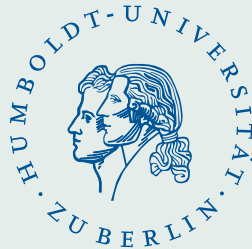# Energy Systems under Uncertainty: Modeling and Computations

W. Römisch

Humboldt-University Berlin
Department of Mathematics

`www.math.hu-berlin.de/~romisch`

Systems Analysis 2015, November 11–13, IIASA (Laxenburg, Austria)

Introduction

Energy systems often contain uncertain parameters, for example, demands, prices, inflows, wind speeds. Hence, an important modeling issue is data driven uncertainty quantification. In most cases, huge data sets are available. We discuss how starting from suitable statistical models sets of scenarios are generated that are representative for the uncertain parameters. The scenarios are then inserted into mathematical models describing the systems.

In this talk we consider the following two examples of energy systems:

- Electricity portfolio management under load-price uncertainty

- Evaluation of gas network capacities under demand uncertainty, validation of nominations and verification of booked capacities (supporting decisions of gas transmission system operators (TSOs) to sell capacity rights to customers)

---

nomination = vector defining the amounts of gas entering and leaving at each entry and exit, respectively
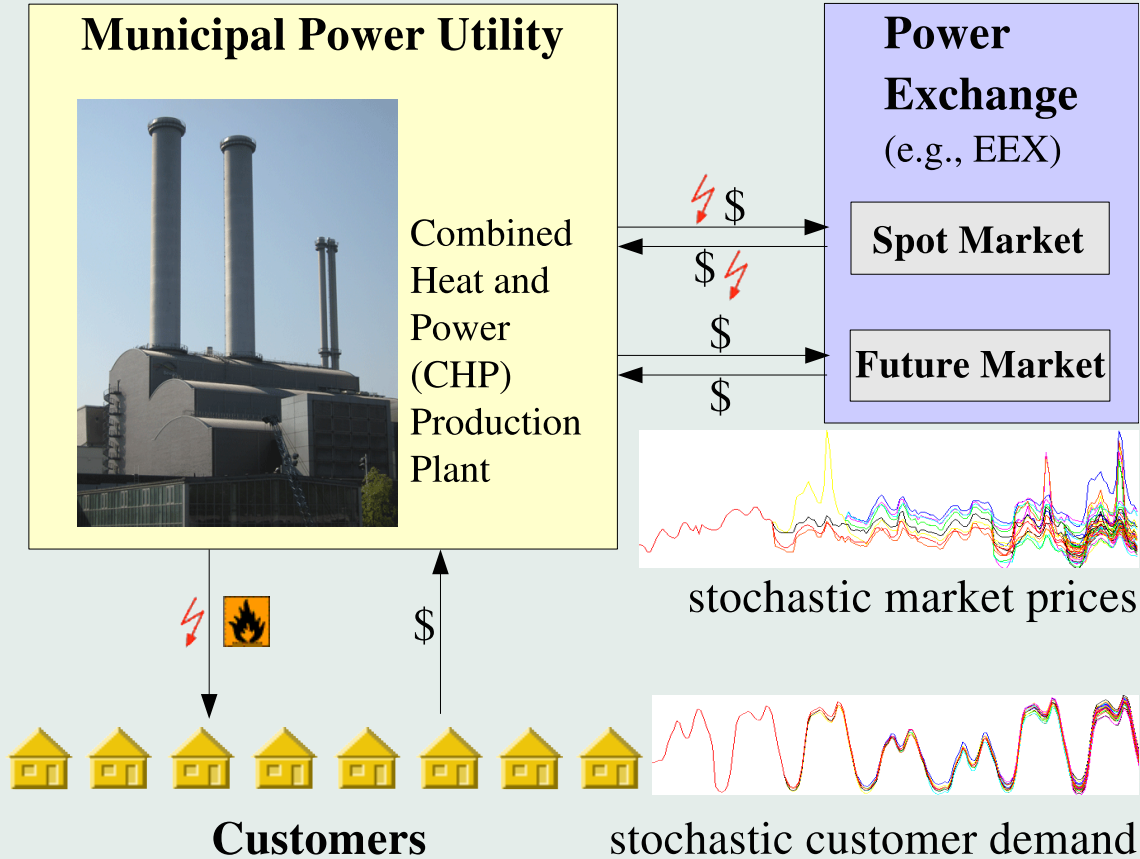**gas network capacity** = determined by the set of all "reasonable" nominations
validation of nominations = answering the question whether particular nominations can be transported
verification of booked capacities = how likely is that all reasonable nominations can be transported?

Mean-Risk Electricity Portfolio Management

(A. Eichhorn, I. Wegner-Specht)

Municipal Power Utility

Combined Heat and Power (CHP) Production Plant

Power Exchange
(e.g., EEX)

Spot Market

Future Market

$

$

$

$

stochastic market prices

Customers

$

stochastic customer demand

We consider the electricity portfolio management of a German municipal electric power company. Its portfolio consists of the following positions:

- Power production (based on company-owned thermal units),

- (yearly) bilateral contracts,

- (physical) (day-ahead) spot market trading (e.g., European Energy Exchange (EEX)) and

- (financial) trading of (monthly) electricity futures.

The time horizon is discretized into hourly intervals. The underlying stochasticity consists in a multivariate stochastic load and price process that will be represented approximately by a finite number of scenarios.

The objective is to **maximize the total expected revenue and to minimize the risk**. The portfolio management model is a large scale (mixed-integer) multi-stage stochastic optimization problem.
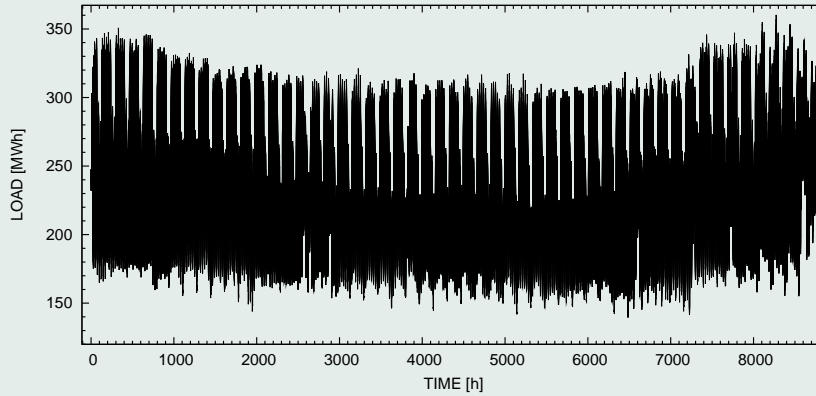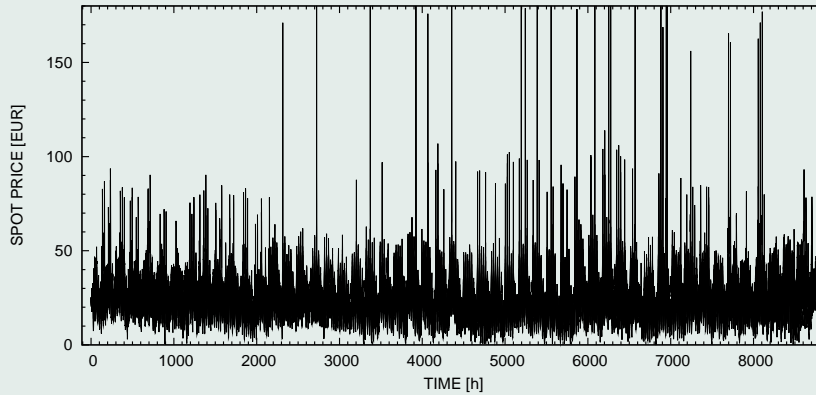
Figure 1: Time plot of load profile for one year

Figure 2: Time plot of spot price profile for one year

Statistical models and scenario trees

For the stochastic input data of the optimization model (here yearly electricity and heat demand, and electricity spot prices), a statistical model is employed.

- cluster classification for the intra-day (demand and price) profiles,

- Three-dimensional linear time series model for the daily average values (deterministic trend functions, a trivariate ARMA model for the (stationary) residual time series),

- Generation of scenarios by computing (Quasi-)Monte Carlo samples from the multivariate normal distribution that corresponds to the ARMA process, and adding on trend functions as well as matched intra-day profiles from the clusters afterwards,

- generation of scenario trees based on monthly recursive scenario reduction applied to a given set of scenarios (Heitsch-Römisch 09).
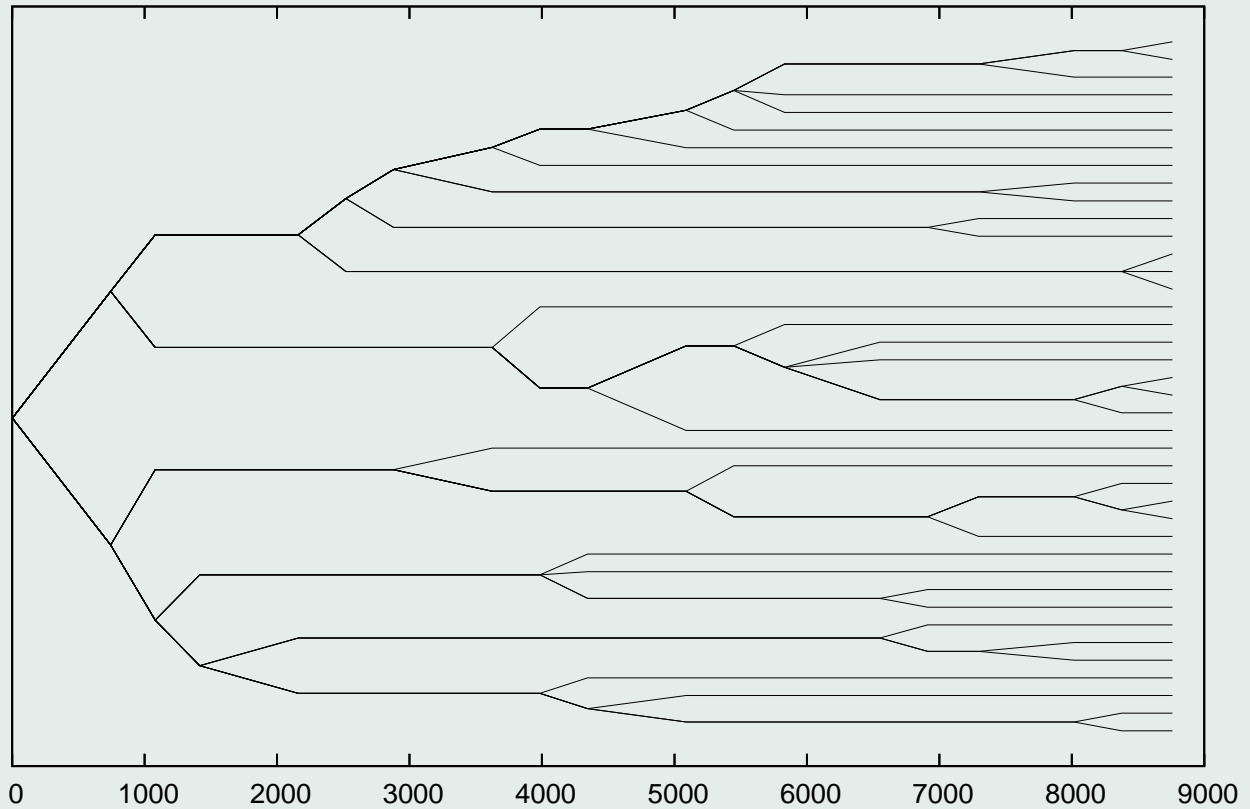
Yearly scenario tree for the trivariate load-price process with monthly branching

Numerical results

Test runs were performed on real-life data of a German municipal power company leading to a linear optimization problem containing $T = 365 \cdot 24 = 8760$ time steps, a scenario tree with 40 demand-price scenarios with about 150.000 nodes. The objective function is of the form

$$\text{Minimize} \quad \gamma \rho(z) - (1 - \gamma)\mathbb{E}(z_T)$$

with some (multiperiod) risk measure ρ with risk aversion parameter $\gamma \in [0, 1]$ ($\gamma = 0$ corresponds to the risk-neutral case).

Risk measure (of aggregation type):

$$\rho(z) = \text{AVaR}_\alpha \left(\min_{j=1,\ldots,J} z_{t_j} \right) = \inf_{r \in \mathbb{R}} \left\{ r + \frac{1}{\alpha}\mathbb{E}[\max\{0, -r - \min_{j=1,\ldots,J} z_{t_j}\}] \right\},$$

where $\{z_t : t = 1, \ldots, T\}$ is the stochastic revenue process and t_j, $j = 1, \ldots, J$, $J = 52$, are the risk measuring time steps. The latter correspond to 11 pm at the last trading day of each week.
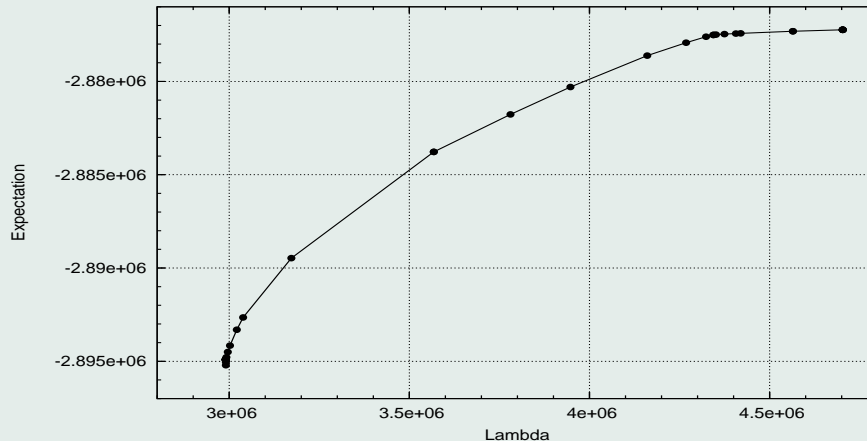
It turns out that the numerical results for the expected maximal revenue and minimal risk

$$\mathbb{E}(z_T^{*\gamma}) \qquad \text{and} \qquad \rho(z_{t_1}^{*\gamma}, \ldots, z_{t_J}^{*\gamma})$$

with the optimal revenue process $z^{*\gamma}$ are (almost) **identical** for $\gamma \in [0.15, 0.95]$ and the risk measure ρ. The efficient frontier

$$\gamma \mapsto \left(\rho(z_{t_1}^{*\gamma}, \ldots, z_{t_J}^{*\gamma}), \mathbb{E}(z_T^{*\gamma}) \right)$$
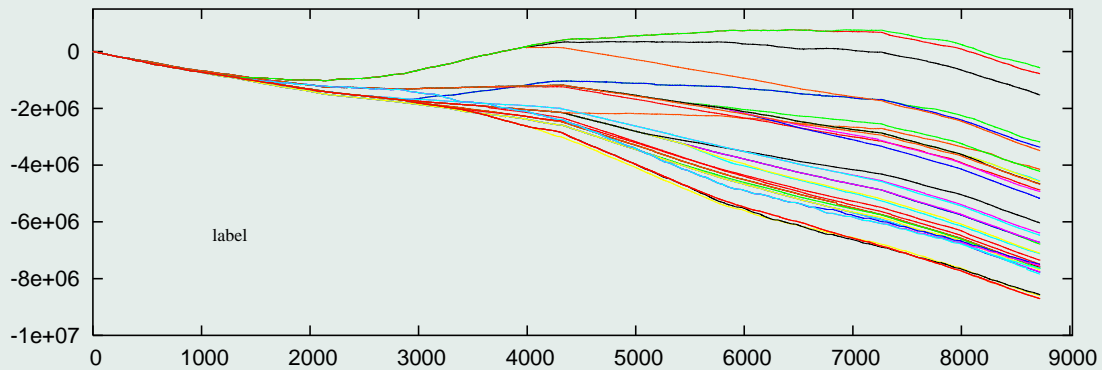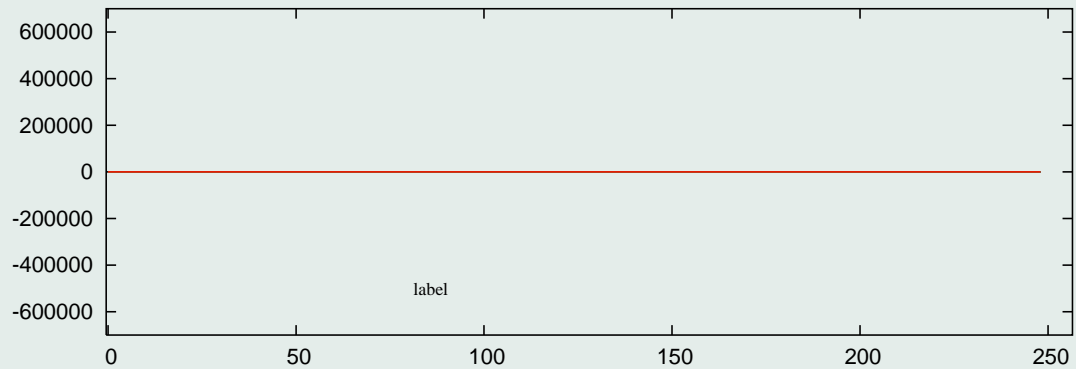
is concave for $\gamma \in [0, 1]$.

Efficient frontier

The LP is solved by CPLEX 9.1 in about 1 h running time on a 2 GHz Linux PC with 1 GB RAM.
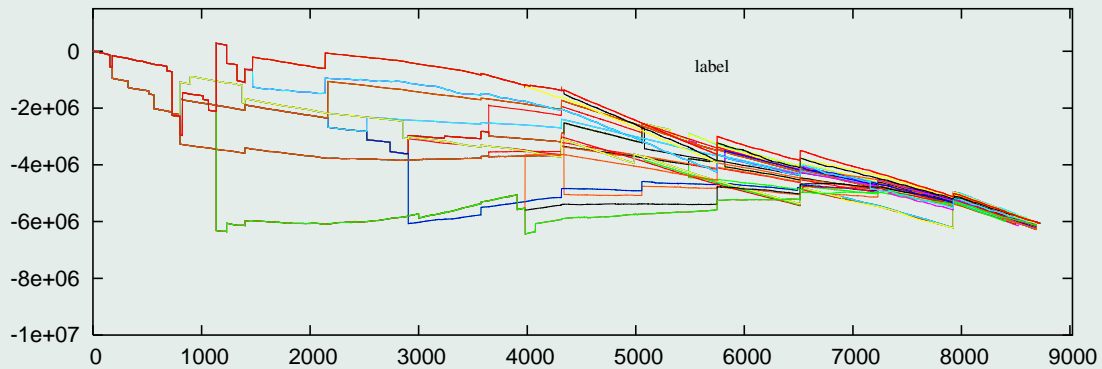
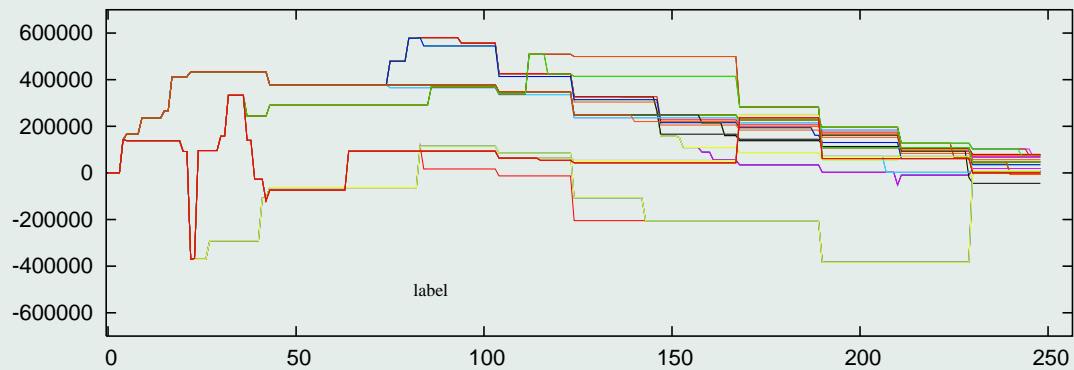Risk aversion costs less than 1% of the expected overall revenue.

Overall revenue scenarios for $\gamma = 0$

Future trading for $\gamma = 0$

Overall revenue scenarios for $\gamma = 0.9$

Future trading for $\gamma = 0.9$

Gas network capacities and validation of nominations

(H. Heitsch, H. Leövey, R. Mirkov, I. Wegner-Specht)

We consider the gas transport network of the company Open Grid Europe GmbH (OGE). It is Germany's largest gas transport company. Such networks consist of intermeshed pipelines which are actuated and safeguarded by active elements (like valves and compressor machines). Here, we consider the stationary state of the network and the isothermal case.

Two different gas qualities are considered: H-gas and L-gas (high and low calorific gas). Both are transported by different networks.

The gas dynamics in a pipe is modeled by the Euler equations, a nonlinear system of hyperbolic partial differential equations. In the stationary and isothermal situation they boil down to nonlinear relations between pressure and flow. Together with models for the active elements, this leads to large systems of nonlinear mixed-integer equations and inequalities.

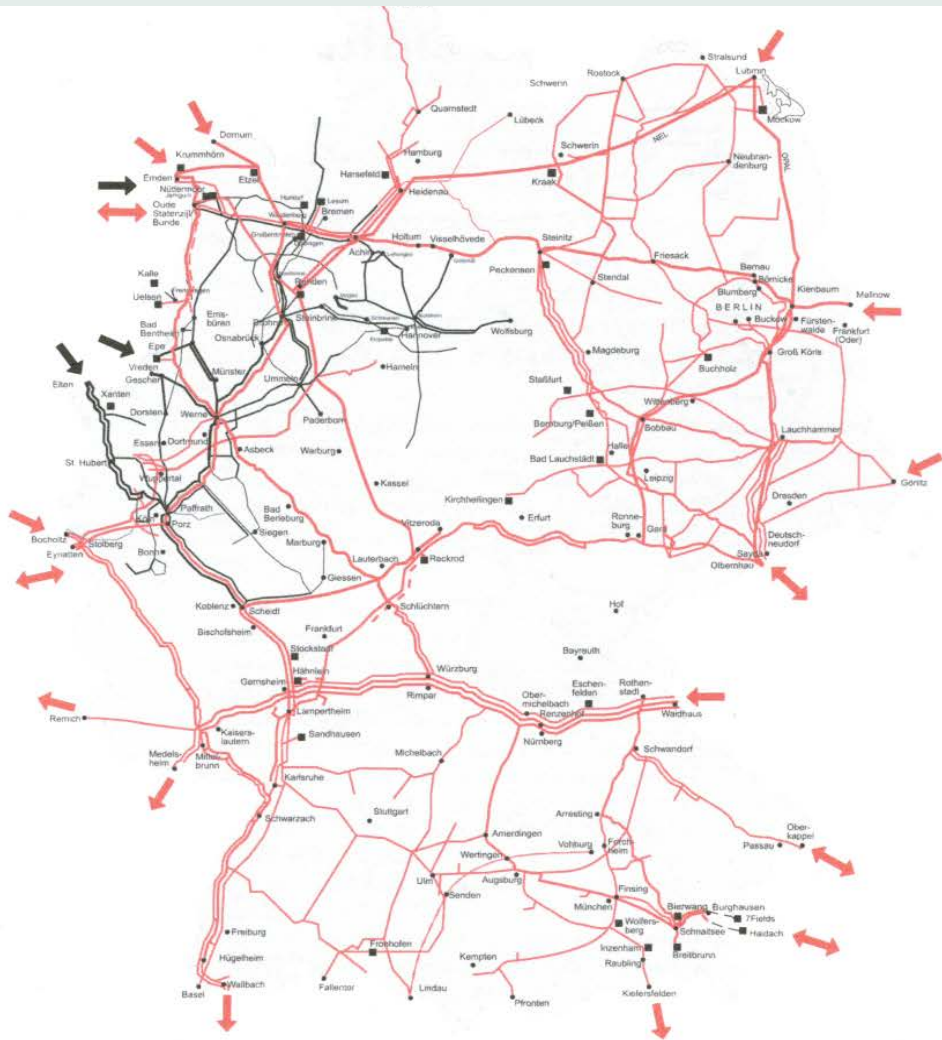**Aim:** Evaluating the capacity of a gas network, validating nominations and verifying booked capacities.

Figure 1.4. *German H-gas (red) and L-gas (black) network systems. The arrows indicate entry and exit nodes. Gas storages are represented by black squares. (Source: OGE.)*
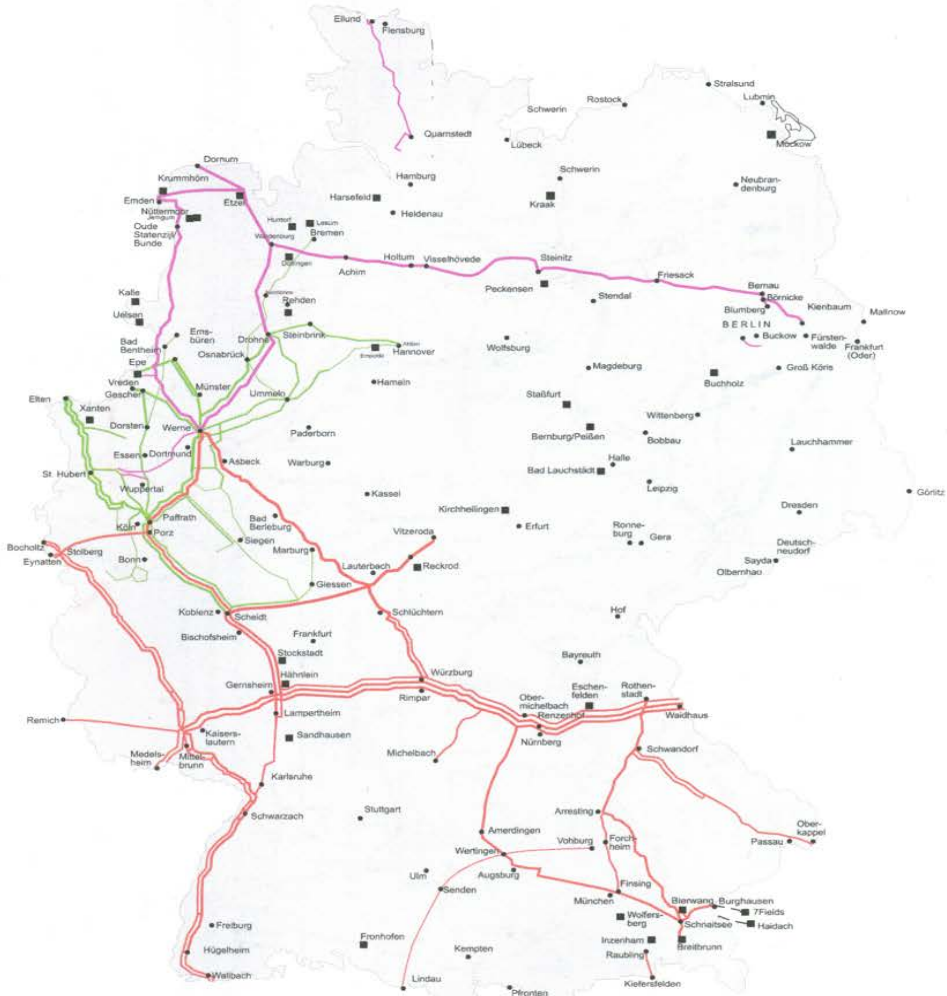
Figure 1.5. *Network systems of OGE. The network model is split into three parts: H-gas north (purple), H-gas south (red), and L-gas (green). (Source: OGE.)*

Statistical data and data analysis

Hourly gas flow data is available at all exit nodes of a given network for a period of eight years. Due to stationary modeling we consider the daily mean gas flow at all exit nodes. Since it depends on the daily mean temperature, we consider a daily reference temperature based on a weighted average temperature taken at different network nodes.

Due to stationary and isothermal modeling we introduce the temperature classes (-15,-4], (-4,-2], (-2,0],..., (18,20], (20,30) and perform a corresponding filtering of all daily mean gas flows at all exit nodes according to the daily reference temperatures. We also check that a reasonable amount of daily mean gas flow data is available for all temperature classes except for (-15,-4]. Another filtering is carried out for day classes (working day, weekend, holiday).
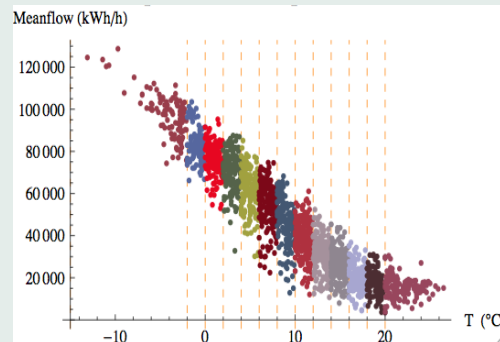
Examples of daily main gas flow at exit nodes as function of the temperature
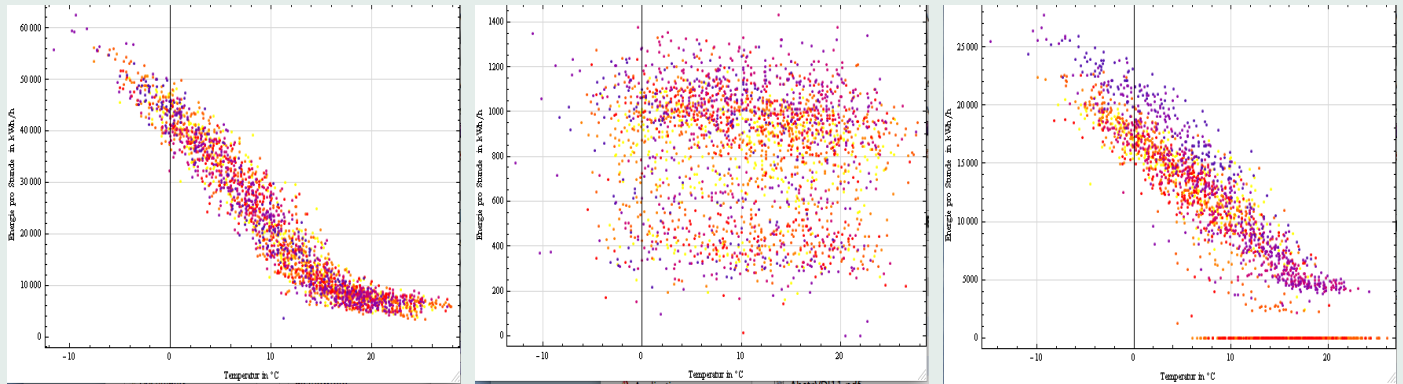
Daily mean gas flow data at exit nodes with municipal power stations, with zero flow (right).
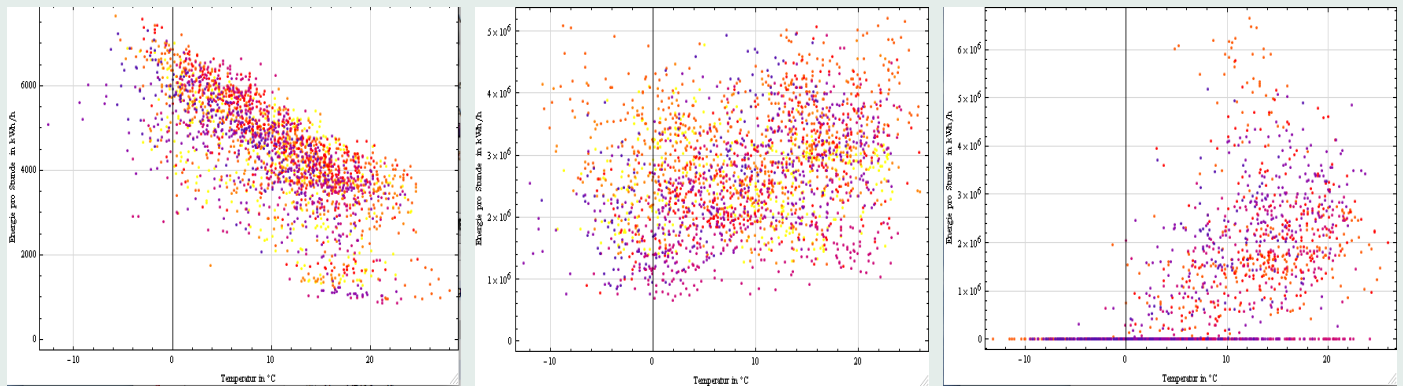
Daily mean gas flow data at exit nodes with company (left), market transition (middle), storage (right).

Univariate distribution fitting

Classes of univariate probability distributions:

- (shifted) uniform distributions

- (shifted) (log)normal distributions

- Zero gas flow appears with empirical probability p at several exit nodes. Hence, we consider the shifted probability distribution function

$$F(x) = p\,F^0(x) + (1 - p)\,F^+(x)$$
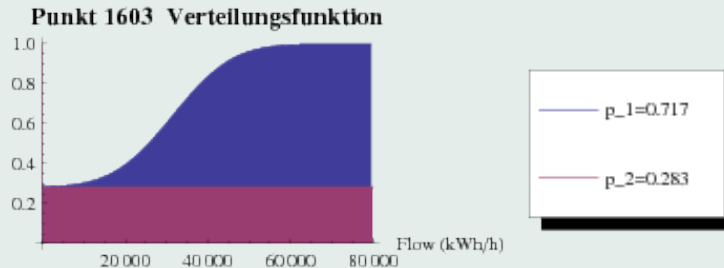
Probability distribution function of a shifted normal distribution at exit 1603

Fitting multivariate normal distributions

- **Multivariate normal distributions** are fitted for exit gas flows that satisfy normality tests and have significant correlations with other exit nodes, i.e., in addition to means and variances, correlations are estimated by standard estimators if sufficient data is available.

- Examples of correlation matrices:
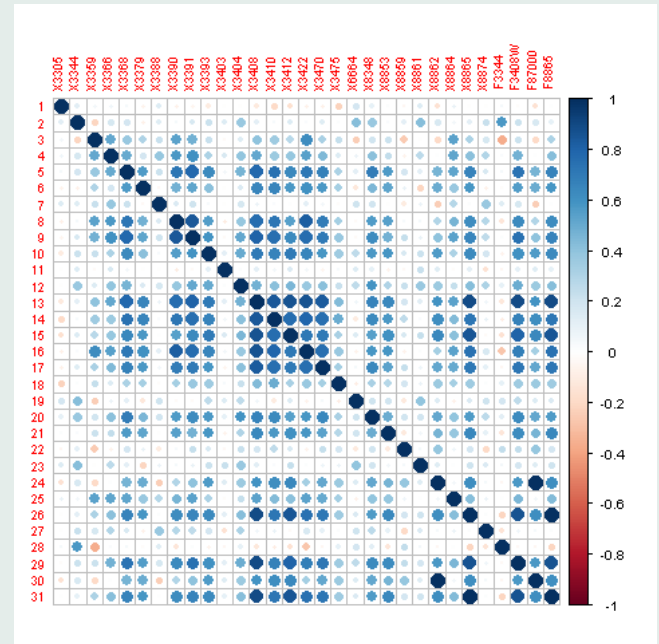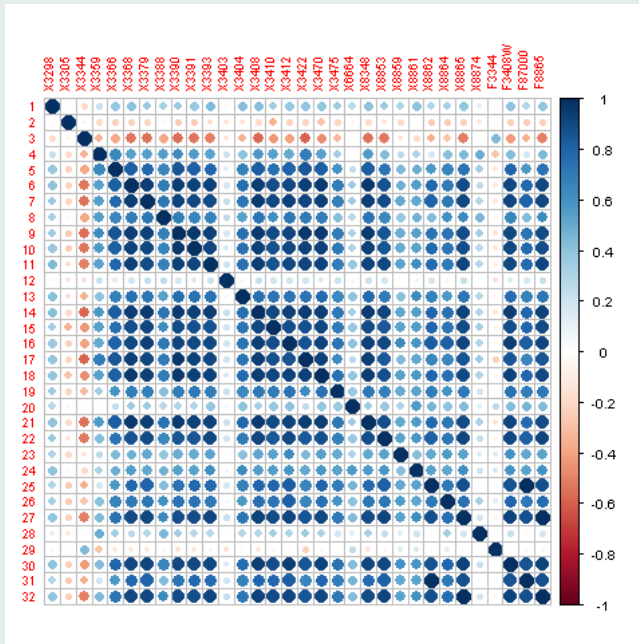
Correlation plots for the temperature classes $(10, 12]$ and $(18, 20]$ in certain areas of the H-gas network.

Forecasting gas flow demand for low temperatures

- Low temperatures require a specific treatment due to lack of data.

- **Idea:** Penalized spline (P-spline) regression with shape constraints.

- **Ansatz:** Least squares regression for all standardized daily maximal gas flows y_i for the temperature t_i at some exit node.

$$\min_{a \in \mathbb{R}^m} \left\{ \sum_{i=1}^{n} (y_i - S(t_i))^2 + \lambda \sum_{j=3}^{m} (\delta^2 a_j)^2 + \kappa \sum_{j=2}^{m} b_j (\delta^1 a_j)^2 \right\},$$
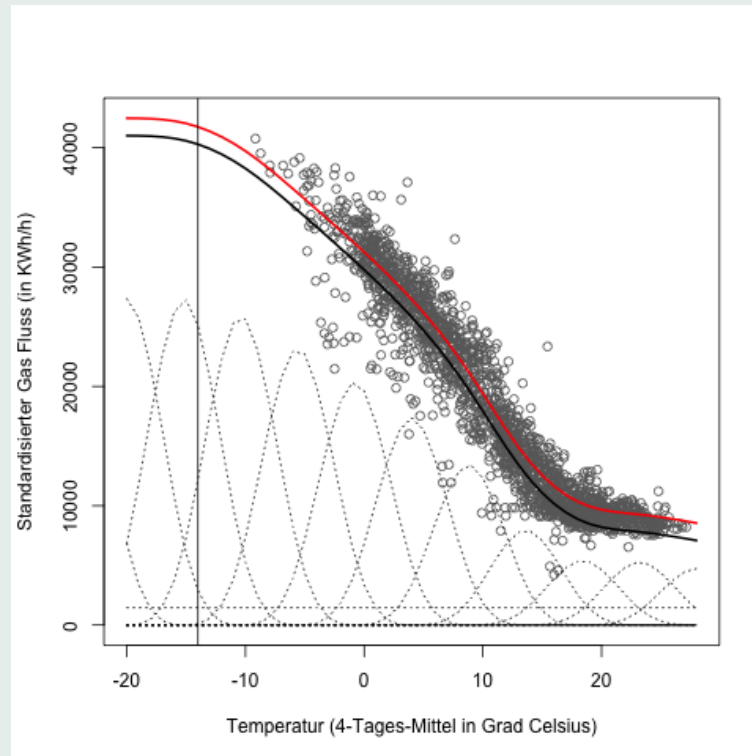
where λ and κ are positive smoothing and shape parameters and

$$S(t) = \sum_{j=1}^{m} a_j B_j(t)$$

is a cubic spline in B-spline basis representation.
Here, it holds $\delta^1 a_j = a_j - a_{j-1}$ und $\delta^2 a_j = a_j - 2a_{j-1} + a_{j-2}$.

- Using the regression model to fit the mean of a univariate normal distribution at $t = -14°C$ and using the variance taken from the temperature class $(-4°C, -2°C]$.

- Example parameters: $\lambda = 2.51$, $\kappa = 100$ und $b_j = 1$.

P-spline regression with flattening asymptotes and comparison with sigmoid regression (red)

Scenario generation

Using randomized Quasi-Monte Carlo methods we determine N samples with probability $\frac{1}{N}$ for the d-dimensional random vector ξ that corresponds to the random gas flows at the d exits of a given network. We proceed as follows:

- We determine N samples η^j of the uniform distribution on $[0,1)^d$ using Sobol' points and perform a componentwise random scrambling of their binary digits using the Mersenne Twister. The scenarios η^j, $j = 1, \ldots, N$, combine favorable properties of both Monte Carlo and Quasi-Monte Carlo methods.

- Determine samples in $\mathbb{R}^d$ by

$$\zeta_i^j = \Phi_i^{-1}(\eta_i^j) \quad (i = 1, \ldots, d; j = 1, \ldots, N)$$

using the univariate distribution function Φ_i of the ith component.

- If a part of the components of ξ has a d-dimensional multivariate normal distribution with mean $m \in \mathbb{R}^s$ and $s \times s$ covariance matrix Σ, we perform a decomposition $\Sigma = A\,A^\top$, where the matrix A preferably corresponds to principal component analysis. Then the s-dimensional vectors

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

are suitable scenarios for this part of the random vector ξ.
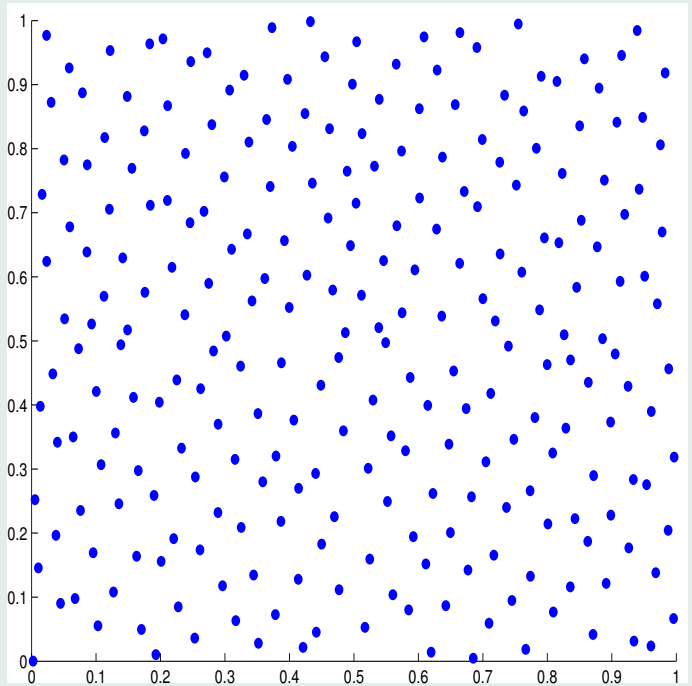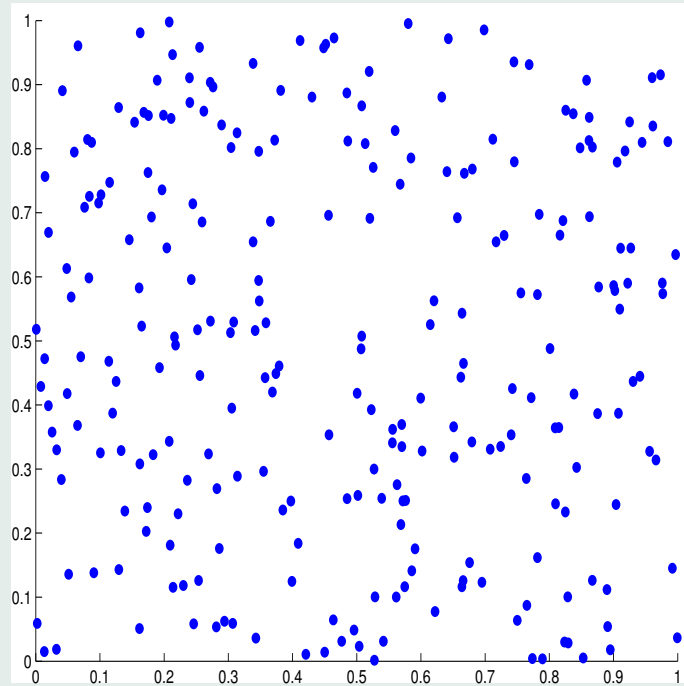
Comparison of $n = 2^7$ Monte Carlo Mersenne Twister points and randomly binary shifted Sobol' points in dimension $d = 500$, projection (8,9)

Optimal scenario reduction

- Given is a large number N of d-dimensional scenarios ξ^i, $i = 1, \ldots, N$, with probabilities $\frac{1}{N}$ and a norm $\|\cdot\|$ on $\mathbb{R}^d$.

- The aim is to select $n \leq N$ scenarios and to determine their new probabilities such that the discrete probability distribution based on the original scenario set is approximated best possible in terms of the so-called Kantorovich distance of probability distributions (optimal scenario reduction).

- First, this requires the solution of the combinatorial optimization problem

$$\min \left\{ \sum_{j \in J} \min_{i \notin J} \|\xi^i - \xi^j\| : J \subset \{1, ..., N\}, \#J = N - n \right\}$$

(called n-median problem) to determine the index set J of deleted scenarios. This problem is $\mathcal{NP}$-hard and may be reformulated as a mixed-integer linear optimization problem. Simple greedy heuristics work well in many cases.

- Secondly, the new probabilities p_i, $i \notin J$, of the selected scenarios are $p_i = \frac{1 + \#J_i}{N}$, where $J_i = \{j \in J : i = i(j)\}$ and $i(j) \in \arg\min_{i \notin J} \|\xi^i - \xi^j\|$, i.e., the probability of any deleted scenario is added to the probability of one of the closest selected scenarios (optimal redistribution).

Illustration:

10^3 samples based on randomized Quasi-Monte Carlo methods are generated and later reduced by scenario reduction to 50 scenarios. The result is shown below where the diameters of the red balls are proportional to the new probabilities.
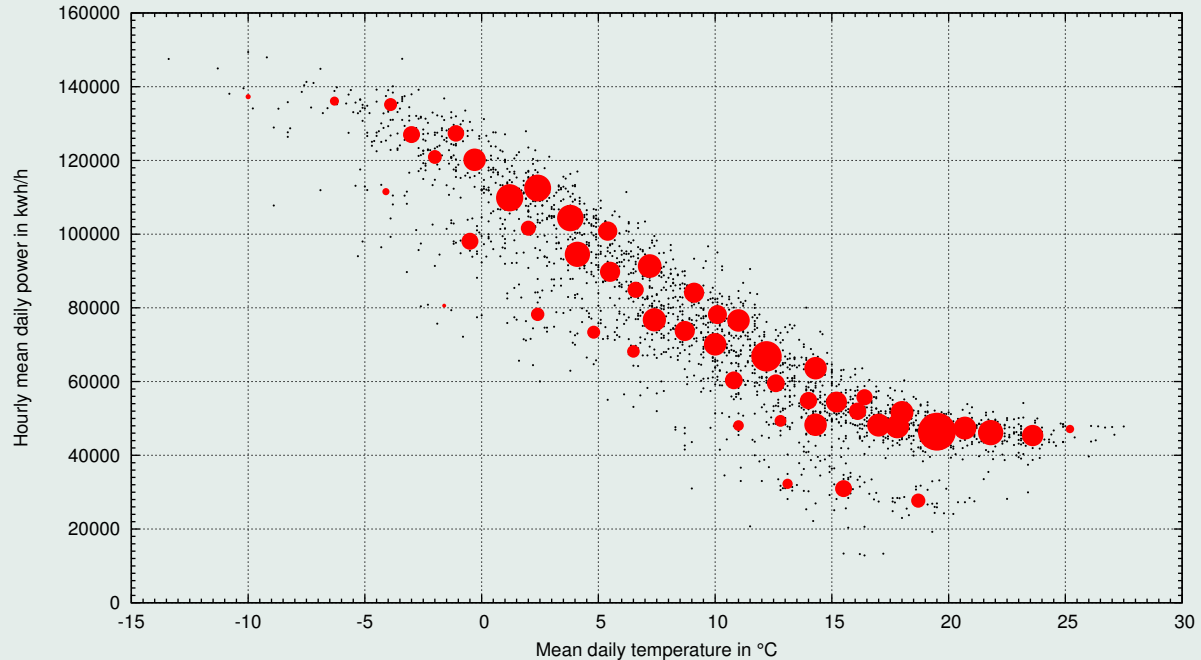
Conclusions

- Filtering exit gas flow data into temperature classes and fitting univariate (shifted) uniform or (log)normal distributions, respectively, and of multivariate (log)normal distributions for groups of exits.

- A univariate normal distribution is assumed for low temperatures. Its mean is fitted by a P-spline regression and its standard deviation is taken from the neighboring temperature class.

- Randomized Quasi-Monte Carlo methods are used to generate a large number of gas flow scenarios at all network exits. Using optimal scenario reduction the large scenario set is represented best possible by a reasonable number of scenarios for computational feasibility studies.

Literature: T. Koch, B. Hiller, M. E. Pfetsch and L. Schewe (Eds.): *Evaluating Gas Network Capacities*, MOS-SIAM Series on Optimization, Philadelphia, 2015.