

# Szenariogenerierung zur Modellierung der stochastischen Auspeiselasten in einem Gastransportnetz

Dr. **T. Koch**, Zuse-Institut Berlin, Dipl.-Math. **H. Leövey**, Dr. **R. Mirkov**,  
Prof. Dr. **W. Römisch**, Dipl.-Math. **I. Wegner-Specht**,  
Institut für Mathematik, Humboldt-Universität Berlin

## Kurzfassung

Wir betrachten ein Gastransportnetz mit mehreren hundert Auspeisepunkten, für die historische Daten vorliegen. Als Grundlage für die Berechnung der technischen Kapazität werden mögliche Szenarien für den Gasfluss an sämtlichen Auspeisepunkten und für alle Temperaturbereiche benötigt. Um den Rechenaufwand vertretbar zu halten, ist es vorteilhaft, wenn die Anzahl der Szenarien nicht zu groß ist. Im Folgenden wird eine Methodik beschrieben, wie unter Verwendung der historischen Daten für jeden betrachteten Temperaturbereich eine Wahrscheinlichkeitsverteilung geschätzt und unter Verwendung dieser Verteilung repräsentative Szenarien nebst der Wahrscheinlichkeiten ihres Auftretens generiert werden.

## 1 Einleitung und Datenbeschreibung

Als Datenbasis für die Generierung von Szenarien dienen historische Daten der Flussmengen an den Auspeisepunkten des H-Gas-Transportnetzes der Open Grid Europe GmbH, die für einen Zeitraum von mehr als sechs Jahren vorliegen. Darin enthalten sind außer den Gasflüssen die meteorologischen Tagesmitteltemperaturen sowie die geographische Lage und Art der Punkte. Im Rahmen einer deskriptiven Datenanalyse wurde eine Typisierung der Datensätze durchgeführt, die einen Überblick über die Vielfalt bzgl. ihrer Temperaturabhängigkeit und ihrer sonstigen Charakteristika gibt. Diese Typisierung findet bei der Filterung der Datensätze zur

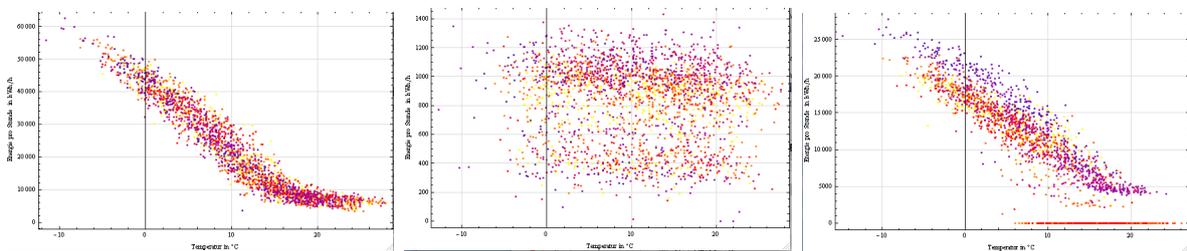


Bild 1: Datensätze von Auspeisepunkten mit Stadtwerken, rechts auch mit Nullauspeisung.

Zuweisung eines Verteilungstyps Anwendung. Nach Einführung von Temperaturklassen werden Wahrscheinlichkeitsverteilungen geschätzt und auf deren Basis Szenarien erzeugt, die für die Berechnung technischer Kapazitäten des Gasnetzes benötigt werden (vgl. [15]).

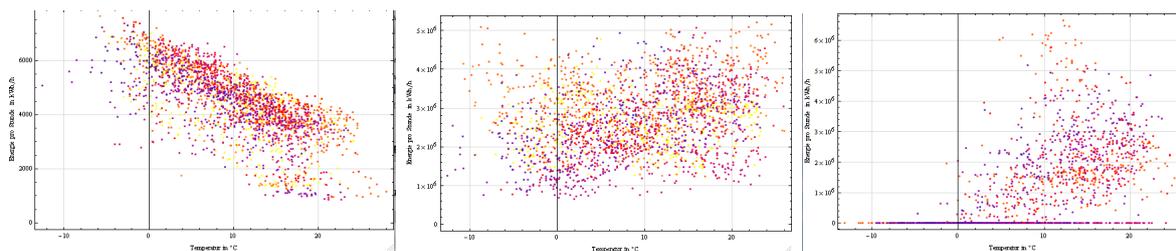


Bild 2: Datensätze von einem Knoten mit Industrieabnehmer (links), Marktübergang (Mitte), Speicher (rechts).

## 2 Schätzung von Wahrscheinlichkeitsverteilungen

Da Gas insbesondere zum Heizen und zur Stromerzeugung verwendet wird, ist der Verbrauch im Allgemeinen stark temperaturabhängig. Entsprechend sind die vertraglich vereinbarten Lieferkapazitäten teilweise an die Temperatur gekoppelt. Dazu wird eine nach einem vertraglich vereinbarten Verfahren ermittelte Referenztemperatur zugrunde gelegt.

Zur Erzeugung statistisch repräsentativer Szenarien, die die Temperaturabhängigkeit des Gasverbrauchs widerspiegeln, gehen wir folgendermaßen vor: Zunächst wird die gewichtete räumlich gemittelte Referenztemperatur bestimmt. Hierzu wird unter Verwendung erprobter Temperaturgewichte ein gemittelter Temperaturvektor aus fünf Temperaturstationen berechnet, vgl. Tabelle 1. Aus der Referenztemperatur wird nun eine exemplarische Zuordnung von Datum

Tabelle 1: Räumliche Temperaturmittelung.

Temperaturstation	Gewichtung
Essen - Bredeney	0.15
Frankfurt am Main	0.40
Flughafen Nürnberg	0.05
Stuttgart - Echterdingen	0.25
Flughafen München Erding	0.15

und räumlicher Tagesmitteltemperatur bereitstellt. Dazu sind geeignete Temperaturintervalle zu bestimmen, die einerseits genügend Daten enthalten und andererseits hinreichend klein sind, um eine temperaturunabhängige Szenariogenerierung zu erlauben. Aus statistischer Sicht sind die folgenden dreizehn Temperaturklassen (in °C) sinnvoll:  $(-15, -4]$ ,  $(-4, -2]$ ,  $(-2, 0]$ ,  $\dots$ ,  $(18, 20]$ ,  $(20, 40)$ . Nun werden für jeden Ausspeiseknoten und für jede Temperaturklasse alle Tage gefiltert, die in das entsprechende Räumliche-Temperaturintervall fallen. Die zugehörigen mittleren Gasflüsse werden als Datenbasis für die Verteilungsschätzungen selektiert, vgl. Bild 3. Überdies wird eine Trennung in Arbeitstage und Wochenendtage durchgeführt, da diesbezüglich Unterschiede bei der Parametrisierung von Verteilungen zu erwarten sind.

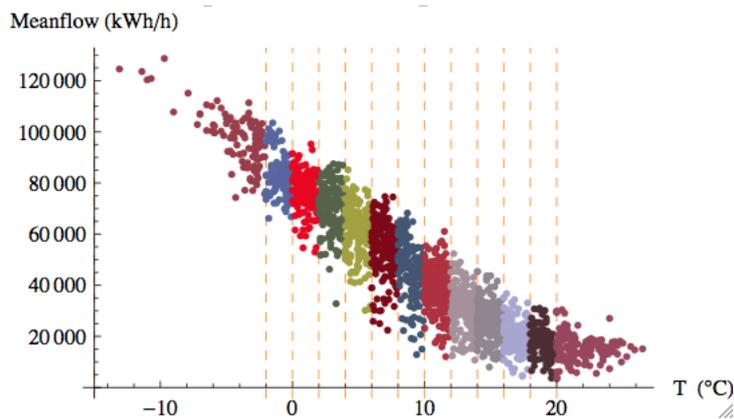


Bild 3: Gasfluss in Temperaturklassen.

Der nächste Schritt besteht in der Anpassung von univariaten Verteilungen für jeden Punkt und jede Temperaturklasse. Im Anschluss daran werden für geeignete Teilmengen von Ausspeisepunkten auch multivariate Verteilungen angepasst. Die so angepassten Verteilungen sind die Basis für die in Abschnitt 4 beschriebene Erzeugung von Szenarien.

Bei zahlreichen Knoten ist der mittlere Gasfluss nicht vollständig positiv, sondern kann auch über längere Zeiträume Null sein. Daher beschränken wir uns innerhalb des Temperaturklassenansatzes darauf, die empirische Häufigkeit der Nullabnahme zu bestimmen und diese als Parameterschätzung in gemischten Verteilungen zu verwenden. Dies legt die Verwendung gemischter Verteilungen innerhalb der Temperaturklassen nahe:

$$F(x) = p F^0(x) + (1 - p) F^+(x), \quad 0 \leq p \leq 1.$$

Dabei bezeichnet  $F^0$  die Funktion mit  $F^0(x) = 1$  für  $x \geq 0$  und  $F^0(x) = 0$  für  $x < 0$ , und  $x$  ist der Gasfluss innerhalb der betrachteten Temperaturklasse. Der stetige Verteilungsanteil für die positive Gasabnahme wird mit  $F^+$  bezeichnet. Es kommen dabei zwei stetige Verteilungen zum Einsatz, die Gleichverteilung für temperaturunabhängige Datensätze und die Normalverteilung für die temperaturabhängigen Datensätze. Bild 4 repräsentiert die gemischte Normalverteilungsfunktion exemplarisch für einen Ausspeisepunkt in der Temperaturklasse  $(6^\circ\text{C}, 8^\circ\text{C}]$ . Ist der Gasfluss innerhalb einer Temperaturklasse vollständig positiv, so erfolgt eine Parametrisierung mit stetigen Verteilungen, d.h.  $p = 0$  und  $F(x) = F^+(x)$ .

Innerhalb der Normalverteilung parametrisieren wir den Erwartungswert tagestypabhängig für Arbeitstage, Wochenendtage und Feiertage und postulieren die Standardabweichung als homogen. Als Schätzer für Erwartungswert und Varianz werden die erwartungstreuen Standardschätzer verwendet.

Für die Schätzung der Kovarianzmatrix einer multivariaten Normalverteilung werden die erwartungswertbereinigten Residuen aus dem univariaten Ansatz mit Trennung nach Tagestyp übernommen. Im multivariaten Fall findet ausschließlich die Normalverteilung Verwendung.

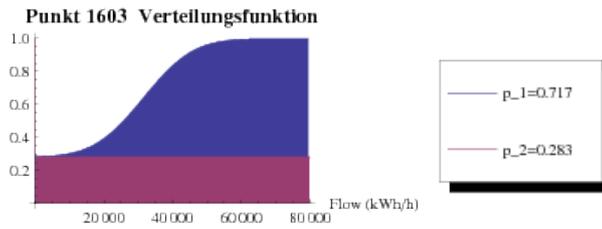


Bild 4: Verteilungsfunktion einer shifted Normalverteilung.

Hier kommen nur Punkte in Betracht, die innerhalb einer Temperaturklasse vollständig positive Gasflüsse aufweisen. Zur Schätzung der Kovarianzmatrix der multivariaten Normalverteilung wird der unverzerrte Standardschätzer nach Pearson verwendet. Die erhaltenen Korrelationsmatrizen für unterschiedliche Temperaturklassen werden in Bild 5 dargestellt.

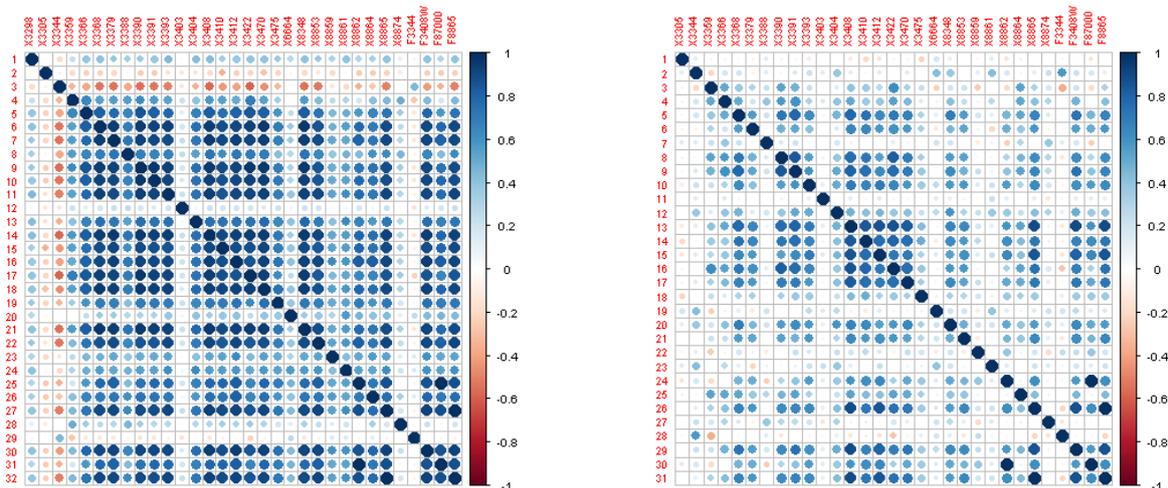


Bild 5: Korrelationsplots für die Temperaturklassen (10, 12] und (18, 20].

### 3 Vorhersage der Gasabnahme im Tiefst-Temperaturbereich

Eine besondere Behandlung erfährt der Tiefst-Temperaturbereich, für den oftmals nicht genügend Daten vorhanden sind. Trotzdem ist dieser Temperaturbereich sehr wichtig, da jeder Gastransporteur verpflichtet ist, die Gasversorgung auch dann noch aufrechtzuerhalten, wenn die Temperatur bis auf die sog. *Auslegungstemperatur* gefallen ist. Die Auslegungstemperatur ist von den örtlichen Klimabedingungen abhängig und liegt in Deutschland meist zwischen  $-12^{\circ}\text{C}$  und  $-16^{\circ}\text{C}$ . Derart niedrige mittlere Tagestemperaturen werden in Deutschland aber sehr selten beobachtet, daher liegen kaum Daten für den Bereich nahe der Auslegungstemperatur vor.

Um den Zusammenhang zwischen den typischen örtlichen Gaslastprofilen und der Temperatur in verschiedenen Wirtschaftssektoren zu beschreiben, schlagen verschiedene Autoren (vgl. z.B. [2], [11]) eine Sigmoidregression vor, die man als ein nichtlineares statistisches Modell interpretieren kann. Mehr zu nichtlinearen Modellen, deren Erweiterung und Anwendung auf die Prognose des Gasflusses nahe der Auslegungstemperatur, findet man in [6] und [17].

Die ungünstigen numerischen Eigenschaften von nichtlinearen Modellen motivieren hier einen speziellen Zugang und die Verwendung von semi-parametrischen statistischen Modellen für die Vorhersage des Gasflusses nahe der Auslegungstemperatur. Basierend auf [4], [5], [6] und [1] wird hier eine sog. P-Splines-Regression (*penalized splines*) mit Form-Restriktionen verwendet. Diese Methodik wurde wegen ihrer Einfachheit und Flexibilität gewählt.

Die Abhängigkeit von standardisierten maximalen täglichen Gaslasten  $y_i$  von der Temperatur  $t_i$ ,  $i = 1, \dots, n$ , wobei  $n = 2343$  der Stichprobenumfang ist, wird durch folgendes Modell beschrieben:

$$y_i = S_{\Delta}(t_i) + \varepsilon_i.$$

Hier bezeichnet  $y_i = y_i^{max} / \bar{y}$  den standardisierten maximalen täglichen Gasfluss und  $\bar{y}$  steht für den empirischen Mittelwert über alle maximalen täglichen Gasflusswerte an einem bestimmten Ausspeisepunkt innerhalb des Netzwerks. Die  $t_i$  geben die räumlich gemittelte Tagesmitteltemperatur an. Eine alternative Möglichkeit wäre, die  $t_i$  als gewichtete Vier-Tages-Mittel zu wählen, d.h.

$$t_i = \sum_{k=0}^3 w_k t_{ik},$$

wobei die  $t_{ik}$ ,  $k = 0, 1, 2, 3$ , die Mittel-Temperaturen der letzten 4 Tage bezeichnen und  $w_k$  die in [13] ermittelten optimalen Gewichte darstellen. Die Funktion  $S_{\Delta}$  ist eine lineare Kombination von sog. B-Splines  $B_j$ ,  $j = 1, \dots, m$ , auf einem Gitter  $\Delta$ . Sie ist in den  $t_i$  gegeben durch

$$S_{\Delta}(t_i) = \sum_{j=1}^m a_j B_j(t_i) + a_{m+1} d_i.$$

Dabei ist  $\varepsilon_i \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2)$ , für  $i = 1, \dots, n$ , ein normalverteilter Fehlerterm, die  $B_j$  sind B-Splines vom Grad  $q$  und das Gitter  $\Delta$  ist äquidistant mit  $m - q$  Teilintervallen. Die Variable  $d_i$ ,  $i = 1, \dots, n$ , zeigt an, ob der Gasfluss an einem Arbeitstag aufgetreten ist, d.h.

$$d_i = \begin{cases} 1 & \text{falls der Tag } i \text{ ein Arbeitstag ist,} \\ 0 & \text{falls der Tag } i \text{ ein Feiertag bzw. am Wochenende ist.} \end{cases}$$

Anstatt wie üblich

$$\sum_{i=1}^n (y_i - S_{\Delta}(t_i))^2$$

zu minimieren, werden Glättungs- und Gestaltungsterme mit den Parametern  $\lambda$  und  $\kappa$  hinzugefügt und die Funktion

$$\sum_{i=1}^n (y_i - S_{\Delta}(t_i))^2 + \lambda \sum_{j=3}^m (\delta^2 a_j)^2 + \kappa \sum_{j=2}^m b_j (\delta^1 a_j)^2$$

minimiert (Methode der kleinsten Quadrate). Hierbei bezeichnet  $\delta^1 a_j = a_j - a_{j-1}$  und  $\delta^2 a_j = a_j - 2a_{j-1} + a_{j-2}$ . Die Variable  $b_j \in \{0, 1\}$  kennzeichnet, für welchen B-Spline  $B_j$  die Formrestriktion für die Ableitung  $dS_{\Delta}(t)/dt$  gelten soll. Diese Formrestriktion wird eingeführt, um zu sichern, dass der Gasabsatz unterhalb der Auslegungstemperatur nicht weiter steigt.

Die P-Splines-Regression erlaubt eine direkte Extrapolation, die für die Vorhersage geeignet ist.

Das angepasste Modell und die Vorhersage werden am Beispiel eines Ausspeiseknotens, der ein Stadtwerk repräsentiert, in Bild 6 illustriert. Dabei werden kubische B-Splines (d.h.  $q = 3$ ) auf einem äquidistanten Gitter mit 11 Stützstellen, den Parametern  $\lambda = 2.51$  und  $\kappa = 100$ , und  $b_1 = b_2 = 1$  verwendet. Für die kubischen B-Splines sei z.B. auf [8, Kap. 6.3.6] verwiesen. Die obere und untere Kurve stellen Gasabnahme an den Arbeitstagen bzw. Wochenenden und Feiertagen dar. Bei der Auslegungstemperatur von  $-14^{\circ}\text{C}$  an einem Arbeitstag wird ein Gasverbrauch von 41750 KWh/h vorhergagt. Die beschriebene Methodik zur Vorhersage wurde

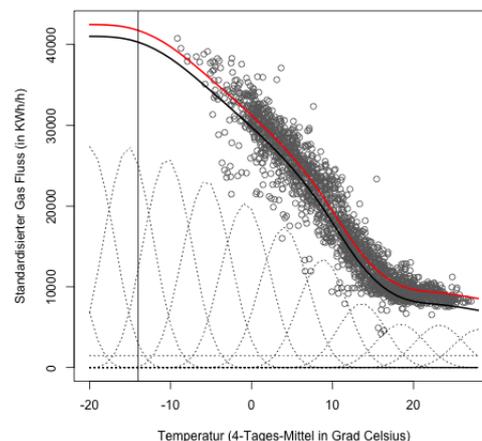


Bild 6: P-Splines-Regression mit abflachender Asymptote.

in R (vgl. [20]) implementiert. Im Tiefsttemperaturbereich wird in jedem Ausspeiseknoten eine univariate Normalverteilung modelliert, deren Mittelwert durch  $\bar{y}S_{\Delta}(t)$  mit  $t = -14^{\circ}\text{C}$  geschätzt und als deren Standardabweichung die für die Temperaturklasse  $(-4^{\circ}\text{C}, -2^{\circ}\text{C})$  geschätzte verwendet wird.

## 4 Erzeugung von Szenarien

Das Ziel besteht darin, eine nicht zu große Zahl  $n$  von Szenarien für die Ausspeiselaast an allen  $d$  Ausspeisepunkten eines Gas-Transportnetzes, für die statistische Daten vorliegen, zu bestimmen. Unter Szenarien verstehen wir dabei Vektoren, deren Komponente  $i$  den Wert für die Ausspeiselaast am Punkt  $i$ ,  $i = 1, \dots, d$ , enthalten. Basis für die Generierung der Szenarien sind die  $d$ -dimensionalen Wahrscheinlichkeitsverteilungen, deren Schätzung im vorhergehenden Kapitel besprochen wurde.

Zur Erzeugung von Szenarien werden traditionell (Pseudo-) Zufallszahlen-Generatoren, d.h. Methoden vom Monte-Carlo-Typ eingesetzt. Obwohl diese Methodik sich deutlich weiterentwickelt hat (vgl. [16]), stellen hohe Dimensionen  $d \approx 300$  oder gar  $d \approx 700$  eine Herausforderung dar. Der Grund dafür soll kurz erläutert werden.

Da jede Methodik damit beginnt, zunächst Szenarien für eine  $d$ -dimensionale Gleichverteilung in  $[0, 1]^d$  zu bestimmen und diese dann später zu transformieren, bezeichne  $G$  zunächst die  $d$ -dimensionale Verteilungsfunktion der Gleichverteilung auf  $[0, 1]^d$ , d.h.

$$G(x) = x_1 x_2 \cdots x_d = \prod_{i=1}^d x_i \quad (x_i \in [0, 1], i = 1, \dots, d).$$

Sind nun  $n$  Szenarien  $\eta^j \in [0, 1]^d$ ,  $j = 1, \dots, n$ , als gleichwahrscheinliche Realisierungen eines auf  $[0, 1]^d$  gleichverteilten Zufallsvektors gegeben und bezeichnet  $G_n$  die  $d$ -dimensionale Verteilungsfunktion eines Zufallsvektors mit diesen  $n$  Realisierungen, so sollte der gleichmäßige Abstand der beiden Verteilungsfunktionen

$$D_n(\eta^1, \dots, \eta^n) = \max_{x \in [0, 1]^d} |G(x) - G_n(x)| = \max_{x \in [0, 1]^d} \left| \prod_{i=1}^d x_i - \frac{1}{n} \sum_{\substack{j=1 \\ \eta^j \leq x}}^n 1 \right|$$

möglichst klein sein. Dabei ist die Ungleichung  $\eta^j \leq x$  komponentenweise zu verstehen und die Summe entspricht der Anzahl der Szenarien, die kleiner oder gleich  $x$  sind. Interessant ist nun das Verhalten von  $D_n(\eta^1, \dots, \eta^n)$  in Abhängigkeit von  $n$ .

Ein klassisches Resultat besagt, dass eine positive Konstante  $C_d$  existiert, so dass

$$D_n(\eta^1, \dots, \eta^n) \geq C_d \frac{(\log n)^{\frac{d-1}{2}}}{n}$$

für alle natürlichen Zahlen  $n \geq 2$  und alle  $\eta^j$ ,  $j = 1, \dots, n$ , gilt. Der logarithmische Term im Zähler der rechten Seite wächst für größere Dimensionen  $d$  des Zufallsvektors mit  $n$  zunächst stark an, bevor für (sehr) große  $n$  der Nenner dafür sorgt, dass die rechte Seite der Ungleichung klein wird.

Seit längerem ist bekannt, dass Folgen  $\eta^1, \eta^2, \dots, \eta^n, \dots$  existieren, für die die Abschätzung

$$(*) \quad D_n(\eta^1, \dots, \eta^n) \leq K_d \frac{(\log n)^d}{n}$$

mit einer geeigneten Konstanten  $K_d$  und für alle natürlichen  $n \geq 2$  gilt (vgl. [18, Kap. 3]). Im Fall, dass die  $\eta_j$  unabhängige und in  $[0, 1)^d$  gleichverteilte Zufallsvektoren sind, also im Fall der Monte-Carlo Methode, ist nicht einmal die Zufallsvariable  $\sqrt{n}D_n(\eta^1, \dots, \eta^n)$  in Wahrscheinlichkeit beschränkt. D.h., die Monte-Carlo-Methode hat deutlich schlechtere Konvergenzeigenschaften als durch (\*) beschrieben.

Da die Konstante  $K_d$  in (\*) dimensionsabhängig ist, ist es darüberhinaus wichtig, solche Folgen zu finden, für die die zugehörige Konstante  $K_d$  klein ist und nicht mit wachsender Dimension  $d$  wächst. Der Schlüssel dazu sind die sog.  $(t, d)$ -Folgen, die z.B. in [18, Kap. 4] diskutiert werden.

Solche  $(t, d)$ -Folgen  $\eta^1, \eta^2, \dots, \eta^n, \dots$  in  $[0, 1)^d$  beruhen darauf, dass für eine gegebene Basis  $b$ , alle  $k \geq 0$  und alle  $m > t$  je  $b^t$  der  $b^m$  Punkte  $\eta^n$  für  $kb^m \leq n < (k+1)b^m$ , in den  $d$ -dimensionalen Intervallen der Form

$$\prod_{i=1}^d [a_i b^{-d_i}, (a_i + 1) b^{-d_i}] \quad (0 \leq a_i < b^m, d_i \geq 0, i = 1, \dots, d)$$

mit Volumen  $\frac{b^t}{b^m}$  liegen.  $(t, d)$ -Folgen erlauben eine Fehlerabschätzung der obigen Form, wobei die Konstante  $K_d = K_d(b, t)$  die Gestalt

$$K_d = K_d(b, t) = \frac{b-1}{2^d d! (\log b)^d} \left(\frac{b}{2}\right)^{d-1} b^t$$

besitzt (siehe [18, Kap. 4.1]). Es existiert eine Reihe von  $(t, d)$ -Folgen, die mit den Namen ihrer Entdecker verbunden sind, wie z.B. Sobol-, Faure- und Niederreiter-(Quasi-Monte Carlo-) Folgen (siehe [3, Kap. 8.1]). Aus der Darstellung von  $K_d(b, t)$  ist ersichtlich, dass diese Konstante für feste Basis  $b$  am kleinsten ist, falls  $t = 0$  gilt. Allerdings existieren  $(0, d)$ -Folgen nur im Fall  $b \geq d$ . Dies würde in unserem Fall zu einer zu großen Basis  $b$  und damit zu großen Anzahlen von Szenarien führen.

Deshalb verwenden wir Niederreiter-Folgen zur Basis  $b = 2$ . Auch diese können jedoch gewisse Gleichverteilungs-Defizite in höheren Dimensionen  $d$  besitzen, wie z.B. auf der linken Seite von Bild 7 illustriert. Es zeigt die 16. und 18. Komponente der ersten 1000 Elemente der Niederreiter-Folge in Dimension  $d = 40$ . Als geeigneter Ausweg erscheint ein sog. *random scrambling* der Folge wie in [19] vorgeschlagen und in [12] implementiert. Owen schlägt dort vor, in allen Komponenten jedes Elementes der Folge eine zufällige Permutation der Ziffern nach dem Dezimalpunkt mittels (Pseudo) Zufallszahlengeneratoren vorzunehmen. Er beweist, dass sich die theoretischen Eigenschaften der  $(t, d)$ -Folge mit Wahrscheinlichkeit 1

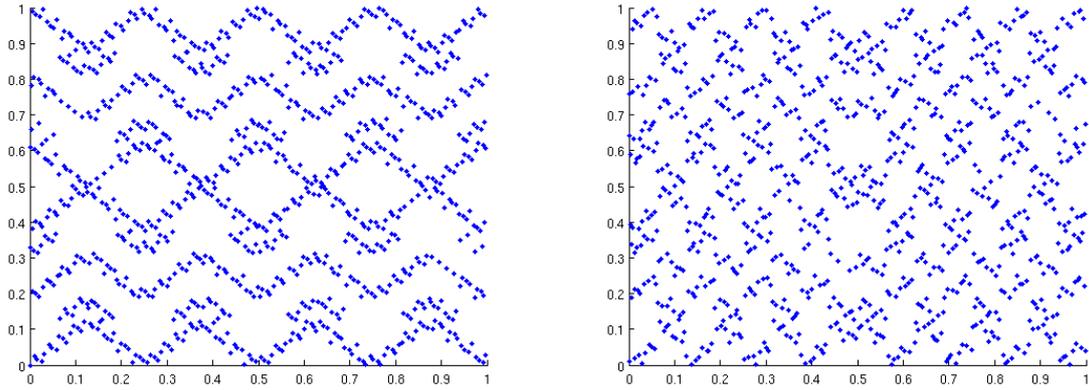


Bild 7: Links: 1 000 Niederreiter-Punkte für  $d = 40$ , Projektion (16,18). Rechts: 1 000 Scrambled-Niederreiter-Punkte für  $d = 40$ , Projektion (16,18).

nicht ändern und dass die zufällig veränderten Elemente der Folge überdies in  $[0, 1)^d$  gleichverteilte Zufallsvektoren darstellen. Nach einer solchen Randomisierung werden also die guten Eigenschaften von Quasi-Monte Carlo Folgen mit den positiven Effekten von Monte-Carlo-Methodiken vereint. Die rechte Seite von Bild 7 illustriert das Ergebnis des scrambling.

Unter Verwendung des Scrambled-Niederreiter-Generators SNIED aus [12] wird nun eine größere Anzahl  $N$ ,  $1\,000 \leq N \leq 10\,000$ , Punkte in  $[0, 1)^d$  erzeugt und im nächsten Schritt auf den  $\mathbb{R}^d$  transformiert, wobei zunächst angenommen wird, dass die  $d$  Komponenten voneinander statistisch unabhängig sind. Dies erfolgt dadurch, dass neue Vektoren  $\xi^j$  durch die Vorschrift

$$\xi_i^j = \Phi_i^{-1}(\eta_i^j) \quad (i = 1, \dots, d, j = 1, 2, \dots, N)$$

erzeugt werden, wobei  $\Phi_i$  die Verteilungsfunktion der  $i$ -ten Komponente des Zufallsvektors ist. Im hier betrachteten Fall sind dies Gleichverteilungen (in gewissen Intervallen) oder (gestutzte) Standard-Normalverteilungen, so dass sich diese Inversion von  $\Phi_i$  schnell bewerkstelligen läßt. Im Fall der Normalverteilung wird die rationale Approximation für  $\Phi_i^{-1}$  aus [7, Kap. 2.3.2] verwendet und anschließend mit einem Newton-Schritt bzgl.  $\Phi_i$  wie in [14] verbessert. Abschließend werden in diesem Transformationsschritt die Korrelationen in den multivariaten Normalverteilungs-Anteilen der Gesamtverteilung berücksichtigt. Hat man z.B. eine  $s$ -dimensionale multivariate Normalverteilung  $\mathcal{N}(m, \Sigma)$  mit Mittelwertsvektor  $m \in \mathbb{R}^s$  und  $s \times s$ -dimensionaler Kovarianzmatrix  $\Sigma$ , so bestimmt man die sog. Cholesky-Zerlegung  $\Sigma = LL^\top$  der Kovarianzmatrix  $\Sigma$  mit einer unteren Dreiecksmatrix  $L$  vom Typ  $s \times s$  (vgl. auch [7, Kap. 2.3]). Dann sind die  $s$ -dimensionalen Vektoren

$$L\xi^j + m \quad (j = 1, 2, \dots, N)$$

geeignete Szenarien für  $\mathcal{N}(m, \Sigma)$  im Quasi-Monte-Carlo-Sinn und die Randomisierung ist verteilt entsprechend  $\mathcal{N}(m, \Sigma)$ .

Da die Anzahl  $N$  der so erhaltenen  $d$ -dimensionalen gleichwahrscheinlichen Szenarien noch sehr groß ist, wird diese abschließend mit Hilfe der Methode der *Szenarioreduktion* (vgl. [9, 10]) auf eine kleinere Anzahl  $n$  (z.B.  $n = 50, 100$ ) reduziert, so dass die verbleibenden  $n$  Szenarien die ursprünglichen  $N$  bestmöglich repräsentieren. Die Wahrscheinlichkeiten der ausgesonderten Szenarien werden dem nächstgelegenen verbleibenden Szenario zugeschlagen. Deshalb besitzen die endgültigen  $d$ -dimensionalen Szenarien  $\xi^j, j = 1, \dots, n$ , unterschiedliche Wahrscheinlichkeiten  $p_j, j = 1, \dots, n$ , für die aber  $\sum_{j=1}^n p_j = 1$  gilt. Die Szenarioreduktion wird in Bild 8 illustriert.

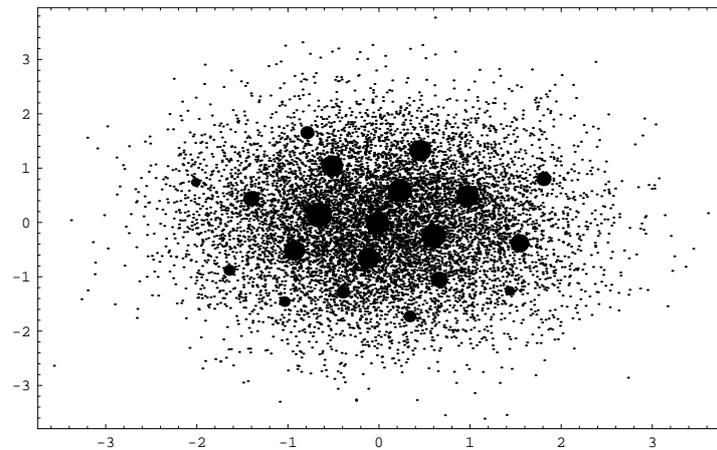


Bild 8:  $N = 10\,000$  Monte-Carlo simulierte Szenarien der zweidimensionalen Standard-Normalverteilung werden auf  $n = 20$  Szenarien reduziert. Die Durchmesser der Kreise sind dabei proportional zu ihren neuen Wahrscheinlichkeiten.

## Danksagung

Diese Arbeit wurde von OGE im Rahmen des Projektes *Forschungskooperation Netzoptimierung (ForNe)* und vom BMWi im Projekt *Technische Kapazität von Gasnetzen* gefördert. Wir bedanken uns beim Kooperationspartner Open Grid Europe GmbH für die Bereitstellung der statistischen Daten.

## Literatur

- [1] K. Bollaerts, H. C. P. Eilers and I. van Mechen: Simple and multiple P-splines regression with shape constraints, *British Journal of Mathematical and Statistical Psychology* 59 (2006), 451–469.
- [2] G. Cerbe: *Grundlagen der Gastechnik*, Hanser Verlag, Leipzig, 2008.
- [3] J. Dick and F. Pillichshammer: *Digital Nets and Sequences. Discrepancy Theory and Quasi-Monte Carlo Integration*, Cambridge University Press, Cambridge, 2010.
- [4] H. C. P. Eilers and B. D. Marx: Flexible smoothing with B-splines and penalties, *Statistical Science* 11 (1996), 89–121.

- [5] H. C. P. Eilers and B. D. Marx: Splines, Knots and Penalties, *Wiley Interdisciplinary Reviews: Computational Statistics* 2 (2010), 637–653.
- [6] H. Friedl, R. Mirkov and A. Steinkamp: Modeling and forecasting gas flow on exits of gas transmission networks, *International Statistical Review* (Special Issue on Energy Statistics), 2011 (erscheint).
- [7] P. Glassermann: *Monte Carlo Methods in Financial Engineering*, Springer, New York, 2004.
- [8] G. Hämmerlin und K.-H. Hoffmann: *Numerische Mathematik*, Springer, Berlin, 1994.
- [9] H. Heitsch: Szenariobaumapproximation für stochastische Optimierungsprobleme in der Energiewirtschaft, in: *Optimierung in der Energiewirtschaft*, VDI-Bericht Nr. 2080, Düsseldorf 2009, 45–60.
- [10] H. Heitsch and W. Römis: A note on scenario reduction for two-stage stochastic programs, *Operations Research Letters* 35 (2007), 731–738.
- [11] M. Hellwig: *Entwicklung und Anwendung parametrisierter Standard-Lastprofile*, Dissertation, Technische Universität München, Fakultät für Energietechnik und Informationstechnik, Institut für Energietechnik, 2003.
- [12] Hee Sun Hong and F. J. Hickernell: Algorithm 823: Implementing scrambled digital sequences, *ACM Transactions of Mathematical Software* 29 (2003), 95–109.
- [13] H. Leövey, W. Römis, A. Steinkamp und I. Wegner-Specht: Modellierung der Gasabnahme als Funktion der Temperatur: Optimierung der Temperaturgewichte, *gwf Gas/Erdgas* (eingereicht).
- [14] G. Marsaglia: Evaluating the normal distribution, *Journal of Statistical Software* 11 (2004), 1–10.
- [15] A. Martin, B. Geißler, C. Hayn, B. Hiller, J. Humpola, T. Koch, A. Morsi, M. E. Pfetsch, L. Schewe, M. Schmidt, R. Schultz, R. Schwarz, J. Schweiger, M. C Steinbach und B. M. Willert: Optimierung technischer Kapazitäten in Gasnetzen, in: *Optimierung in der Energiewirtschaft*, VDI-Bericht, Düsseldorf, 2011.
- [16] M. Matsumoto, T. Nishimura: Mersenne Twister: A 623-dimensionally equidistributed uniform pseudo-random number generator, *ACM Transactions on Modeling and Computer Simulation* 8 (1998), 3–30.
- [17] R. Mirkov and H. Friedl: Nonlinear and Spline regression models for forecasting gasflow on exits of gas transmission networks, in: Proceedings of the 26th International Workshop on Statistical Modelling (D. Conesa et al. eds.), ADEIT, University of Valencia, 2011, 394–399.
- [18] H. Niederreiter: *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia, 1992.
- [19] A. B. Owen: Randomly permuted  $(t, m, s)$ -nets and  $(t, s)$ -sequences, in: *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, Lecture Notes in Statistics, Vol. 106, Springer, New York, 1995, 299–317.
- [20] R Development Core Team: *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, 2008 (<http://www.R-project.org>).