrelationskoeffizienten 0.5 Latin Hypercube Sampling

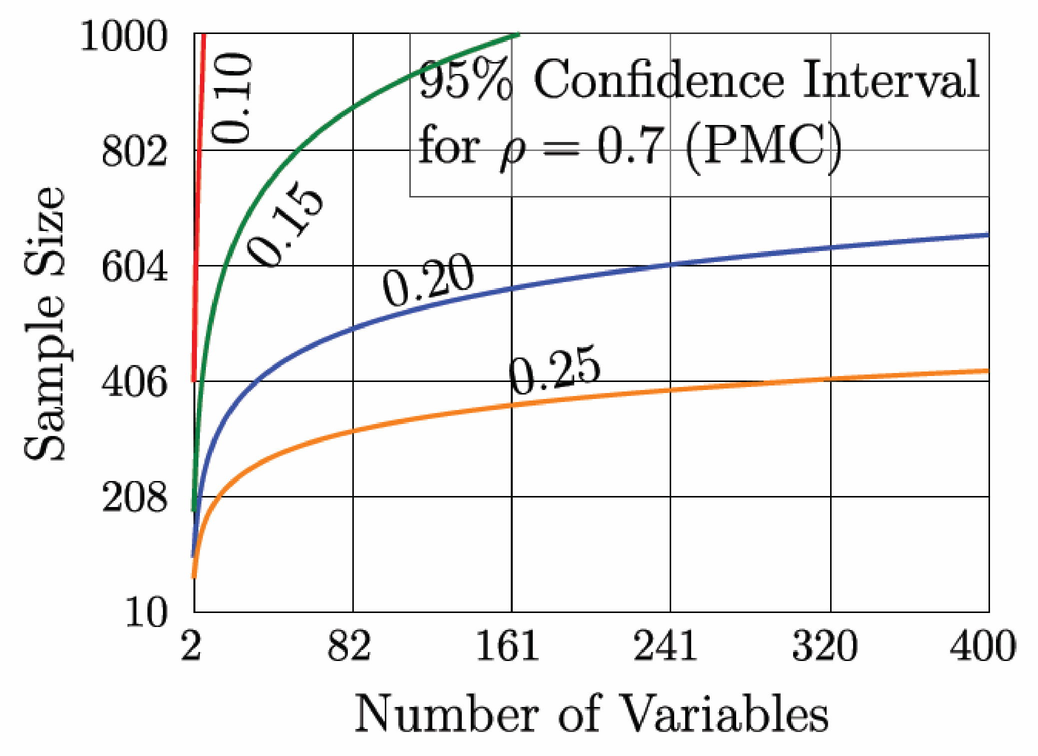

Abbildung 6 Konfidenzintervalle (Konfidenzlevel 95%) für M-Korrelationskoeffizienten 0.7 Monte Carlo Sampling

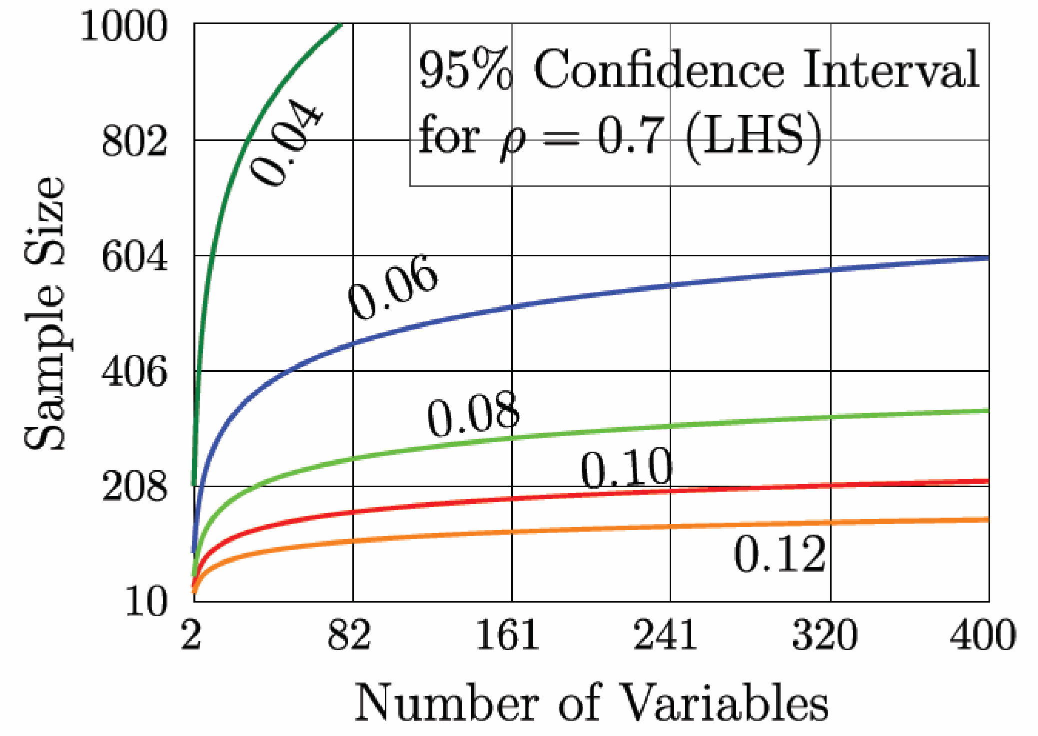

Abbildung 7 Konfidenzintervalle (Konfidenzlevel 95%) für M-Korrelationskoeffizienten 0.7 Latin Hypercube Sampling

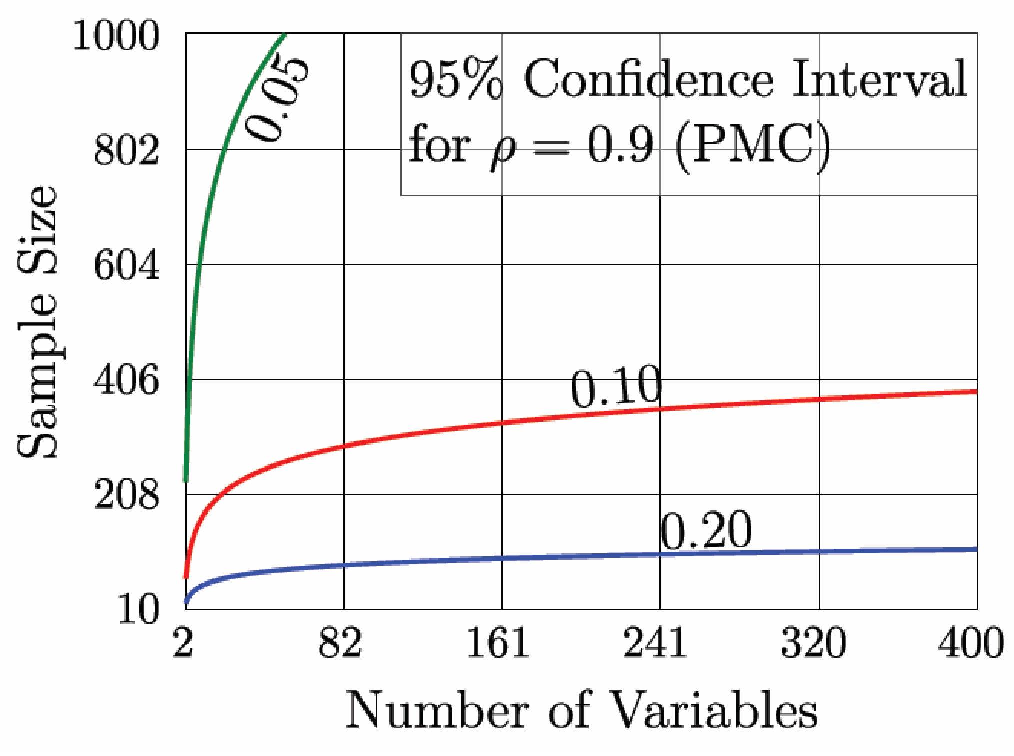

Abbildung 8 Konfidenzintervalle (Konfidenzlevel 95%) für M-Korrelationskoeffizienten 0.9 Monte Carlo Sampling

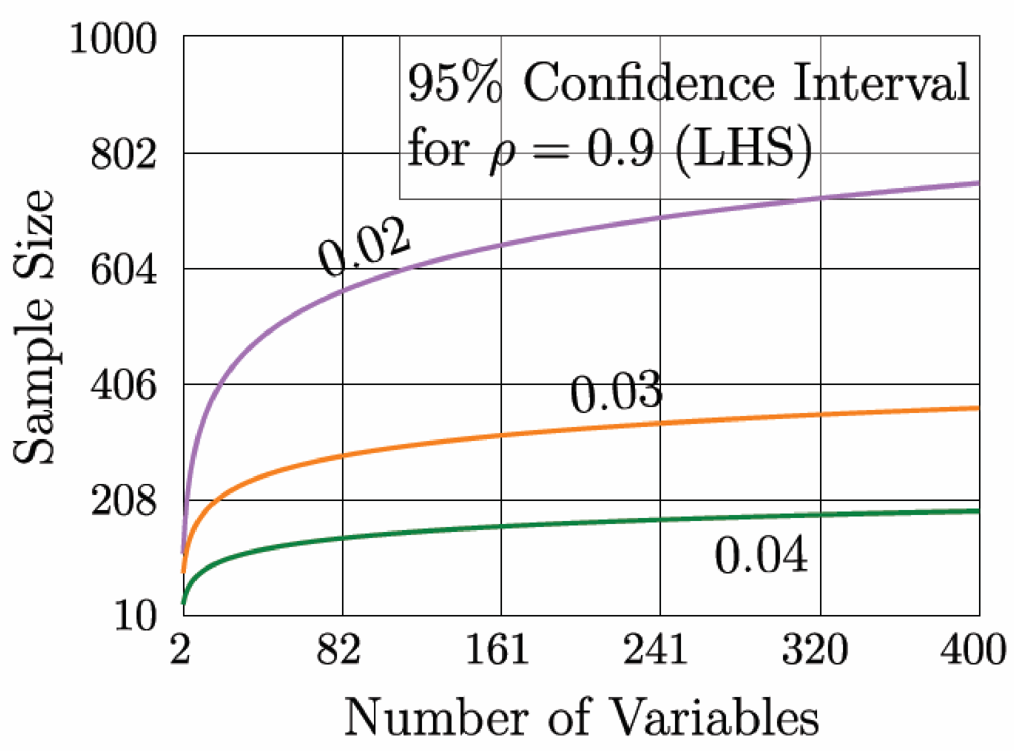

Abbildung 9 Konfidenzintervalle (Konfidenzlevel 95%) für M-Korrelationskoeffizienten 0.9 Latin Hypercube Sampling

Abbildungen 4 bis 9 zeigt die Abhängigkeit die Konfidenzintervalle der Schätzung von M Korrelationskoeffizienten aus m Eingangs- und Ergebnisgrößen (Number of Variables)von der Anzahl der Stichproben (Sample Size). Die Konfidenzintervalle wurden hier nach Kap. 3.2 mittels Fischer's z-Transformation und Bonferroni Korrektur ermittelt. Dabei wurde ein Faktor von 12 in der Stichprobengröße zwischen Monte Carlo Sampling und Latin Hypercube Sampling berücksichtigt.

3.4 Bestimmtheitsmaße

Konfidenzintervalle von Korrelationskoeffizienten geben die Vertrauensbereiche der Schätzung an. Aus Sicht praktischer Aufgabenstellungen sind Bestimmtheitsmaße weitere wichtige Kriterien über die Aussagekraft der Korrelationsanalyse. Das Bestimmtheitsmaß ist ein quantitatives Maß wie viel der Variation einer Größe mittels der gefundenen Korrelationen zu den Eingangsgrößen erklärt werden kann. Das Bestimmtheitsmaß kann dabei Werte zwischen 0 und 1 beziehungsweise 0 bis 100% annehmen

Das Bestimmtheitsmaß einer linearen Beziehung zwischen zwei unkorrelierten Zufallsgrößen ist als das Quadrat des Korrelationskoeffizienten definiert.

$$R_{ij}^2 = \frac{S_{ij}^2}{S_{ii} S_{jj}} \qquad \text{(Gl. 3-5)}$$

Bestimmtheitsmaße werden allgemeiner (z.B. in optiSLang) aus den Korrelationen der beobachteten Ausgangsgrößen z und der über Regression von gefitteten Werten $x_j = z^{(k)}(x_i)$ ermittelt.

Wird das Bestimmtheitsmaß einer Ergebnisgröße bezüglich aller n-Eingangsgrößen ermittelt, ist eine quantitative Aussage gefunden, wie viel der Variation der Ergebnisgröße mit der zugrunde liegenden Regressionshypothese erklärt werden kann. Existiert also bei linearer Regressionshypothese eine nennenswerte Abweichung zu 100 %, existieren entweder (außerhalb linearer Zusammenhänge) nennenswerte weitere Korrelationen (z.B. quadratische Zusammenhänge, Clusterungen) oder die Schätzung der Korrelationskoeffizienten ist zu ungenau oder die Variation der Ergebnisgrößen enthält nennenswert Rauschen, z.B. aus dem CAE-Berechnungsprozess oder dem Extraktionsprozedere der Ergebnisgröße.

Bei der Schätzung der Bestimmtheitsmaße von mehreren Eingangsvariablen zu einer Ausgabegröße können sich die Fehler des Schätzers vor allem bei kleiner Stichprobenanzahl und kleinen Korrelationskoeffizienten nennenswert summieren. Diese Korrelationskoeffizienten und dazugehörigen Bestimmtheitsmaße kleiner Korrelationskoeffizienten (z.B. < 0.3) unterliegen bei kleinen Stichproben großen statistischen Fehlern und erhöhen nur scheinbar die Bestimmtheit des Modells. Zur Prüfung dieses Sachverhaltes kann ein so genannter $adjusted\ R_j^2$ herangezogen werden:

$$adjusted\ R_j^{\ 2} = \sum_{i=1}^{n} \left(R_i^{\ 2} - (1 - R_i^{\ 2}) \right) \left(\frac{k-1}{n-k} \right) \qquad \text{Gl. 3-6}$$

Ein deutlich kleinerer $adjusted\ R_j^{\ 2}$ als R_j^2 zeigt dabei an, dass zahlreiche nicht signifikante Eingangsgrößen beim Bestimmtheitsmaß berücksichtigt worden sind. Zeigen die Bestimmtheitsmaße $R_j^{\ 2}, adjusted\ R_j^{\ 2}$ große Unterschiede an, ist davon auszugehen, dass die Schätzung der Bestimmtheitsmaße nicht vertrauenswürdig ist und sie sollte für einen kleineren Satz von signifikanten Eingangsvariablen wiederholt werden. Die Signifikanz der Eingangsvariablen wird dabei anhand der Größe des Korrelationskoeffizienten festgestellt.

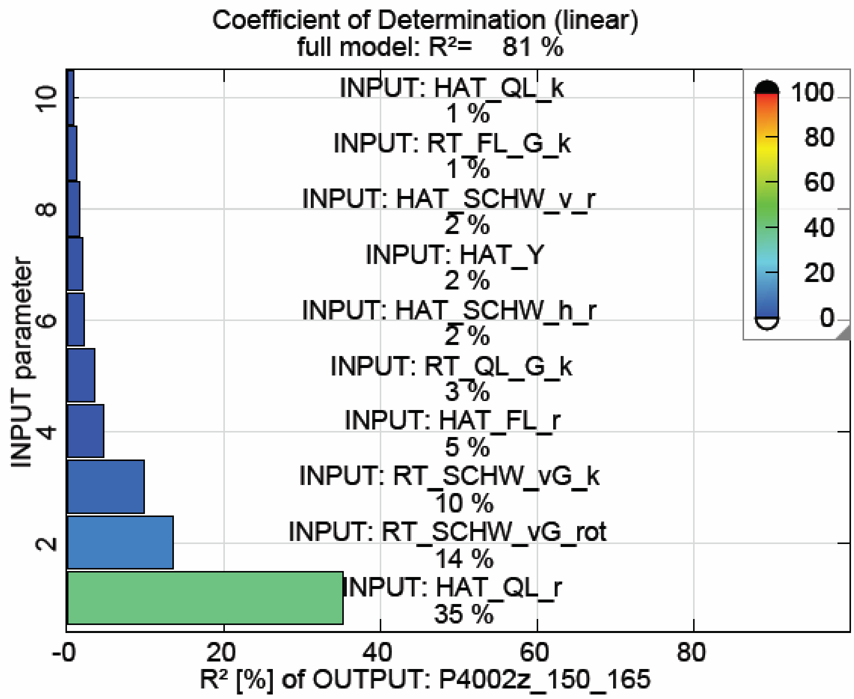

Abbildung 10 Bestimmtheitsmaße einer Ergebnisgröße sowie Bestimmtheitsmaße der Anteile einzelner streuender Eingangsvariablen

3.5 Quadratische Korrelationskoeffizienten

In der Literatur wird immer wieder darauf hingewiesen, dass die Feststellung geringer linearer Korrelationen kein hinreichender Grund für die Annahme ist, dass keine wichtigen Zusammenhänge zwischen den Variablen existieren. Geht man zum Beispiel davon aus, dass zwei Variablen einen ideal quadratischen Zusammenhang haben, ist ihr linearer Korrelationskoeffizient 0.0 und ihr quadratischer Korrelationskoeffizient 1.0. Deshalb wurde die Korrelationsanalyse in optiSLang auf quadratische Regressionshypothese erweitert.

Quadratische Korrelationskoeffizienten werden in der Regel über quadratische Regression der Antwortgrößen x_j zu den Eingangsgrößen x_i ermittelt.

$$x_j = A_i + B_i x_i + C_i x_i^2 = z(x_i) \qquad \text{Gl. 3-7}$$

Die Regressionskoeffizienten A_i, B_i, C_i werden standardmäßig über Minimierung der Fehlerquadrate über die Stichproben $x_i^{(k)}, x_j^{(k)}, k = 1...N$ ermittelt. Die quadratischen Korrelationskoeffizienten ergeben sich danach durch Einsetzen der gefitteten Werte $z^{(k)}$ an Stelle von $x_i^{(k)}$ sowie der $x_j^{(k)}$ in Gleichung 3.3. Die Werte der Korrelationskoeffizienten gegenüber der quadratischen Funktion 3.7 enthalten dabei sowohl den linearen als auch den quadratischen Anteil und liegen zwischen 0.0 und 1.0. Die Bestimmtheitsmaße werden analog linearer Regressionshypothese aus Gleichung 3-6 ermittelt. Da die Regression in Gleichung 3-7 nicht einfach die Rollen von x_i und x_j vertauschen kann, ist die sich ergebende Matrix der quadratischen Korrelationskoeffizienten nicht symmetrisch.

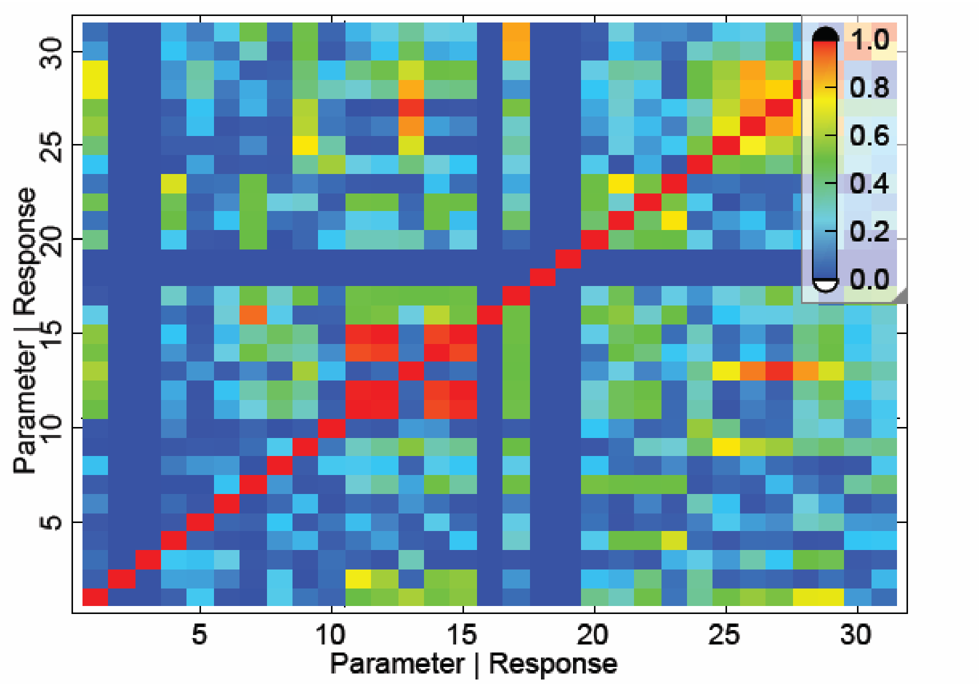

Abbildung 11 Matrix quadratischer Korrelationen

4 Zusammenfassung

Aus den Studien und bisherigen Erfahrungen mit rechnerischen Robustheitsbewertungen lässt sich für praktische Aufgabenstellung folgende Empfehlung ableiten:

a) Verwendung eines Latin Hypercube Sampling

b) Ermittlung der Anzahl notwendiger Durchrechnungen für die Schätzung von erwarteten Korrelationskoeffizienten mit einem gewünschten Vertrauensbereich. Werden zum Beispiel lineare Korrelationskoeffizienten von 0.50 mit einem Konfidenzintervall 0.12 für ein Konfidenzlevel von 95% für alle Korrelationskoeffizienten gesucht, kann unter Berücksichtigung der Anzahl der Eingangs- und Ergebnisgrößen aus Abbildung 12 die notwendige Anzahl von Stichproben abgeschätzt werden. So würde sich für eine Anzahl von 40 Eingangs- und Ergebnisgrößen eine notwendige Anzahl von 200 Stichproben ergeben. Damit würde ein Korrelationskoeffizient von 0.50 mit 95-prozentiger Sicherheit innerhalb eines Fehlerintervalls der Größe ±0.06 geschätzt werden.

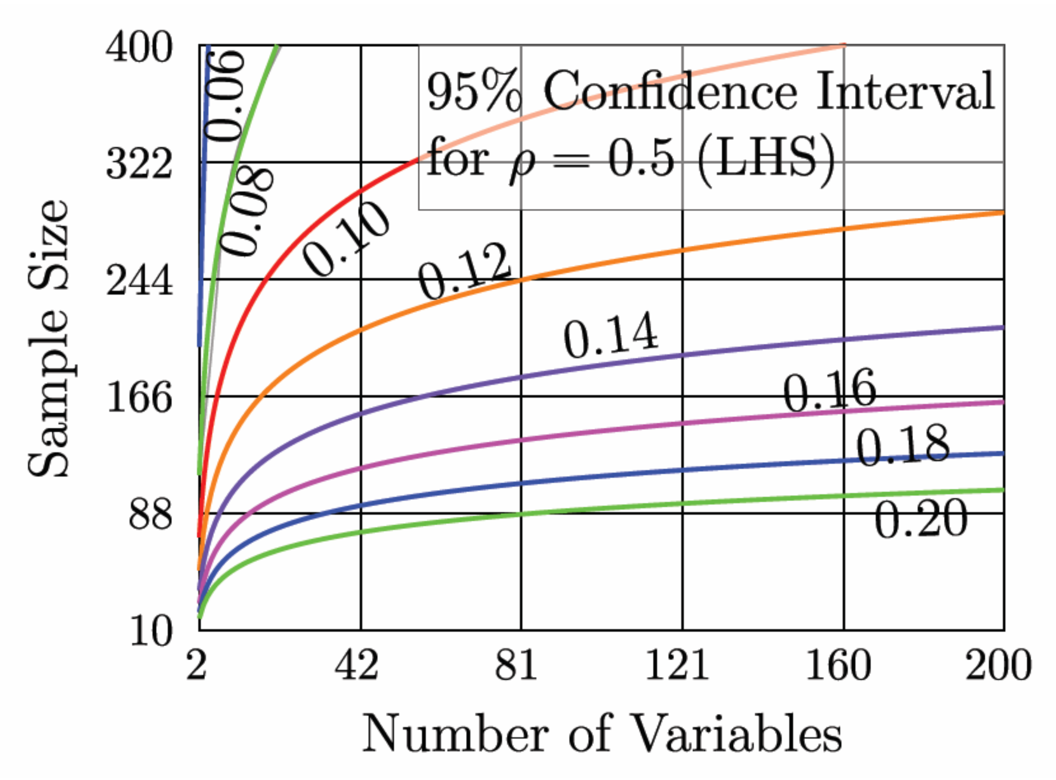

Abbildung 12 Konfidenzintervalle (Konfidenzlevel 95%) für M-Korrelationskoeffizienten der Größe 0.5, Latin Hypercube Sampling

c) Nach der Durchrechnung können in folgender Reihenfolge die statistischen Maße zur Bewertung herangezogen werden:
   a. Variationsbreite und Variationskoeffizient. Überschreitet die Variationsbreite der Ergebnisgrößen zulässige Werte beziehungsweise sind die Variationskoeffizienten wichtiger Ergebnisgrößen signifikant größer als die mit diesen Ergebnisgrößen verknüpften Eingangsgrößen, sollten mittels

Bestimmtheitsmaßen und Korrelationskoeffizienten die sensiblen Eingangsgrößen identifiziert werden.

b. Bestimmtheitsmaße wichtiger Ergebnisgrößen linearer Korrelationshypothese. Ist das Bestimmtheitsmaß (R^2, *adjusted* R^2) hoch (> 80 bis 100%) kann über lineare Korrelationen die Variation der betreffenden Ergebnisgröße in ausreichendem Maße erklärt werden und die Korrelationen zu den wichtigsten Eingangsgrößen können identifiziert werden und in den dazugehörigen Anthill-Plots auf Plausibilität geprüft werden.

c. Sind die Bestimmtheitsmaße linearer Korrelation kleiner als 80% sollten für diese Ergebnisgrößen die quadratischen+linearen Korrelationskoeffizienten und Bestimmtheitsmaße geprüft werden. Aufgefundene signifikante quadratische Korrelationen sollten in den dazugehörigen Anthill-Plots auf Plausibilität geprüft werden.

d. Kann die Variation der Ergebnisgröße nicht in ausreichendem Maße durch lineare und quadratische Korrelationen (geringe Bestimmtheitsmaße R^2 < 80%) erklärt werden, sind im Zweifelsfall alle Anthill Plots der Eingangsgrößen zur betreffenden Ergebnisgröße auf Nichtlinearitäten, wie Clusterungen oder Verzweigungen zu prüfen.

e. Können keine signifikanten linearen und quadratischen Korrelationen beziehungsweise Nichtlinearitäten aufgefunden werden, sollten weitere potentielle Ursachen der Variation der Ergebnisgröße überprüft werden. Resultieren geringe Bestimmtheiten von Ergebnisgrößen aus dem Anregen von Verzeigungsproblemen durch numerischen Rauschen, sollten die mit diesen Verzweigungsproblemen verbundenen physikalischen streuenden Eingangsgrößen identifiziert und ins Modell integriert werden. Ist die Bestimmtheit wichtiger Ergebnisgröße dann immer noch gering, kann es sinnvoll sein, die Konstruktion bezüglich der Ergebnisgröße bestimmter auszulegen. Wenn die Variation der Ergebnisgröße in einem hohen Maße von Approximationsfehlern der CAE-Berechnung beeinflusst wird oder die Extraktion der Ergebnisgrößen Streuungen (z.B. durch Abhängigkeiten von Ausgabezeitschritten oder Filtern) erzeugt oder geringe Bestimmtheiten aus Unzulänglichkeiten der Modelle resultieren, sollten diese Unzulänglichkeiten behoben werden.

Literatur

[1] U. Bourgund, C. Bucher: „Importance Sampling Procedure Using Design Point (ISPUD) – a Users Manual, Bericht Nr. 8-86, Institut für Mechanik, Universität Innsbruck, 1986

[2] I. Bronstein; K. Semendjajew; G. Musiol: Taschenbuch der Mathematik, Vieweg Verlag, 5. Auflage 2000

[3] C. Bucher: Adaptive sampling-an iterative fast Monte Carlo procedure. Structural Safety, 5(2):119–126, 1988.

[4] C. Bucher; U. Bourgund: A fast and efficient response surface approach for structural reliability problems. Structural Safety, 7:57–66, 1990.

[5] C. Bucher, Y. Schorling; W. A. Wall: SLang–the structural language, a tool for computational stochastic structural analysis. In S. Sture, editor, Engineering Mechanics, Proceedings of the 10th Conference, pages 1123–1126. ASCE, 1995.

[6] V. Bayer; C. Bucher: Importance sampling for first passage problems of nonlinear structures. Probabilistic Engineering Mechanics, 14:27–32, 1999.

[7] M. Macke, C. Bucher: Importance sampling for randomly excited dynamical systems. Journal of Sound and Vibration, (268):269–290, 2003.

[8] optiSLang - the Optimizing Structural Language Version 2.1, DYNARDO, Weimar, 2005, www.dynardo.de

[9] PAPULA, L.: Mathematik für Ingenieure und Naturwissenschaftler, Band 3 Vektoranalysis, Wahrscheinlichkeitsrechnung, Mathematische Statistik, Fehler- und Ausgleichsrechnung, Vieweg Verlag, 2001

[10] J. Unger, D. Roos: Investigation and benchmarks of algorithms for reliability analysis, Proceedings Weimarer Optimierungs- und Stochastiktage 1.0, Dezember 2004, Weimar, www.dynardo.de

[11] Will, J.; Möller, J-St.; Bauer, E.: Robustheitsbewertungen des Fahrkomfortverhaltens an Gesamtfahrzeugmodellen mittels stochastischer Analyse, VDI-Berichte Nr.1846, 2004, S.505-527

[12] Will, J.; Baldauf, H.: Robustheitsbewertungen bezüglich der virtuellen Auslegung passiver Fahrzeugsicherheit; Proceedings Weimarer Optimierung- und Stochastiktage 2.0, 2005, Weimar, Germany

[13] Will, J.; Bucher, C.; Ganser, M.; Grossenbacher, K.: Berechnung und Visualisierung statistischer Maße auf FE-Strukturen für Umformsimulationen; Proceedings Weimarer Optimierung- und Stochastiktage 2.0, 2005, Weimar, Germany