

---

# A B S C H L U S S B E R I C H T

## Verbundprojekt **Technische Kapazitäten von Gasnetzen:** Teilprojekt **Untersuchung der stochastischen Ausspeiselasten**

Zuwendungsempfänger: HU Berlin  
Projektleiter: Römisch  
Förderkennzeichen: 0328006E  
Laufzeit: 01.07.2009 bis 31.12.2012  
Berichtszeitraum: 2009–2012

---

Industriepartner: OPEN GRID EUROPE (OGE, vorher E.ON GT)

Anzahl der Ergebnisdarstellungen:  
Veröffentlichungen: 5                      Konferenzbeiträge: 10  
Diplomarbeiten: 1                         Dissertationen: –  
Habilitationsschriften: –

---

Das diesem Bericht zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Wirtschaft und Technologie aufgrund eines Beschlusses des Deutschen Bundestages unter dem Förderkennzeichen 0328006E gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

## 1 Personelle Besetzung der Projektstelle

Die Projektstelle BAT IIa/2 war im Zeitraum Januar 2010 bis Dezember 2012 durchgängig mit Dipl.-Math. Hernan Leövey besetzt. Die Dissertation von Herrn Leövey ist in Vorbereitung und wird 2013 eingereicht. Wesentliche Ergebnisse dieses Teilprojektes und auch Untersuchungen zu Quasi-Monte Carlo Methoden für Modelle der Quantenphysik (siehe [26]) werden Bestandteil seiner Dissertation.

## 2 Wissenschaftliche Ergebnisse des Vorhabens

Im folgenden geben wir in den Kapiteln 2.1–2.3 einen Überblick über die wissenschaftlichen Ergebnisse des Teilprojektes im gesamten Berichtszeitraum. Dabei wird am Ende jedes Kapitels der Wert der Ergebnisse zur Erweiterung des Verständnisses und zur Optimierung von technischen Kapazitäten in realen Gastransportnetzen dargestellt.

### Inhalt:

- 2.1 Modellierung der Gasabnahme als Funktion der Temperatur: Optimierung der Temperaturgewichte 2
- 2.2 Vorhersage der Gasabnahme im Tiefst-Temperaturbereich 11
- 2.3 Szenariogenerierung zur Modellierung der stochastischen Ausspeiselasten in einem Gastransportnetz 14

## 2.1 Modellierung der Gasabnahme als Funktion der Temperatur: Optimierung der Temperaturgewichte

Es wurde eine neue Methode zur optimalen Bestimmung der Temperaturgewichte in Modellen der Gasausspeisung an Exits von Gasnetzwerken entwickelt. Diese Modelle basieren auf gewichteten Temperaturmitteln unter Verwendung von Tagesmitteltemperaturen der letzten 4–6 Tage. In der Regel werden für diesen speziellen Anwendungsfall Sigmoidmodelle bzw. lineare Modelle (eingeschränkt auf gewisse Temperaturbereiche) verwendet. Für beide Modellklassen wurde gezeigt, dass sich die resultierenden Optimallösungen deutlich von den in der Praxis verwendeten Gewichten (siehe etwa [4]) unterscheiden. Die numerischen Ergebnisse basieren auf den Beispieldaten des Gasnetzbetreibers Open Grid Europe (OGE). Die mathematische Methodik ist allgemein auf solche uns ähnliche Problemstellungen anwendbar. Die Verwendung der mittels Optimierung generierten Temperaturgewichte verbessert die Modellgüte wesentlich.

### 2.1.1 Einleitung

Der Gasverbrauch ist in der Regel temperaturabhängig. Die Art der funktionalen Abhängigkeit wird in der Literatur hauptsächlich über Funktionen sigmoiden Typs, vgl. [4] für eine umfassende Einführung in die Problematik, oder mittels linearer Funktionen (eingeschränkt auf gewisse Temperaturbereiche) beschrieben. Letzteres wird aktuell in der Praxis bevorzugt, daher wird diese Modellklasse in die vorliegende Analyse aufgenommen. In einigen Fällen ist die Temperaturabhängigkeit sehr schwach ausgeprägt, hier kommen lineare Modelle zur Anwendung. Prinzipiell beschränken wir uns innerhalb der Modellklasse sigmoiden Typs auf das Basis-Sigmoidmodell. Es ist wohlbekannt, dass die Korrelation des Gasverbrauchs mit der Tagesmitteltemperatur weniger stark ausgeprägt ist als die Korrelation des Gasverbrauchs mit einem gewichteten Temperaturmittel unter Verwendung einer gewissen Anzahl von Tagesmitteltemperaturen der vergangenen Tage. Die Ursache liegt in der verzögerten Reaktion des Gasverbrauchs auf plötzliche Temperaturänderungen auf Grund der Wärmespeicherkapazität von mehreren Tagen der in Deutschland typischen Gebäude in massiver Bauweise (vgl. [11]). Dies legt die Integration der Temperaturen der vergangenen Tage als erklärende Variablen in das Modell nahe. In der Praxis ist die Verwendung spezieller Temperaturgewichte für dieses Problem üblich. Dabei handelt es sich um die normalisierten vier ersten Terme der geometrischen Reihe unter der impliziten Annahme, dass der Einfluss länger zurückliegender Tagestemperaturen abnimmt. Die fixierten Gewichte für vier aufeinanderfolgende Tage sind beispielsweise in [4] gegeben als

$$w_1 = \frac{8}{15}, w_2 = \frac{4}{15}, w_3 = \frac{2}{15}, w_4 = \frac{1}{15}.$$

Das resultierende Temperaturmittel ist somit definiert als

$$\bar{t}_{(day)} = w_1 t_{(day)} + w_2 t_{(day-1)} + w_3 t_{(day-2)} + w_4 t_{(day-3)},$$

wobei  $t_{(day)}$  die Tagesmitteltemperatur des aktuellen Tages sowie  $t_{(day-1)}, \dots$  entsprechend die Tagesmitteltemperaturen der davorliegenden Tage bezeichnen. Inner-

halb der weiteren Beschreibung verwenden wir die Abkürzung  $t_{(d)}$  für die Temperatur am Tag  $d$ . In unserer Verallgemeinerung fassen wir diese Temperaturgewichte als Optimierungsvariablen auf und fügen keine künstliche Positivitätsrestriktion für diese Variablen hinzu. Diese Restriktion kann auf Wunsch der Anwender allerdings ergänzt werden. Die numerischen Ergebnisse in Abschnitt 2.1.4 werden zeigen, dass die optimalen Gewichte in der Regel positive Werte aufweisen. Wir nehmen, ebenso wie im klassischen Fall - vgl. [4] - an, dass die Gewichte normiert sind. Falls alle Optimallösungen positiv sind, können sie als Wahrscheinlichkeitsverteilung über den betrachteten Tagen aufgefasst werden. Auf diese Weise sind die Gewichte als relativer Beitrag der vergangenen Tagestemperaturen interpretierbar. Die Verwendung der neuen optimalen Gewichte trägt zur Verbesserung der Modellgüte sowie der Vorhersagegüte der künftigen Entnahmemengen durch die Gasverbraucher bei.

### 2.1.2 Regressionsprobleme: Sigmoidale und lineare Modelle

Wir betrachten zunächst das klassische Modell vom sigmoiden Typ für die Gasentnahme an einem Exit, wobei vier Parameter  $\alpha, \beta, \gamma, \delta$  zu bestimmen sind. Das Modell ist gegeben durch

$$f_0(\theta_1, \theta_2, \theta_3, \theta_4, \bar{t}, X) = \mathbb{E}(X) \left( \frac{\theta_1}{1 + \left(\frac{\theta_2}{\bar{t}-40}\right)^{\theta_3}} + \theta_4 \right), \quad (1)$$

wobei  $\mathbb{E}(X)$  den Erwartungswert bzw. den empirischen Mittelwert der vorliegenden Gasverbrauchsdatenvektor  $X$  bezeichnet, und  $\bar{t}$  das gewichtete Standardtemperaturmittel. Für eine allgemeine Beschreibung sigmoidaler Wachstumsmodelle und Strategien zu Bestimmung von Startwerten für die nichtlinearen Regressionsprobleme verweisen wir auf [55, Kap. 7.3].

In Abbildung 1 ist das typische Verhalten der Gasauspeisung an einem Exit eines deutschen Gasnetzes dargestellt.

Zur Lösung dieses nichtlinearen Regressionsproblems betrachten wir das zugehörige Kleinste-Quadratmittel-Problem. Wir verweisen darauf, dass an dieser Stelle weitere Normen - wie die  $L_1$ -Norm - ohne Einschränkung der allgemeinen Methodik verwendet werden können. Innerhalb des Kleinste-Quadratmittel-Problems basiert das Modell auf den Temperaturmitteln  $\bar{t}_{\min}, \dots, \bar{t}_{\max}$  und den zugeordneten Gasverbrauchswerten. In der Regel sind nun eine Anzahl von  $d_{\bar{t}_i}$  Quadratterme derselben Mitteltemperatur  $\bar{t}_i$  zugeordnet, weisen jedoch unterschiedliche Gasflusswerte  $X(\bar{t}_i, j)$  auf, für  $1 \leq j \leq d_{\bar{t}_i}$ .

In der Statistik werden nun üblicherweise die vorliegenden Daten in kleine Intervalle des Definitionsbereiches von  $\bar{t}$  gruppiert, um die verschiedenen Beobachtungen innerhalb der Intervalle zu Klassen zusammenzufassen. Die Terme innerhalb einer Klasse werden anhand varianzbasierter Kriterien gewichtet. Vereinfachend arbeiten wir innerhalb dieser Analyse mit gleichgewichteten Klassen. Das Kleinste-Quadratmittel-Problem läßt sich nun folgendermaßen formulieren:

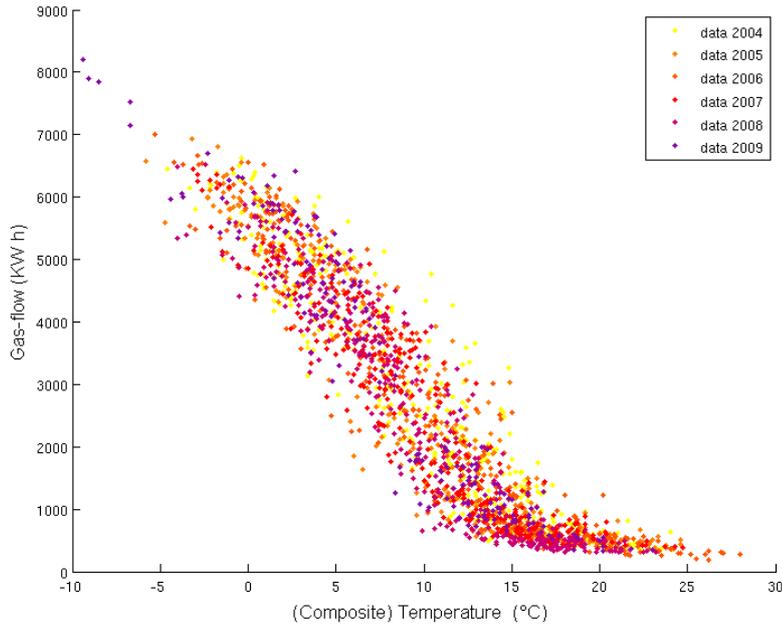


Abbildung 1: Gasflus in einem typischen Exit eines deutschen Gasnetzes

$$\min_{\theta} \sum_{\bar{t}_i=\bar{t}_{min}}^{\bar{t}_{max}} \sum_{j=1}^{d_{\bar{t}_i}} (f_0(\theta_1, \theta_2, \theta_3, \theta_4, \bar{t}_i, X) - X(\bar{t}_i, j))^2, \quad (2)$$

$$\theta \in \Theta, \quad d_{\bar{t}_i} = \text{Anzahl der Daten in der Temperaturklasse } \bar{t}_i, \quad (3)$$

wobei  $\Theta$  die zulässige Menge gemäß den Modellrestriktionen bezeichnet. Ein Beispiel einer Lösung eines Kleinste-Quadratmittel-Problems mit fixierten Temperaturgewichten ist in Abbildung 2 dargestellt.

Wir betrachten zusätzlich das klassische skalierte lineare Modell:

$$f_1(a, b, \bar{t}, X) = E(X) (a\bar{t} + b) \quad (4)$$

Hier beschränken wir die Modellierung auf einen Teil des Definitionsbereiches von  $\bar{t}$ . Innerhalb der Implementierung betrachten wir lediglich Daten von Tagen mit einer mittleren Temperatur zwischen  $-5$  und  $12$  Grad Celsius und führen damit eine Restriktion für das gewichtete Temperaturmittel  $\bar{t}$  ein. Der Grund liegt in der andersartigen Modellierung der Gasauspeisung bei sehr tiefen Temperaturen. In diesem Bereich weisen die Daten bzw. die resultierenden Residuen oftmals eine weitaus höhere Varianz auf. Die geringe Datenbasis für die seltenen sehr tiefen Temperaturen kommt erschwerend hinzu. In diesem Zusammenhang sei auf die Ergebnisse des nächsten Kapitels verwiesen.

Das entsprechende lineare Kleinste-Quadratmittel-Problem ist gegeben durch:

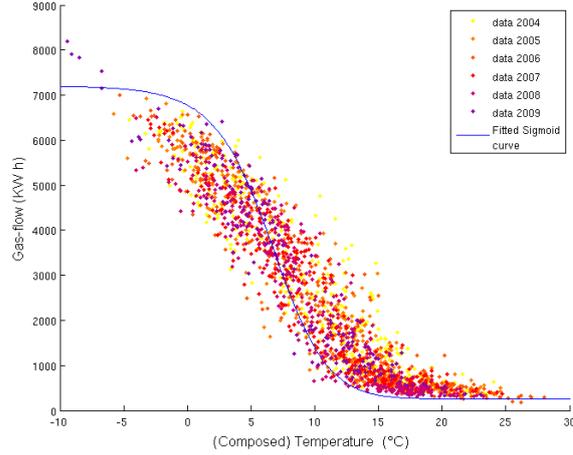


Abbildung 2: Sigmoid-Modell für die Gasflussdaten in einem Exit

$$\min_{(a,b)} \sum_{\bar{t}_i = \bar{t}_{min}^I}^{\bar{t}_{max}^I} \sum_{j=1}^{d_{\bar{t}_i}} (f_1(a, b, \bar{t}_i, X) - X(\bar{t}_i, j))^2, \quad (5)$$

wobei  $\bar{t}_{min}^I$  ( $\bar{t}_{max}^I$ ) die minimale (maximale) zusammengesetzte Temperatur nach Beschränkung der Gasflussdaten auf jene Tage mit einer Tagesmitteltemperatur aus dem Intervall  $I$  bezeichnet. In unserem Anwendungsfall ist das Temperaturintervall als  $I = [-5, 12]$  definiert. Wiederum repräsentiert  $d_{\bar{t}_i}$  die Anzahl der Daten in der Temperaturklasse  $\bar{t}_i$ .

### 2.1.3 Verallgemeinerung und Lösungsmethoden

Wir suchen zunächst die optimalen Gewichte für das Sigmoidmodell als Lösung eines verallgemeinerten nichtlinearen Kleinste-Quadratmittel-Problems. Wir betrachten

$$\bar{t}_{(d)} = \sum_{l=1}^4 \omega_l t_{(d-l+1)}. \quad (6)$$

Nun können wir die Zielfunktion umschreiben, vgl. (2), in der Form

$$\min_{(\theta, \omega)} \sum_{d=d_1}^{d_N} (f_0(\theta_1, \theta_2, \theta_3, \theta_4, \bar{t}_{(d)}, X) - X_d)^2, \quad (7)$$

Wobei  $d_1$  den ersten Tag,  $d_N$  den letzten Tag,  $N$  die Dimension von  $X$  und  $X_d$  einen Messwert am Tag  $d$  bezeichnen.

Wir integrieren zudem die Modellrestriktionen, welche in Abschnitt 2.1.2 beschrieben wurden. Nunmehr ist die Gesamtheit der Restriktionen gegeben durch:

$$(\theta, \omega) \in \Theta \times \Omega, \quad \sum_{l=1}^4 \omega_l = 1, \quad (8)$$

wobei die Menge  $\Theta \times \Omega$  die neuen Modellrestriktionen beschreibt. In analoger Weise könnte man zusammengesetzte Temperaturen basierend auf einer beliebigen festen Anzahl  $M$  der Temperaturen der vorangegangenen Tage betrachten.

$$\bar{t}_{(d)} = \sum_{l=1}^M \omega_l t_{(d-l+1)}. \quad (9)$$

In der Tat zeigen numerische Experimente, dass die Varianz der Residuen signifikant abnimmt, sofern man die Temperaturen der letzten  $M = 5$  oder  $M = 6$  Tage in die Modellierung aufnimmt. Im Gegensatz dazu nimmt die Residuenvarianz für  $M = 2$  Tage signifikant zu.

Eine *zeitabhängige* Methodik für die Bestimmung der optimalen Temperturgewichte wäre bei täglichem bzw. wöchentlichem Datenupdate ebenfalls denkbar. In der Praxis werden allerdings üblicherweise für einen Zeitraum von mehreren Monaten bzw. einem Jahr feste Datenmengen aus vergangenen Jahren verwendet, daher ist der hier vorgeschlagene Zugang geeigneter.

Für die Lösung dieses nichtlinearen restringierten Kleinste-Quadratmittel-Problems verwenden wir Methoden vom Gauss-Newton-Typ mit dem SQP-Ansatz, welcher oftmals in der Literatur empfohlen wird, vgl. [49, 53]. In der letzten Instanz kommen Optimierungstechniken zum Einsatz, welche beispielsweise in [52] umfassend beschrieben sind. Die Verwendung geeigneter Löser wie NLPLSQ [52] von Prof. K. Schittkowski garantieren im Vergleich zu anderen kommerziellen Lösern sehr zufriedenstellende Resultate. Die gleiche Verallgemeinerungsmethodik 6 innerhalb des linearen Regressionsproblems erzeugt ein lineares Kleinste-Quadratmittel-Problem mit verschiedenen Variablen und einer neuen linearen Restriktion. Die Lösung dieser Probleme ist als Lösung einer linearen Gleichungssystem gegeben. Für diesen Zweck steht eine Vielzahl von kommerziellen Lösern zur Verfügung.

#### 2.1.4 Numerische Resultate

Die Optimierungsprobleme werden für fixierte bzw. variable Temperaturgewichte mit den in 7 beschriebenen Restriktionen gelöst. Wir betrachten die H- bzw. L-Gasnetze eines großen deutschen Gasnetzbetreibers. Wir wenden die Verallgemeinerung auf jeden Exit des Netzes an und verwenden die Bezeichnung (LC). Die Datenbasis umfasst mehr als 600 Exit-Knoten. Zusätzlich analysieren wir die Gesamtausspeisung aller Exit-Knoten und bezeichnen dieses Teilproblem mit (GB). Damit wird dem Wunsch des Industriepartners nach deutschlandweit einheitlichen Temperaturgewichten Rechnung getragen. Die numerischen Resultate basieren auf Daten der Jahre 2004–2009 in stündlicher Diskretisierung und beziehen sich auf das Tagesmaximum des Gasflusses. Die gleiche Analyse ist problemlos für die Tagesmittelwerte durchführbar. Alle Rechnungen wurden auf Standard-PC's durchgeführt. Nachfolgend werden die Ergebnisse für die Temperaturgewichte aufgeführt.

##### Resultate für das H-Netz mit verallgemeinertem Sigmoidmodell

$\omega_1$	$\omega_2$	$\omega_3$	$\omega_4$	Punktwahl
0.4912	0.1555	0.0962	0.2569	(GB)
0.4306	0.2658	0.0754	0.2283	Durchschnittswerte von (LC)

### Resultate für das L-Netz mit verallgemeinertem Sigmoidmodell

$\omega_1$	$\omega_2$	$\omega_3$	$\omega_4$	Punktwahl
0.4288	0.2826	0.0853	0.2031	(GB)
0.4250	0.2623	0.0780	0.2347	Durchschnittswerte von (LC)

### Resultate für 5 und 6 Temperaturgewichte von (GB) für das H-Netz mit verallgemeinertem Sigmoidmodell

$\omega_1$	$\omega_2$	$\omega_3$	$\omega_4$	$\omega_5$	$\omega_6$	Punktwahl
0.47	0.11	0.20	-0.005	0.21		(GB) 5-Tage
0.45	0.12	0.16	0.06	0.04	0.13	(GB) 6-Tage

### Resultate für 5 und 6 Temperaturgewichte von (GB) für das L-Netz mit verallgemeinertem Sigmoidmodell

$\omega_1$	$\omega_2$	$\omega_3$	$\omega_4$	$\omega_5$	$\omega_6$	Punktwahl
0.41	0.25	0.16	-0.03	0.19		(GB) 5-Tage
0.39	0.25	0.14	0.04	-0.0098	0.17	(GB) 6-Tage

### Resultate für das H-Netz mit verallgemeinertem linearen Modell

$\omega_1$	$\omega_2$	$\omega_3$	$\omega_4$	Punktwahl
0.6988	0.0064	0.0628	0.2318	(GB)
0.4991	0.2094	0.0570	0.2345	Durchschnittswerte von (LC)

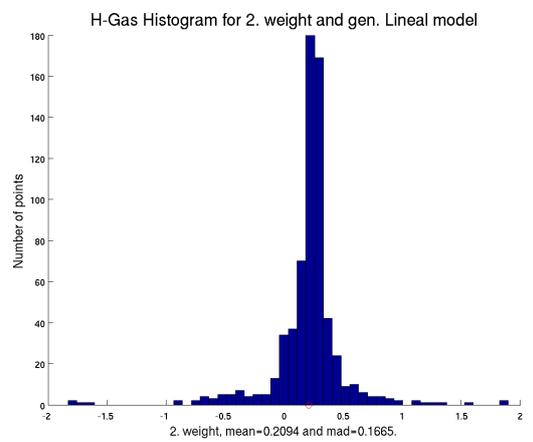
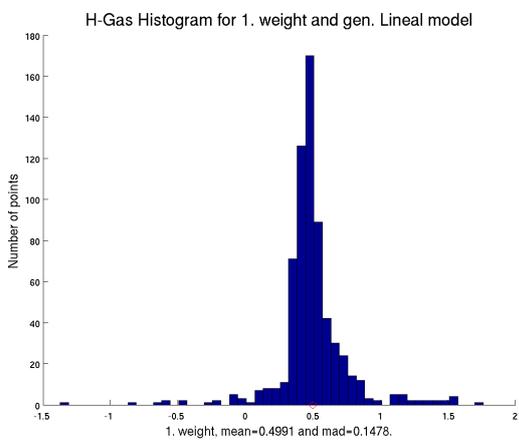
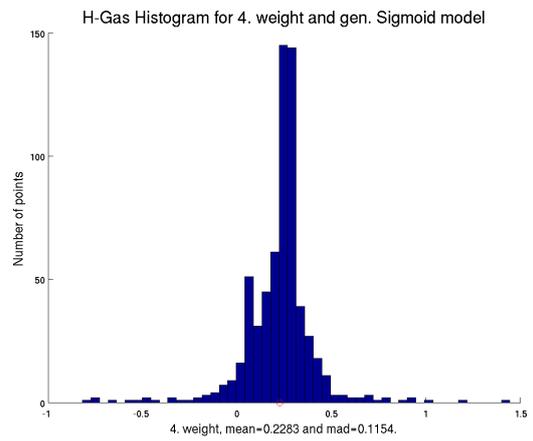
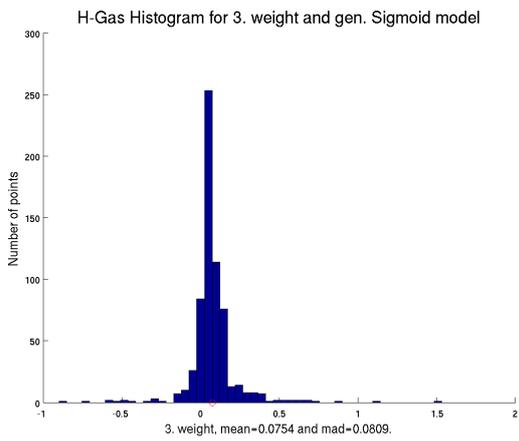
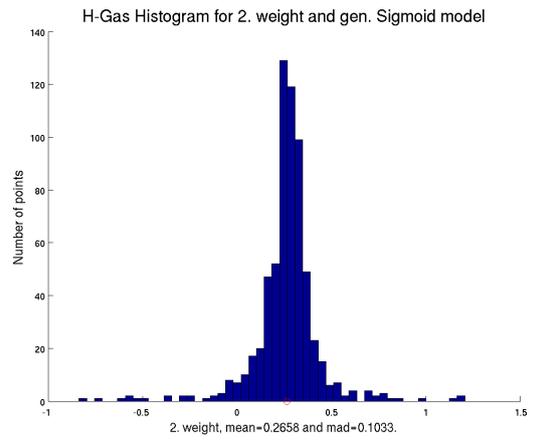
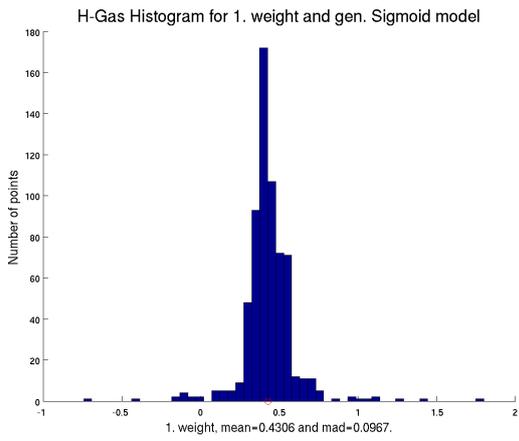
### Resultate für das L-Netz mit verallgemeinertem linearen Modell

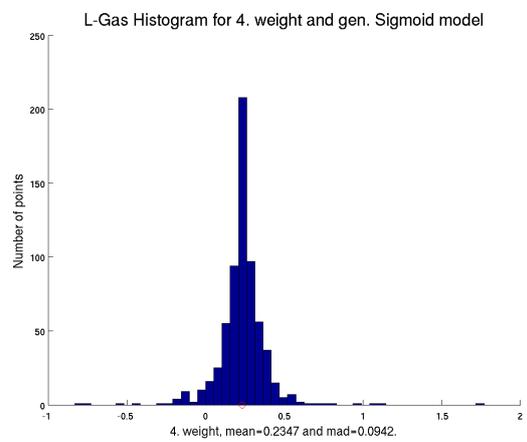
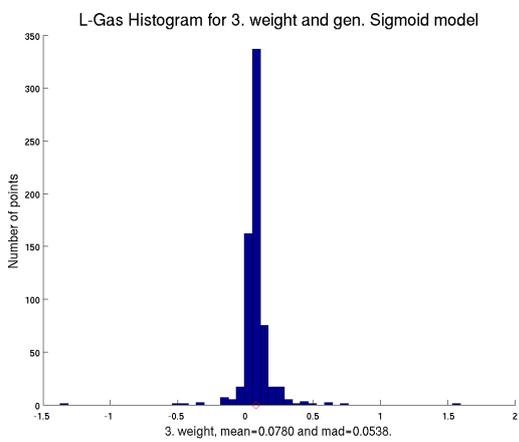
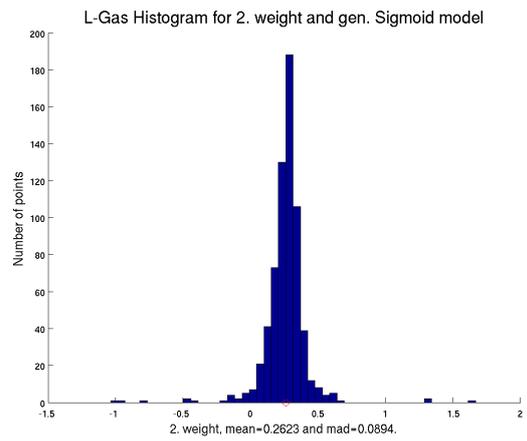
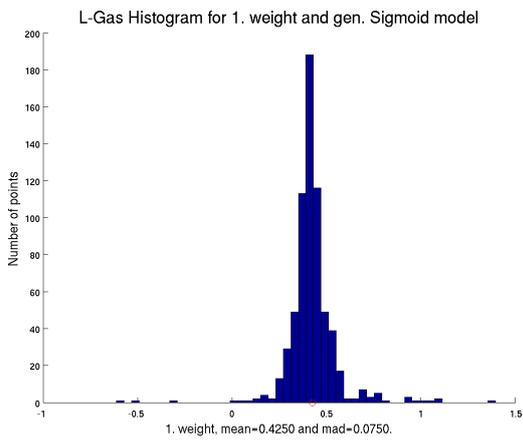
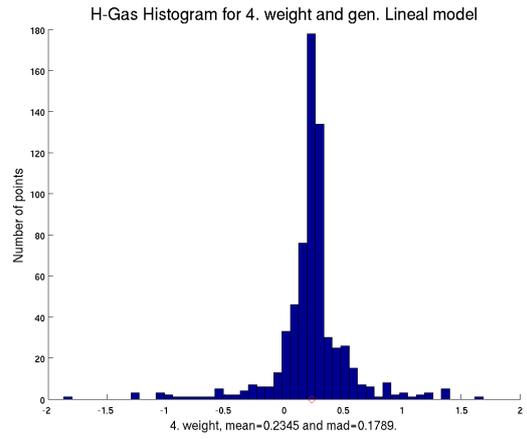
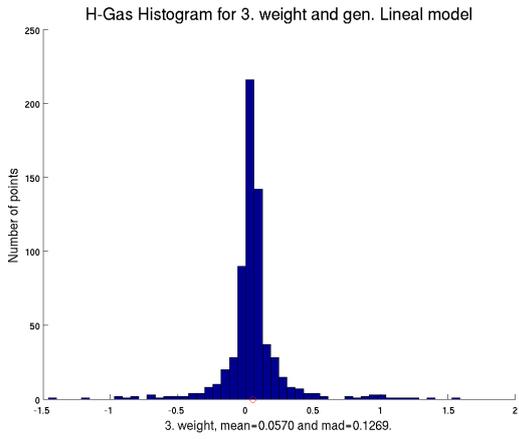
$\omega_1$	$\omega_2$	$\omega_3$	$\omega_4$	Punktwahl
0.4410	0.2337	0.0775	0.2475	(GB)
0.4912	0.2625	0.0646	0.1818	Durchschnittswerte von (LC)

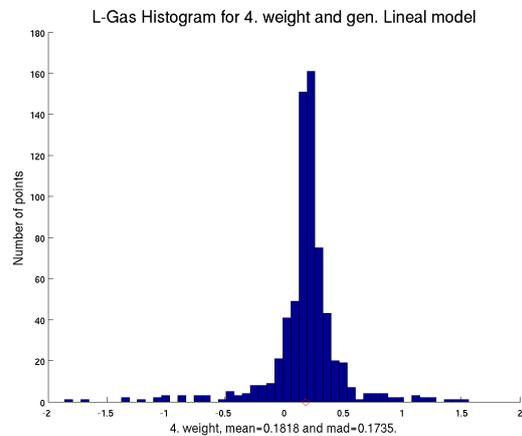
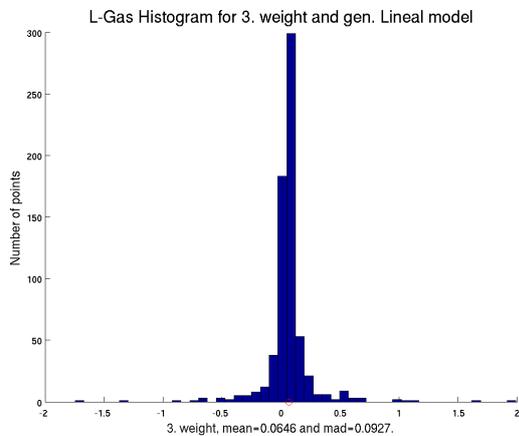
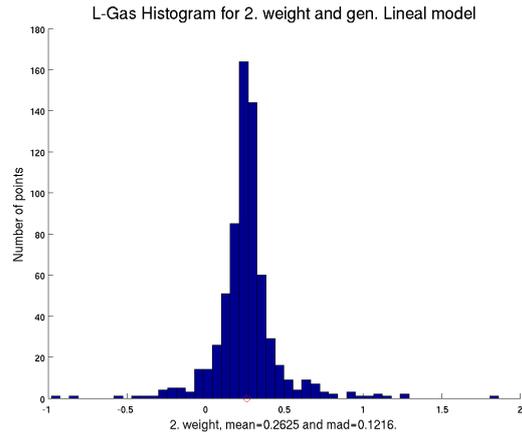
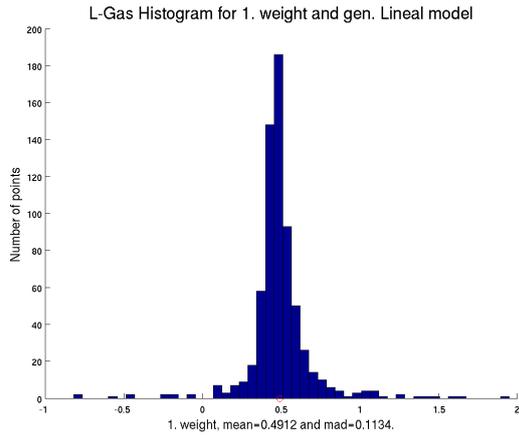
Die Resultate für (GB) zeigen bezogen auf die skalierte Zielfunktion etwa 1 bis 2 Prozent kleinere Optimalwerte im Vergleich zu den Optimalwerten des Problems mit fixierten Gewichten. Die Resultate für 5 und 6 Temperaturgewichte im sigmoiden Fall wurden hinzugefügt, um die mögliche Änderung der Gewichte bei Erhöhung der Anzahl der betrachteten Tage im Temperaturmittel zu zeigen. Die Optimalwerte bei 5 oder 6 allgemeinen Gewichten lagen in allen Fällen etwa 2 Prozent unter den Optimalwerten des Problems mit 4 allgemeinen Gewichten. Die lokalen Resultate in jedem Exit des Netzes weisen innerhalb der skalierten Zielfunktion oftmals kleinere Optimalwerte auf. Wir zeigen im folgenden Abschnitt ihre Histogramme, vgl. 2.1.5.

#### 2.1.5 Histogramme der Gewichte

Die folgenden Histogramme zeigen die Resultate für die optimalen Temperaturgewichte im H- und L-Netz. Dabei bezeichnet "weight. mean" das gewichtete Mittel und "MAD" die median absolute deviation.







## 2.1.6 Zusammenfassung

Die beschriebene verallgemeinerte Methodik generiert deutlich geeignetere Temperaturgewichte. Der beste Weg, diese Verallgemeinerung anzuwenden, ist die lokale Variante für jeden Exit-Knoten im Gasnetz. Allerdings werden in der Praxis lediglich vier einheitliche Gewichte für das gesamte Netz verwendet. Die Histogramme zeigen, dass die Verwendung der Mittelwerte der lokalen Optimallösungen eine gute Option für eine fixierte Menge von Gewichten sein könnte, obgleich einige Gewichte eine relativ hohe mittlere absolute Abweichung aufweisen.

Nach jedem Update der Datenbasis kann eine Neubestimmung einer festen Anzahl von optimalen Temperaturgewichten vorgenommen werden. Die numerischen Resultate zeigen eine signifikante Verbesserung bei Verwendung von allgemeinen Temperaturgewichten in sigmoiden bzw. linearen Modellen gegenüber den in der Literatur verwendeten Standardgewichten  $w_1 = 0.533\bar{3}$ ,  $w_2 = 0.266\bar{6}$ ,  $w_3 = 0.133\bar{3}$  und  $w_4 = 0.66\bar{6}$ . Die Implementierung dieser verallgemeinerten Technik verursacht keinen erhöhten numerischen Aufwand im Vergleich zu den üblichen Regressionsverfahren verbessert aber die Modellierung der Gasausspeisungen und damit das Verständnis der technischen Kapazitäten in realen Gasnetzen wesentlich.

Diese Thematik wurde Ende Juni 2011 mit der Einreichung der Publikation [34] abgeschlossen. Herr Leövey hat an dieser Thematik von Beginn seiner Tätigkeit gearbeitet. Die Entwicklung der Methodik, der Software und die numerischen Tests wurden wesentlich von ihm realisiert.

## 2.2 Vorhersage der Gasabnahme für Tiefsttemperaturen

### 2.2.1 Einleitung und P-Spline Regression

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. [4], [20]) eine Sigmoidregression vor wie in (1). Mehr zu nichtlinearen Modellen, deren Erweiterung und Anwendung auf die Prognose des Gasflusses nahe der Auslegungstemperatur, findet man in [10].

Die ungünstigen numerischen Eigenschaften von nichtkonvexen Modellen, insbesondere die mögliche Berechnung nur lokaler Lösungen, motivieren nun einen speziellen Zugang und die Verwendung von semi-parametrischen statistischen Modellen für die Vorhersage des Gasflusses nahe der Auslegungstemperatur. Basierend auf [63], [8], [9], [10] und [2] wird hier eine sog. P-Splines-Regression (*penalized splines*) mit Form-Restriktionen verwendet. Diese Methodik wurde wegen ihrer Einfachheit und Flexibilität gewählt. Überdies sind Sigmoid-Kurven im Bereich tiefer Temperaturen zu niedrig und bei hohen Temperaturen nicht niedrig genug. Deshalb wird das Verhalten des mittleren Gasflusses durch die P-Splines wesentlich günstiger dargestellt.

Die Abhängigkeit von standardisierten maximalen täglichen Gaslasten  $y_i$  von der Temperatur  $t_i$ ,  $i = 1, \dots, n$ , wobei  $n = 2\,343$  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 und die  $\varepsilon_i \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$ ,  $i = 1, \dots, n$ , sind unabhängige identisch normalverteilte Fehlerterme. Eine alternative Möglichkeit wäre, die  $t_i$  als gewichtete Vier-Tages-Mittel wie im vorherigen Kapitel zu wählen.

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

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

Dabei sind die  $B_j$  B-Splines vom Grad  $q$  und das Gitter  $\Delta$  ist äquidistant mit  $m - q$  Teilintervallen.

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

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

minimiert (Methode der kleinsten Quadrate). Hierbei bezeichnet  $\delta^2$  den Differenzenquotienten zweiter Ordnung, d.h.,  $\delta^2 a_j = a_j - 2a_{j-1} + a_{j-2}$ . Die Variable  $b_j \in \{0, 1\}$  kennzeichnet, für welchen B-Spline  $B_j$  die Form-Restriktion 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 Formrestriktion wurde in [2] vorgeschlagen, um die Methode der kleinsten Quadrate ohne Restriktionen und die Zurückführung auf die Lösung der Normalengleichung anwenden zu können. Da dies wegen der schlechten Kondition der Matrizen diskutabel ist, führen wir die linearen Restriktionen

$$\frac{dS_{\Delta}}{dt}(t_{\min}) = \frac{dS_{\Delta}}{dt}(t_{\max}) = 0, \quad S_{\Delta}(t) \geq 0.$$

ein und bestimmen die Spline-Koeffizienten  $a_j$  durch Lösung eines linearen kleinste Quadrate-Problems mit linearen Restriktionen. Dafür existiert spezialisierte Software, aber auch jeder Löser für quadratische Optimierungsprobleme mit linearen Restriktionen ist anwendbar (z.B. CPLEX).

Für die robuste Wahl des Glättungsparameters  $\lambda$  werden verschiedene Kriterien in [33] diskutiert. Im Projekt wurde die Methode der absolute cross validation (ACV)(siehe 3.2 in [33]) verwendet, die im vorliegenden Vorhersage-Modell sehr gute Ergebnisse lieferte.

Das angepasste Modell und die Vorhersage werden am Beispiel eines Ausspeiseknotens, der ein Stadtwerk repräsentiert, in Bild 3 illustriert. Dabei werden kubische B-Splines (d.h.  $q = 3$ ) auf einem äquidistanten Gitter verwendet (zur Veranschaulichung kubischer B-Splines sei auf [15, Kap. 6.3.6] verwiesen) und P-Splines und Sigmoid-Fits verglichen.

### 2.2.2 Zusammenfassung

Die vorgestellte Methodik der P-Spline-Regression wurde bis Ende 2011 teilweise im begleitenden Industrieprojekt mit OGE entwickelt. Herr Leövey hat im Jahr

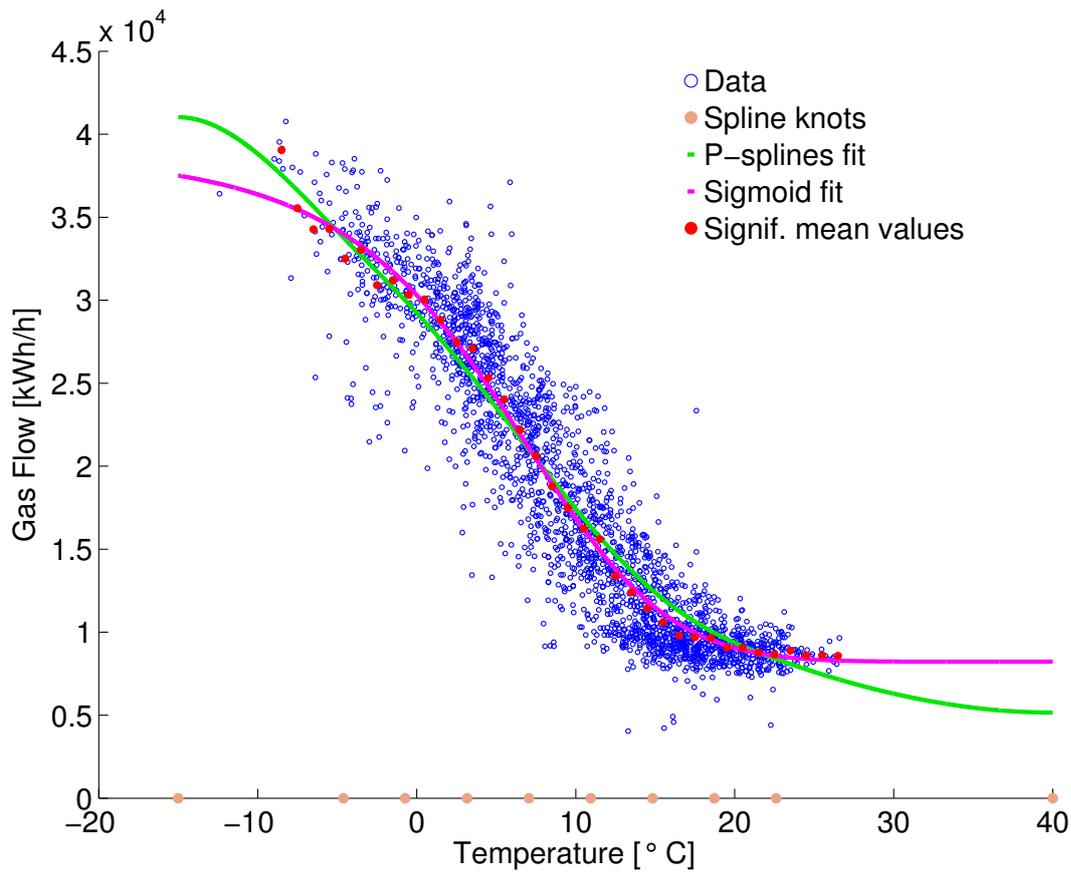


Abbildung 3: P-Splines und Sigmoid-Fits des mittleren Gasflusses an einem Exit mit  $m = 9$  kubischen B-Splines und  $t_{\min} = -15^\circ\text{C}$ ,  $t_{\max} = 40^\circ\text{C}$

2012 an einer Verbesserung der Methodik, deren Implementierung und Testung gearbeitet. Insbesondere hat er die linearen Restriktionen (Ableitung des Splines am Rand gleich 0 und Nichtnegativität des Splines) einbezogen und die Wahl des Glättungsparameters mittels cross validation entwickelt. Diese Weiterentwicklungen erforderten auch eine Veränderung des Lösungsverfahrens und eine Neuimplementierung nebst Testung.

## 2.3 Szenariogenerierung zur Modellierung der stochastischen Ausspeiselasten in einem Gastransportnetz

Wir betrachten ein Gastransportnetz mit mehreren hundert Ausspeisepunkten, für die historische Daten vorliegen. Als Grundlage für die Berechnung der technischen Kapazität werden mögliche Szenarien für den Gasfluss an sämtlichen Ausspeisepunkten und für alle Temperaturbereiche benötigt. Um den Rechenaufwand vertretbar zu halten, ist es vorteilhaft, wenn die Anzahl der Szenarien nicht zu groß ist. Nachfolgend 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.

### 2.3.1 Einleitung und Datenbeschreibung

Als Datenbasis für die Generierung von Szenarien dienen historische Daten der Flussmengen an den Ausspeisepunkten der H-Gas- und L-Gas-Transportnetze der Firma 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

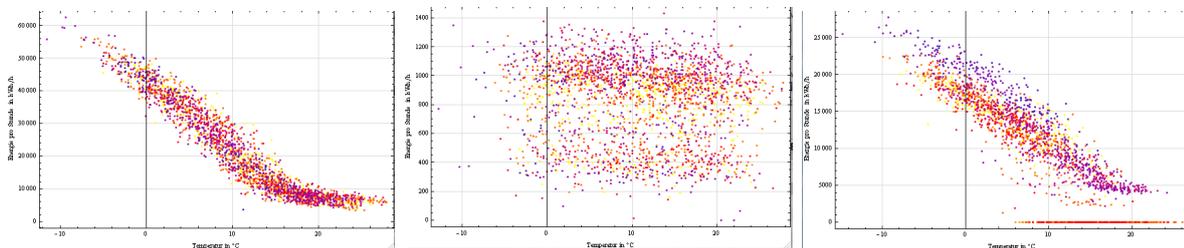


Abbildung 4: Datensätze von Ausspeisepunkten mit Stadtwerken, rechts auch mit Nullausspeisung.

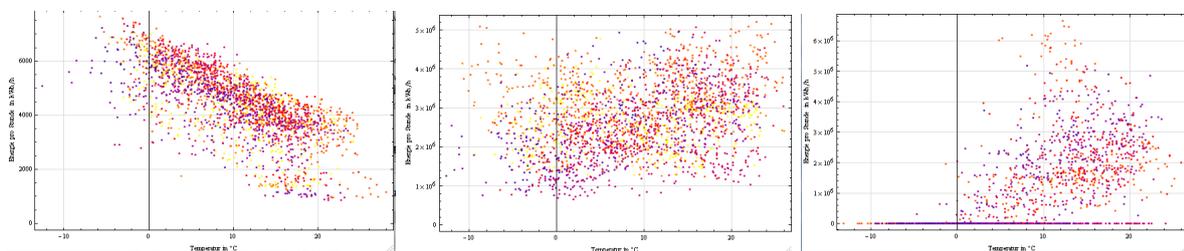


Abbildung 5: Datensätze von einem Ausspeisepunkt mit Industrieabnehmer (links), Marktübergang (Mitte), Speicher (rechts).

bei der Filterung der Datensätze zur 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. [37]).

### 2.3.2 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 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 6. Überdies wird eine Trennung in Arbeitstage und Wochenendtage durchgeführt, da diesbezüglich Unterschiede bei der Parametrisierung von Verteilungen zu erwarten sind.

Tabelle 1: Räumliche Temperaturmittelung.

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

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 2.3.4 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

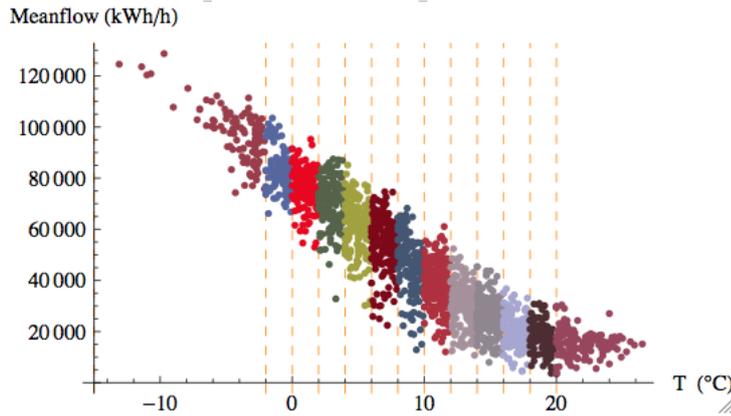


Abbildung 6: Gasfluss in Temperaturklassen.

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 drei stetige Verteilungen zum Einsatz, die Gleichverteilung für temperaturunabhängige Datensätze, die Lognormalverteilung und die Normalverteilung für die temperaturabhängigen Datensätze. Bild 7 repräsentiert die gemischte Normalverteilung exemplarisch für einen Ausspeisepunkt in der Temperaturklasse ( $6^\circ\text{C}, 8^\circ\text{C}$ ]. Ist der Gasfluss innerhalb einer Temperaturklasse vollständig

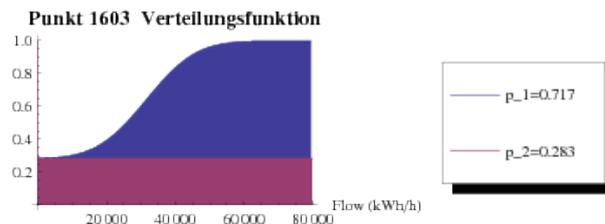


Abbildung 7: Verteilungsfunktion einer shifted Normalverteilung.

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.

Im Tiefsttemperaturbereich wird in jedem Ausspeiseknoten eine univariate Normalverteilung modelliert, deren Mittelwert durch  $\bar{y} = S_\Delta(t)$  mit  $t = -16^\circ\text{C}$  geschätzt (vgl. Kapitel 2.2) und als deren Standardabweichung die für die Temperaturklasse ( $-4^\circ\text{C}, -2^\circ\text{C}$ ] geschätzte verwendet wird.

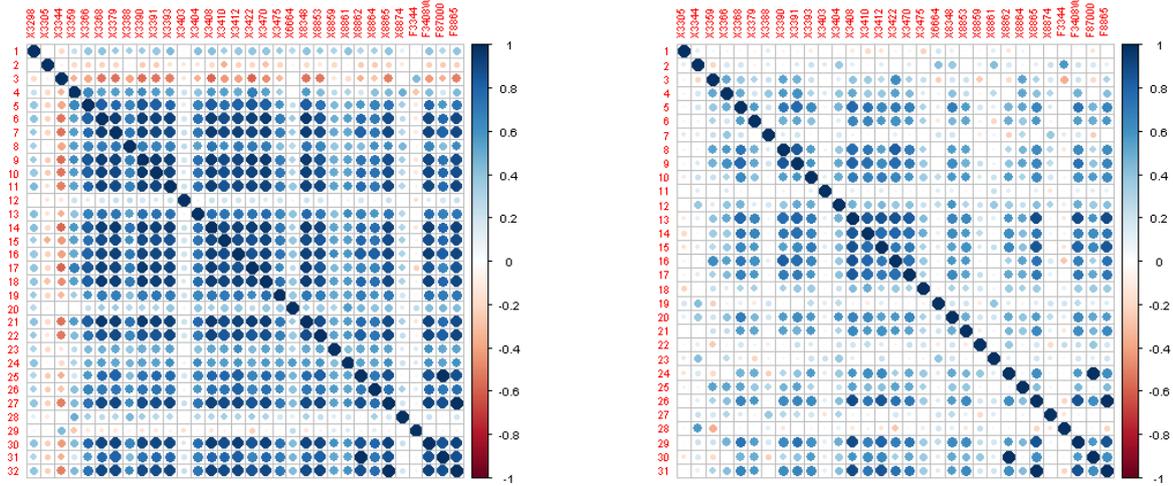


Abbildung 8: Korrelationsplots für die Temperaturklassen (10, 12] und (18, 20].

Für die Schätzung der Kovarianzmatrix einer multivariaten (Log)Normalverteilung werden die erwartungswertbereinigten Residuen aus dem univariaten Ansatz mit Trennung nach Tagestyp übernommen. Im multivariaten Fall findet ausschließlich die (Log)Normalverteilung Verwendung. 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 8 dargestellt.

### 2.3.3 ANOVA Dekomposition von Funktionen und effektive Dimension

Die ANOVA-Dekomposition (”Analysis of Variance”) wurde zuerst als Werkzeug in der Statistik (siehe [24]) vorgeschlagen und erst später von Sobol’ [57] zur Untersuchung von numerischen Integrationsmethoden verwendet.

Wir betrachten eine Dichtefunktion  $\rho$  auf  $\mathbb{R}^d$  der Form

$$\rho(\xi) = \prod_{i=1}^d \rho_i(\xi_i).$$

und wie in [14] die gewichteten  $\mathcal{L}_p$  Räume über  $\mathbb{R}^d$ , i.e.,  $\mathcal{L}_{p,\rho}(\mathbb{R}^d)$ , mit der Norm

$$\|f\|_{p,\rho} = \begin{cases} \left( \int_{\mathbb{R}^d} |f(\xi)|^p \rho(\xi) d\xi \right)^{\frac{1}{p}} & \text{if } 1 \leq p < +\infty, \\ \text{ess sup}_{\xi \in \mathbb{R}^d} \rho(\xi) |f(\xi)| & \text{if } p = +\infty. \end{cases}$$

Es sei  $D = \{1, \dots, d\}$  und  $f \in \mathcal{L}_{1,\rho}(\mathbb{R}^d)$ . Die Projektion  $P_k$ ,  $k \in D$ , ist definiert durch

$$(P_k f)(\xi) := \int_{-\infty}^{\infty} f(\xi_1, \dots, \xi_{k-1}, s, \xi_{k+1}, \dots, \xi_d) \rho_k(s) ds \quad (\xi \in \mathbb{R}^d).$$

Für  $u \subseteq D$  bezeichnen wir mit  $|u|$  die Kardinalität von  $u$ , mit  $-u$  die Menge  $-u = D \setminus u$ , und die verallgemeinerte Projektion

$$P_u f = \left( \prod_{k \in u} P_k \right) (f),$$

wobei unter dem Produktzeichen die Komposition der Projektionen  $P_k$  zu verstehen ist. Die Funktion  $P_u f$  ist konstant bzgl. aller  $\xi_k$ ,  $k \in u$ .

Die ANOVA Dekomposition von  $f \in \mathcal{L}_{1,\rho}(\mathbb{R}^d)$  besitzt dann die Form [58, 32]

$$f = \sum_{u \subseteq D} f_u$$

wobei  $f_u$  nur von  $\xi^u$ , i.e., von den Variablen  $\xi_j$  mit  $j \in u$  abhängt. Der ANOVA Term  $f_u$  besitzt die Eigenschaft  $P_j f_u = 0$  für alle  $j \in u$  und kann durch die Beziehungen

$$f_\emptyset = I_{d,\rho}(f) = P_D(f) \quad \text{and} \quad f_u = P_{-u}(f) - \sum_{v \subseteq u} f_v.$$

definiert werden. In der Arbeit [32] wurde gezeigt, dass  $f_u$  auch direkt mit Hilfe von Projektionen dargestellt werden kann

$$f_u = \sum_{v \subseteq u} (-1)^{|u|-|v|} P_{-v} f = P_{-u}(f) + \sum_{v \subset u} (-1)^{|u|-|v|} P_{u-v}(P_{-u}(f)),$$

wobei die Projektionen  $P_{-u}$  and  $P_{u-v}$  Integration bzgl.  $\xi_j$ ,  $j \in D \setminus u$  und  $j \in u \setminus v$ , bedeuten. Diese zweite Darstellung motiviert, dass  $f_u$  im wesentlichen so glatt wie  $P_{-u} f$  ist [14, Theorem 2].

Das folgende Resultat ist bereits länger bekannt (siehe [58]).

**Aussage:**

Falls  $f \in \mathcal{L}_{2,\rho}(\mathbb{R}^d)$ , so sind die ANOVA Terme  $\{f_u\}_{u \subseteq D}$  orthogonal in  $\mathcal{L}_{2,\rho}(\mathbb{R}^d)$ .

Für die Varianz  $\sigma(f)$  von  $f$  gilt deshalb

$$\sigma^2(f) = \sum_{\emptyset \neq u \subseteq D} \sigma_u^2(f) \quad \text{mit} \quad \sigma_u^2(f) = \|f_u\|_{2,\rho}^2 \quad (u \subseteq D)$$

Die Quotienten  $\frac{\sigma_u^2(f)}{\sigma^2(f)} \in [0, 1]$ , wobei  $\sigma(f) > 0$  vorausgesetzt wird, sind dann ein Indikator für die Bedeutung der Variablengruppe  $\xi^u$  in der Funktion  $f$ .

In [42, 35] wurden diese Quotienten genutzt, um die Dimensionsverteilung einer Funktion zu definieren. Die Dimensionsverteilung einer Funktion  $f$  im Superpositions- bzw. im Abschneide-Sinn ist die Wahrscheinlichkeitsverteilung  $\nu_S$  bzw.  $\nu_T$  definiert durch

$$\nu_S(s) := \nu_S(\{s\}) = \sum_{|u|=s} \frac{\sigma_u^2(f)}{\sigma^2(f)} \quad \text{bzw.} \quad \nu_T(s) = \sum_{\max\{j:j \in u\}=s} \frac{\sigma_u^2(f)}{\sigma^2(f)} \quad (s \in D).$$

In der Arbeit [35] werden auch der Mittelwert, die Varianz und weitere Momente der Dimensionsverteilungen  $\nu_S$  bzw.  $\nu_T$  berechnet und untersucht.

Für kleines  $\varepsilon \in (0, 1)$  ( $\varepsilon = 0.01$  wird in einer Reihe von Publikationen empfohlen) ist die *effektive Superpositions- bzw. Abschneide-Dimension*  $d_S(\varepsilon) \in D$  bzw.  $d_T(\varepsilon) \in D$  als das  $(1 - \varepsilon)$ -Quantil von  $\nu_S$  bzw.  $\nu_T$  definiert, d.h.

$$\begin{aligned} d_S(\varepsilon) &= \min\{s \in D : \nu_S(u) \geq 1 - \varepsilon, |u| \leq s\} \\ d_T(\varepsilon) &= \min\{s \in D : \nu_T(\{1, \dots, s\}) \geq 1 - \varepsilon\}. \end{aligned}$$

Es gilt  $d_S(\varepsilon) \leq d_T(\varepsilon)$  und (siehe [58])

$$\max \left\{ \left\| f - \sum_{|u| \leq d_S(\varepsilon)} f_u \right\|_{2,\rho}, \left\| f - \sum_{u \subseteq \{1, \dots, d_T(\varepsilon)\}} f_u \right\|_{2,\rho} \right\} \leq \sqrt{\varepsilon} \sigma(f).$$

Ist die effektive Superpositions-Dimension  $d_S(\varepsilon)$  einer Funktion  $f$  klein, so ist das in aller Regel ein Hinweis darauf, dass QMC Methoden Monte Carlo Methoden bei der Integration dieser Funktion überlegen sind (vgl. auch [61]). Algorithmen zur Berechnung der effektiven Dimensionen werden z.B. in [58, 59] diskutiert, wobei die effektive Abschneide-Dimension deutlich einfacher zu berechnen ist.

Im Kontext der Erzeugung von Szenarien für die stochastischen Ausspeiselasten in einem Gastransportnetz sind relevanten Funktionen in Abhängigkeit von  $\xi \in \mathbb{R}^d$  (den Ausspeisungen) gerade die Flüsse und Drücke im Gasnetzwerk. Da diese Funktionen Lösungen von Systemen von Gleichungen und Ungleichungen sind, die durch glatte (unendlich oft differenzierbare) Funktionen beschrieben werden, sind sie *stückweise glatt* und gehören also nicht zu den Funktionenräumen  $\mathcal{W}_{2,\text{mix}}^{(1,\dots,1)}([0, 1]^d)$  nach Transformation von  $\mathbb{R}^d$  auf  $[0, 1]^d$ , die in Kapitel 2.3.4 definiert werden.

In [13, 14] und unseren Arbeiten [17, 48] wird aber für wichtige Teilklassen von stückweise glatten Funktionen, nämlich solchen der Form  $\max\{0, g(\cdot)\}$  (mit glatter Funktion  $g$ ) und für stückweise lineare Funktionen unter gewissen Voraussetzungen, gezeigt, dass alle ANOVA Terme bis auf den Term  $f_D$  höchster Ordnung glatt sind, falls nur alle Marginaldichten  $\rho_k$ ,  $k \in D$ , glatt sind. Ist nun überdies die effektive Superpositions-Dimension  $d_S(\varepsilon)$  dieser Funktionen kleiner als  $d$  bzw. am besten so klein wie möglich, so ist der einzige nichtglatte ANOVA Term (sehr) klein und kann vernachlässigt werden.

Aus den Arbeiten [3, 58, 59, 60, 62] ist bekannt, dass im Fall (log)normalverteilter Zufallsvektoren die Art und Weise der Zerlegung der Kovarianzmatrix  $\Sigma$ , d.h. die Wahl der Matrix  $A$ , so dass  $\Sigma = AA^\top$ , einen bedeutenden Einfluss auf die effektive Dimension haben kann. Insbesondere die Standardwahl von  $A$  über die Cholesky Zerlegung erweist sich als sehr ungünstig, da dadurch die effektive Dimension nicht oder kaum reduziert wird. Als günstig erweist sich in vielen Fällen die sog. Hauptkomponentenanalyse ("principal component analysis" (PCA)). Dabei wird  $A = (\sqrt{\lambda_1}u_1, \dots, \sqrt{\lambda_d}u_d)$  gewählt, wobei  $\lambda_1 \geq \dots \geq \lambda_d \geq 0$  die Eigenwerte von  $\Sigma$  in absteigender Ordnung und  $u_i$ ,  $i = 1, \dots, d$ , die orthogonalen Eigenvektoren von  $\Sigma$  sind. Viele Autoren berichten in diesem Fall von einer enormen Reduktion der effektiven Abschneide-Dimension (siehe auch [17]). Dies wird im nachfolgenden Kapitel 2.3.4 berücksichtigt.

### 2.3.4 Erzeugung von Szenarien

Das Ziel besteht darin, eine nicht zu große Zahl  $n$  von Szenarien für die Ausspeiselast an allen  $d$  Ausspeisepunkten eines Gas-Transportnetzes, für die statistische Daten vorliegen, zu bestimmen. Unter Szenarien verstehen wir dabei  $d$ -dimensionale Vektoren, deren  $i$ -te Komponente einen möglichen Wert für die Ausspeiselast am Punkt  $i$ ,  $i = 1, \dots, d$ , darstellt. 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. [39]), 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

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

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. [40, Kap. 3]). Man nennt

$$Q_{n,d}(f) = \frac{1}{n} \sum_{j=1}^n f(\eta^j)$$

Quasi-Monte Carlo Methode, falls die Folge  $(D_n(\eta^1, \dots, \eta^n))$  gegen 0 konvergiert.

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. Es gilt nämlich

$$\mathbb{E} \left( \left| \int_{[0,1]^d} f(x) dx - \frac{1}{n} \sum_{j=1}^n f(\eta^j) \right|^2 \right)^{\frac{1}{2}} = \frac{\sigma(f)}{\sqrt{n}},$$

wobei  $\sigma^2(f) = \int_{[0,1]^d} (f(x) - \int_{[0,1]^d} f(y) dy)^2 dx$ .

Es ist inzwischen gelungen, solche Folgen  $(\eta^j)$  zu finden, dass für geeignete lineare normierte Räume  $\mathbb{F}_d$  von Funktionen und für jedes  $\delta \in (0, \frac{1}{2}]$  eine nur von  $\delta$  und nicht von der Dimension  $d$  abhängige Konstante  $C(\delta)$  existiert, so dass

$$(**) \quad \sup_{f \in \mathbb{B}_d} \left| \int_{[0,1]^d} f(x) dx - \frac{1}{n} \sum_{j=1}^n f(\eta^j) \right| \leq C(\delta) n^{-1+\delta}$$

gilt. Hierbei ist  $\mathbb{B}_d = \{f \in \mathbb{F}_d : \|f\|_d \leq 1\}$ , d.h. die Einheitskugel im Raum  $\mathbb{F}_d$ . Die Konvergenzordnung in (\*\*) ist "fast"  $O(\frac{1}{n})$  und damit etwa quadratisch besser als  $O(\frac{1}{\sqrt{n}})$  für Monte-Carlo und ebenfalls unabhängig von der Dimension. Statt 10.000 Monte Carlo Samples können also gut 100 geeignet gewählte QMC Samples auch für große Dimensionen verwendet werden, falls  $f$  zu  $\mathbb{F}_d$  gehört.

Als geeignete Funktionenräume  $\mathbb{F}_d$  haben sich sog. kernreproduzierende Hilberträume von Funktionen auf  $[0, 1]^d$  erwiesen (siehe [23, 5, 6]). Ist nämlich  $K : [0, 1]^d \times [0, 1]^d \rightarrow \mathbb{R}$  ein Kern mit der Eigenschaft  $K(\cdot, y) \in \mathbb{F}_d$  und  $\langle f, K(\cdot, y) \rangle = f(y)$  für alle  $y \in [0, 1]^d$  und  $f \in \mathbb{F}_d$ , und das Integral

$$I_d(f) = \int_{[0,1]^d} f(x) dx$$

ein lineares stetiges Funktional auf  $\mathbb{F}_d$ , so gilt

$$\sup_{f \in \mathbb{B}_d} |I_d(f) - Q_{n,d}(f)| = \sup_{\|f\|_d \leq 1} |\langle f, h_n \rangle| = \|h_n\|_d,$$

wobei  $\langle \cdot, \cdot \rangle$  das Skalarprodukt und  $\|\cdot\|_d$  die Norm in  $\mathbb{F}_d$  bezeichnen. Die Funktion  $h_n \in \mathbb{F}_d$  hat die Form

$$h_n(x) = \int_{[0,1]^d} K(x, y) dy - \frac{1}{n} \sum_{i=1}^n K(x, \eta^i) \quad (\forall x \in [0, 1]^d)$$

und ist also vollständig durch den Kern  $K$  bestimmt.

Ein wichtiges Beispiel für einen kernreproduzierenden Hilbertraum ist der gewichtete Tensorprodukt Sobolev-Raum (siehe [56])

$$\mathbb{F}_d = \mathcal{W}_{2,\gamma,\text{mix}}^{(1,\dots,1)}([0, 1]^d) = \bigotimes_{i=1}^d W_{2,\gamma}^1([0, 1])$$

mit der gewichteten Norm  $\|f\|_\gamma^2 = \langle f, f \rangle_\gamma$  und dem gewichteten Skalarprodukt

$$\langle f, g \rangle_\gamma = \sum_{u \subseteq \{1, \dots, d\}} \prod_{j \in u} \gamma_j^{-1} \int_{[0,1]^{|u|}} \frac{\partial^{|u|}}{\partial x^u} f(x^u, 1^{-u}) \frac{\partial^{|u|}}{\partial x^u} g(x^u, 1^{-u}) dx^u,$$

Die Funktionen in diesem Raum besitzen also partielle Ableitungen der Form

$$\frac{\partial^d f}{\partial x_1 \cdots \partial x_d}$$

im Sinne von Sobolev. Der Kern hat die Gestalt

$$K_{d,\gamma}(x, \tilde{x}) = \prod_{j=1}^d (1 + \gamma_j [1 - \max\{x_j, \tilde{x}_j\}]) \quad (x, \tilde{x} \in [0, 1]^d).$$

Um die Fehlerabschätzung (\*\*\*) zu erhalten, müssen die nichtnegativen Gewichte ( $\gamma_j$ ) die Bedingung erfüllen

$$\sum_{j=1}^{\infty} \gamma_j^{\frac{1}{2(1-\delta)}} < \infty.$$

Quasi-Monte Carlo Methoden, die Fehlerabschätzungen der Form (\*\*\*) ermöglichen, basieren auf randomisierten  $(t, d)$ -Folgen und lattice rules.

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 [40, 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 [5, 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 Sobol' Folgen mit  $b = 2$ . Eine ausgezeichnete extrem hochdimensionale Implementierung von Sobol' Folgen liegt mit [27] vor und wird im folgenden verwendet. Es ist überdies oft vorteilhaft die besten Eigenschaften von MC und QMC Methoden zu kombinieren (siehe [41]). Die einfachste Form der Randomisierung von solchen Folgen ist digital-shifting. Dabei wird eine random shift  $\Delta \in [0, 1)^d$  zu jedem Punkt der determinierten Menge addiert indem die Operationen über dem Ring  $\{0, 1, \dots, b - 1\}$  ausgeführt werden.

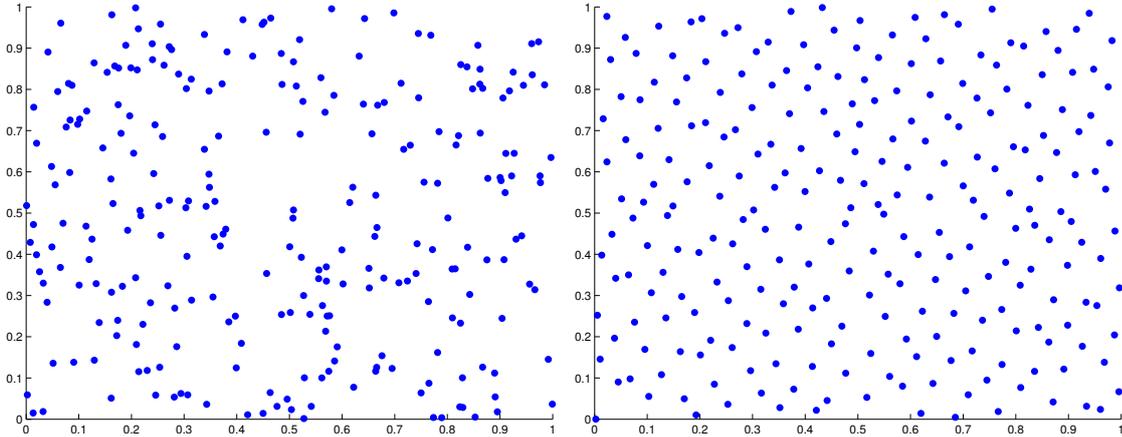


Abbildung 9: Links:  $N = 256$  Monte Carlo Mersenne-Twister-Punkte für  $d = 500$ , Projektion (8,9). Rechts:  $N = 256$  RQMC random digitally-shifted Sobol'-Punkte für  $d = 500$ , Projektion (8,9).

Unter Verwendung des random digital-shifting von Sobol' Folgen wird nun eine größere Anzahl  $N$ ,  $1\,000 \leq N \leq 10\,000$ , Punkte  $\eta^j$  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 (Marginal) 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 [12, Kap. 2.3.2] verwendet und anschließend mit einem Newton-Schritt bzgl.  $\Phi_i$  wie in [36] 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 PCA-Zerlegung  $\Sigma = AA^\top$  der Kovarianzmatrix  $\Sigma$  (vgl. Kapitel 2.3.3). Dann sind die  $s$ -dimensionalen Vektoren

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

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  $\xi^j$  im allgemeinen noch groß ist, wird diese abschließend mit Hilfe der Methode der *optimalen Szenarioreduktion* (vgl. [7, 16, 18, 19]) auf eine kleinere Anzahl  $n$  (z.B.  $n = 50, 100$ ) reduziert, so dass die verbleibenden  $n$  Szenarien die ursprünglichen  $N$  bestmöglich repräsentieren. Mathematisch bedeutet dies, dass die beste Approximation  $Q_J$  basierend auf Szenarien  $\xi^j$ ,  $j \in J$ , mit  $|J| = n$  und Wahrscheinlichkeiten  $q_j$ ,  $j \in J$ , zu einer Wahrscheinlichkeitsverteilung  $P$  mit  $N$  gleichwahrscheinlichen Szenarien bzgl. eines geeigneten metrischen Abstandes  $\mathbb{D}$  von Wahrscheinlichkeitsmaßen bestimmt wird. Da die zu lösenden Buchungsvalidierungs- und Kapazitätsmaximierungsprobleme stochastische nichtlineare gemischt-ganzzahlige Zulässigkeits- bzw. Optimierungsprobleme sind, erscheinen sog. Diskrepanzen und insbesondere die Kolmogorov-Metrik, d.h. der gleichmäßige Abstand der multivariaten Verteilungsfunktionen von  $P$  und  $Q_J$ ,

$$\mathbb{D}(P, Q_J) = \sup_{x \in \mathbb{R}^d} \left| \frac{1}{N} \sum_{i=1, \xi^i \leq x}^N 1 - \sum_{j \in J, \xi^j \leq x} q_j \right|$$

als geeignete Abstände (siehe [47]). Allerdings zeigen die Ergebnisse von [21, 22], dass die optimale Szenario-Reduktion bzgl. der Kolmogorov-Metrik für größere Dimensionen  $d$  wie im vorliegenden Fall gegenwärtig noch nicht in geeigneter Zeit lösbar ist. Deswegen verwenden wir den  $L_1$ -Wasserstein Abstand (siehe [44])

$$W_1(P, Q_J) = \inf \left\{ \sum_{i=1}^n \sum_{j \in J} \eta_{ij} \|\xi^i - \xi^j\| : \eta_{ij} \geq 0, \sum_{i=1}^n \eta_{ij} = q_j, \sum_{j \in J} \eta_{ij} = \frac{1}{N} \right\}$$

von  $P$  und  $Q_J$ . Die optimalen Wahrscheinlichkeiten  $q_j$ ,  $j \in J$ , bei gegebener Indexmenge  $J$  können durch Lösung des Problems

$$\min \left\{ W_1(P, Q_J) : q_j \geq 0, j \in J, \sum_{j \in J} q_j = 1 \right\} \quad (10)$$

(optimale Neuverteilung) explizit bestimmt werden (siehe [7]) zu

$$q_j = \frac{1}{N}(1 + |I(j)|), \text{ wobei } I(j) = \left\{ i \in \{1, \dots, N\} \setminus J : i \in \arg \min_{i \notin J} \|\xi^i - \xi^j\| \right\} \quad (j \in J).$$

Die Wahrscheinlichkeiten der ausgesonderten Szenarien werden bei der Neuverteilung also zu der des nächstgelegenen verbleibenden Szenario addiert. Das Infimum von (10) ist gerade

$$D_J = \frac{1}{N} \sum_{i \notin J} \min_{j \in J} \|\xi^i - \xi^j\|$$

und das zu lösende Problem der besten Approximation hat die Gestalt

$$\min \{ D_J : J \subseteq \{1, \dots, N\}, |J| = n \}.$$

Dieses kombinatorische Optimierungsproblem ist in der Literatur als *n-median Problem* bekannt und NP-schwer (siehe [28]).

Seien nun  $c_{ij} = \|\xi^i - \xi^j\|$ ,  $i, j = 1, \dots, N$ , die Abstände der Szenarios  $\xi^i$  und  $\xi^j$  und  $y_i \in \{0, 1\}$ ,  $i = 1, \dots, N$ , zeige an, ob Szenario  $\xi^i$  gestrichen wird oder nicht.

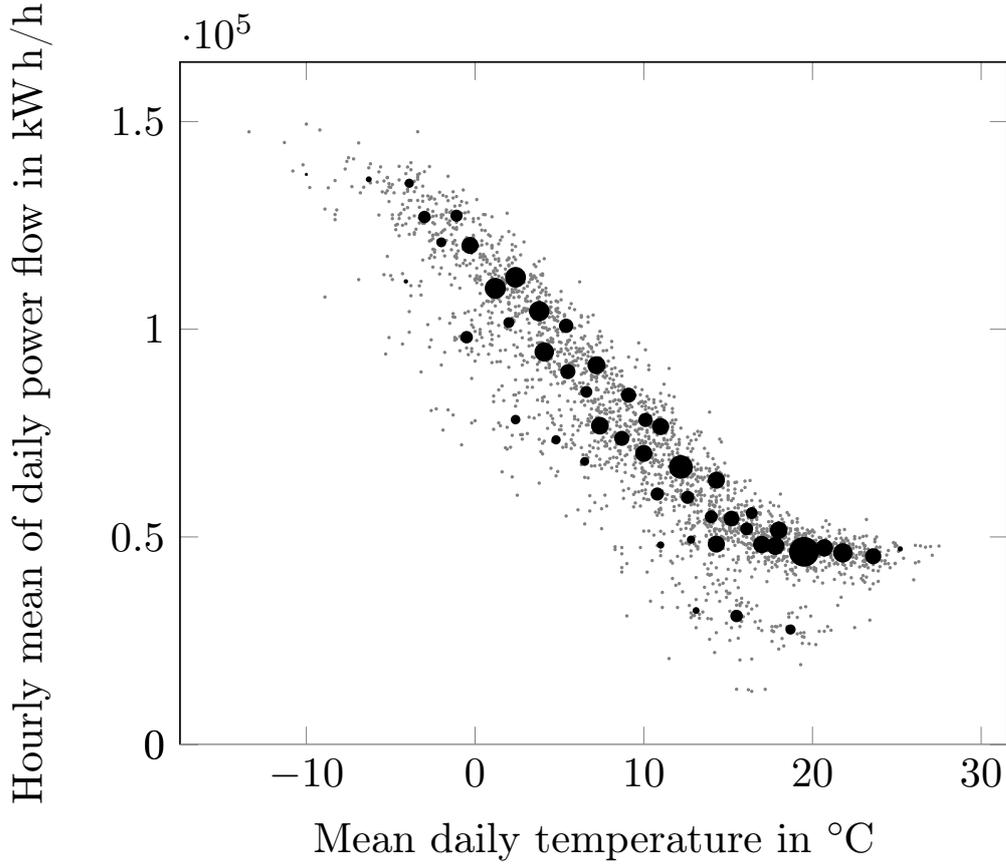


Abbildung 10: Optimale Szenarioreduktion angewandt um eine Menge von  $N = 2340$  gleichwahrscheinlichen temperaturabhängigen mittleren täglichen Gasfluss-Szenarios an einem typischen Exitknoten auf  $n = 50$  Szenarios mit neuen Wahrscheinlichkeiten zu reduzieren. Die Größe der Kreise ist proportional zur Wahrscheinlichkeit der verbliebenen Szenarios.

Das Problem der optimalen Reduktion von  $N$  auf  $n$  Szenarios erlaubt dann die Formulierung (siehe auch [1])

$$\min \left\{ \frac{1}{N} \sum_{i,j=1}^N c_{ij} x_{ij} : 0 \leq x_{ij} \leq 1, y_i \in \{0, 1\}, x_{ij} \leq y_i, \sum_{\substack{i=1 \\ i \neq j}}^N x_{ij} + y_j = 1, \sum_{i=1}^N y_i = n \right\}$$

als (großes) gemischt-ganzzahliges lineares Optimierungsproblem (mit  $N^2$  stetigen und  $N$  binären Variablen) und ermöglicht deshalb dessen Lösung mit Standard-Software (wie z.B. CPLEX). Ist nämlich  $J \subset \{1, \dots, N\}$  die Indexmenge mit  $y_j = 1$ ,  $j \in J$ , so folgt sofort, dass  $x_{ij} = 0$ ,  $i \notin J$ ,  $j = 1, \dots, N$ , und  $x_{ij} = 0$ ,  $i = 1, \dots, N$ ,  $i \neq j$ ,  $j \in J$ , gilt und  $\sum_{i \in J} x_{ij} = 1$ ,  $j \notin J$ . Daraus ergibt sich die untere Schranke

$$D_J = \frac{1}{N} \sum_{j \notin J} \min_{i \in J} c_{ij} \leq \frac{1}{N} \sum_{i \in J} \sum_{j \notin J} c_{ij} x_{ij} = \frac{1}{N} \sum_{i,j=1}^N c_{ij} x_{ij},$$

die für

$$0 \leq x_{ij} = \frac{\min_{i \in J} c_{ij}}{n c_{ij}} \leq 1 \quad (i \in J, j \notin J)$$

angenommen wird.

Alternativ sei auf heuristische Algorithmen, approximative Algorithmen basierend auf semidefiniten Optimierung und ein für großes  $N$  geeignetes branch-cut-and-price Verfahren (siehe [1, 43, 46]) verwiesen. Wir verwenden oft einfache Vorwärts- bzw. Rückwärts-Greedy Heuristiken mit abschließender optimaler Neuverteilung, die für viele Zwecke ausreichen. Dabei kommt die Vorwärts-Variante zum Einsatz, wenn  $n < \frac{N}{4}$  (siehe [18]). Die optimale Szenarioreduktion wird in Abbildung 10 illustriert.

### 2.3.5 Zusammenfassung

Die Arbeiten zur Datenanalyse und zur Schätzung von Wahrscheinlichkeitsverteilungen erfolgten im begleitenden Industrieprojekt mit OGE.

Die Generierung von Gasabnahme-Szenarien war die zentrale Aufgabenstellung des Teilprojekts. An dieser Thematik hat Herr Leövey während des gesamten Zeitraums gearbeitet. Anfangs wurde die Szenarien-Generierung noch mit Hilfe der Monte Carlo Methodik (und dem Mersenne Twister als (Pseudo-) Zufallszahlengenerator) realisiert (siehe HU Vortrag auf dem Kapazitäts-Workshop in Bonn, Nov. 2010). Die hohen Dimensionen der Szenarien und die Notwendigkeit der Verwendung einer nur geringen Anzahl von Szenarien bei guter Approximationsgüte führten dann aber über ein tiefgehendes Literaturstudium zu randomisierten Quasi-Monte Carlo Methoden, die von einer Dimensionsreduktion begleitet werden.

Herr Leövey hat sowohl an der Entwicklung der immer detaillierter werdenden Methodiken als auch an deren komplexen Implementierung gearbeitet. Wesentliche Weiterentwicklungen wurden im zweiten Halbjahr 2011 und 2012 erreicht. Die aktuelle Methodik verwendet neueste Entwicklungen der QMC-Forschung.

## 3 Veröffentlichungen

Die oben genannten wissenschaftlichen Ergebnisse wurden in den 5 Publikationen [17], [26], [30], [34] und [48] (siehe Literatur), der Diplomarbeit [45] sowie in den Kapiteln 12 und 13 der in Vorbereitung befindlichen Monographie [29] veröffentlicht und in 10 Vorträgen auf folgenden nationalen und internationalen Konferenzen vorgestellt:

- IMA-Workshop Computing with uncertainty, Minneapolis (USA), 18.-22.10.2010.  
W. Römisch: Generating and handling scenarios in stochastic programming.
- Workshop Kapazitätsberechnung, Bundesnetzagentur, Bonn, 8.11.2010.  
W. Römisch: Stochastische Eigenschaften des Abnehmerverhaltens.
- 9. VDI-Fachtagung Optimierung in der Energiewirtschaft, Nürtingen, 22./23.11.2011.  
W. Römisch: Szenariogenerierung zur Modellierung der stochastischen Auspeiselasten in einem Gastransportnetz.

- The Tenth International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing (MCQMC 2012), February 13–17, 2012, Sydney, Australia.  
H. Leövey: *Fast evaluation of mixed derivatives and calculation of optimal weights for integration.*  
W. Römisch: *QMC methods for stochastic programs: ANOVA decomposition of integrands.*
- 9th International Conference on Computational Management Science (CMS2012), April 18–20, 2012, London, Imperial College.  
W. Römisch: *Towards Quasi-Monte Carlo scenario generation in stochastic programming.*
- 3rd Conference on Optimization Methods and Software, May 13–17 2012, Chania, Crete, Greece.  
H. Leövey: *Automatic evaluation and bounding of cross-derivatives: Optimal weights for high dimensional integration.*
- 17th International Conference on Mathematical Methods in Economy and Industry, 24. 28. Juni 2012, Berlin-Schmöckwitz.  
H. Leövey: *Efficient computation of optimal weights for lattice rules in high dimensional integration.*
- International Symposium on Mathematical Programming, August 19–24, 2012, Berlin.  
W. Römisch: *Are Quasi-Monte Carlo methods efficient for two-stage stochastic programs?*
- Workshop 2 on Optimization Under Uncertainty, IMS, December 10–14, 2012, National University of Singapore, Singapore.  
H. Leövey: *Efficient computation of weights for lattice rules in high dimensional integration.*

## 4 Zusammenarbeit mit dem Industriepartner

Die Zusammenarbeit mit dem Industriepartner E.ON GT/OGE war im gesamten Projektzeitraum vorbildlich. In einer Reihe von Treffen und den Workshops alle 3 Monate erhielten wir alle notwendigen Informationen und viele Hinweise zur erfolgreichen Bearbeitung des Projektes.

## 5 Zusammenarbeit mit wissenschaftlichen Stellen

Das Projektvorhaben ist Teil eines Verbundes mit dem ZIB Berlin und den Universitäten Braunschweig, Duisburg-Essen, Erlangen und Hannover. Bei verschiedensten Treffen mit den anderen Projektleitern erfolgte ein intensiver Austausch zu den inhaltlichen Zielen und zur Abstimmung der Arbeiten.

(Prof. Dr. W. Römisch, Teilprojektleiter)

# Literatur

- [1] P. Avella, A. Sassano, I. Vasil'ev: Computational study of large-scale  $p$ -median problems, *Mathematical Programming* 109 (2007), 89–114.
- [2] 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.
- [3] R. E. Caflisch, W. Morokoff and A. Owen: Valuation of mortgage backed securities using Brownian bridges to reduce effective dimension, *Journal of Computational Finance* 1 (1997), 27–46.
- [4] G. Cerbe: *Grundlagen der Gastechnik* (7. Auflage), Carl Hanser Verlag, München, 2008.
- [5] J. Dick and F. Pillichshammer: *Digital nets and sequences. Discrepancy theory and quasi-Monte Carlo integration*, Cambridge University Press, Cambridge, 2010. (ISBN: 978-0-521-19159-3)
- [6] J. Dick, F. Y. Kuo and I. H. Sloan: High-dimensional integration – the Quasi-Monte Carlo way, *Acta Numerica* 2014, Cambridge University Press, 2014 (to appear).
- [7] J. Dupačová, N. Gröwe-Kuska and W. Römisch: Scenario reduction in stochastic programming: An approach using probability metrics, *Math. Program.* 95 (2003), 493–511.
- [8] H. C. P. Eilers and B. D. Marx: Flexible smoothing with B-splines and penalties, *Statistical Science* 11 (1996), 89–121.
- [9] H. C. P. Eilers and B. D. Marx: Splines, Knots and Penalties, *Wiley Interdisciplinary Reviews: Computational Statistics* 2 (2010), 637–653.
- [10] 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), 80.1 (2012), 24–39.
- [11] B. Geiger und M. Hellwig: Entwicklung von Lastprofilen für die Gaswirtschaft, em Gewerbe, Handel und Dienstleistungen, München, 2002.
- [12] P. Glasserman: *Monte Carlo Methods in Financial Engineering*, Springer, New York, 2004.
- [13] M. Griebel, F. Y. Kuo and I. H. Sloan: The smoothing effect of the ANOVA decomposition, *Journal of Complexity* 26 (2010), 523–551.
- [14] M. Griebel, F. Y. Kuo and I. H. Sloan: The smoothing effect of integration in  $\mathbb{R}^d$  and the ANOVA decomposition, *Mathematics of Computation* (to appear).
- [15] G. Hämmerlin und K.-H. Hoffmann: *Numerische Mathematik*, Springer, Berlin, 1994.

- [16] H. Heitsch: Szenariobaumapproximation für stochastische Optimierungsprobleme in der Energiewirtschaft, in: *Optimierung in der Energiewirtschaft*, VDI-Bericht Nr. 2080, Düsseldorf 2009, 45–60.
- [17] H. Heitsch, H. Leövey, W. Römisch: Are Quasi-Monte Carlo algorithms efficient for two-stage stochastic programs? *Stochastic Programming E-Print Series* 5 (2012) ([www.speps.org](http://www.speps.org)) and submitted.
- [18] H. Heitsch and W. Römisch: Scenario reduction algorithms in stochastic programming, *Comp. Optim. Appl.* 24 (2003), 187–206.
- [19] H. Heitsch and W. Römisch: A note on scenario reduction for two-stage stochastic programs, *Operations Research Letters* 35 (2007), 731–738.
- [20] M. Hellwig: *Entwicklung und Anwendung parametrisierter Standard-Lastprofile*, Dissertation, Technische Universität München, Fakultät für Energietechnik und Informationstechnik, Institut für Energietechnik, 2003.
- [21] R. Henrion, C. Küchler and W. Römisch: Discrepancy distances and scenario reduction in two-stage stochastic mixed-integer programming, *Journal of Industrial and Management Optimization* 4 (2008), 363–384.
- [22] R. Henrion, C. Küchler and W. Römisch: Scenario reduction in stochastic programming with respect to discrepancy distances, *Comp. Optim. Appl.* 43 (2009), 67–93.
- [23] F. J. Hickernell: A generalized discrepancy and quadrature error bound, *Mathematics of Computation* 67 (1998), 299–322.
- [24] W. Hoeffding: A class of statistics with asymptotically normal distribution, *Annals of Mathematical Statistics* 19 (1948), 293–325.
- [25] K. Jansen, H. Leovey, A. Nube, A. Griewank and M. Müller-Preussker: A first look at quasi-Monte Carlo for lattice field theory problems, Proceedings of the Conference on Computational Physics (CCP2012). [arXiv:1211.4388].
- [26] K. Jansen, H. Leövey, A. Nube, A. Griewank and M. Müller-Preussker: Quasi-Monte Carlo methods for lattice systems: a first look, Submitted (2013), [arXiv:1302.6419].
- [27] S. Joe and F. Y. Kuo: Remark on Algorithm 659: Implementing Sobol’s quasi-random sequence generator, *ACM Transactions on Mathematical Software* 29 (2003), 49–57.
- [28] O. Kariv, L. Hakimi: An algorithmic approach to network location problems, *Operations Research* 37 (1979), 539–560.
- [29] Torsten Koch (Ed.): *From Simulation to Optimization: Evaluating Gas Network Capacities*, Monograph, to be published by Society of Industrial and Applied Mathematics (SIAM), 2013.

- [30] T. Koch, H. Leövey, R. Mirkov, W. Römisch, I. Wegner-Specht: Szenariogenerierung zur Modellierung der stochastischen Ausspeiselasten in einem Gastransportnetz, in *Optimierung in der Energiewirtschaft*, VDI-Berichte 2157, VDI-Verlag, Düsseldorf, 2011, 115–125.
- [31] F. Y. Kuo, Ch. Schwab and I. H. Sloan: Quasi-Monte Carlo methods for high-dimensional integration: the standard (weighted Hilbert space) setting and beyond, *ANZIAM Journal* 53 (2011), 1–37.
- [32] F. Y. Kuo, I. H. Sloan, G. W. Wasilkowski and H. Woźniakowski: On decomposition of multivariate functions, *Mathematics of Computation* 79 (2010), 953–966.
- [33] Jong Soo Lee and D. D. Cox: Robust smoothing: Smoothing parameter selection and applications to fluorescence spectroscopy, *Comput. Stat. Data Anal.* 54 (2010), 3131–3143.
- [34] H. Leövey, W. Römisch, A. Steinkamp, I. Wegner-Specht: Modellierung der Gasabnahme als Funktion der Temperatur: Optimierung der Temperaturgewichte, *gwf-Gas-Erdgas* 11 (2011), 778–785.
- [35] R. Liu and A. B. Owen: Estimating mean dimensionality of analysis of variance decompositions, *Journal of the American Statistical Association* 101 (2006), 712–721.
- [36] G. Marsaglia: Evaluating the normal distribution, *Journal of Statistical Software* 11 (2004), 1–10.
- [37] 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.
- [38] J. Matoušek: On the  $L_2$ -discrepancy for anchored boxes, *Journal of Complexity* 14 (1998), 527–556.
- [39] 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.
- [40] H. Niederreiter: *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia, 1992.
- [41] 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, 1995, 299–317.
- [42] A. B. Owen: The dimension distribution and quadrature test functions, *Statistica Sinica* 13 (2003), 1–17.

- [43] Jiming Peng and Yu Wei: Approximating k-means-type clustering via semidefinite programming, *SIAM Journal on Optimization* 18 (2007), 186–205.
- [44] S. T. Rachev: *Probability Metrics and the Stability of Stochastic Models*, Wiley, Chichester, 1991.
- [45] H. Reckziegel: Quasi-Monte Carlo Methoden mit Dimensionsreduktion für zweistufige stochastische Produktionsmodelle, Diplomarbeit, Institut für Mathematik, Humboldt-Universität Berlin, 2013.
- [46] M. G. C. Resende and R. F. Werneck: A hybrid heuristic for the p-median problem, *Journal of Heuristics* 10 (2004), 59–88.
- [47] W. Römisch: Stability of Stochastic Programming Problems, in: *Stochastic Programming* (A. Ruszczyński and A. Shapiro eds.), Handbooks in Operations Research and Management Science, Volume 10, Elsevier, Amsterdam 2003, 483–554.
- [48] W. Römisch: ANOVA decomposition of convex piecewise linear functions, in: *Proceedings of MCQMC12 (Sydney 2012)*, Springer (to appear).
- [49] K. Schittkowski: Solving nonlinear least squares problems by a general purpose SQP method, in *Trends in Mathematical Optimization* (K.-H. Hoffmann et al. eds.), Birkhäuser, Boston 1988, 295–309.
- [50] K. Schittkowski: EASY-FIT: A software system for data fitting in dynamic systems, *Structural and Multidisciplinary Optimization*, 23 (2002), 153–169.
- [51] K. Schittkowski: *Numerical Data Fitting in Dynamical Systems – A Practical Introduction with Applications and Software*, Kluwer, 2002.
- [52] K. Schittkowski: NLPLSQ: A Fortran implementation of an SQP-Gauss-Newton algorithm for least squares optimization - user’s guide, version 1.0, Report, Department of Computer Science, University of Bayreuth, 2007.
- [53] H. Schwetlick, W. Schellong, and V. Tiller: Gauss-Newton-like methods for nonlinear least squares with equality constraints - local convergence and applications to parameter estimation in implicit models, *Statistics* 16 (1985), 167–178.
- [54] H. Schwetlick and V. Tiller: Nonstandard scaling matrices for trust region Gauss-Newton methods, *SIAM J. Scient. Statist. Comput.* 10 (1989), 654–670.
- [55] G.A.F. Seber and C.J. Wild: *Nonlinear Regression*, Wiley Series in Probability and Statistics, 2003.
- [56] I. H. Sloan and H. Woźniakowski: When are Quasi Monte Carlo algorithms efficient for high-dimensional integration, *Journal of Complexity* 14 (1998), 1–33.
- [57] I. M. Sobol’: *Multidimensional Quadrature Formulas and Haar Functions*, Nauka, Moscow, 1969 (in Russian).

- [58] X. Wang and K.-T. Fang: The effective dimension and Quasi-Monte Carlo integration, *Journal of Complexity* 19 (2003), 101–124.
- [59] X. Wang and I. H. Sloan: Why are high-dimensional finance problems often of low effective dimension, *SIAM Journal Scientific Computing* 27 (2005), 159–183.
- [60] X. Wang and I. H. Sloan: Brownian bridge and principal component analysis: towards removing the curse of dimensionality, *IMA Journal of Numerical Analysis* 27 (2007), 631–654.
- [61] X. Wang and I. H. Sloan: Low discrepancy sequences in high dimensions: How well are their projections distributed ? *Journal of Computational and Applied Mathematics* 213 (2008), 366–386.
- [62] X. Wang and I. H. Sloan: Quasi-Monte Carlo methods in financial engineering: An equivalence principle and dimension reduction, *Operations Research* 59 (2011), 80–95.
- [63] E. J. Wegman and I. W. Wright: Splines in Statistics, *Journal of the American Statistical Association* Vol. 78, Number 382 (1983), 351–365.