University of Twente | Gasunie Transport Services

A stochastic approach to allocation of capacity in a gas transmission network

Student:
D.I. van Huizen

Supervisors:
Prof. dr. R.J. Boucherie
Drs. H. Dijkhuis
Dr. J.J. Steringa

June 4, 2015

Abstract

A new model for the allocation of capacity in the gas transmission network of Gasunie Transport Services is proposed. The model includes a stochastic approach to predict the supply and demand on different network points more realistically. It also includes the effects of the balancing regime on the supply and demand and a simple way to determine the flows through the network. The current method of GTS mostly uses maximum or minimum contract utilization by the shippers. The model may be useful to determine oversell capacity based on predictions, which is a step forward because the oversell capacity of GTS is now mainly based on realizations.

Contents

1 Introduction

This graduation report was written at the Network Configuration Department of Gasunie Transport Services. GTS is a subsidiary of N.V. Nederlandse Gasunie and is the Transmission System Operator in the Netherlands. GTS provides for the gas transport in the Netherlands; they do not own any gas or gas field. Since a couple of years European regulations state that parties that want to transport gas via the network of GTS have to independently contract rights to inject and extract gas from the network. This is called decoupled entry and exit and leads to numerous new planning issues.

Shippers are the owners of gas and gas fields. When they want to transport gas in the Netherlands they can only use the network of Gasunie Transport Services. GTS offers contracts for network capacity at every network point to shippers. A shipper sends in its transport request every hour. A transport request is an amount of entry or exit for every point on the network at every hour. A shipper can only request to inject or extract an amount of gas at a network point that is lower than the capacity it contracted. The gas that is injected in the network is transported via the pipelines. For GTS no possibility exists to influence the transport tasks, other than handing out contracts. All transport requests that are submitted have to be fulfilled by GTS. GTS transports energy, not specific molecules. A shipper that injects gas into the network does not get its own molecules back necessarily. By injecting gas in the network the shippers obtains the right to extract this same amount somewhere else in the network. Although it costs time for the gas to flow through the network, the location of injecting gas does not influence the location where it can be extracted.

The network of GTS consists out of two separate networks, one for H-gas and one for G-gas. H-gas is gas with a higher quality, it contains more energy per cubic meter than G-gas. Shippers can inject G-gas in the G-gas network and extract H-gas out of the H-gas network, as long as the energy they inject and extract is approximately in balance. GTS has the legal task to keep the network operating. When the shippers in total are too much out of balance (all the shippers injected a lot more than they extracted, or vice versa) GTS does a balancing action. In a balancing action GTS buys gas to inject into the network or sells gas to extract from the network to keep the network operating. Shippers that are responsible for the imbalance in the network are penalized for this. Therefore shippers must react and anticipate on the balance position of the network in their transport requests.

Besides network points and pipelines the gas transmission network contains a number of compressor stations. Through these compressor stations GTS can regulate the pressure in the network. This is highly important for multiple reasons. Gas flows from high to low pressure. When the pressure is too high on some points in the network this can be unsafe. Furthermore could it become impossible for shippers to inject gas into the network. A pressure that is too low can give operational problems. When a shipper receives gas at a network point with a too low pressure, it could become impossible to transport it further into other networks. Besides the amount they can enter or exit from the network, the contracts that GTS offers to shippers on network points contain a pressure range within the shipper has the right to inject or extract gas.

The planning department of GTS is used to deal with many uncertainties to calculate entry and exit capacities, but uncertainties in supply and demand so far have not always been fully taken into account. To find severe transport situations one usually deals with maximum or minimum contract utilization of all the shippers, while it is rare that all shippers request what they maximally can at the same time. A first approach to simulate the network using this type of uncertainty in supply and demand is given in this graduation project.

2 Problem description and literature study

First the problem is described, followed by a brief description of all the chapters in this report. In the second section the literature study is presented.

2.1 Problem description

Shippers that wish to transport gas from A to B have to independently reserve entry capacity at A and exit capacity at B. By doing these reservations, shippers have a tendency to take an operational margin. In general shippers do not always use the total capacity that they reserve. At the planning department of GTS this is not taken fully taken into account. Most of the time one deals with maximum or minimum contract utilization. This can lead to unused capacity in the network that is not sold.

In "Balans tussen gevraagde en beschikbare transportcapaciteit" [10], a memorandum of GTS, the question arose what the flows through the network would be if the entry and exit behavior of the shippers followed a certain pattern. There is a deeper question behind this question, and this deeper question is also the main research question of this report: "Can a simulation of the gas transmission network be made with the uncertainty in supply and demand of gas taken fully into account instead of just minimum or maximum contract utilization?"

The main research question breaks down into three subquestions:

- What are the flows through the network given a transport task?
- What is the distribution of supply and demand of gas at all the network points?
- What is the influence of the balancing regime on the supply and demand of gas at all the network points?

If the answers to all the three subquestions are known, a more realistic simulation of the gas transmission network will be possible. First, in chapter 3 it is explained in more detail how the Dutch gas transmission network operates. In chapter 4 a Newton model is used to determine the flows through the network, when entries and exits on all the network points are known for a certain time slot.

The pattern of entry and exit of gas for an individual shipper consists of two parts. The first part is that a shipper must be approximately in balance. Shippers can independently enter gas into and exit gas from the network; however, the difference between the amount they enter and exit must not be too large. GTS forces the shippers to be "approximately" in balance. How shippers react to the balancing regime of GTS is described in chapter 5.

The second part is distribution of entry and exit at every network point of every shipper. The network points are categorized by the markets they serve, because these different markets lead to a different utilization of the reserved capacities (chapter 6 to 8). A model that includes the uncertainty has been made to predict the supply of gas in the (near) future for some of the network points. Due to time restrictions not all the types of network points could be analyzed. The choice was made to analyze the network points in the domestic market and the border exit points. The reason to choose for these points is that their capacities may be used for overselling.

Knowing the behavior of the shippers at the network points, the potentially unused capacity in the network can be determined. This capacity could be offered as oversell capacity (chapter 9). Currently GTS bases its oversell capacity mainly on realizations. In this way GTS does not have detailed insight in the exact risk that would be taken when all the oversell capacity would be sold. That is why we only looked at the network points which can be used for overselling. The border network points have the highest need for oversell capacity. Therefore we looked at the interior domestic market network points, from which capacity can be shifted to a border exit point, and the border network points itself. By predicting the entry and/or exit at a network point, an upper bound for the risk taken can be found.

To make a more realistic simulation of the complete network also the behavior (of the shippers) at all the other network points must be known. The analysis of other network points has been left for further research.

2.2 Literature study

The decoupling of entry and exit capacity all over the European Union leads to new planning issues. Before the decoupling, there already was need to determine the flows through the network, especially for networks that contain a ring of pipelines. In [14] the flows through a network have been determined with a maximum flow algorithm. In [6] a gas transmission network has been optimized with a linear program. Finding the flows through the network was part of the optimization. In the literature, optimizing a gas transmission network often reduces to finding the ideal settings of the compressor stations and valves in the network to complete a transport task. In [1] was suggested that the flow through the network can be determined with the use of a Newton method.

The balancing regime of GTS is unique in the world [3]. In [3] the effect of the balancing system on the shippers is described when GTS performs a balancing action. The fact that shippers can anticipate on the imbalance position of the network is neglected. To our best knowledge, no specific research was done on this topic before. Even at GTS itself the behavior of individual shippers based on the balancing regime was never modeled.

A lot more research has been done on the demand of gas. In [8] the relation between demand and the outside temperature is researched. In [7] the yearly demand of gas in Turkey is predicted with an autoregressive model. In [5] more information on autoregressive models was found. The demand of electricity, which depends on the temperature, was predicted with an autoregressive model in [2].

3 The gas transmission network

The Dutch gas transmission network has length of over 15000 km and transports approximately 125 billion m^3 natural gas every year. There are about 50 entry points and 1100 exit points. Approximately 100 shippers use the gas transmission network. In this chapter the most important information on the operation of the network is given.

3.1 Types of gas

The network consists of two separate networks, one for H-gas and one for G-gas. H-gas is high caloric gas and has a higher quality than G-gas. A higher quality means that it needs more oxygen per cubic meter gas to incinerate. This H-gas is produced in e.g. Norway, Russia and the Middle East. H-gas enters the country in the northeast by pipelines or is shipped in via the port of Rotterdam. Liquid gas can enter the network there via the LNG terminal. H-gas is mainly used in the industry. A significant part of all the H-gas that flows into the Netherlands is transported to the United Kingdom or the rest of western Europe. G-gas is produced in Groningen and in the North Sea. This gas is used in the domestic market and the smaller industries, mainly in the Netherlands and Belgium. A mixture of $\frac{1}{3}^{rd}$ H- and $\frac{2}{3}^{rd}$ G-gas is L-gas. L-gas has a quality in between H- and G-gas and is transported abroad to be used in the domestic market. There is no L-gas network in the Netherlands. The L-gas that is transported is mixed at the border.

3.2 Types of network points

In the total network of GTS in the Netherlands are close to 1200 network points. These network points can be classified in different types. First there is a difference between the G-gas and the H-gas network points. The supply and demand for points connected to the G-gas network usually follows a seasonal pattern. In the colder months of the year the allocations at these points are higher than in the warmer months, because G-gas is mainly used for heating.

The supply of this H-gas usually does not follow a seasonal pattern, because it is not used for heating. This gas has, besides the interior exit points, exits on the East, South and Northwest border. Points in the network that can be entries or exits are storages. Storages are empty gas fields or salt caverns that can be used to store gas. Usually the empty gas fields are filled in the summer and emptied in the winter. The salt caverns usually are filled during the night and emptied in the day time.

Figure 3.1: The gas transmission network in the Netherlands. In yellow the H-gas network, in black the G-gas.

3.3 Bookings, nominations and allocations

Shippers send transport requests to GTS. Before they can send a transport request, they have to reserve capacity for a time slot. This is called a booking. There exist year, quarter, month and day bookings. Network capacity for (small) industries and households can be requested at

GTS. Bookings on gas storages and border network points can be done at auctions. For all these network points exists a different auction. At these auctions shippers can book a constant amount of capacity per hour for the total contract duration. Once a shipper has a booking, it can do a nomination. This nomination is a transport request. In a nomination a shipper tells GTS a few hours in advance exactly the amount of gas it wants to inject or extract at a network point during a certain hour. Of course this nomination can not be larger than the booking. Due to operational circumstances, the exact amount of gas that is entered or exited can vary a little from the nomination. Therefore, afterwards, a shipper gets allocated the exact amount of gas that has flown through a network point and belongs to his portfolio. Usually the allocation is equal to the nomination.

Capacity is sold firm by GTS. A shipper that reserves capacity has to be able to use all of this capacity. It can not be (partly) turned off. GTS has the legal obligation to offer the whole technical capacity of a network point for sale. When this capacity is sold out at a network point GTS can choose to offer interruptible capacity. The security of supply of this capacity varies between 85 and 95%.

GTS has the legal obligation to deliver gas to domestic market with a security of supply of 100% when the outside daily average temperature is higher than minus 9 degrees Celsius. Although normally the shippers do the bookings, on domestic market network points it is GTS that does this. The price of these bookings is charged to the shippers, based on a mixture of the size of the market they serve and historical user data.

3.4 Balance in the network

The gas transmission network always contains gas. When the network contains the optimal amount of gas, the network is in balance. The position of the network, which is the difference between the actual amount of gas in the network and the optimal amount, should stay between set limits to keep the network operating. GTS makes the shippers responsible for this. Every hour, at 15 minutes past the whole hour, GTS predicts what the position of the network will be at the end of the hour, based on the nominations the shippers did. Whenever it is predicted that the network position will be out the position limits, GTS buys or sells the difference between the predicted position and the position limit at the gas spot market. The shippers that caused this imbalance are charged for this transaction.

4 Flows through the network

In the memorandum "Balans tussen gevraagde en beschikbare transport capaciteit" [10] was need for a model to determine the flow through a gas transmission network, that contains a ring of pipelines, when the entry and exists amounts on all the network points are known. A model, that includes a Newton model, is developed to do this. After the flow is determined, the pressure on every point in the network can be obtained also. The active elements in the network, such as compressor stations, are ignored.

Gas flows due to a difference in pressure in the network. Shippers have to inject or extract gas wit a certain pressure, based on the location of the network point. GTS has to maintain this pressure in the network. The pressure drop in a pipeline p depends on numerous factors. The main factors are the input pressure ($P_{p,in}$), the length (L_p) of the pipeline, its diameter (D_p) and the transported volume (Q_p). All the other factors, such as pipe coarseness and temperature, are included in a constant factor c_p. The relation between input and output pressure ($P_{p,out}$) is:

$$P_{p,in}^2 - P_{p,out}^2 = \frac{c_p \cdot L_p \cdot Q_p^2}{D_p^5}. \quad (4.0.1)$$

P is in bar, L and D are in meters and Q is m^3/h. Pipelines that together form a cycle in the network are called a ring. When there are no rings in the network, the flow through the network is defined by the fact that all the gas that enters a node, also has got to leave it. When there are one or more rings in the network, an additional requirement to define the flow is that there can be no pressure drop over a ring. There can be no pressure drop over a ring, because there can only be one pressure at a point in the network. If a gas molecule would flow exactly one round over the ring, it must have its original pressure back and consequently the pressure drop over the ring is zero.

4.1 Gas transmission network

A gas transmission network consists of nodes n that are connected by pipelines p. On some places in the network two nodes are directly connected by more than one pipeline. This group of pipelines is seen as one pipeline p with the properties of the group of pipelines that connect the nodes. Gas can enter or leave the network at the nodes. The gas is shipped via the pipelines. Gas can not be conserved at a node. Node balance is required, all the gas that enters a node has to leave it also. Furthermore the pressure drop over a ring in the network is zero. With those two requirements the flow over a gas transmission network is fixed. [15]

All pipelines are given a defined direction. Positive flow means that the gas flows in the direction of definition. Negative flows means the opposite. The transport task is F_n. Negative F_n means entry of gas at node n, positive F_n represents exit.

An example of a gas transmission network is presented in figure 4.1. There are three nodes in figure 4.1 where node balance is required.

$$F_A = Q_{AB} - Q_{CA}, \quad (4.1.1)$$

$$F_B = Q_{BC} - Q_{AB}, \quad (4.1.2)$$

$$F_C = Q_{CA} - Q_{BC}. \quad (4.1.3)$$

There is one ring in figure 4.1. The pressure drop over this ring must be zero. Out of report [15] follows that:

$$\frac{c_{AB} L_{AB} Q_{AB}|Q_{AB}|}{D_{AB}^5} + \frac{c_{BC} L_{BC} Q_{BC}|Q_{BC}|}{D_{BC}^5} + \frac{c_{CA} L_{CA} Q_{CA}|Q_{CA}|}{D_{CA}^5} = 0. \quad (4.1.4)$$

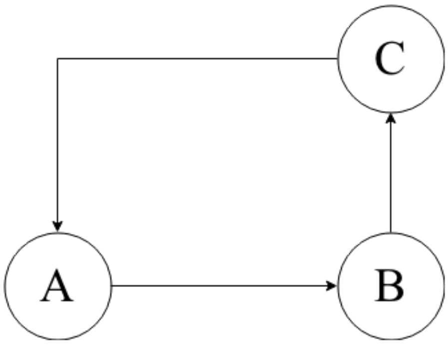

Figure 4.1: Network with a ring and defined directions on the pipelines.

4.2 Newton method to approximate the flow

The set of equations 4.1.1 till 4.1.4 must be solved to determine the flows through the network of figure 4.1. A way to get an approximation of the flows is to use the Newton Method [1]. The Newton Method is an iterative method that approximates the roots of a system of equations. Store the flows over all the pipelines in vector $Q = [Q_{AB}\ Q_{BC} \cdots Q_m]^T$. H is the set of equations that must be approximated, $H(Q) = [H_A(Q)\ H_B(Q) \cdots H_r(Q)]^T$. The first equations in H require node balance, the last equations require that the pressure drop over a ring must be zero. J_H is the Jacobian of H. Q^0 is an initial guess for Q. Solving equation 4.2.1 for Q^{i+1} should be repeated until $Q^i \approx Q^{i+1}$ to find the flows through the network.

$$J_H(Q) = \begin{bmatrix} \frac{\partial H_A(Q)}{\partial Q_{AB}} & \frac{\partial H_A(Q)}{\partial Q_{BC}} & \cdots & \frac{\partial H_A(Q)}{\partial Q_m} \\ \frac{\partial H_B(Q)}{\partial Q_{AB}} & \frac{\partial H_B(Q)}{\partial Q_{BC}} & \cdots & \frac{\partial H_B(Q)}{\partial Q_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial H_r(Q)}{\partial Q_{AB}} & \frac{\partial H_r(Q)}{\partial Q_{BC}} & \cdots & \frac{\partial H_r(Q)}{\partial Q_m} \end{bmatrix},$$

$$J_H(Q^i)(Q^{i+1} - Q^i) = -H(Q^i). \tag{4.2.1}$$

To solve 4.2.1 for Q^{i+1}, the inverse of $J_H(Q^i)$ must be found. However, the dimension of vector H (and the number of rows of J) is the number of nodes plus the number of rings in the network and the number of columns in J is the number of pipelines in the network. The number of nodes plus the number of rings is always strictly larger than the number of pipelines in the network. This last statement is proven in lemma 1. That means that $J_H(Q^i)$ is not a square matrix and therefore has no inverse.

Lemma 1: A gas transmission network that consists of N nodes and R rings, consists of $P < N+R$ pipelines.
Proof: The statement can be proven by mathematical induction. First think of the most basic gas transmission network. That consists of two nodes that are connected by one pipeline, obviously there is no ring in this network. So, $N + R = 2$ and $P = 1 < N + R$. For the basic step the statement holds.
Now think about any gas transmission network after k expansions of the most basic network, with $N(k)$ nodes, $R(k)$ rings and $P(k)$ pipelines. When this network is expanded, one pipeline is added ($P(k+1) = P(k)+1$). This pipeline has to start at a node that was already part of the network, because otherwise this new pipeline would not be part of the existing gas transmission network. This pipeline can lead to a node that was not already part of the network or one that already was. In the first case a node is added to the network and not a new ring ($N(k+1) = N(k)+1$ and $R(k+1) = R(k)$). In the second case, one or more rings are added and not a new node

$(N(k+1) = N(k)$ and $R(k+1) \geq R(k)+1)$. Therefore: $N(k)+R(k)+1 \leq N(k+1)+R(k+1)$.

$$\begin{aligned}
P(k) &< R(k)+N(k) \\
P(k)+1 &< R(k)+N(k)+1 \\
P(k+1) &< R(k)+N(k)+1 \leq R(k+1)+N(k+1) \\
P(k+1) &< R(k+1)+N(k+1)
\end{aligned}$$

Q.E.D.

The Jacobian is $J_H \in \mathbb{R}^{r\times m}$ with $r > m$. Therefore, the generalized left inverse of J_H, which is $(J_H^T J_H)^{-1} J_H^T$, should be used to find Q^{i+1} in 4.2.1.

$$\begin{aligned}
J_H(Q^i)(Q^{i+1}-Q^i) &= -H(Q^i), \\
(J_H^T(Q^i)J_H(Q^i))^{-1}J_H^T(Q^i)J_H(Q^i)(Q^{i+1}-Q^i) &= -(J_H^T(Q^i)J_H(Q^i))^{-1}J_H(Q^i)^T H(Q^i), \\
Q^{i+1} &= Q^i - (J_H^T(Q^i)J_H(Q^i))^{-1}J_H(Q^i)^T H(Q^i).
\end{aligned}$$

The repetition of $Q^{i+1} = Q^i - (J_H^T(Q^i)J_H(Q^i))^{-1}J_H(Q^i)^T H(Q^i)$ should be stopped when Q^i and Q^{i+1} are "approximately" the same. When MATLAB is used to do the calculations, programming $Q^{i+1} = Q^i - J_H \backslash H(Q^i)$ is much more efficient. For the network of figure 4.1 Q and $H(Q)$ are:

$$\begin{aligned}
Q &= \begin{bmatrix} Q_{AB} \\ Q_{BC} \\ Q_{CA} \end{bmatrix}, \\
H(Q) &= \begin{bmatrix} Q_{AB} - Q_{CA} - F_A \\ Q_{BC} - Q_{AB} - F_B \\ Q_{CA} - Q_{BC} - F_C \\ \frac{c_{AB}L_{AB}Q_{AB}|Q_{AB}|}{D_{AB}^5} + \frac{c_{BC}L_{BC}Q_{BC}|Q_{BC}|}{D_{BC}^5} + \frac{c_{CA}L_{CA}Q_{CA}|Q_{CA}|}{D_{CA}^5} \end{bmatrix} = 0.
\end{aligned}$$

The absolute value of Q_p can also be written as $\sqrt{Q_p^2}$. That makes derivative: $\frac{\partial(Q_p|Q_p|)}{\partial Q_p} = \frac{\partial(Q_p\sqrt{Q_p^2})}{\partial Q_p} = \sqrt{Q_p^2} + \frac{Q_p^2}{\sqrt{Q_p^2}} = 2|Q_p|$. The Jacobian of H(Q) for the network of figure 4.1 is then:

$$J_H(Q) = \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ \frac{2c_{AB}L_{AB}|Q_{AB}|}{D_{AB}^5} & \frac{2c_{BC}L_{BC}|Q_{BC}|}{D_{BC}^5} & \frac{2c_{CA}L_{CA}|Q_{CA}|}{D_{CA}^5} \end{bmatrix}. \tag{4.2.2}$$

4.3 Pressure

When the flows through a network are known, the difference between the squares of the pressures at all the ends of the pipelines are also known by equation 4.0.1. P_{in} in equation 4.0.1 is the pressure at that end of a pipeline where the gas enters this pipeline. When the flow trough a pipeline is in the direction of the definition of the pipeline, then this is the end of the pipeline where the pipeline starts. When the flow is negative, then gas enters the pipeline where this pipe ends by definition. The definition of the pipelines of the network of figure 4.1 can be seen in the first rows of the Jacobian of H (equation 4.2.2). Split J_H in parts K^T and R. Let matrix K describe the mapping of the network and let R be the matrix with derivatives from the functions in H that describe the flow over a ring. In matrix K, rows represent nodes and columns represent pipelines.

$$J_H = \begin{bmatrix} K^T \\ R \end{bmatrix}$$

The number of rows in K is equal to the number of pipelines in the network. The number of rows in R is equal to the number of rings. K has as many columns as there are nodes network. The definition of K is that if entry $(i,j) = 1$ then pipeline j starts by definition in node i and if $(i,j) = -1$ then pipeline j ends by definition in node i .

Q is the vector with all the flows over the pipelines. Let L be a vector with lengths of pipelines, D a vector with diameters of pipelines and c a vector with the pressure constant for every pipeline. The difference in the square of the pressure between the starts (by definition) of the pipeline and ends of the pipelines (vector S) is then:

$$S = Q \circ |Q| \circ c \circ L \circ D^{-5},$$

with $\circ$ being the Hadamard product. The Hadamard product is an operation for two matrices of the same size that produces a matrix of this same size with in entry (i,j) the multiplication of the $(i,j)^{th}$ entries of the two input matrices.

From K it is known which pipeline connects which nodes. With the knowledge of the pressure drop over the pipelines, the difference between the squares of the pressure in the nodes can be known. All the equations that should be fulfilled for the network of figure 4.1 are:

$$P_A^2 - P_B^2 = S_{AB}, \tag{4.3.1}$$
$$P_B^2 - P_C^2 = S_{BC}, \tag{4.3.2}$$
$$P_C^2 - P_A^2 = S_{CA}. \tag{4.3.3}$$

Let $\hat{P}$ be the vector with the squares of the pressure of the nodes. The equations 4.3.1 till 4.3.3 can be described in matrix form.

$$\underbrace{\begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix}}_{K} \underbrace{\begin{bmatrix} P_A^2 \\ P_B^2 \\ P_C^2 \end{bmatrix}}_{\hat{P}} = \underbrace{\begin{bmatrix} S_{AB} \\ S_{BC} \\ S_{CA} \end{bmatrix}}_{S} \tag{4.3.4}$$

$$\hat{P} = (K^T K)^{-1} K^T S \tag{4.3.5}$$

Since in 4.3.4 K and S are known, $\hat{P}$ can be calculated easily in 4.3.5. In this example K is a square matrix, but this is clearly not the case for any network. Therefore, the generalized inverse is used to calculate $\hat{P}$. When $\hat{P}$ is known, the pressure in one node in the network should be known to calculate the pressure in every point. This is because from $\hat{P}$ can be seen what the difference between the squares of the pressure is between all nodes. The gas transmission network is namely connected. So for every pair of nodes is the difference in squares of the pressure known from $\hat{P}$. That is, for arbitrary nodes i and j, $\hat{P}_i - \hat{P}_j$.

Solving $\hat{P}$ in this way will lead to a solution where the entry in $\hat{P}$ that corresponds to the node with the lowest pressure in the network is 0. At this node in the network, the pressure can be set at for example the delivery pressure, called Z. From there, the pressure in all the points in the network can be calculated. Let P be the vector with the pressure of all the nodes. From 4.0.1 follows that:

$$P = \sqrt{S + Z^2}$$

For networks with a ring is $K^T K$ a singular matrix. The matrix K stores in that case more information than is strictly needed, because it does not matter for the pressure of the gas in a node on a ring from which directions this gas comes. So when one pipeline of a ring is ignored, only one way to determine the pressure in the nodes on this ring exists. One row of matrix K

should be removed for each ring in the network. This must be a row that represents a pipeline that belongs to a ring. Simultaneously the corresponding entry in S must be removed. For rings that are connected should this be done more careful, because from every ring in the network exactly one pipeline must be removed.

After performing all these calculations the flows through the network (vector Q) and the pressures on all the nodes in the network (vector P) are known for transport task F.

5 Balance in the network

A shipper does not have to inject the same amount of gas into the network as it extracts every hour. However, the combined amount of entry and exit of gas of all the shippers should not be too much out of balance, because this would give operational problems to the network. GTS has a balancing regime to force the shippers to maintain individual balance and total balance of the network. This system, as well as the way the individual shippers react in their nominations to this balancing system, is worked out in this chapter. The reaction of the individual shippers to the balancing regime has not been modeled before at GTS.

The gas transmission network must always contain a certain amount of gas to keep the network operating. With this exact amount of gas in the network, the network is in balance. The position of the network (the difference between the actual amount of gas in the network and the ideal amount) must be within a range. Therefore shippers do not necessarily have to be perfectly in balance. However, GTS holds all the shippers responsible for the total network position [3]. The individual position of a shipper is the difference between the amount of gas it has injected into and extracted from the network in its total history. The total network position is the sum of the positions of the individual shippers. When the network position is outside the set limit, the shippers that cause the problems are held responsible and the shippers that help solve the problem can profit from this.

The shippers that have the same imbalance position, in sign, as the network are called the causers. Shippers that have the opposite network position, in sign, are the helpers. GTS will not take any action when the position of the shippers as total (which is the position of the network) stays between the set position limits. Every hour, at 15 minutes past the whole hour, GTS determines what the position of the network at the end of the hour will be, based on the all the nominations of the shippers. If it is expected that this position will be outside the set limit, GTS will sell or buy as much gas as is needed at the gas spot market to bring the network position within the position limits. The price of this gas, which is some kind of penalty, will be charged to all the shippers that are causers. The shippers that are the causers will pay the total price of this gas proportionally to their individual position. All the shippers have real time insight in the position of the network and their own position.

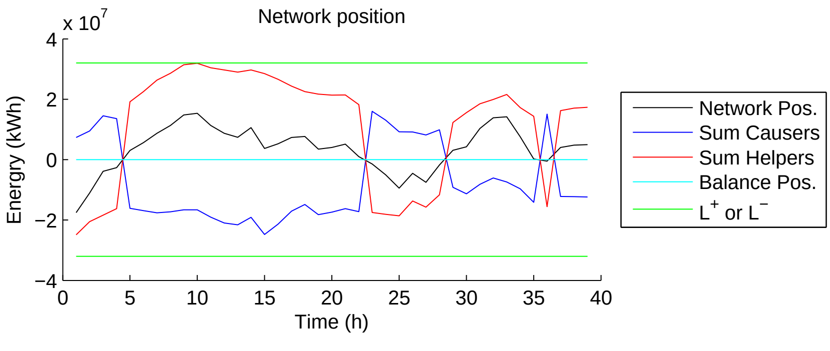

Figure 5.1: Historical overview of the network position on December 14 and 15 2014. L^+ and L^- are the position limits of the network.

5.1 Active balancing

A shipper has the obligation to maintain its own balance position. The easiest way for a shipper to change its own position is buying or selling gas at the Title Transfer Facility (TTF), because no operational interventions are needed for this. The TTF is a virtual market place where gas that is already injected in the system can be bought or sold by shippers. [13] Such a trade of gas only influences the position of the individual shippers that are involved in this trade. It does not change the network position. To actively change the network position, operational interventions are the only option. That involves a real change in entry or exit at one or more network points.

Lowering or increasing the nomination for balancing on most of the network points is problematic. Behind all the interior exit points, except the gas storages, is a market. Lowering the nominations at such points is not the first thing that a shipper would do, because they have contracts to supply gas to these markets. Increasing the nominations for balancing at exit points with a market behind it is impossible, because the gas can not be used there. Furthermore, the network points on the border of the county have a market as well. Adjusting the nominations at these points for balancing can give operational problems at the other side of the border.

Gas storages are exit or entry points that can be used for balancing. The large gas field in Groningen can be used for balancing also. Small changes in allocations from hour to hour here are not a heavy intervention. However, there is only one shipper that can enter there: Gasterra has the legal task to sell the gas that is produced in Groningen. This is one of the largest entry points in the network. The smaller gas fields in the north of the country have to produce a constant amount of gas, for operational reasons. Therefore, balancing on these points is impossible.

5.2 The SBS

The balance position of the network is displayed on the SBS (Systeem Balans Signaal). The SBS can fluctuate to some extent before GTS intervenes. When the SBS is outside the position limits (L^+ and L^-, with $L^+ = |L^-|$), GTS will intervene. The position of the total network at hour i is SBS_i. The individual position of a shipper s is $B_{s,i}$. This $B_{s,i}$ is the difference between the amount of energy a shipper has ever injected to and extracted from the network. The sum of the positions of the individual shippers is the total network position. The network points are called n. A shipper sends a transport task (nominations) every hour to GTS. No difference is made between allocations and nominations, they are assumed to be the same. The nomination of shipper s at hour i on a network point n is $F_{s,i,n}$. A positive number F represents an exit of the network, a negative number an entry. The sum of all nominations in one hour of a shipper is its deviation in network position from the last hour due to operations. Furthermore a shipper can change its individual position by buying or selling gas already injected in the network at the TTF. The allocation of a shipper at the TTF is $T_{s,i}$. The SBS is updated once every hour.

$$\begin{aligned} SBS_i &= \sum_s B_{s,i} \ \forall i, \\ B_{s,i} &= B_{s,i-1} + T_{s,i} + \sum_n F_{s,i,n} \ \forall s, i. \end{aligned}$$

On every hour the shippers are divided into two subgroups, the causers and the helpers. The causers (set C_i) are the shippers that have in sign the same individual position as the network position at the end of the hour. The other shippers are the helpers (set H_i). The sets C_i and H_i can change from hour to hour. When the SBS is outside the position limits, the causers will be penalized for that.

$$\begin{aligned} s \in C_i \quad \text{if} \quad \frac{B_{s,i}}{|B_{s,i}|} &= \frac{SBS_i}{|SBS_i|}, \\ s \in H_i \quad \text{if} \quad \frac{B_{s,i}}{|B_{s,i}|} &= -\frac{SBS_i}{|SBS_i|}. \end{aligned}$$

5.3 Allocation at network points of shippers

Shippers earn money by trading gas. This is the main reason they nominate on network points. Besides this reason, a shipper can also change its nomination for balancing. For a shipper there are two types of balancing. A shipper can think it is too much out of balance and therefore adjusts its nomination or a shipper can participate in a balancing action of GTS. Therefore the the nomination of a shipper consists of three parts. These are:

- $\tilde{F}$ is the amount of gas a shipper nominates by the supply and demand of gas.
- $\bar{F}$ is the amount of gas a shipper nominates to adjust its individual position.
- $\hat{F}$ is the amount of gas a shipper nominates for a balancing action of GTS.

That makes the total nomination of a shipper in an hour:

$$F_{s,i,n} = \tilde{F}_{s,i,n} + \bar{F}_{s,i,n} + \hat{F}_{s,i,n} \ \forall s, i, n.$$

Prior to every hour a shipper sends its transport request to GTS. This only consists of the parts $\tilde{F}$ and $\bar{F}$. A shipper can not know prior to an hour if there will be a balancing action by GTS. This part of the nomination is determined during the hour, when there is a balancing action or not.

5.4 Shipper behavior based on the balancing regime

The goal of the balancing regime is that shippers keep the position of the network within position limits. To do this, they have to react on the position of the network and on their individual position. In the ideal world the individual position of every shipper would always be zero. However, due to market forces, this does not work out for every shipper. The most advantageous position of a shipper is having a position close to zero and being helper. This is the most advantageous position, because a helper can never get a penalty and furthermore, a small imbalance in the sum of allocations on network points and TTF trades in the next hour do not make this shipper a causer immediately. Being a helper while being much out of balance is not an instant problem, but when the sign of the SBS changes, the shipper risks a large penalty immediately. On the other hand, being a causer is never profitable.

The easiest way for a shipper to change its individual position is doing a trade at TTF. Helpers that are too much out of balance can offer a part of their position there. They are assumed not to offer their total position, because the most advantageous position is being a small helper. The causers can buy what is offered. For a causer it is always profitable to improve its position on TTF. A TTF trade does not require operational interventions that can hurt the contracts between shippers for the supply and demand of gas and it reduces the probability of a penalty. Therefore, causers should always be willing to buy what is offered at TTF by the helpers.

The absolute values of the individual position of a shipper for the cases when it is a helper and it is a causer are plotted in figure 5.2. The form of these plots is similar for every shipper. A few things can be seen from the plots. First notice the peaks: apparently a shipper can always actively adjust its position when it wants to, because when the shipper is more out of balance than usual, the individual position of a shipper heads back to zero. The second observation is that the peaks are higher when the shipper is a causer than when the shipper is a helper. Also the average absolute value of the individual position of every shipper is higher when a shipper is a causer than when it is a helper. This may look strange, because a shipper should not be willing to be more out of balance when it is a causer than when it is a helper, but there can be a reasonable explanation for this. The ideal position for a shipper is being a small helper. When a helper is "too much" out of balance it offers a part of its position on TTF. There should always be a (group of) causers that is willing to buy this, because it improves their individual position(s). Consequently a helper can become a small helper when it wants to. It is impossible for all the causers to recover all their positions via TTF. To totally recover, a causer has to do an operational intervention. Operational interventions are usually more expensive than trades on TTF. Therefore it is assumed that a shipper only tries to adjust its individual position with operational interventions when it is a causer and "too much" out of balance.

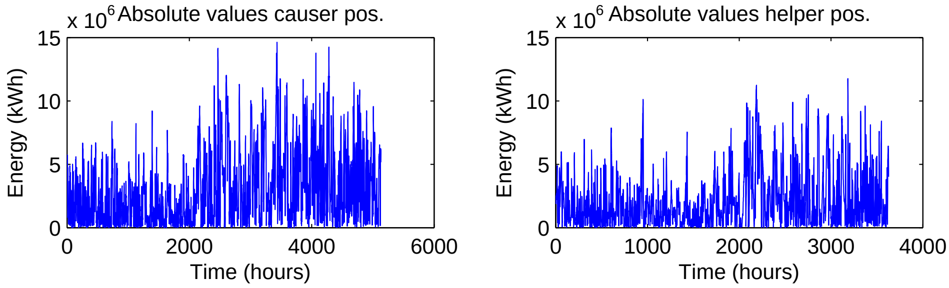

Figure 5.2: The absolute values of the individual position of a shipper when it was a causer (left plot) and when it was a helper (right plot) in 2014. In this year the shipper was 5132 hours a causer and 3628 hours a helper.

The set $\hat{H}_i$ is the set of helpers that is offering a part of their position at the TTF. A shipper that is a helper offers something at TTF when it expects to be "too much" out of balance in the next hour: $|B_{s,i-1} + \sum_n \tilde{F}_{s,i,n}| > \tau_s L^+$, with $0 < \tau_s < 1$. The part of the individual position that a shipper offers is γ_s, with $0 < \gamma_s < 1$:

$$s \in \tilde{H}_i \quad \text{if:} \quad |B_{s,i-1} + \sum_n \tilde{F}_{s,i,n}| > \tau_s L^+, s \in H_i,$$

$$T_{s,i} \quad = \quad -\gamma_s(B_{s,i-1} + \sum_n \tilde{F}_{s,i,n}) \; \forall i, s \in \tilde{H}_i.$$

The absolute value of the sum of the individual positions of all the causers is always larger than the absolute value of the sum of the individual positions of all the helpers. The causers are assumed to buy everything at the TTF that is offered by the helpers, because it is an easy and relative cheap way to improve their position. It is not known in advance which shipper is going to buy what amount. Therefore, the shippers that are most out of balance are assumed to be most willing to buy at the TTF, because they have the hardest reason to balance their position. Everything that is offered by the helpers is assumed to be divided over the causers proportionally to their individual balance position:

$$T_{s,i} \quad = \quad \frac{B_{s,i-1}}{\sum_{s \in C_i} B_{s,i-1}} (\sum_{s \in \tilde{H}} \gamma_s (B_{s,i-1} + \sum_n F_{s,i,n})) \; \forall i, s \in C_i.$$

A causer is too much out of balance when its expected individual position at the end of the hour due to regular nominations on network points and nominations on TTF : $|B_{s,i-1} + \sum_n \tilde{F}_{s,i,n} + T_{s,i}|$ is larger than $\eta_s L^+$, with $0 < \eta_s < 1$. The set of causers that is too much out of balance is $\hat{C}_i$:

$$s \in \tilde{C}_i \quad \text{if} \quad |B_{s,i-1} + \sum_n \tilde{F}_{s,i,n} + T_{s,i}| > \eta_s L^+, s \in C_i.$$

The shippers that are causers and too much out of balance are going to do an operational intervention. When the shipper has a booking on a gas storage or the production field in Groningen (set $\hat{N}$), it is assumed the operational intervention is done there proportionally to the bookings of the shipper. When the shipper does not have this possibility, it is assumed it goes to the spot market to buy capacity at a network point out of the set $\hat{N}$. It is not known in advance which of the shippers that have bookings on $\hat{N}$ are going to offer gas. Therefore, when a shipper does not have a booking on $\hat{N}$, it is assumed that this gas is allocated proportionally to all the bookings of all the shippers that have bookings on $\hat{N}$.

Everything the shippers out of $\hat{C}_i$ can not adjust via the TTF, will be adjusted via the gas spot market. $\bar{F}_{s,i,n}$ is the amount the causers are going to allocate with operational interventions. The exit bookings on a network point of a shipper in an hour are $W^+_{s,i,n}$ and the entry bookings are $W^-_{s,i,n}$. The amount a shipper wants to nominate for individual balancing is $\bar{A}_{s,i}$. A causer does not necessarily make up totally for its imbalance, because this can be expensive. It can also make up for a part of its position and reevaluate in the next hour. A causer that is too much out of balance is assumed to balance part ν_s of its individual position, with $0 < \nu_s < 1$.

$$\bar{A}_{s,i} = -\nu_s(B_{s,i-1} + \sum_n \tilde{F}_{s,i,n} + T_{s,i}) \ \forall i, s \in \hat{C}.$$

The nomination on every network for every shipper to adjusts it own position is:

$$\bar{F}_{s,i,n} = \begin{cases} \frac{W^+_{s,i,n}}{\sum_{n\in\hat{N}} W^+_{s,i,n}} \bar{A}_{s,i} & \text{if: } n \in \hat{N}, \sum_{n\in\hat{N}} W^+_{s,i,n} > 0, \bar{A}_{s,i} > 0 \\ \frac{W^-_{s,i,n}}{\sum_{n\in\hat{N}} W^-_{s,i,n}} \bar{A}_{s,i} & \text{if: } n \in \hat{N}, \sum_{n\in\hat{N}} W^-_{s,i,n} < 0, \bar{A}_{s,i} < 0 \\ \frac{\sum_s W^+_{s,i,n}}{\sum_s \sum_{n\in\hat{N}} W^+_{s,i,n}} \bar{A}_{s,i} & \text{if: } n \in \hat{N}, \sum_{n\in\hat{N}} W^+_{s,i,n} = 0, \bar{A}_{s,i} > 0 \\ \frac{\sum_s W^-_{s,i,n}}{\sum_s \sum_{n\in\hat{N}} W^-_{s,i,n}} \bar{A}_{s,i} & \text{if: } n \in \hat{N}, \sum_{n\in\hat{N}} W^-_{s,i,n} = 0, \bar{A}_{s,i} < 0 \\ 0 & \text{otherwise} \end{cases}$$

5.5 Balancing regime

Every hour, at 15 minutes past the whole hour, GTS predicts what the network position will be at the end of the hour. This prediction is made of the current position of the network and all the nominations of the shippers for the next hour. The nomination a shipper sends in is: $\hat{F}_{s,i,n} + \bar{F}_{s,i,n}$. When GTS predicts that the SBS at the end of the hour will be outside the position limits, a balancing action will take place. The prediction of the SBS by GTS is E_i and the amount that will be balanced is A_i. Information on the TTF trades are not included in this prediction, because the sum over the TTF trades of all the shippers within one hour is always zero. The prediction of the GTS is:

$$E_i = SBS_{i-1} + \sum_n \sum_s (\hat{F}_{s,i,n} + \bar{F}_{s,i,n}) \ \forall i.$$

Then is determined if and how much there will be balanced:

$$A_i = \begin{cases} -(|E_i| - L^+)\frac{E_i}{|E_i|} & \text{if: } |E_i| > L^+ \\ 0 & \text{if: } |E_i| \leq L^+ \end{cases}$$

When there is a balancing action, it is not known in advance which shippers are going to offer gas. Therefore the gas A_i is allocated to the network points in $\hat{N}$, proportionally to the bookings on those network points. The shippers that will have this gas allocated are the causers, proportionally to their individual positions.

$$\hat{F}_{s,i,n} = \begin{cases} \frac{\sum_s W^+_{s,i,n}}{\sum_s \sum_{n\in\hat{N}} W^+_{s,i,n}} \frac{B_{s,i}}{\sum_{s\in C_i} B_{s,i}} A_i & \text{if: } n \in \hat{N}, \sum_s \sum_{n\in\hat{N}} W^+_{s,i,n} > 0, A_i > 0, s \in C_i \\ \frac{\sum_s W^-_{s,i,n}}{\sum_s \sum_{n\in\hat{N}} W^-_{s,i,n}} \frac{B_{s,i}}{\sum_{s\in C_i} B_{s,i}} A_i & \text{if: } n \in \hat{N}, \sum_s \sum_{n\in\hat{N}} W^-_{s,i,n} < 0, A_i < 0, s \in C_i \\ 0 & \text{otherwise} \end{cases}$$

With that information the three parts of the total nomination of a shipper on every network point are known and the behavior of shippers based on the balancing regime is modelled.

6 Analysis of the domestic market

The demand of gas in the domestic market follows the same pattern every year. The utilizations of the bookings, that are done by GTS in this market, are used to make a model to predict the demand. The average hour demand over a day at a network point is predicted with an autoregressive model, based on the average hour demand of the preceding days. First will be tested what the distribution of the demand is. Afterwards is determined how the demand at a network relates to the demand at other domestic market network points.

The domestic market is a typical market with temperature influences. On cold days the demand for gas is higher than on warm days. A booking, allocation and utilization plot as in figure 6.1 is very usual for shippers at these kind of network points. In practice, the capacity at these points is not really booked by the shippers. The booked capacity is actually capacity that is reserved by GTS for these points, because GTS has the legal task to supply gas to the domestic market with a security of supply of 100%.

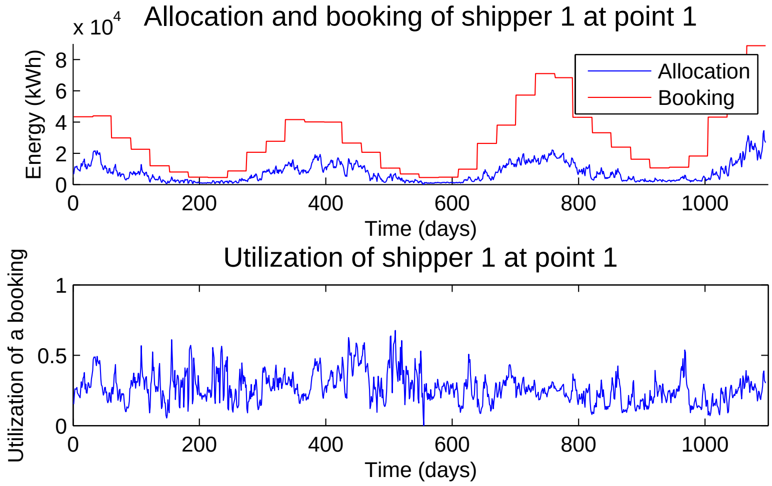

Figure 6.1: The average allocation per hour over a day, the bookings and average utilization of the booking over a day of shipper 1 at network point 1 between January 2012 and Decemer 2014.

From the plot of 6.1 can be seen clearly that every winter much more gas is demanded than every summer. However it looks like the utilization over the whole period has a constant mean. In other words, the weather influences the demand, but not the utilization of a booking. Bookings are done in advance and with the additional information that the utilization follows a probability distribution the demand can be predicted.

Fitting a probability distribution for the utilization over such a long period of time can be difficult, because it is unlikely that any will fit for too much data. However fitting a probability distribution on too less data is also problematic, since any reasonable distribution can not be rejected. From the histograms of figure 6.2 for the data of figure 6.1 it looks like the utilization could be described by a log-normal distribution. A log-normal distribution is a distribution whose logarithm is normally distributed. When the utilizations of the bookings can be described by a log-normal distribution, the properties of the normal distribution can be used in the analysis of the behavior at interior domestic market network points. The logarithm of the utilization is tested for log-normality by a Chi-square goodness-of-fit test. The Chi-square goodness-of-fit test places the utilizations of the bookings into bins and then compares the expected and observed counts for these bins. The expected counts are estimated from the data. The null hypothesis (H_0) is that the utilizations can be described by a normal distribution. The alternative hypothesis (H_1) is that the utilization does not come from a normal distribution. For a test with k bins, N_i observations in bin i and E_i expected observations:

$$X = \sum_{i=1}^{k} \frac{(N_i - E_i)^2}{E_i},$$

and the probability is determined that the distance between expected and observed utilizations is at least as large as X when the sample data would be normally distributed (equation 6.0.1). Under the null hypothesis, X has approximately a Chi-square distribution with $k-1$ degrees of freedom.

$$P(\chi^2(k-1) > X|H_0). \tag{6.0.1}$$

When the probability of equation 6.0.1 is larger than a certain number, usually taken 5%, the null hypothesis is not rejected.

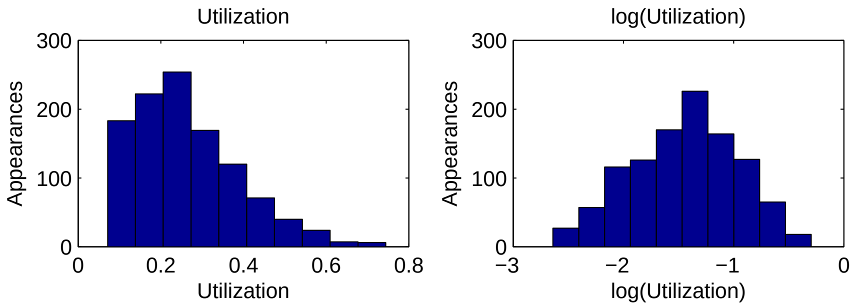

Figure 6.2: Histograms of the utilization of a booking and the logarithm of the utilization of a booking as in figure 6.1.

The data of the cluster of interior domestic market network points Hilvarenbeek-Zeeland (see table 9.1) is tested for periods of six months in 2012 and 2014. For every day in this period on every network point is tested if the utilizations of the bookings in the six months prior to this day could be described by a log-normal distribution. Log-normality was not rejected by the Chi-square goodness-of fit test in 60% of the tested periods when a significance level of 5% is used. For all those days it is assumed that utilization follows a log-normal distribution with a constant mean and variance. We want to describe the logarithm of the utilization by an autoregressive model, which is common for processes that are temperature related. In an autoregressive model the value of the output depends on the previous values of the output.

6.1 Autoregressive model

The utilizations of the bookings at time step n are called R_n. Let U_n be the logarithm of the utilization of the bookings minus the mean of the logarithm of the utilization (6.1.1). By definition U_n is a process with mean 0, which is going to be used in the analysis. U can be described by an autoregressive model of order p.

$$U_n = log(R_n) - \hat{\mu}_{log(R)}, \tag{6.1.1}$$

$$U_n = \sum_{i=1}^{p} \alpha_i U_{n-i} + \beta_n. \tag{6.1.2}$$

β_n is a random number generated from a normal distribution with parameters that have yet to be determined. α_i is a factor for the dependence of U_n and U_{n-i}. U_n is a process with expected value 0. The process U can be measured. The properties of α and β will be determined from the measurements of U. The estimators of the mean, variance and covariance of U are assumed to be the mean, variance and covariance of U. $U^{(a)}$ is the U of network point a.

$$\begin{aligned}
\hat{\mu}_U &= \frac{1}{N}\sum_{n=1}^{N} U_n, \\
\hat{\sigma}_U^2 &= \frac{1}{N-1}\sum_{n=1}^{N} U_n^2 - \hat{\mu}_U^2, \\
\hat{\sigma}_{U^{(a)}U^{(b)}} &= \frac{1}{N-1}\sum_{n=1}^{N} (U_n^{(a)} - \hat{\mu}_{U^{(a)}})(U_n^{(b)} - \hat{\mu}_{U^{(b)}}).
\end{aligned}$$

There is looked at the estimator of the auto covariance to determine the order of the model. The auto covariance (C_{UU}, estimated with $\hat{C}_{UU}$) is the covariance of pairs of points of a string of numbers.

$$\begin{aligned}
cov(U_t, U_s) &= E[(U_t - E[U_t])(U_s - E[U_s])], && (6.1.3)\\
&= E[U_t U_s], && (6.1.4)\\
&= C_{UU}(t-s). && (6.1.5)
\end{aligned}$$

For the numbers k for which $\hat{C}_{UU}(k)$ is "almost zero", there exists no real relationship between U_n and U_{n-k}. The order of the autoregressive model can be the number of time steps back for which there is a relationship. Determining the order of the model is done in section 7.1. The parameters α from equation 6.1.2 can be determined.

$$\begin{aligned}
U_n &= \sum_{i=1}^{p} \alpha_i U_{n-i} + \beta_n, && (6.1.6)\\
U_n U_{n-l} &= \sum_{i=1}^{p} \alpha_i U_{n-i} U_{n-l} + \beta_n U_{n-l}, && (6.1.7)\\
E[U_n U_{n-l}] &= \sum_{i=1}^{p} \alpha_i E[U_{n-i} U_{n-l}] + E[\beta_n U_{n-l}]. && (6.1.8)
\end{aligned}$$

Note that there is no relation between U_i and β_j when $i < j$, because the value of β_j does not influence the value of U_i, when j is later in time than i. Therefore the covariance between U_i and β_j is zero when $i < j$. The mean of both U and β is zero. If $i < j$, then:

$$\begin{aligned}
cov(\beta_j, U_i) &= E[(\beta_j - E[\beta_j])(U_i - E[U_i])], \\
&= E[\beta_j U_i] = 0.
\end{aligned}$$

When $l > 0$ in equation 6.1.8, then $E[\beta_n U_{n-l}] = 0$. If $l > 0$, then:

$$\begin{aligned}
E[U_n U_{n-l}] &= \sum_{i=1}^{p} \alpha_i E[U_{n-i} U_{n-l}], && (6.1.9)\\
\hat{C}_{UU}(l) &= \sum_{i=1}^{p} \alpha_i \hat{C}_{UU}(l-i). && (6.1.10)
\end{aligned}$$

The equation 6.1.10 is known as the Yule-Walker equation [5]. The set of equations in 6.1.10 can be written in matrix form and solved. Note that $C_{UU}(l) = C_{UU}(-l)$.

$$\begin{bmatrix} \hat{C}_{UU}(1) \\ \hat{C}_{UU}(2) \\ \vdots \\ \hat{C}_{UU}(p) \end{bmatrix} = \begin{bmatrix} \hat{C}_{UU}(0) & \hat{C}_{UU}(-1) & \cdots & \hat{C}_{UU}(p-1) \\ \hat{C}_{UU}(1) & \hat{C}_{UU}(0) & \cdots & \hat{C}_{UU}(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{C}_{UU}(p-1) & \hat{C}_{UU}(p-2) & \cdots & \hat{C}_{UU}(0) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix}.$$

Now the α's are known, the distribution of β in equation 6.1.2 has to be determined. β has to be normally distributed with mean 0. The variance of U_n is estimated from the data. From there, the estimator of the variance of β is determined.

$$U_n = \sum_{i=1}^{p} \alpha_i U_{n-i} + \beta_n, \tag{6.1.11}$$

$$var(U_n) = var(\sum_{i=1}^{p} \alpha_i U_{n-i} + \beta_n), \tag{6.1.12}$$

$$\hat{\sigma}_U^2 = \sum_{i=1}^{p} \alpha_i^2 \hat{\sigma}_U^2 + 2\sum_{j=1}^{p-1} \sum_{k=j+1}^{p} \alpha_j \alpha_k \hat{C}_{UU}(k-j) + \hat{\sigma}_\beta^2, \tag{6.1.13}$$

$$\hat{\sigma}_\beta^2 = (1 - \sum_{i=1}^{p} \alpha_i^2)\hat{\sigma}_U^2 - 2\sum_{j=1}^{p-1} \sum_{k=j+1}^{p} \alpha_j \alpha_k \hat{C}_{UU}(k-j). \tag{6.1.14}$$

The individual utilizations U of the bookings are now modeled. Simulated number of U can be converted to a simulation of the utilization of the bookings R.

$$\begin{aligned} U_n &= log(R_n) - \hat{\mu}_{log(R)}, \\ e^{U_n} &= e^{log(R_n) - \hat{\mu}_{log(R_n)}}, \\ R_n &= e^{U_n + \hat{\mu}_{log(R)}}. \end{aligned}$$

The correlation between the utilizations on different network points is up to this point not taken into account. To do this, the covariance between the different U_ns must be found. These covariances must be used to determine the numbers β. Assume two different network points (a and b) with both their own U_n ($U_n^{(a)}$ and $U_n^{(b)}$). The covariance between these points ($\sigma_{U_n^{(a)} U_n^{(b)}}$) can be estimated from the data ($\hat{\sigma}_{U_n^{(a)} U_n^{(b)}}$). With this information the estimator of the covariance between $\beta^{(a)}$ and $\beta^{(b)}$ can be determined. First $\sigma_{\beta_n^{(b)} \beta_n^{(a)}}$ is determined to later determine $\hat{\sigma}_{\beta_n^{(b)} \beta_n^{(a)}}$. The number $p(a)$ is the order of the autoregressive model for network point a.

$$\sigma_{U_n^{(a)} U_n^{(b)}} = E[(U_n^{(a)} - E[U_n^{(a)}])(U_n^{(b)} - E[U_n^{(b)}])], \tag{6.1.15}$$

$$= E[U_n^{(a)} U_n^{(b)}], \tag{6.1.16}$$

$$= E[(\sum_{i=1}^{p(a)} \alpha_i^{(a)} U_{n-i}^{(a)} + \beta_n^{(a)})(\sum_{k=1}^{p(b)} \alpha_k^{(b)} U_{n-k}^{(b)} + \beta_n^{(b)})]. \tag{6.1.17}$$

$$= E[\sum_{i=1}^{p(a)} \sum_{k=1}^{p(b)} \alpha_i^{(a)} \alpha_k^{(b)} U_{n-i}^{(a)} U_{n-k}^{(b)}] + E[\beta_n^{(b)} \sum_{i=1}^{p(a)} \alpha_i^{(a)} U_{n-i}^{(a)}] + \cdots$$

$$\cdots E[\beta_n^{(a)} \sum_{k=1}^{p(b)} \alpha_k^{(b)} U_{n-k}^{(b)}] + E[\beta_n^{(b)} \beta_n^{(a)}] \tag{6.1.18}$$

The different parts of equation 6.1.18 are treated separately.

$$E[\sum_{i=1}^{p(a)} \sum_{k=1}^{p(b)} \alpha_i^{(a)} \alpha_k^{(b)} U_{n-i}^{(a)} U_{n-k}^{(b)}], \tag{6.1.19}$$

$$= \sum_{i=1}^{p(a)} \sum_{k=1}^{p(b)} \alpha_i^{(a)} \alpha_k^{(b)} E[U_{n-i}^{(a)} U_{n-k}^{(b)}], \tag{6.1.20}$$

$$= \sum_{j=1}^{min(p(a),p(b))} \alpha_j^{(a)} \alpha_j^{(b)} E[U_j^{(a)} U_j^{(b)}] + \sum_{i=1}^{p(a)} \sum_{k=1, k\neq i}^{p(b)} \alpha_i^{(a)} \alpha_k^{(b)} E[U_{n-i}^{(a)} U_{n-k}^{(b)}]. \tag{6.1.21}$$

The $E[U_j^{(a)}U_j^{(b)}]$ in equation 6.1.21 is equal to the covariance of $U_n^{(a)}$ and $U_n^{(b)}$. That shows equation 6.1.16.

$$\begin{aligned}
\sigma_{U_{n-i}^{(a)}U_{n-k}^{(b)}} &= E[U_{n-i}^{(a)} - E[(U_{n-i}^{(a)}])(U_{n-k}^{(b)} - E[U_{n-k}^{(b)}])], && (6.1.22)\\
&= E[U_{n-i}^{(a)}U_{n-k}^{(b)}], && (6.1.23)\\
&= C_{U^{(a)}U^{(b)}}(k-i). && (6.1.24)
\end{aligned}$$

The $C_{U^{(a)}U^{(b)}}(k-i)$ in 6.1.24 is the covariance between pairs of points of $U^{(a)}$ and $U^{(b)}$ that differ $k-i$ time steps. It is assumed that $C_{U^{(a)}U^{(b)}}(k-i)$ is constant over time. Now equation 6.1.21 can be written as:

$$\begin{aligned}
&\sum_{j=1}^{min(p(a),p(b))} \alpha_j^{(a)}\alpha_j^{(b)}E[U_j^{(a)}U_j^{(b)}] + \sum_{i=1}^{p(a)}\sum_{k=1,k\neq i}^{p(b)} \alpha_i^{(a)}\alpha_k^{(b)}E[U_{n-i}^{(a)}U_{n-k}^{(b)}], && (6.1.25)\\
= &\sum_{j=1}^{min(p(a),p(b))} \alpha_j^{(a)}\alpha_j^{(b)}\sigma_{U_n^{(a)}U_n^{(b)}} + \sum_{i=1}^{p(a)}\sum_{k=1,k\neq i}^{p(b)} \alpha_i^{(a)}\alpha_k^{(b)}C_{U^{(a)}U^{(b)}}(k-i). && (6.1.26)
\end{aligned}$$

There is no relation between $\beta_i^{(a)}$ and $U_j^{(b)}$ when $i > j$, because the number β is generated independently from a U that lies further in history. Therefore the covariance between $\beta_i^{(a)}$ and $U_j^{(b)}$ when $i > j$ is zero. That implies that, when $i > 0$, then:

$$\begin{aligned}
\sigma_{\beta_n^{(b)}U_{n-i}^{(a)}} &= E[(\beta_n^{(b)} - E[\beta_n^{(b)}])(U_{n-i}^{(a)} - E[U_{n-i}^{(a)}])], && (6.1.27)\\
&= E[\beta_n^{(b)}U_{n-i}^{(a)}] = 0. && (6.1.28)
\end{aligned}$$

From equation 6.1.28 can be seen that $E[\beta_n^{(b)}\sum_{i=1}^{p(a)}\alpha_i^{(a)}U_{n-i}^{(a)}] = 0$ and $E[\beta_n^{(a)}\sum_{k=1}^{p(b)}\alpha_k^{(b)}U_{n-k}^{(b)}] = 0$. At last can be shown that $E[\beta_n^{(b)}\beta_n^{(a)}]$ is the covariance between $\beta_n^{(b)}$ and $\beta_n^{(a)}$.

$$\begin{aligned}
\sigma_{\beta_n^{(b)}\beta_n^{(a)}} &= E[(\beta_n^{(b)} - E[\beta_n^{(b)}])(\beta_n^{(a)} - E[\beta_n^{(a)}])], && (6.1.29)\\
&= E[\beta_n^{(b)}\beta_n^{(a)}]. && (6.1.30)
\end{aligned}$$

The equations 6.1.30, 6.1.28 and 6.1.26 show that:

$$\sigma_{U_n^{(a)}U_n^{(b)}} = \sum_{j=1}^{min(p(a),p(b))} \alpha_j^{(a)}\alpha_j^{(b)}\sigma_{U_n^{(a)}U_n^{(b)}} + \sum_{i=1}^{p(a)}\sum_{k=1,k\neq i}^{p(b)} \alpha_i^{(a)}\alpha_k^{(b)}C_{U^{(a)}U^{(b)}}(k-i) + \sigma_{\beta_n^{(b)}\beta_n^{(a)}}.$$

That leads to the conclusion, that the covariance between $\beta^{(a)}$ and $\beta^{(b)}$ is:

$$\sigma_{\beta_n^{(b)}\beta_n^{(a)}} = (1 - \sum_{j=1}^{min(p(a),p(b))} \alpha_j^{(a)}\alpha_j^{(b)})\sigma_{U_n^{(a)}U_n^{(b)}} - \sum_{i=1}^{p(a)}\sum_{k=1,k\neq i}^{p(b)} \alpha_i^{(a)}\alpha_k^{(b)}C_{U^{(a)}U^{(b)}}(k-i). \quad (6.1.31)$$

Since the exact $\sigma_{U_n^{(a)}U_n^{(b)}}$ and $C_{U^{(a)}U^{(b)}}$ are not known, the estimators are used to determine the covariance between $\beta^{(a)}$ and $\beta^{(b)}$:

$$\hat{\sigma}_{\beta_n^{(b)}\beta_n^{(a)}} = (1 - \sum_{j=1}^{min(p(a),p(b))} \alpha_j^{(a)}\alpha_j^{(b)})\hat{\sigma}_{U_n^{(a)}U_n^{(b)}} - \sum_{i=1}^{p(a)}\sum_{k=1,k\neq i}^{p(b)} \alpha_i^{(a)}\alpha_k^{(b)}\hat{C}_{U^{(a)}U^{(b)}}(k-i). \quad (6.1.32)$$

With the knowledge of equations 6.1.32 and 6.1.14 an estimator of the covariance matrix can be created for the βs of an autoregressive model for a cluster of interior domestic market network points. The estimator of the covariance matrix, the knowledge of the αs and the data of last days can be used to do simulation for the average demand per hour for each of the coming days.

7 Simulating the demand in the domestic market

In chapter 6 a theoretic model is developed to predict the average hour allocation in the domestic market. Before a simulation can be made of the demand in the domestic market, some model choices must be made. First the order of the autoregressive model must be determined. When that is done must be checked if the covariance matrix that is estimated is positive semidefinite. The correlated normally distributed random numbers are generated with a Cholesky factorization. For a Cholesky is required that a matrix is positive semidefinite. Since it are the estimators of the variance and covariance that are assumed to be the variance and covariance it is not by definition true that the matrix that is assumed to be the covariance matrix is positive semidefinite. When this is not the case the nearest positive semidefinite to the estimated covariance matrix will be found and used as covariance matrix. At last must the day and night effect in demand of gas be modeled. This effect is taken in account in the last paragraph of this chapter.

7.1 Order of the model

To find the order of the autoregressive for the demand in the domestic market, the influence of the demand time step $i-k$ on the demand in time step i should be researched. The αs in the autoregressive model (equation 6.1.2) can be found by estimating the auto covariance of the utilizations of the bookings for a period prior to a certain day on a interior domestic market network point and then using the Yule-Walker equation 6.1.10. There is looked at December 30^{th} 2014 for all the points in the cluster Hilvarenbeek-Zeeland (see table 9.1). For the data of the network point Dongen log-normality was rejected for a period of six months prior to this day. Therefore, the parameters of Dongen where not determined. We only want to take the steps back into account that have an influence on the current time step. Therefore, the parameters are determined for different model orders. Due to the noise in the data an α for large amount of steps back in time will probably never be 0. When α "is as large as the noise", it is assumed that that amount of steps back in time has no relationship with the current time step anymore.

Network Point		1	2	3	4	5	6	7	8
p=1	α_1	0,78	0,87	0,83	0,87	0,82	0,86	0,88	0,82
p=2	α_1	0,89	0,84	1,07	1,02	1,03	1,13	1,04	1,06
	α_2	-0,14	-0,04	-0,29	-0,17	-0,25	-0,32	-0,19	-0,3
p=3	α_1	0,92	0,87	1,14	1,07	1,09	1,19	1,07	1,14
	α_2	-0,32	-0,28	-0,52	-0,44	-0,49	-0,53	-0,38	-0,59
	α_3	0,21	0,26	0,21	0,26	0,24	0,19	0,19	0,27

Table 7.1: The αs for different order of the model for 8 different network points in the cluster Hilvarenbeek-Zeeland

Besides the parameters for the orders 1,2 and 3 of the autoregressive model are also the parameters for all the network points for order 30 determined. The results of this are plotted in figure 7.1. The parameters in the first time step are clearly the largest and the parameters in the second time step are the second largest in absolute value. However, the absolute value of the parameters in time step 3 are not necessarily larger than some of the parameters further back in time. Therefore it is assumed that the parameters in time step 3 are more caused by noise and than by a relationship between the utilization in time step 0 and time step 3. The demand in the domestic market mainly depends on the outside temperature. Therefore it seems unreasonable to assume that there is a bigger relationship in the demand between two days that lie further apart than two days that are closer apart.

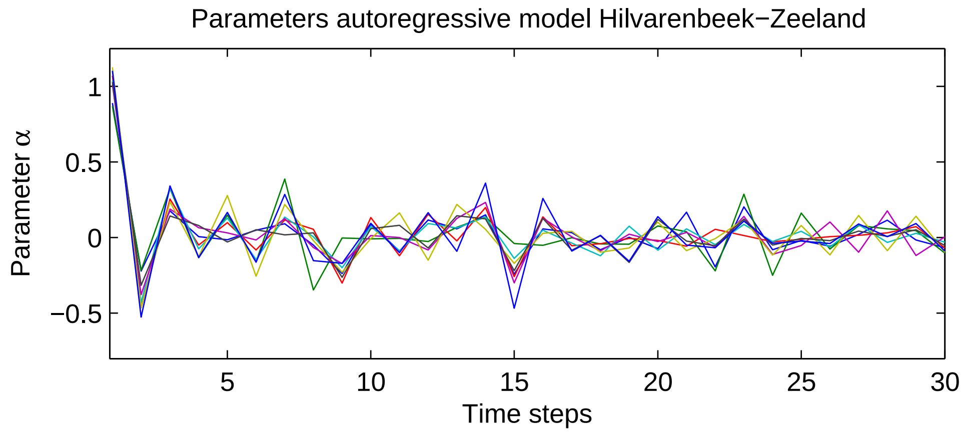

Figure 7.1: The parameters of the autoregressive model of order 30 of the network points in the cluster Hilvarenbeek-Zeeland on December 30^{th} 2014.

7.2 Numerical errors

The covariance matrix C for the βs in chapter 6 is made of the estimators of the variance and covariance of the Us. The difference between the real numbers and estimators can give computational problems. It can occur that the matrix C made from the estimators, that is assumed to be the covariance matrix, is not positive semidefinite. This gives a problem by generating correlated normally distributed random numbers, since correlated normally distributed random numbers β are generated with the help of a Cholesky factorization [11]. To make a Cholesky factorization of C, it is required that C is positive semidefinite. A Cholesky factorization creates an upper or lower triangular matrix $\tilde{C}$ out of positive semi definite matrix C (7.2.2), with the property that $C = \tilde{C}^T\tilde{C}$ [12]. The numbers $\hat{\beta}$ are generated from covariance matrix $C \in \mathbb{R}^{n\times n}$.

$$\tilde{C} = chol(C), \tag{7.2.1}$$
$$C = \tilde{C}^T\tilde{C}, \tag{7.2.2}$$
$$Z \in \mathbb{R}^{n\times 1}, \tag{7.2.3}$$
$$Z_i \sim N(0,1)\ \forall i, \tag{7.2.4}$$
$$\hat{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}, \tag{7.2.5}$$
$$\hat{\beta} = \tilde{C}Z. \tag{7.2.6}$$

The $\hat{\beta}$ in 7.2.6 has distribution $N(0, \tilde{C}^T\tilde{C})$. Therefore it is required that $\tilde{C}^T\tilde{C} = C$, because it is required that $\tilde{\beta} \sim N(0, C)$. In 7.2.6 matrix $\tilde{C}$ is lower triangular.

When C is not positive semidefinite, must be searched for the nearest symmetric positive semidefinite matrix to C. The nearest symmetric positive semidefinite matrix in Frobenius norm can be found by an algorithm in [9]. The Frobenius norm of C $||C||_F$ is:

$$||C||_F = (\sum_{i,j} c_{i,j}^2)^{\frac{1}{2}}.$$

First express C in the symmetric part A of C and the skew-symmetric part B.

$$C = A + B = \frac{C + C^T}{2} + \frac{C - C^T}{2}.$$

Then should be searched for a symmetric positive semidefinite matrix $\hat{C}$ that is nearest to C in Frobenius norm. Find $\hat{C}$ such that:

$$||C - \hat{C}||_F$$

is minimized.

Assume matrix $\hat{C}$ is symmetric positive semidefinite, it follows that $(A - \hat{C}) = (A - \hat{C})^T$ and $B = -B^T$ and:

$$\begin{aligned} ||C - \hat{C}||_F^2 &= ||A + B - \hat{C}||_F^2, &(7.2.7)\\ &= ||A - \hat{C}||_F^2 + ||B||_F^2. &(7.2.8) \end{aligned}$$

The problem of minimizing 7.2.7 is reduced to minimizing $||A - \hat{C}||_F^2$. The eigenvalues of A are λ_i. The spectral decomposition of A is: $A = ZSZ^T$ with unitary matrix Z ($Z^T Z = I$) and diagonal matrix of the eigenvalues S ($S = diag(\lambda_i)$). Choose $Y = Z^T \hat{C} Z$. A lower bound for $||A - \hat{C}||_F^2$ can be found:

$$\begin{aligned} ||A - \hat{C}||_F^2 &= ||Z^T (A - \hat{C}) Z||_F^2,\\ &= ||S - Y||_F^2,\\ &= \sum_{i \neq j} y_{ij}^2 + \sum_i (\lambda_i - y_{ii})^2,\\ &\geq \sum_{\lambda_i < 0} (\lambda_i - y_{ii})^2 \geq \sum_{\lambda_i < 0} \lambda_i^2. \end{aligned}$$

Note that all the diagonal elements of Y (y_{ii}) are non-negative, because $\hat{C}$ is positive semidefinite. Therefore also $Z^T \hat{C} Z = Y$ is positive semidefinite. Now choose Y=diag(d_i), with:

$$d_i = \begin{cases} \lambda_i & \text{if: } \lambda_i \geq 0\\ 0 & \text{if: } \lambda_i < 0 \end{cases}$$

The expression S-Y is then a diagonal matrix with only the negative eigenvalues of A.

$$||S - Y||_F^2 = ||diag(\lambda_i) - diag(d_i)||_F^2 = \sum_{\lambda_i < 0} \lambda_i^2.$$

The lower bound of $||A - \hat{C}||_F^2$ is attained for Y=diag(d_i). Consequently:

$$\hat{C} = Z diag(d_i) Z^T.$$

is the closest symmetric positive semidefinite matrix to C in Frobenius norm. The distance from C to $\hat{C}$ is:

$$\delta_F(C - \hat{C})^2 = \sum_{\lambda_i < 0} \lambda_i(A)^2 + ||B||_F^2.$$

Obivously $\hat{C}$ is positive semidefinite, because diag(d_i) is a diagonal matrix with nonnegative entries and therefore nonnegative eigenvalues.

7.3 Daily pattern

Every day the demand of gas follows about the same pattern, see figure 7.2. In the morning there is peak in demand. Later in the morning the demand drops. The demand increases in the start of the evening again and drops in the night. There can be calculated what percentage of the total daily capacity is demanded in every single hour on average over a period of time for a network point or a region. With that information and the analysis of the average demand per hour on a day, the anticipated demand per hour can be approximated.

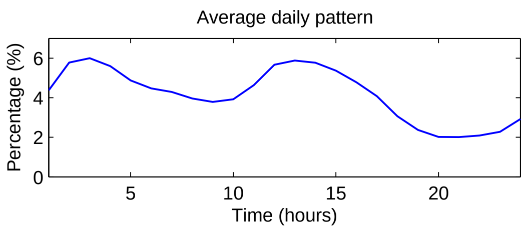

Figure 7.2: Average daily pattern in 2013 for exit points in the domestic market in the cluster Hilvarenbeek-Zeeland (table 9.1). On the vertical axis the percentage of daily capacity that is used in a certain hour on the day is displayed. A gas day starts by definition at 6 o'clock in the morning.

8 Analysis of H-gas border allocation

The demand and supply of gas on a H-gas border point follows a totally different pattern than in the G-gas market. The influence of the weather is less, because this gas is mainly used in the industries. In previous researches of GTS has never been looked at the individual shipper behavior at these kind of network points. This is something that is done in the analysis in this chapter. The (most of the) shippers in this market have a tendency to allocate the same amount of gas for some consecutive hours. This behavior is modeled by creating a Markov chain for the behavior of every individual shipper for every network point.

8.1 Operational levels

Opposite to the behavior of shippers in the domestic market, it is in the H-gas market very common to allocate the same amount of gas in two consecutive hours at a border network point. This assumes that shippers can set their allocation for a certain amount of time. Usually the allocation of a shipper at a border point stays at a level for a certain amount of time before changing to another level. The time that a shipper operates on a certain level can depend on the level itself. Between 2012 and 2014 individual shippers allocated in 78% of the consecutive hours the same amount of gas on H-gas border network points.

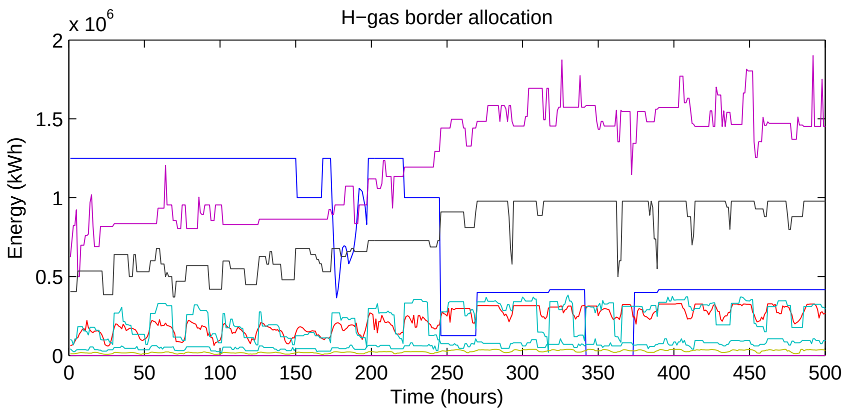

Figure 8.1: Every line in the plot represents the hour allocation of a shipper at one H-gas border network point in a period of 500 hours in 2014. Most of the shippers have a tendency to allocate the same amount for a period of time.

Figure 8.1 shows the allocation of different shippers at the same border network point. All shippers seem to show total different behavior. We want to describe this individual shipper behavior. The sum of the allocations of all the individual shippers on a network point is the allocation of the network point in total. Not all the shippers operate on the same (number of) levels. In figure 8.2 are three examples of allocations by three different shippers plotted.

The three shippers from figure 8.2 operate on a different number of levels. The most left shipper usually allocates zero. Sometimes it allocates more, but it turns back to allocating 0 afterwards. The average residence time of operating on level 0 is much longer than on any of the other levels. The middle shipper in figure 8.2 operates on the levels 0 kWh and $-3 \cdot 10^5$ kWh. The most right shipper operates on multiple levels.

A reason that shippers operate on levels on H-gas border points is that this gas is for customers relative far away. For a shipper it is the easiest to recover for fluctuations in the demand on locations closer to the markets, because the gas that flows from the Netherlands to southern Europe crosses multiple gas transmission networks of different operators. Another reason is that every border

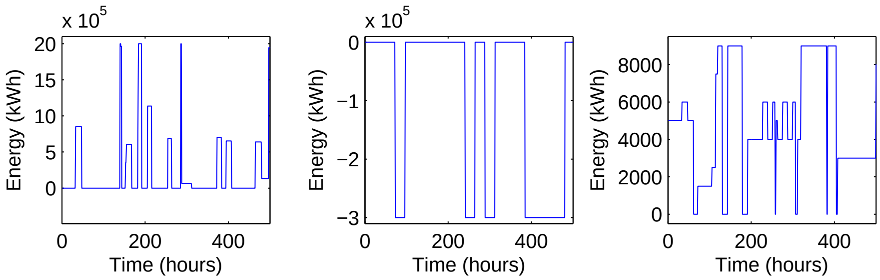

Figure 8.2: Different types of allocation behavior of shippers at H-gas border points. The left shipper operates mainly on one level, the middle on two and the right on multiple levels.

network point has one balancing shipper. The balancing shipper will have the total difference between all the nominations and the total allocation on a network point allocated to its portfolio. That means that for every shipper, except the balancing shipper, the nominations and allocations are equal by definition. These shippers do never have to correct in their nominations because of the difference between their nomination and allocation. Furthermore, the demand of H-gas is more steady than the demand of G-gas. H-gas is mainly used in the industry, with machines that need a constant amount of gas to keep operating.

8.2 Markov chain of shipper behavior

Every hour again a shipper sends in it transport request. So every hour a shipper can change its level of operation or not. Therefore it is assumed that the behavior of a shipper can be described with a discrete time Markov chain. The states of the Markov chain are the level of operations of a shipper. The transition probabilities are the probabilities of changing from one operational level to the other. For the shipper whose allocation is plotted in the middle of plot 8.2, its Markov chain is as in figure 8.3. The Markov Chain has two states, because there are two levels on which the shipper operates. The transition probabilities are obtained from the allocations in history. Whenever this shipper at this network point allocated 0 kWh in an hour, it also allocated 0 kWh in the following hour 98,8% of the time and $-3 \cdot 10^5$ kWh in the following hour in 1,2% of the time. The residence time in the state $-3 \cdot 10^5$ kWh was a little smaller.

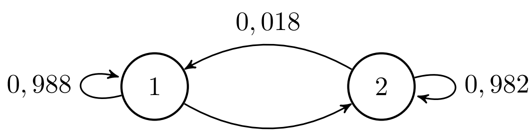

Figure 8.3: Markov chain representation of the behavior of the middle shipper of figure 8.2. State 1 is an allocation of 0 kWh, state 2 is an allocation of $-3 \cdot 10^5$ kWh.

A Markov chain can only have a finite amount of states, while the shipper can take an infinite amount of operational levels. A shipper can choose to allocate any number between its exit and its entry booking. Therefore, the total range of possible allocations of every shipper is divided into k smaller ranges, see figure 8.4 for an example. The shipper of the plot in figure 8.4 has an exit booking of 4000 kWh and no entry booking (0 kWh entry capacity). The range of 0-4000 kWh is divided into 10 equal smaller ranges of 400 and the allocations are rounded off to the closest boundary of two of the smaller ranges. From the rounded off allocations can a Markov chain be created with 11 states. If some of the, in this case 11, operational levels has never been used by a shipper, then the Markov chain has unreachable states. In that way a Markov chain is created for every shipper with a matching amount of states that is reached.

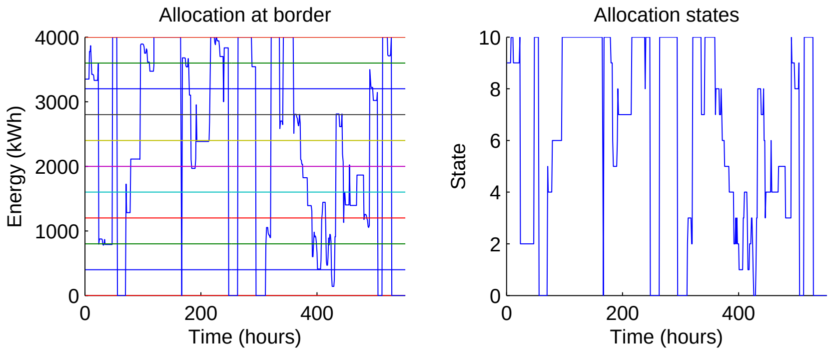

Figure 8.4: In the left plot the border allocation for a shipper, with the total range of allocations divided into ten smaller ranges. After rounding the allocations, the allocation plot is as the plot on the right. Of this right plot a Markov chain is created for the behavior of this shipper at this network point.

Two thing are important by analyzing the behavior of a shipper at an H-gas border network point. The first thing is that the behavior of a shipper probably changes when it does a new booking. It is assumed that a shipper makes a booking with an idea to serve a certain market. When a shipper changes its booking, it is likely that a shipper has other interests than before. When a shipper has other interests, it can show other behavior. Therefore the knowledge of the behavior of the shipper is reset when it does not continue with its current booking. The second important thing is that there must have been enough hours of allocation in the period of the current booking to make a reliable estimation of the transition probabilities.

For every individual shipper at every H-gas border network point an analysis of its behavior in the period of the current booking can be made, when the period of the current booking of the shipper already lasts long enough and the shipper tends to allocate the same amount of gas in consecutive hours. The sum of the allocations of the individual shippers is the total allocation at a network point. The shippers which behavior can not be analyzed by the model of this chapter, can be assigned their maximum allocation in the near future. In this way at least an upper bound of the total allocation at a network point in the near future can be found.

9 Oversell capacity

The general principle to offer oversell capacity is explained first in this chapter. After that, for two border points in the network the amount of oversell capacity that can be offered is determined. The first network point is the L-gas exit at Hilvarenbeek, the second is the H-gas exit in 's-Gravenvoeren. For these points is chosen to show the procedure, because these points lie in the two different networks and the way to determine the oversell capacity there is totally different. For the point of Hilvarenbeek is first an upper bound created for the allocation one day in advance in a cluster of network points from which the capacity can totally be shifted to Hilvarenbeek. After the upper bound of the allocation for the cluster is created, the upperbound for the allocation at the network point itself is found. The reason to do this in this order, is that it uses the same technique (the model of chapter 6 and 7), but to determine the upper bound for the network point itself some calculations must be done in advance. At last an upper bound for the allocation one day ahead on 's-Gravenvoeren is determined. From the upper bounds of the allocations at the network points the amount of oversell capacity that can be offered on the one day ahead auction is determined.

9.1 Overselling in general

A flow chart of the system of overselling can be seen in figure 9.1. When oversell capacity on a network point is determined, first is investigated if the data at this network point follows a certain pattern. First the procedure for network points in the G-gas market is explained. The data of the last six months of the utilization of the bookings at all the individual interior network points of a cluster are tested for log-normality with a Chi-square goodness-of-fit test. When the hypothesis of log-normality is rejected, no oversell capacity at this point will be sold. The original reservation for this point will stay. For the cluster of points for which log-normality was not rejected the parameters of the model to predict them will be determined, see chapter 6. The allocation of the cluster will be simulated and used to predict the allocation in the following days. The simulations are used to create a one-sided confidence interval that gives an upper bound for the average hour allocations in the coming days. With the knowledge of the daily pattern in gas demand, the upper bound for the maximum allocation in a single hour is created. The room that is left between the maximum hour allocation in the simulated period and the original reservation in this period is the capacity that can be offered for overselling over the period.

For an H-gas border point is a similar procedure, but on a shipper level. For every individual shipper at one H-gas border network point will be tested if it has a significant amount of equal consecutive hour allocations. Furthermore the period of the current booking of a shipper has to be long enough to make a reliable approximation of the Markov chain of its allocations. When one of these two requirements is not met for a shipper on a H-gas border network point, no oversell capacity will be sold on its booking. For all the shippers on an H-gas border that do meet the requirements, a Markov chain of the levels of allocation will be created. These Markov chains will be used to simulate the allocation of every individual shipper in the coming days. These simulations are used to create a one-sided confidence interval of the maximum hour allocation in the coming days. The difference between the upper bound of the maximum hour allocation for the total group of shippers at one network point and the total booking of the individual shippers is the capacity that can be offered for overselling.

9.2 Domestic market behind Hilvarenbeek

When gas flows through the compressor station in Ravenstein it is the last time the gas is compressed before it flows to the provinces of Noord-Brabant and Zeeland. Low caloric gas in the south of the Netherlands is used in the domestic market, smaller industries and for transport to Belgium. The export station to Belgium is in Hilvarenbeek. All the capacity of the network points

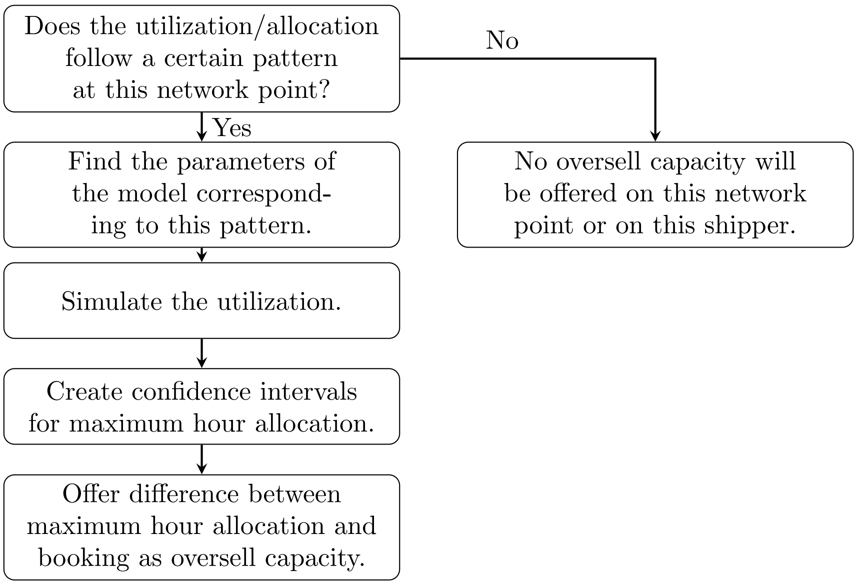

Figure 9.1: The system of overselling

that are downstream of Hilvarenbeek can be shifted to Hilvarenbeek in total or even more than that, because the delivery pressure of the gas of this export station is the same as the delivery pressure of a network point in the domestic market. A part of the capacity of the points that lies upstream Hilvarenbeek and downstream Ravenstein can be shifted to Hilvarenbeek. The cluster of network points that lies downstream of Hilvarenbeek is called Hilvarenbeek-Zeeland, see table 9.1. This cluster is analyzed. Capacity at network points in the domestic market is reserved by GTS a relatively long time in advance. With the technique of chapter 6 a better approximation of the allocation on a short term can be made. With that technique and the data of the allocations of the last time an approximation can be made for an upper bound of the allocation of tomorrow.

Cluster HZ (Hilvarenbeek-Zeeland)
Hoogerheide
Steenbergen
Sprundel
Schoondijke
Breda
Bergen op Zoom
Dongen
Etten-Leur
Gilze

Table 9.1: The cluster of network points in the domestic market in the southwest of the Netherlands.

9.2.1 Cluster Hilvarenbeek-Zeeland

The average hour utilization over a day in the cluster Hilvarenbeek-Zeeland is 30%. We want to determine how much of the booked capacity in the cluster can be shifted to Hilvarenbeek on the short term. There is focused on the one day ahead auction. The are U_is determined for all the network points n in the cluster from the utilizations $R_{i,n}$ for the day i. N is the sample size.

$$U_{i,n} = log(R_{i,n}) - \hat{\mu}_{log(R_n)},$$
$$\hat{\mu}_{log(R_n)} = \frac{1}{N}\sum_{i=1}^{N} log(R_{i,n}).$$

First it is tested for every network point if utilizations of the last six months could be described by a log-normal distribution. A Chi-square goodness-of-fit test is performed for the utilizations of every individual network point. When this test does not reject normality for the logarithm of the Us, it is assumed that the utilizations are a log-normal process with constant mean and variance. December 31^{st} in the year 2014 is the day that is focused on to show the procedure. The oversell capacity on the one day ahead auction is determined one day in advance, which is on December 30^{th} 2014. On that day, the most recent available data is the data of December 29^{th}. Log-normality of the utilizations for the six months prior to that day was rejected only for the network point Dongen out of the cluster if a significance level of 0,05 is used. On the network point Dongen no oversell capacity will be offered for December 31^{st} 2014. The capacity that originally was reserved for this network point stays reserved. For the rest of the network points is the estimator of the variance determined.

$$\hat{\sigma}^2_{U_n} = \frac{1}{N-1}\sum_{i=1}^{N} U^2_{i,n}. \tag{9.2.1}$$

Note that the mean of all the Us is zero by definition. This estimator of the variance (equation 9.2.1) is assumed to be the variance of U. For the cluster HZ the estimators of the variances of U of the individual network points are:

$$\begin{bmatrix} \hat{\sigma}^2_{Hoogerheide} \\ \hat{\sigma}^2_{Steenbergen} \\ \hat{\sigma}^2_{Sprundel} \\ \hat{\sigma}^2_{Schoondijke} \\ \hat{\sigma}^2_{Breda} \\ \hat{\sigma}^2_{BergenopZoom} \\ \hat{\sigma}^2_{Etten-Leur} \\ \hat{\sigma}^2_{Gilze} \end{bmatrix} = \begin{bmatrix} 0,095 \\ 0,113 \\ 0,092 \\ 0,124 \\ 0,106 \\ 0,102 \\ 0,072 \\ 0,143 \end{bmatrix}$$

The auto covariance vectors are estimated with a covariance estimator:

$$\hat{C}_{UU}(j) = \frac{1}{N-1}\sum_{i=1}^{N}(U_i U_{i-j}).$$

For example the estimator of the auto covariance vector for Hoogerheide is:

$$\hat{C}_{UU(Hoogerheide)} = \begin{bmatrix} C_{UU(Hoogerheide)}(-1) \\ C_{UU(Hoogerheide)}(0) \\ C_{UU(Hoogerheide)}(1) \end{bmatrix} = \begin{bmatrix} 0,070 \\ 0,095 \\ 0,07 \end{bmatrix}. \tag{9.2.2}$$

In 9.2.2 are only $C_{UU}(-1)$ till $C_{UU}(1)$ displayed, because these are the only important numbers. As is explained in chapter 7, the desired order of the autoregressive model for the domestic market on this day is 2. Then these three numbers are enough to determine the parameters of the model. That is also the next step in the procedure. With the Yule-Walker equations (6.1.10) the parameters can be determined for the autoregressive model for this cluster.

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 0,85 & 0,83 & 1,05 & 1,08 & 0,97 & 1,13 & 1,01 & 0,97 & 1,09 \\ -0,16 & 0,025 & -0,29 & -0,25 & -0,22 & -0,33 & -0,19 & -0,25 & -0,28 \end{bmatrix}. \tag{9.2.3}$$

With the αs, the estimators of the variances and auto covariances of the Us and equation 6.1.14 the variances of the corresponding βs are determined. At last the covariance is determined between the Us of two different network points in one clusters. The estimator of this covariance, which is assumed to be the covariance, between network points a and b is:

$$\hat{C}_{U^{(a)}U^{(b)}} = \frac{1}{N-1}\sum_{i=1}^{N} U_i^{(a)}U_i^{(b)}. \tag{9.2.4}$$

Now the matrix C with estimators of the variance and covariance of the βs at every network point can be determined. For the cluster HZ at December 29^{th} it is:

$$C = \begin{bmatrix} 0,042 & 0,024 & 0,032 & 0,030 & 0,036 & 0,032 & 0,034 \\ 0,024 & 0,031 & 0,015 & 0,015 & 0,021 & 0,025 & 0,010 \\ 0,032 & 0,015 & 0,027 & 0,027 & 0,027 & 0,026 & 0,029 \\ 0,030 & 0,015 & 0,026 & 0,029 & 0,026 & 0,022 & 0,025 \\ 0,036 & 0,021 & 0,027 & 0,026 & 0,037 & 0,028 & 0,028 \\ 0,032 & 0,025 & 0,024 & 0,022 & 0,028 & 0,027 & 0,021 \\ 0,034 & 0,010 & 0,029 & 0,025 & 0,028 & 0,021 & 0,037 \end{bmatrix}$$

This is all the information that is needed to simulate the total demand for gas in the cluster Hilvarenbeek-Zeeland.

$$\begin{aligned} U_{i,n} &= \alpha_{1,n}U_{i-1,n} + \alpha_{2,n}U_{i-2,n} + \beta_{i,n}\ , \\ R_{i,n} &= e^{U_{i,n}+\hat{\mu}_{log(R_n)}}. \end{aligned}$$

The simulated Us are transformed to utilizations of the booking R. The multiplication of the utilizations with the bookings gives the average day allocation. Every day the demand of gas follows about the same pattern. There is a peak in the demand in the morning and in the afternoon. In the time in between the demand of gas is stable. The average percentage of the daily demand that is used per specific hour is used to determine the demand of every hour. The daily pattern in this cluster in 2013 can be seen table 9.2.

Average use of day capacity in an hour							
Hour	part day cap.	Hour	part day cap.	Hour	part day cap.	Hour	part of day cap.
6	0,044	12	0,043	18	0,059	0	0,024
7	0,058	13	0,040	19	0,058	1	0,020
8	0,060	14	0,038	20	0,054	2	0,020
9	0,056	15	0,039	21	0,048	3	0,021
10	0,049	16	0,046	22	0,041	4	0,023
11	0,045	17	0,057	23	0,031	5	0,029

Table 9.2: Per hour the part of the total day capacity that is used in single hour in the cluster HZ in 2013. A gas day starts by definition at 6 o'clock in the morning.

To determine the the unused capacity for the day of tomorrow, two days ahead must simulated. This is because the data of today is not fully known yet. The data of yesterday is the most recent available data to predict the allocation of tomorrow. Therefore, two days ahead of yesterday must simulated to determine the unused capacity for tomorrow.

The booking with the shortest time is a day booking. In a day booking is for every hour on a gas day (from 6 o'clock in the morning to 6 o'clock the next morning) the same amount reserved at a network point. Due to the daily pattern of gas demand the utilization of this reservation is fluctuating. Therefore is investigated what the smallest amount of unused capacity is in an hour per simulation in the next two days. This is capacity, from which is assumed that it can be forfeited for this cluster, is shiftable. The allocations for December 30^{th} and December 31^{st} 2014 of every individual network point in the cluster HZ are simulated. The sum of the individual allocations of the network points is the sum of the total allocation of the cluster. In figure 9.2 the upper bounds for the different confidence intervals of every hour are plotted for the two days. Hour 74 (8 o'clock in the morning of December 31^{st} 2014) has the highest upper bounds in the coming two days. The capacity that is not used in this hour is seen as the shiftable capacity for this day. See table 9.3 for the capacity that is reserved for the individual network points in the cluster.

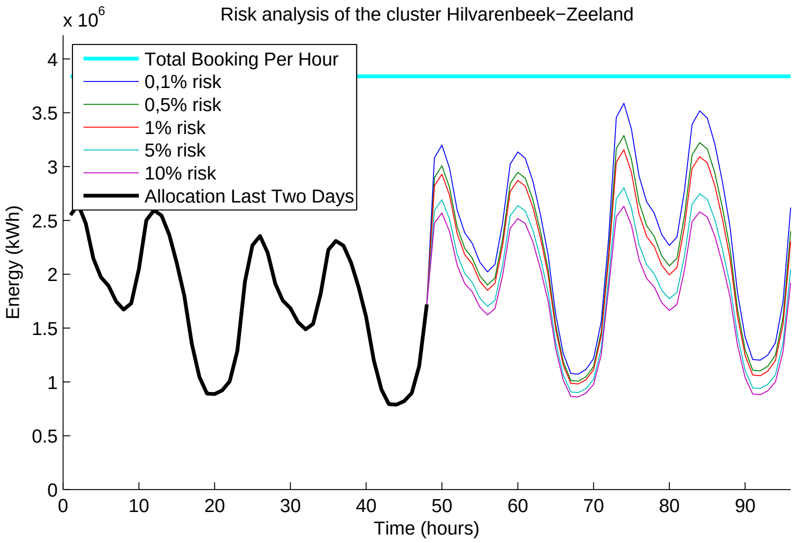

Figure 9.2: Upper bounds of the one-sided confidence intervals of the allocations on December 30^{th} and 31^{st} 2014 for the cluster HZ. Hour 0 is December 28^{th} 2014 6 o'clock in the morning.

Upper bounds for allocations with a certain risk						
Network Point	**Booking**	**Risk**				
		0,1%	**0,5%**	**1%**	**5%**	**10%**
Hoogerheide	278395	279230	240190	223870	188100	171230
Steenbergen	289086	393470	340600	323940	280780	258290
Sprundel	344989	278190	247110	234890	204750	203750
Schoondijke	253199	218890	178760	168000	141840	132900
Breda	455825	453200	388690	367710	307890	280840
Bergen op Zoom	1073211	930760	816300	778760	653830	621070
Dongen	427778	427778	427778	427778	427778	427778
Etten-Leur	527450	496090	440800	416850	355890	329120
Gilze	187896	176220	161130	148730	122280	110140
Total	**3837829**	**3664786**	**3293006**	**3167216**	**2810616**	**2685826**
Shiftable capacity	-	**173043**	**544823**	**670613**	**1027213**	**1152003**
Shiftable cap. (%)	-	**5**	**14**	**17**	**27**	**30**

Table 9.3: Capacity in kWh per hour that should be reserved per individual interior network point by planning with a certain risk for December 31^{st} 2014, based on the information that was available on December 29^{th} 2014. Dongen failed the test for log-normality.

For December 31^{st} $3,8 \cdot 10^6$ kWh per hour is reserved for the cluster HZ. This is the sum of all the reservations of the individual network points. Table 9.3 shows that if GTS takes 0,1% risk in the planning, 5% of the total capacity of cluster HZ can be forfeited. Not every individual network point contributes the same to this number. For example in Steenbergen we would like to reserve more capacity for December 31^{st} than originally is booked. On all the other points is significant room to shift capacity.

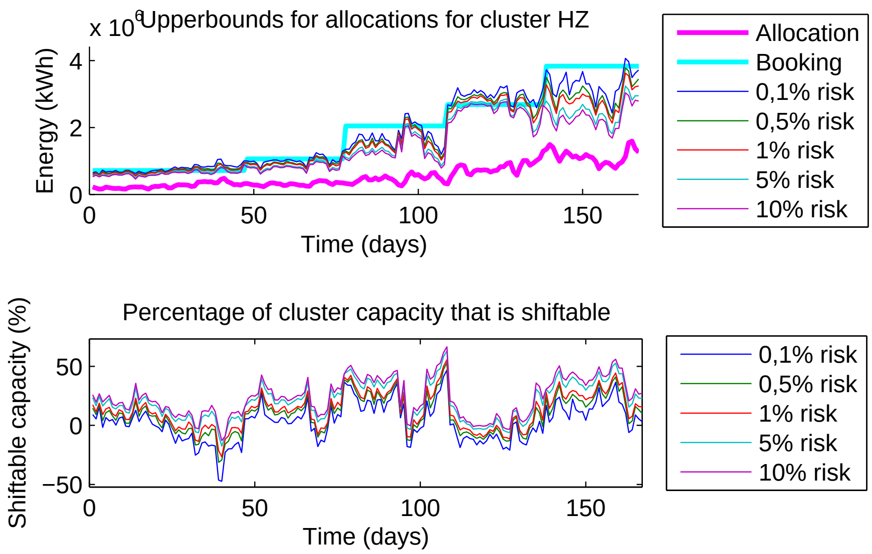

Figure 9.3: In the upper plot the capacity that should be reserved when planning two days ahead with uncertainty over the last six months of 2014. In the lower plot the percentage of the reserved capacity that can be shifted to Hilvarenbeek. The allocation shown, is the average hour allocation over a day.

In figure 9.3 the amount of capacity that would be reserved one day in advance when there would be planned with uncertainty is plotted for the last six months of 2014. Note the difference between the day capacity in the allocations and the maximum hour capacity in the bookings. When there is planned with 10% risk does that mean that there is a 10% probability that in at least one hour of the next day the allocation is higher than $\frac{1}{24}^{th}$ of the total day booking. The average hour allocation over a day, which is plotted in figure 9.3, can in such a case be significantly lower than the estimated upper bound, because of the peaks in demand over a day. Note that for the cluster Hilvarenbeek-Zeeland between day 110 and 130 in figure 9.3 the utilization of the bookings in a lot of the points did not seem log-normally distributed. In that time, on most of the points, no shiftable capacity is determined.

The technical capacity at Hilvarenbeek is around $3 \cdot 10^7$ kWh per hour. When for the 31^{st} of December would have been planned with 0,5% risk, 544823 kWh per hour G-gas capacity could be shifted from the cluster to Hilvarenbeek. Since it is L-gas that flows over the border here, which consists for $\frac{2}{3}^{th}$ out of G-gas, $\frac{3}{2} \cdot 544823 = 817235$ kWh per hour extra capacity on Hilvarenbeek can be offered for December 31^{st}. This is nearly 3% of the technical capacity of Hilvarenbeek. It is required then that this H-gas capacity is also available. In reality more capacity could be shifted to Hilvarenbeek, because of the fact that the cluster lies downstream Hilvarenbeek is ignored. Furthermore is ignored that gas can be buffered in the gas transmission network. There can be anticipated on the peaks in the demand. To determine the consequences of these two effects extensive calculations must be performed. These are not included in this project.

9.3 L-gas border Hilvarenbeek

Via the network point Hilvarenbeek L-gas flows from the Netherlands to Belgium. L-gas is in Belgium and France used in the domestic market. The form of the plot (figure 9.4) of the allocations is familiar to that of a domestic market network point in the Netherlands. However, the plot of the bookings does not look like the reservations of GTS at the interior network points. The bookings are more constant over the year. This is because shippers can get economies of scale when they book for longer periods of time. The firm capacity at Hilvarenbeek was totally sold between 2003 and 2013.

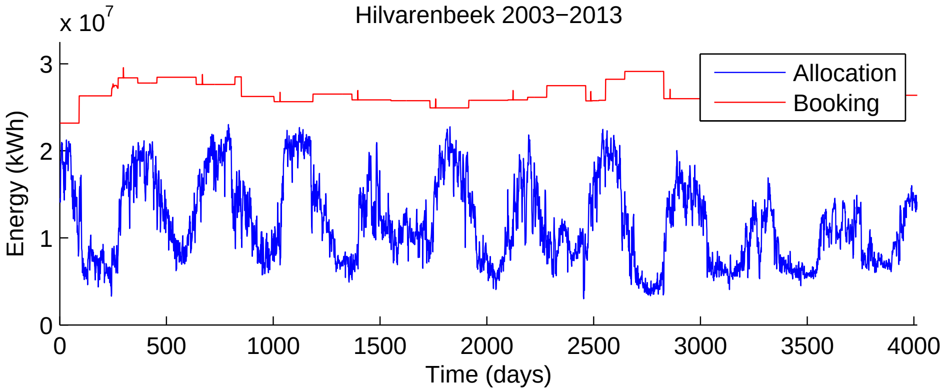

Figure 9.4: Average hour allocation per day and hour bookings on the border point Hilvarenbeek between 2003 and 2013.

The allocations and utilizations of the individual shippers can not be seen independently of each other at Hilvarenbeek. Although the technical capacity from Belgium to the Netherlands is zero, the option exists to book and nominate firm capacity in this direction. This is called back haul capacity. All the gas that is nominated in this direction is subtracted from the nominations in the forward direction. That is the amount that in reality flows in the forward direction. Shippers use this option to send gas to each others portfolio. Therefore the allocations of shippers on Hilvarenbeek are not always independent and the total firm booked and allocated capacity is analyzed.

Projecting the allocations on the bookings, as is done on the interior network points in the domestic market, is not an option, because of the flat pattern of the bookings. The allocations can be corrected for the period of the year. The average utilization of the total booking per period of the year is determined for the period 2003-2010. The average utilization on a day in the year is seen as the average utilization in a period of four weeks around the specific day over the years 2003-2010. For example: the average utilization on May 15^{th} is the average utilization between May 1^{st} and 29^{th} in the years 2003 until 2010. See figure 9.5 for the results.

The average utilization of the bookings on day i is $\hat{A}_i$. The utilizations in the last three years in the same time period as day i are used to determine the parameters for the autoregressive model to simulate the corrected utilizations ($U_i/\hat{A}_i$) for day i in the current year . The year 2013 is seen as the current year to show the procedure. To use the same model as in chapter 6 the mean of the corrected utilization ($\mu_{U/\hat{A}}$) has to be 0. $X_i^{(j)}$ (equation 9.3.1) is the vector with the corrected utilizations of the year j and the two years before, minus the average corrected utilization over the corresponding period that is used to determine the variance of the corrected utilization. The corrected utilization is assumed to be a process with a constant mean and variance. When the behavior of the shippers is the same as in an average year, the mean of the corrected utilizations would be one. In a warm period, when less gas is demanded, the mean is probably lower than

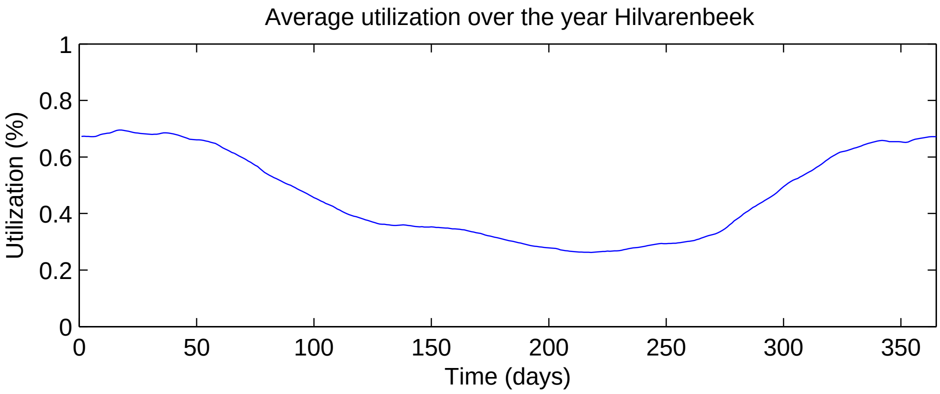

Figure 9.5: Average utilization per period of the year in Hilvarenbeek between 2003 and 2010.

one. In a cold period the mean is higher.

$$X_i^{(j)} = \begin{bmatrix} U_{i-31}^{(j-2)}/\hat{A}_{i-31} - \frac{1}{31}\sum_{l=i-31}^{i-1} U_l^{(j-2)}/\hat{A}_l \\ \vdots \\ U_{i-1}^{(j-2)}/\hat{A}_{i-1} - \frac{1}{31}\sum_{l=i-31}^{i-1} U_l^{(j-2)}/\hat{A}_l \\ U_{i-31}^{(j-1)}/\hat{A}_{i-31} - \frac{1}{31}\sum_{l=i-31}^{i-1} U_l^{(j-1)}/\hat{A}_l \\ \vdots \\ U_{i-1}^{(j-1)}/\hat{A}_{i-1} - \frac{1}{31}\sum_{l=i-31}^{i-1} U_l^{(j-1)}/\hat{A}_l \\ U_{i-31}^{(j)}/\hat{A}_{i-31} - \frac{1}{31}\sum_{l=i-31}^{i-1} U_l^{(j)}/\hat{A}_l \\ \vdots \\ U_{i-1}^{(j)}/\hat{A}_{i-1} - \frac{1}{31}\sum_{l=i-31}^{i-1} U_l^{(j)}/\hat{A}_l \end{bmatrix}, \tag{9.3.1}$$

$$\hat{\sigma}^2_{X_i^{(j)}} = \sum_{k=1}^{93} (X_{i,k}^{(j)})^2. \tag{9.3.2}$$

The covariance matrix for the border exit Hilvarenbeek only consists of its own variance. The relations with other network points are not taken into account. In the vector X (equation 9.3.1) is for every month of every year the average corrected utilization in a month subtracted from the corrected utilizations on the individual days, instead of subtracting the average corrected utilization of the three years from all the utilization of the days that are taken into account. This is to have a more fair comparison of the corrected utilizations in the same periods over the three years.

For the year 2013 is checked if the corrected utilization of all the shippers could be normally or log-normally distributed. The Chi-square goodness-of-fit test rejects normality for 20% of the days in 2013 spread over the whole year, when a significance level of 0,05 is used. It rejects log-normality more often with the same significance level. Therefore, it is assumed that the corrected utilization is normally distributed over the whole year to show the procedure. An order 2 autoregressive model is used to simulate the corrected utilizations. The parameters are determined as in chapter 6. A one-sided confidence interval is created to find the upper bound of the corrected utilization and consequently the allocations in the next two days.

The allocation that is simulated is the average hour allocation per day. Although there is not a clear day and night pattern, the allocation is usually not constant over a day. In the period 2003-2013 in 10% of the days the maximum hour allocation was 4,43% or less of the total day allocation, on 20% the days 4,5% or less, etc. . All the numbers can be seen in table 9.4. These numbers are used to estimate an upper bound for the maximum hour allocation out of the simulation of the average hour allocation.

Percentage of days	Max max hour allocation (% day all.)	Percentage of days	Max max hour allocation (% day all.)
10	4,43	60	4,75
20	4,5	70	4,83
30	4,55	80	4,95
40	4,61	90	5,15
50	4,68	100	8,03

Table 9.4: Percentage of the days between 2003 and 2012 where the maximum hour allocation on a day is lower or equal than a certain percentage of the total day allocation.

9.3.1 GTS oversell capacity

GTS determines the oversell capacity directly from the difference between the realizations and the technical capacity. In table 9.5 can be seen how much oversell capacity GTS offers. The OSBB is the percentage of the difference between the technical capacity (TBG) and the maximum hour allocation in the period in the last column, that is offered on the different auctions as oversell capacity. There is a maximum and minimum percentage of the technical capacity that is offered by GTS as oversell capacity.

Auction	OSBB (%)	Min TBG (%)	Max TBG (%)	Period
Day	100	5	20	1 day (yesterday)
Month	50	5	20	2 times 1 month (last and second last year)
Quarter	30	0	15	2 times 3 months (last and second last year)
Year	15	0	10	24 months

Table 9.5: For the mentioned auctions is the maximum hour allocation determined in the period of the last column. The percentage OSBB of the difference between the maximum hour allocation in this period and the technical capacity is offered as oversell capacity at the auction. However, GTS will never offer more than the percentage Max TBG of the technical capacity and never less than the percentage Min TBG of the technical capacity. [4]

9.3.2 Oversell capacity on the one day auction in 2013

The oversell capacity that GTS would have offered on the one day auction in 2013 can be compared to the oversell capacity that would have been offered after an analysis by the autoregressive model. The results of this for the last six months of 2013 can be seen in figure 9.6.

A one-sided 99,9% confidence interval for the maximum hour allocation in the next two days is determined for all the days of the last six months of 2013. The difference between the technical capacity and the upper bound of this confidence interval, is the capacity that could be offered for overselling with 0,1% risk. A risk of 0,1% in this case means that there is a probability of 0,1% that the shippers that have booked firm capacity are going to use more than what is reserved for them (the green line in the upper plot of figure 9.6) on at least one hour in the next day. The upper bound for the maximum hour allocation on the next day that GTS uses is determined by table 9.5.

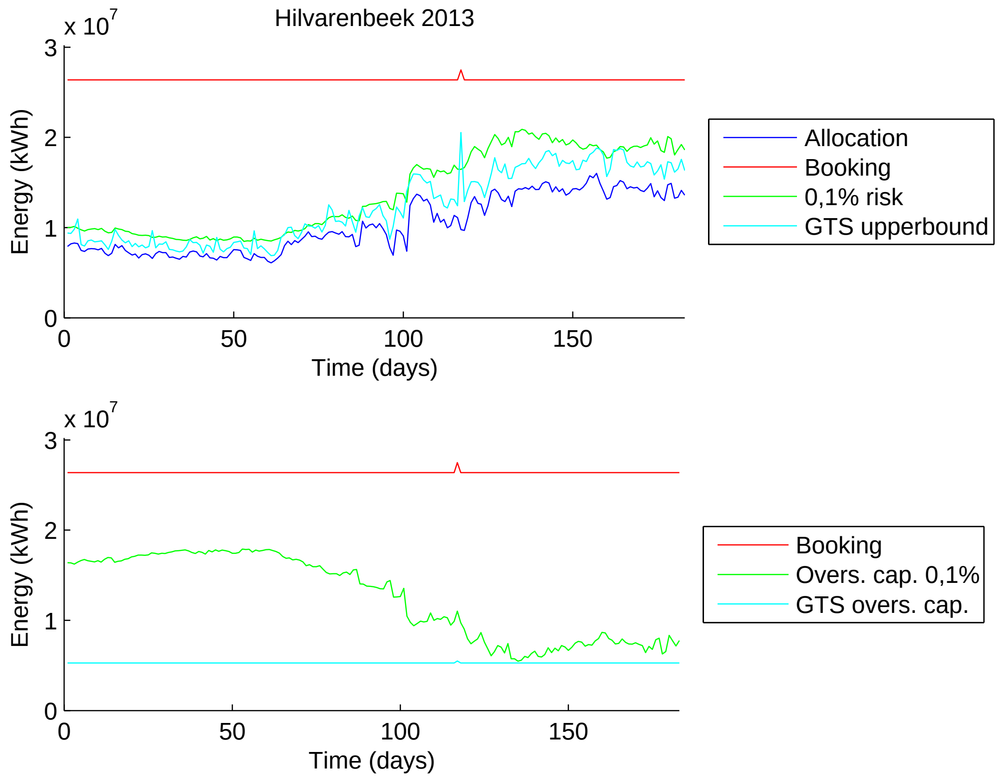

Figure 9.6: In the upper figure are the average hour allocation per day and the bookings plotted for the last six months of 2013. Furthermore is the upper bound for the allocation of the day after plotted as determined by GTS via table 9.5 and a 0,1% risk upper bound for the allocation as determined via the model. In the lower plot can be seen how much oversell capacity GTS would offer and how much would be offered with 0,1% risk by the model.

In the last six months of 2013 there was not a single hour with the allocation surpassing 80% of the bookings at Hilvarenbeek. Since only a maximum of 20% of the technical capacity (and in this case the booking) is offered as oversell capacity, the oversell capacity on the one day ahead auction of the last six months of 2013 was always 20% of the bookings. The 0,1% risk line in the upper plot of figure 9.6 lies for nearly all the days above the upper bound for the total allocations that GTS takes, but also stays under de 80% of the bookings mark all the time. The cyan colored line in the upper plot of figure 9.6 is both the upper bound of the total allocation that GTS takes for the next day and the maximum hour allocation on the day before. However, the cyan line is not as smooth as the dark blue line. An explanation for the odd looking peaks in the cyan line could not be found.

A major difference between the oversell capacity that is determined on the export point itself and on the domestic market cluster Hilvarenbeek-Zeeland is that shippers that book on the border exit itself could, for whatever reason, differ from their behavior in the past. This is (nearly) impossible in the domestic market. A shipper could for example choose to fill a gas storage in Belgium with L-gas and therefore use its total booking. This is impossible in the domestic market, because it is fully known what is behind these points.

It can be seen from figure 9.6 that the upper bound of the maximum hour allocation of tomorrow that is predicted is nearly always less than the upper bound GTS uses. This is because GTS sees the maximum hour allocation of today as an upper bound of the maximum hour allocation of tomorrow. Except of some days with odd looking peaks in the maximum hour allocation, GTS would always offer more oversell capacity than could be done with 0,1% risk, if GTS did not have the policy to offer a maximum of 20% of the technical capacity.

9.4 H-gas border 's-Gravenvoeren

's-Gravenvoeren is an H-gas exit border station in the south of the province Limburg. A total of 23 shippers operate on this network point, which together booked a total of $16,8 \cdot 10^6$ kWh per hour exit and $1,52 \cdot 10^6$ kWh per hour entry capacity on December 31^{st} 2014. Although gas can only flow in the forward direction, the backward capacity is sold firm. The forward direction was totally sold out on this day. 22 of the 23 shippers allocated in more than 60% of the consecutive hours the same allocation for the period of their current booking. For these shippers it is assumed that they intentionally set their allocation for a period of time and that their behavior can be approximated with a discrete time Markov chain. For every individual shipper such a model is created. The sum of the individual results is the result of the network point as total. The single shipper that did not allocate the same amount in that much consecutive hours (0,4%) is neglected for the analysis. On its booking no oversell capacity will be offered.

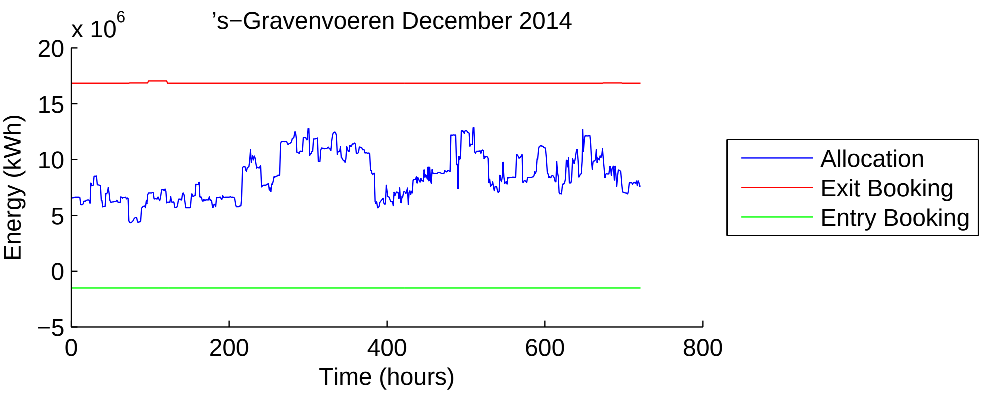

Figure 9.7: The total bookings and hour allocations of shippers on 's-Gravenvoeren in December 2014.

Allocation 's-Gravenvoeren December 31^{st} 2014		
Confidence interval (%)	**Max hour allocation (10^6 kWh)**	**Oversell cap. exit (10^6 kWh)**
99,9	14,9	1,9
99	14,0	2,8
95	13,2	3,6
90	12,9	3,9

Table 9.6: Upperbounds for the maximum hour allocation on December 31^{st} 2014 for the network point 's-Gravenvoeren with corresponding oversell capacity for the one day auction that could be offered. Shippers booked a total of 16,8 $\cdot 10^6$ kWh per hour capacity on this day.

For every shipper the range between its exit and entry booking is divided into 50 smaller ranges. All the allocations on hours of days with the same bookings as December 31^{th} are rounded to the closest border of two of the smaller ranges, see figure 8.4 for an example. The borders of the ranges are the states of the Markov chain for the individual shipper. The realizations of the total period with the same booking as on December 31^{st} are used as transition probabilities between the states. The allocation for December 31^{st} 2014 is simulated. One-sided confidence intervals for the hour with the most allocation are created. The difference between the upper bound of the confidence interval for the hour with the most allocation and the bookings in exit direction can be offered for overselling in the forward direction. The results for the one day ahead auction can be seen in table 9.6. It is assumed that the behavior of a shipper is known when it has at least 7 preceding days with the same booking.

From the table 9.6 can be seen that $1,9 \cdot 10^6$ kWh energy per hour oversell capacity can be offered at the network point in the exit direction on December 31^{st} 2014, with 0,1% risk that in one or more hours of the day the allocations surpass the reserved amount of $14,9 \cdot 10^6$ kWh per hour. That means that 11% of the technical capacity can be offered with that risk. Obviously, by taking more risk, more capacity can be offered. Because the maximum hour allocation on December 29^{th} did not exceed 80% of the technical capacity, GTS would offer 20% of the technical capacity for overselling on December 31^{st} (see table 9.5 for the details).

The same analysis as for December 31^{st} 2014 is done for every day of the last six months of 2014. The results are plotted in figure 9.8. The green line, that represents the upper bound of the 99% confidence interval for the maximum hour allocation, fluctuates heavily sometimes. The main reason for that, besides the fluctuation in allocation, is the appearance of new bookings. For the shippers that change their booking is after 7 days assumed to be known what their behavior is. Up to that time no oversell capacity is offered on their bookings. Their total individual booking is then reserved for them for the next day. The same holds for day bookings. When a shipper does a day booking on top of its month, quarter or year booking, its behavior is unknown. Their total bookings, the original bookings plus the day bookings, will be reserved for them. It is probably not realistic to assume that when a shipper does a day booking on top of its original booking, its behavior over the original booking is the same as without an additional booking, because the original booking does not suffice anymore. The oversell capacity that can be offered with 1% risk in forward direction is the difference between the red and the green line in the upper plot of figure 9.8.

In almost none of the hours of the last six months of 2014 was the total booking of all the shippers at 's-Gravenvoeren higher than 80% of the the technical capacity. GTS would have offered 20% of the technical capacity for overselling on nearly all days, which is $3,36 \cdot 10^6$ kWh per hour. By taking 1% risk, the capacity that can be offered for overselling fluctuates between 0% and 40% of the technical capacity. At the time of around 800 hours some new month bookings are done. This can immediately be seen the oversell capacity, which is almost zero at that point and a week afterwards.

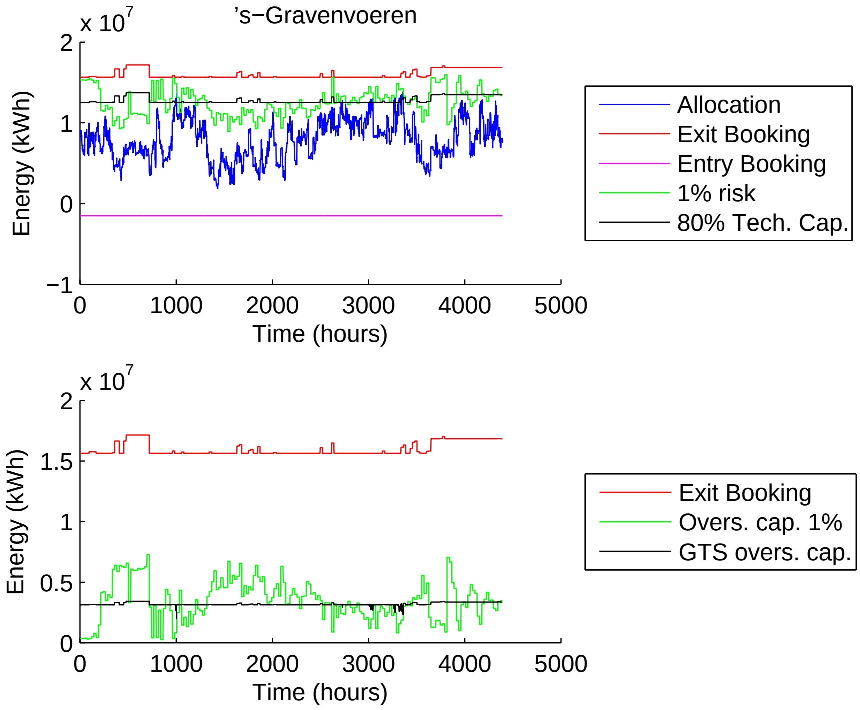

Figure 9.8: The total bookings and allocations of all the shippers in the last six months of 2014, together with a the upper bound of a one-sided 99% confidence interval for the maximum hour allocation on a day in the upper plot. In the lower plot the corresponding oversell capacity that can be offered at the one day ahead auction by GTS and by taking 1% risk.

10 Conclusions, recommendations and remarks

The end of this report is a brief retrospect of what was achieved and where there is room for further improvement. First the conclusions are presented and then some recommendations are made.

10.1 Conclusions

A number of conclusions can be made on the different topics that were researched in this report.

- When the entry and exit in a time slot at every network point is known, the flows through the network can be determined by use of a Newton model.

If the active elements (such as compressor stations) in a gas transmission network are neglected, the flows through the network can be obtained by running a Newton model that solves all the equations for node balance and ring balance in the network. To find the corresponding pressures at all the network points, the pressure at one point in the network must be known, or chosen. Knowledge of the pressure in one point and knowledge of the flows trough the network, fixes the pressures in the rest of the network.

- Anticipation and reaction of shippers on the balancing regime can be modeled by looking at the reaction of the shippers to their individual position.

Shippers mainly care about their individual position, which they want to have close to zero. When a shipper is a causer and too much out of balance, it will do an operational intervention to adjust its individual position. Helpers will not intervene operationally, they will sell a part of their position on the TTF when they are too much out of balance.

- The allocation on interior domestic market network points can be predicted by making use of the fact that the utilization of the bookings is not seasonally dependent and is usually log-normally distributed.

The average day utilization on most of the interior network points has a log-normal distribution for the majority of the time. This, plus the fact that there is a high correlation between interior domestic market points, can be used to predict the allocation at an interior domestic market network point or a cluster of these points.

- The allocation behavior of individual shippers at H-gas border network points can be modeled by a discrete time Markov chain.

Shippers on H-gas border network points usually allocate the same amount for consecutive hours, before switching to another level. The states of the Markov chain are the levels of operation per hour, with the transition probabilities obtained from the switching behavior in the past.

- The capacity that can be offered for overselling on the L-gas border of Hilvarenbeek based on predictions is approximately the same as when based on realizations.

The oversell capacity that would be offered by GTS based on realizations on the border exit Hilvarenbeek itself is slightly higher on the one day ahead auction than what would have been offered based on predictions (if the policy of GTS that only a maximum of 20% of the technical capacity is offered at a network point is ignored). However, the possibility of shifting capacity from somewhere else in the network to Hilvarenbeek is not taken into account by GTS. That compensates for the extra offered capacity.

- The capacity that can be offered for overselling based on predictions on the H-gas border 's-Gravenvoeren fluctuates around the amount GTS offers.

The total allocation at the network point 's-Gravenvoeren rarely surpasses the mark of 80% of the technical capacity. Therefore GTS usually offers 20% of the technical capacity as oversell capacity on the one day ahead auction. At the moment that a lot of shippers change their booking, the system for overselling based on predictions offers little oversell capacity, because little is known about the behavior of the shippers then. At the time when more is known about the behavior of the individual shippers, usually more than 20% of the technical capacity would be offered as oversell capacity.

10.2 Recommendations

A number of recommendations can be done based on the research in this report.

- The behavior of the allocation at other types of network points should be analyzed.

To simulate the gas transmission network more knowledge is needed for the points that have not been analyzed here. Examples are the gas production in Groningen and gas storages.

- Additional research could be devoted to the behavior of a virtual entry point in the network.

All the G-gas that flows to Hilvarenbeek and further downstream comes from Groningen. The behavior of a virtual entry point just before Hilvarenbeek can be analyzed. When the behavior of this virtual entry point is known, a more realistic simulation of a part of the G-gas network that is downstream of Hilvarenbeek can be made.

- Research should be done on the parameters in the model of the balancing regime.

Approximations for the parameters used in the model of the balancing regime for the behaviors were not yet found in this report. These will be useful in a simulation of (a part of) the network.

- In the model for the balancing regime it is neglected that shippers are more willing to balance when the system is a lot out of balance. This could be researched.

For the behavior of the shippers it is assumed that they only take their individual position and the sign of the network position into account. Shippers should be more willing to balance when they are a causer and the system is much out of balance. This has not been taken into account yet.

- To determine an upper bound for the allocations at Hilvarenbeek, an autoregressive model with a seasonal component would probably give more accurate results.

In the allocations of Hilvarenbeek the temperature plays an important role. An adjustment for seasonality to the model of the domestic market was used to determine an upper bound for the allocations at this border point. However, inserting a real seasonal component would describe the allocation probably better.

11 References

[1] Ralf Gollmer Armin Fugenschuh, Bjorn Geissler et al. Mathematical optimation for challenging network planning problems in unbundled liberalized gas markets. *Energy Systems*, pages 449–473, 2013. doi:10.1007/s12667-013-0099-8.

[2] Jungwei Lu C. Bennet, R. Stewart. Novel autoregressive basis structure model for short-term forecasting of customer electricity demand. *TENCON Spring Conference, 2013 IEEE*, 1: 62–67, 2013. doi:10.1109/TENCONSpring.2013.6584418.

[3] Zofia Lukszo Gerard P.J. Dijkema Catalin Bucura, Alfredas Chmieliauskas. Modelling the new balancing regime of the natural gas market in the netherlands. *Networking, Sensing and Control (ICNSC), 2011 IEEE International Conference on*, pages 68–73, 13 April 2011. doi:10.1109/ICNSC.2011.5874925.

[4] Altinus Meijer Frauke Huning, Jan Albert Laverman. *Methodiek vaststellen jaar- en kwartaalveilingen.* Memorandum, Gasunie Transport Services b.v., 2014.

[5] Will Gersch. Estimation of the autoregressive parameters of a mixed autoregessive moving-average time series. *Proceedings of the IEEE*, 65:979–980, June 1977. doi:10.1109/TAC.1970.1099560.

[6] Sebnem Dilaveroglu Halit Uster. Optimization for design and operation of natural gas transmission networks. *Applied Energy*, 133:56–69, 2014. doi:10.1016/j.apenergy.2014.06.042.

[7] Nil Aras Haydar Aras. Forecasting residential natural gas demand. *Energy sources*, 26: 464–472, 2004. doi:10.1080/00908310490429740.

[8] Ansgar Steinkamp Herwig Friedl, Radoslava Mirkov. Modelling and forecasting gas flow on exits of gas transmission networks. *International Statistical Review*, 80:24–39, 2012. doi:10.1111/j.1751-5823.2011.00171.x.

[9] Nicholas J. Higham. Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra and its Applications*, 103:103–118, 1988. doi:10.1016/0024-3795(88)90223-6.

[10] Jan Albert Laverman Jan Willem van Dijk, Jarig Steringa et al. *Balans tussen gevraagde en beschikbare transportcapaciteit.* Memorandum, Gasunie Transport Services b.v., 2013.

[11] Averill M. Law. *Simulation Modeling and Analysis.* 4th edition, 2007. ISBN 978-0-07-298843-7.

[12] Charles F. Van Loan and Gene H. Golub. *Matrix Computations.* 2006. ISBN 978-0801854149.

[13] Nederlandse Gasunie N.V. *TTF.* URL `http://www.gasunietransportservices.nl/en/products-services/entry-exitcapacity/ttf`.

[14] V. Kopustinskas P. Praks, S. Contini. Monte carlo and fault tree aproaches in reliability applications of gas transmission network. *Electrical Power Engineering (EPE), Proceedings of the 2014 15th International Scientific Conference on*, 1:69–74, 12-14 May 2014. doi:10.1109/EPE.2014.6839421.

[15] Dick van Huizen. *Identifying the most severe transport situations in a gas transmission network.* Internship Thesis, University of Twente, 2014.