Zuse Institute Berlin

Takustr. 7
14195 Berlin
Germany

MILENA PETKOVIC, YING CHEN, INKEN GAMRATH, UWE GOTZES, NATALIA SELINI HADJIDIMITRIOU, JANINA ZITTEL, THORSTEN KOCH

A Hybrid Approach for High Precision Prediction of Gas Flows

ZIB Report 19-26 (Dec 2019)

Zuse Institute Berlin
Takustr. 7
14195 Berlin
Germany

Telephone: +49 30-84185-0
Telefax: +49 30-84185-125

E-mail: bibliothek@zib.de
URL: http://www.zib.de

ZIB-Report (Print) ISSN 1438-0064
ZIB-Report (Internet) ISSN 2192-7782

A Hybrid Approach for High Precision Prediction of Gas Flows

Milena Petkovic[a,], Ying Chen[b], Inken Gamrath[a], Uwe Gotzes[c], Natalia Selini Hadjidimitriou[d], Janina Zittel[a], Thorsten Koch[a,e]

[a] *Zuse Institute Berlin*
[b] *National University of Singapore*
[c] *Open Grid Europe*
[d] *University of Modena and Reggio Emilia*
[e] *Technische Universtität Berlin*

Abstract

About 23% of the German energy demand is supplied by natural gas. Additionally, for about twice this amount Germany serves as a transit country. Thereby, the German network represents a central hub in the European natural gas transport network. The transport infrastructure is operated by so-called *transmissions system operators* or TSOs. The number one priority of the TSOs is to ensure security of supply. However, the TSOs have only very limited knowledge of the intentions and planned actions of the shippers (traders).

Open Grid Europe (OGE), one of Germany's largest TSO, operates a high-pressure transport network of about 12.000 km length. With the introduction of peak-load gas power stations, being able to predict in- and out-flow of the network is of great importance to ensure the necessary flexibility and security of supply for the German Energy Transition ("Energiewende").

In this paper we introduce a new hybrid forecast method applied to gas flows at the boundary nodes of a transport network. The new method employs optimized feature minimization and selection. We use a combination of an FAR, LSTM DNN and mathematical programming to achieve robust high quality forecasts on real world data for different types of network nodes.

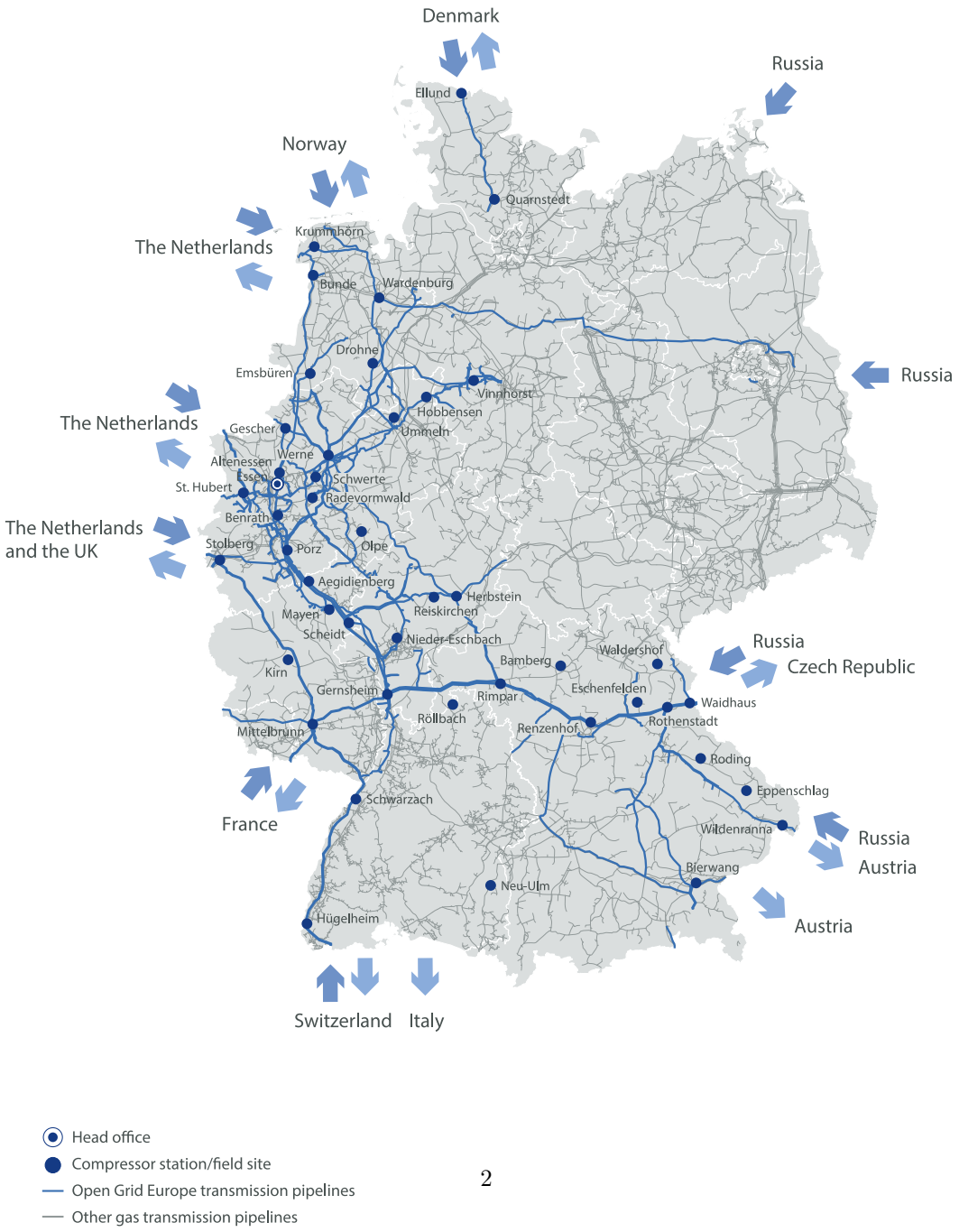
Keywords: Gas Forecast, Time series, Hybrid Method, FAR, LSTM, Mathematical Optimisation

1. Introduction

About 23% of the German (and European) energy demand is met by natural gas. Additionally, for about twice the amount Germany serves as a transit country. Thereby, the German network represents a central hub in the European natural gas transport network. In light of the German Energy transition ("Energiewende") with an increasing share of renewable energy sources as well as the envisioned international transition towards substantially less fossil fuels and related greenhouse gas emissions, the importance of natural gas will increase even more. A critical task of gas power plants is to deliver electricity in peak load situations, where electricity from renewable energy sources is not sufficient to cope with the demands. From the gas network point of view this leads to huge gas demands on very short notice.

Email address: `petkovic@zib.de` (Milena Petkovic)

Figure 1: Map of the gas transmission network operated by Open Grid Europe



The gas network is operated by so-called *transmissions system operators* or TSOs. Their duty is to operate the long distance transport network and to ensure that all demands are met. Until 2005, gas transport and supply in Europe was provided by a handful of companies, owning and operating the natural gas transmission system to do so. To establish a European gas market, in 2005, the European Union legislated that gas trading and transport had to be done by mutual independent companies to ensure discriminatory-free access to the transport network for all traders. Before this, TSOs were part of an integrated organization and could plan the network operation and expansion together with the traders. Since then, they are independent and need to plan under the uncertainty regarding the gas-flow situations resulting from trading.

Although on the contractual level all gas transports of a market area have to be balanced, this needs only to be achieved on average over time. Some outflow might actually only be balanced by an inflow at a later time.

Despite of these challenges for the TSOs they need to meet all transport demands. It is the obligation of the TSO to monitor the situation, foresee possible shortages and react accordingly to ensure safety of supply. Since changes in gas networks happen rather slowly it is therefore extremely important to have accurate forecasts on the demands and supply of the network to be able to react on time.

We collaborated with one of the biggest German TSOs, operating a gas network with pipe length of about 12,000 km in total (see Figure 1), to improve their hourly forecasts for demand and supply. We aim to:

- Predict as *precisely* as possible the *average hourly gas flows* for the upcoming *gas day*, i.e., from 6am to 6am, just before the start of the gas day (at about 5:59am);
- The prediction needs to be *appropriate for all different types of nodes* ranging from connections to other networks or countries to industrial users or municipal consumers, leading to very diverse data characteristics;

To reach these goals, we investigated real data from the transport network operated by Open Grid Europe. We propose a powerful and robust hybrid forecast model that benefits from the combination of state of the art forecasting approaches and optimisation, leading to improved forecast accuracy. We interpreted the most important features that our model automatically selects.

In the following, the next subsections present nomenclature and an overview of related work. Section 2 describes the data we used in this study. Section 3 gives details on the proposed models. Section 4 describes the evaluation methodology and the evaluation of computational experiments. Finally, we draw some conclusions.

1.1. Nomenclature

Open Grid Europe GmbH	OGE
Municipal Power Stations	MUN
Power Stations and Industry	IND
Storage	STO
Transfer Points to other Networks	NET
Mathematical Programming	MP
Linear Program	LP
Mixed Integer Linear Programming	MILP
Functional Autoregressive	FAR
Long Short Term Memory	LSTM
Hybrid model	HYB
Heating Day Degree	HDD
Baseline forecast(persistance)	BAS

1.2. Related work

Models on natural gas demand forecasting are mainly focused on long term issues. There are quite some publications regarding electricity demand forecasting, (see, e.g. [41], [23], [1], [25]) but electricity behaves very differently from gas. A survey on models to predict natural gas consumption published between 1949 and 2010 is presented by [36] who evidences that only a few works are focused on hourly gas flow prediction. A more recent survey ([45]) considers 187 papers published between 2002 and 2017. The authors point out that the majority of works provide daily predictions and recognize that neural networks are the most used models. The authors also show that, on the considered period, most of the works were performed at an aggregated level (i.e. country or city) and only three papers proposed models to forecast the hourly gas consumption. In [43], two neural networks were tested to forecast natural gas consumption based on historical data and environmental variables. The authors found a better prediction accuracy when using the multi-layer perceptron compared to the radial basis function. In [42], a model similar to radial basis neural network was proposed to predict gas consumption in a distribution system. In this work input variables were selected using a genetic algorithm. Residential hourly gas consumption was predicted with neural networks by [10]. In this work, the heating degree-hour method which considers the gap between outdoor and indoor temperature was considered. The best hyper-parameters configuration consisted of 29 neurons, a feed-forward backpropagation algorithm and tangent, sigmoid and linear functions for the input, hidden and output layers respectively. Similarly, [39] proposed neural networks to forecast residential natural gas demand. The proposed network consisted of a multi-layer perceptron with one hidden layer. The input features included calendar (i.e. month, day of the month, day of the week, hour) and weather (temperature) information. The authors found that average prediction error was higher during the winter months because gas flow was higher. More recently, [15] compared several machine learning models to predict residential natural gas hourly demand and found that recurrent neural network and linear regression were the most accurate models. The prediction results of monthly gas consumption of residential buildings using Extreme Learning Machine (ELM), artificial neural networks (ANNs) and genetic programming (GP) was presented by [16]. The ELM is characterized by higher training speed compared to backpropagation and it was found to perform better, in terms of RMSE, compared to the other two techniques. In [18] the authors set up a two stages methodology to predict daily gas consumption of utility companies.

In the first stage, two NNs are run in parallel to produce daily forecasts; in the second stage, a nonlinear transformation of some features of the input vector is performed. The combination of the two stages is based on several methods such as average forecast, recursive least squares, etc. The results show that the mix between the two forecasters has higher accuracy although the combination of the two models increases the complexity. Overall, these works show that the consumer profile is very important when forecasting gas flow. In this regard, [32] identified seventeen groups profiles, based on their historical consumption and predicted daily gas demand. The overall prediction was obtained from the combination of the single predictions.
The backpropagation algorithm optimized with a genetic algorithm was implemented by [47] to increase the training speed and to achieve a global minimum. The authors predict next day gas loads based on temperature and weather conditions. Furthermore, the authors tested the algorithm on a three years real dataset recorded in Shanghai to predict one month and a half gas load. Similarly, [48] propose a recurrent neural network to predict daily gas flow. The Output-Input-Hidden Feedback-Elman neural network takes into account, not only the hidden nodes feedbacks but also considers the output nodes feedbacks. The results improved compared to these obtained with standard Elman network. However, the authors recognize that further research is needed to forecast gas demand during holidays. In [3], an adaptive network-based fuzzy inference system (ANFIS) consisting of a neural network integrated with fuzzy logic was proposed to forecast short term natural gas demand. The main advantage of this model was its ability to handle uncertainty, noise and non-linearity in the data and, compared to standard neural network models, provided more accurate results. Wavelet transform has been deployed by [38] to decompose the hourly gas demand time series and Bi-LSTM and LSTM are optimized using genetic algorithm. The model was applied to winter data on which it has shown good prediction accuracy. Several static and adaptive models have been tested by [30] for short-term gas consumption forecast (random-walk, temperature correlation model, linear regression model, ARX, adaptive (recursive) linear auto-regressive model (RARX), neural network (NN), Recurrent NN, Support Vector Regression). They found that the best performance was obtained by the RARX of order 3. Furthermore, they found that nonlinear models such as neural networks and support vector machines had a lower generalization capacity compared to linear models. Finally, they concluded that the adaptive models overall performed better than static models.

The traditional approaches are regression and econometric models. In this regard, the performance of non linear mixed effects, ARIMAX and ARX models to predict gas consumption of 62 residential and small commercial customers was assessed by [8]. The authors forecast daily consumption of an entire month based on the previous 18 months. The time series included zero flows and missing data which were excluded for the training process. The prediction performance was similar in terms of daily mean absolute error which was close to zero for all the tested models. Thus, the authors propose to combine multiple models although they recognize that this might be a difficult task because of increased computational complexity. Multiple linear regression has been proposed by [34] who predicted annual gas consumption based on socio-economic variables (GDP and inflation in the case of Turkey) that have been selected based on their statistical significance. Based on the forecast, the authors propose alternative energy policies.

A hybrid model formed by a grey model and an autoregressive integrated moving average model has been proposed by [46] to predict monthly shale gas production. The authors conclude that the results of the combined model are more accurate than the single linear and non linear models.

In [26], Multivariate Adaptive and Conic Multivariate Adaptive Regression Splines were proposed to predict residential daily gas demand. The two models provided better results in terms of

prediction errors (MAE and RMSE) compared to these obtained with linear regression and neural networks. In [35], the nonlinear characteristics of the natural gas consumption is modeled with several Grey models that are compared to predict the yearly natural gas consumption in China. Non linear programming and genetic algorithm have been proposed by [12] to predict natural gas consumption in the residential and commercial sectors on a yearly basis. Similarly, [17] proposed the breeder hybrid algorithm which consists of three steps for natural gas flow demand forecast. In the first stage, the coefficients of a non linear regression model are estimated. Successively, the estimates are improved using genetic algorithm. Finally, the optimized coefficients are deployed as initial solutions for the simulated annealing. Nearest neighbor and local regression were proposed by [5] to predict gas flow in a small gas network with a 15 minutes resolution. The authors evidence the importance of environmental variables such as the temperature. Their method allowed to detect anomalies and the consumption patterns based on one year historical data. In the literature, there are also combinations of several methods to predict one day-head natural gas consumption. In [27], the time series was decomposed into low-frequency and high frequency components using Wavelet transform. In a second step, the genetic algorithm and Adaptive Neuro-Fuzzy Inference System were deployed to predict each of the decomposed time series. The output was finally fed into a feed-forward neural network to refine the prediction. The research was focused on different types of natural gas distribution points. The authors obtained better prediction results using the data of distribution points located near the city center. Neural networks have been also compared to the performance of autoregressive models. In [40], for instance, short term natural gas consumption in Turkey was predicted using SARIMAX model and Neural Networks (Multilayer and Radial Basis) and multivariate regression. They found that SARIMAX had better prediction performance. The temperature correlation model, proposed by [37], was compared with several configurations of ARX, stepwise regression, Support Vector Regression and neural network. The author found that SVR and NN performed better on the training set, while high order ARX model performed better on the test set. Support Vector Regression has been deployed with false neighbours filtered approach to predict short term natural gas consumption ([49]). The local predictor was based on the nearest neighbour approach so that the Euclidean distance between the training and test data and the neighbour filter was applied to determine the validity of the predicted values based on the exponential separation rate. The authors obtained better performance prediction compared to ARIMA, neural networks and Support Vector Regression.

Overall, the analyzed literature shows that there are few works that are focused on the comparison between methods to predict hourly gas flow of different types of nodes in a gas network or combining the advantages of different forecasting methods to a hybrid model for hourly gas flows. Therefore, we propose a hybrid model based on optimisation and machine learning and compare its results to four different models to predict hourly gas flow. To address the heterogeneity of the time series for the different node types we compare results obtained for four different types of nodes.

2. Data

We consider high-resolution natural gas inflows and outflows in the high- pressure gas pipeline network operated by Open Grid Europe GmbH (OGE).

The gas transmission network has more then 1000 boundary nodes which can be classified into four different groups:

Table 1: Number of nodes and percentage of flow

Type	#	flow(%)
Network	34	72.95
Municipal	726	16.68
Industry	234	4.26
Storage	14	6.11

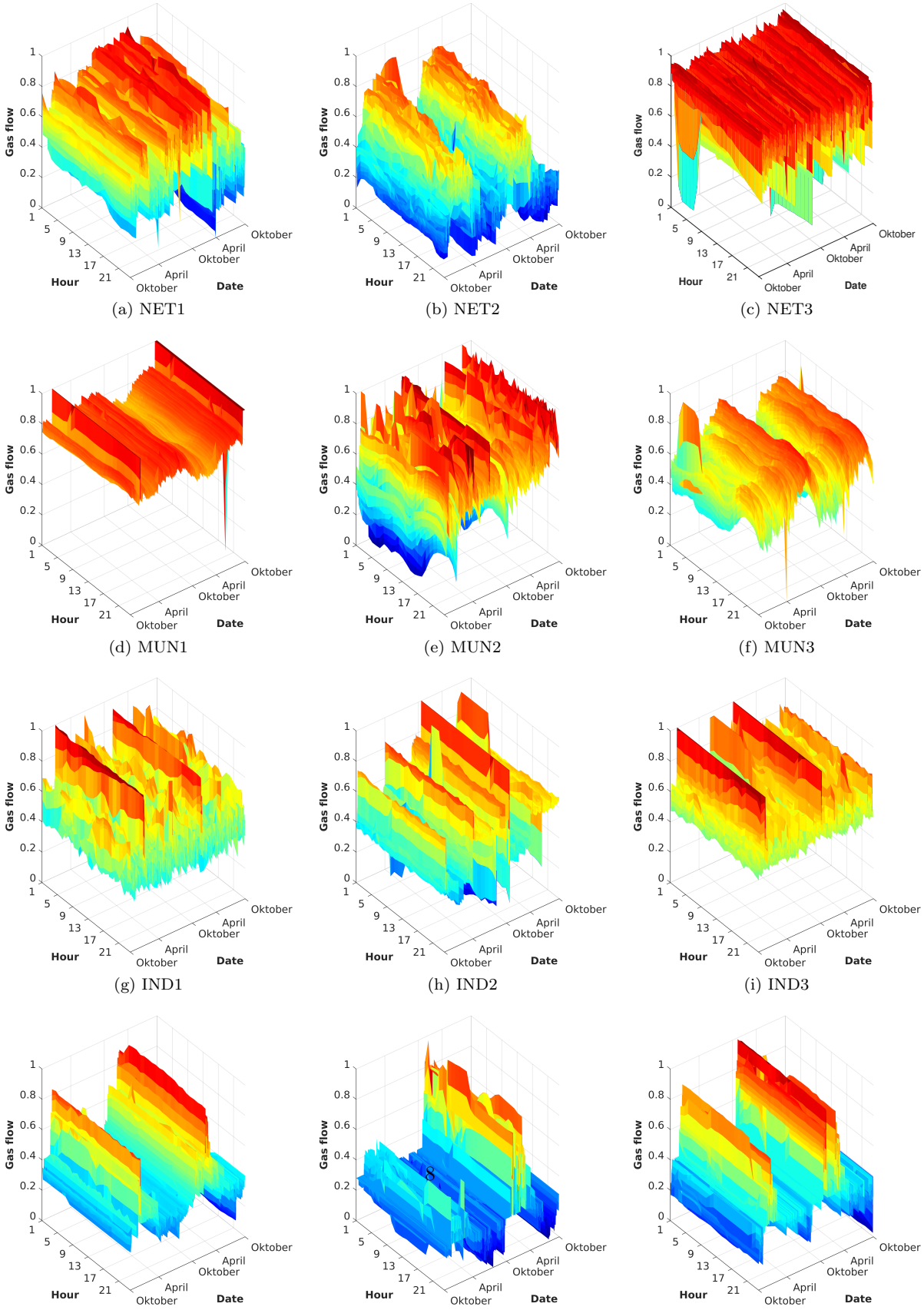
- Network Transfer Points (NET) are large nodes with natural gas imported and exported to other networks mostly outside Germany. These can be entries and/or exits.
- Municipal Utility nodes (labeled MUN) serve residential and small commercial constituents and are always only exits. They are often temperature dependent, exhibit daily and seasonal patterns, and simultaneously are influenced by weekends/holidays.
- Industry and Power Stations (IND) represent electricity generation and factory production nodes. These are also always exits and naturally exhibit weekly patterns due to working routines.
- Storage nodes (STO) usually have a large number of zero flow hours with some substantial, often constant transfer in between. The nodes can always both be entries and exits.

While, in principle, we know the above classification, it is proved not to be reliable regarding the behavior of the nodes, so we will not use this information as part of the forecast, but just to explain certain behavior.

Table 1 depicts the number of nodes belonging to the different groups and percentage of gas flow explained by each group.

As illustration, we carefully selected three nodes for each type. The three network nodes we selected occupy 22% of the whole network flow. The municipal nodes are considered important by the TSO. The Industry nodes selected represent power plants and play a key role in energy generation with high renewable energy shares, as they are fast to start and can produce the necessary energy in times of peak demand. For the representative nodes from the Storage group we selected the most frequently used nodes in the observed period.

Figure 2 shows normalized (to the range of [0, 1]) flows of the nodes considered in this study.



(a) NET1 (b) NET2 (c) NET3

(d) MUN1 (e) MUN2 (f) MUN3

(g) IND1 (h) IND2 (i) IND3

(j) STO1 (k) STO2 (l) STO3

Figure 2: Normalized flow of selected nodes

For each node, the gas flows are measured hourly. Additionally, we were given the average daily temperatures measured at the nodes. Some statistical properties of selected nodes are given in Table 2. As can be seen from Figure 2 and in the Table 2 some nodes have continuous flow, while other are active only occasionally. Storage nodes have the highest percentage of zero flows. For the ones considered in this study hours with zero flow amount for 26% to 53% of the time. Network nodes show the highest variability and are always inflows. Industry nodes are always outflows and are clearly not temperature dependent. Municipal nodes are usually temperature dependent and have a strong daily, weekly and seasonal patterns.

Table 2: Properties of nodes used in the study. We denote inflows as positive and outflows as negative values

Node	Type	mean	std	min	max	zeros (%)
NET1	Network	13546	3177	4063	21092	0.00
NET2	Network	7752	3792	0	19113	0.01
NET3	Network	26241	3506	0	32593	0.06
MUN1	Municipal	-171	48	-763	0	2.68
MUN2	Municipal	-591	230	-1658	0	0.85
MUN3	Municipal	-111	35	-299	0	0.02
IND1	Industry	-161	32	-334	0	0.01
IND2	Industry	-141	41	-305	0	0.43
IND3	Industry	-121	23	-302	0	1.23
STO1	Storage	227	2469	-8013	11850	25.88
STO2	Storage	-277	2606	-5453	13972	52.88
STO3	Storage	150	3377	-9940	15017	39.51

3. Methods

Many research studies showed that combining forecasts improves accuracy relative to individual forecasts [20],[9]. In this Section we will first present three different individual forecasting methods; Functional AutoRegressive (FAR), Long Short-Term Memory Network(LSTM) and Mathematical Programming(MP) model. Then we will propose a hybrid model(HYB) based on MP method which is using output of two other forecasting models, FAR and LSTM, as additional inputs (features).

3.1. Functional AutoRegressive (FAR) model

In this section, we consider the hourly gas flows as discrete observations in continuous daily flow curves and predict the continuous curves over time with Functional AutoRegressive (FAR) dynamics. [7] developed Functional AutoRegressive (FAR) model and Yule-Walker estimator for one series of functional time-dependent data ([6], Kim, [13], [2], [19]). [24] proposed a maximum likelihood estimation with Fourier expansions. Chen extended the maximum likelihood estimation with stochastic variations and employed an adaptive approach so that the predictive model can be directly applicable for both stationary and non-stationary situations. We detail the FAR setup and show how to obtain the maximum likelihood estimator of the functional parameters, with the help of Fourier expansion and sieve.

Our interest is to model the dynamic dependence of gas flow curves over time. Let $\left\{X_t(\tau)\right\}_{t=1}^{n}$ denote a sequence of random curves over a time domain $\tau \in [0,1]$. The curve $X_t(\tau)$ is a continuous

function taken values in the Hilbert space $\mathcal{H}$ endowed with its Borel σ-algebra $B_{\mathcal{H}}$ at time point t. Suppose that there exists a Hilbert-Schmidt operator ρ_t that can be written as $\rho_t = \sum_j \lambda_{j,t} < \cdot, e_j > f_j$, where $\{e_j\}$ and $\{f_j\}$ are orthonormal bases of $\mathcal{H}$, $< \cdot, \cdot >$ is the inner product derived to Hilbert-Schmidt norm $|| \cdot ||$, and $\{\lambda_j, t\}$ is a real sequence such that $\sum_j \lambda_{j,t}^2 < \infty$, see [7] for the theory of Hilbertian autoregressive model. The functional autoregressive model of order 1, i.e. the FAR(1) model, is defined as:

$$X_t(\tau) - \mu(\tau) = \int_0^1 K(\tau - s)[X_{t-1}(s) - \mu(s)]ds + \varepsilon_t(\tau), \quad \tau \in [0, 1], \tag{1}$$

where $\mu(\tau)$ is the time-dependent mean function of $X_t(\tau)$. The AR operator is represented as a convolution kernel operator, which is one implementable form of the Hilbert-Schmidt operator. This choice of convolution kernel operator is very common in the study of functional linear processes and autoregressive processes, see [31], [24] and [22]. The kernel $K \in L^2([0, 1])$ is an even function with $\|K\|_2 < 1$ and $\| \cdot \|_2$ denotes the standard L^2 norm. The innovation $\varepsilon_t(\tau)$ is a strong $\mathcal{H}$-white noise with zero mean and finite second moment $E\|\varepsilon(\tau)\|^2 < \infty$ Assume the Fourier coefficients of the innovation function $\varepsilon_t(\tau)$, denoted as $a_0(\varepsilon_t), a_k(\varepsilon_t)$ and $b_k(\varepsilon_t)$, are independent and identically Gaussian distributed with mean zero and variance σ_k^2, we define a transition density as follows:

$$\begin{aligned} g(X_t, X_{t-1}, \rho) &= \frac{2\pi^{-(2m_n+1)/2}}{\sigma_0 \prod_{k=1}^{m_n} \sigma_k^2} \cdot \exp\Big\{ -\frac{1}{2\sigma_0^2}(a_{t,0} - p_0 - c_0 a_{t-1,0})^2 \\ &\quad - \sum_{k=1}^{m_n} \frac{1}{2\sigma_k^2}\Big[(b_{t,k} - q_k - \frac{1}{\sqrt{2}} c_k b_{t-1,k})^2 + (a_{t,k} - p_k - \frac{1}{\sqrt{2}} c_k a_{t-1,k})^2\Big]\Big\}, \end{aligned}$$

and thus the conditional log-likelihood $L(X_1, ..., X_n; \rho)$ is

$$\begin{aligned} L(X_1, ..., X_n; \rho) &= \log\Big\{ \prod_{t=2}^{n} g(X_t, X_{t-1}, \rho)\Big\} \\ &= -\frac{(2m_n + 1)(n - 1)}{2} \log 2\pi - (n - 1)\log\sigma_0 - (n - 1)\sum_{k=1}^{m_n} \log\sigma_k^2 \\ &\quad - \frac{1}{2\sigma_0^2} \sum_{t=2}^{n} (a_{t,0} - p_0 - c_0 a_{t-1,0})^2 \\ &\quad - \sum_{t=2}^{n} \sum_{k=1}^{m_n} \frac{1}{2\sigma_k^2}\Big\{(b_{t,k} - q_k - \frac{1}{\sqrt{2}} c_k b_{t-1,k})^2 + (a_{t,k} - p_k - \frac{1}{\sqrt{2}} c_k a_{t-1,k})^2\Big\}. \end{aligned}$$

We implement the FAR modelling to forecast the daily gas flow curves. The gas flow forecast h-step ahead, denoted as $\hat{X}_{t+h}(\tau)$ is directly based on the h-step ahead forecast of the mean function and the kernel operator:

$$\hat{X}_{t+h}(\tau) = \hat{\mu}(\tau) + \int_0^1 \hat{K}(\tau - s)[X_t(s) - \hat{\mu}(s)]ds \tag{2}$$

For each forecast point, we estimate the Fourier coefficients to obtain the estimated mean function $\hat{\mu}(\cdot)$ and kernel operator $\hat{K}(\cdot)$ and the fitted model is then used to compute $h = 1-$ and 2-step ahead forecasts of the gas flow curves.

3.2. Long Short-Term Memory Network

In this section, we use Long Short-Term Memory Network (LSTM) to predict gas flow based on the previous 24-hours. LSTM are a special types of Recurrent Neural Networks (RNN) that have been introduced in the eighties (i.e. [33], [11]) to model time interrelations by allowing connection between hidden units with a time delay ([28]). At each iteration, the hidden state vector receives the input vector and its previous hidden state. The hidden state vector can therefore be seen as a representation of time sequences ([29]).

Long Short-Term Memory (LSTM) networks, proposed by [14], include a memory gate that controls what passes through the network and what is blocked so that some of the information that is feedback to the network is remembered and some other is forgotten. An additional gate keeps memory and filters out what has to be forgotten. At each time step, the network memorizes the information and filters out what is not relevant for the prediction. Finally, another set of gates ignores what is irrelevant. The formulation of the LSTM, as presented in [21], consists of three layers that are called gates: the input (3), forget (4) and output gates (7), respectively:

$$i = sig(W_i \dot{[} h_{t-1}, x_t] + b_i) \tag{3}$$

$$f = sig(W_f \dot{[} h_{t-1}, x_t] + b_f) \tag{4}$$

Where h_{t-1} is the hidden state computed at time $t-1$ which is calculated based on previous hidden state, h_t. Each of the gate has a sigmoid function so that the values range between 0 and 1.

The decision on which information will be stored into a cell is based on the input layer (3) and on a hyperbolic tangent (or sigmoid) function assigned to the layer that returns the set of candidate values, $\hat{C}$ (5):

$$\hat{C}_t = tanh(W_C \dot{[} h_{t-1}, x_t] + b_C) \tag{5}$$

To update the cell state C_{t-1} into C_t, the old state is multiplied by the forget gate and added to the new candidate values (6):

$$C_t = f_t \circ C_{t-1} + i \circ \hat{C}_t \tag{6}$$

Finally, the output is obtained by passing the cell state C_t to a rectified linear function (or hyperbolic tangent) function to decide which part of the information is passed to the output 7. Moreover, the cell state is multiplied by the output of the relu (or tanh) gate (8).

$$o_t = sig(W_o[h_{t-1}, x_t] + b) \tag{7}$$

$$h_t = tahn(C_t) \circ o_t \tag{8}$$

Thanks to this architecture, LSTM has the ability to look back several time steps and, thus, to improve the predictions. Also recurrent neural networks can look time steps back but the problem they incur is called vanishing or exploding gradient for which the results either become very large or small.

The LSTM deployed to forecast gas flow consisted of one single layer and an early stop function with patience set to four. This means that the training of the network stops as soon as the value of

Table 3: Parameters of the LSTM

Node	Neurons	Dropout	Activation	Activation Output	Batch size
NET1	55	0	Sigmoid	Relu	2
NET2	70	0	Sigmoid	Tanh	243
NET3	30	0	Sigmoid	Relu	2
MUN1	40	0.7	Sigmoid	Relu	2
MUN2	30	0.7	Tanh	Relu	2
MUN3	30	0.7	Sigmoid	Relu	2
IND1	70	0.5	Tanh	Relu	12
IND2	70	0.7	Tanh	Relu	10
IND3	70	0.7	Tanh	Relu	324
STO1	70	0.7	Sigmoid	Relu	48
STO2	80	0.7	Tanh	Relu	80
STO3	80	0.7	Tanh	Relu	80

loss function remains the same after four iterations. In this work, different types of parameters are manually selected to forecast gas flow depending on the type of node and based on trial and error.

The configuration of the parameters of the network for each node is reported in Table 3. The most influential parameter is the batch size that is the number of training examples that are utilized in each iteration. The higher the value of this parameter the faster is the training of the network. Only two types of activation functions are selected for the hidden state, depending on the node: hyperbolic tangent function (tanh) or sigmoid. The transfer function of the output gate selected for all nodes, except for one network node, is the Rectified Linear Unit (Relu). Overall for storage nodes that are characterized by high variability between negative and positive values and by a high number of hours with zero flows, the batch size was set between 48 and 80 and the number of neurons between 70 and 80.

3.3. Mathematical Programming (MP) for time series forecasting

In this section, we use Linear Programs (LP) together with Mixed Integer Linear Programs (MILP) for prediction of the flows - supplies and demands of the gas network. Given a set of measurements $m_{d,h} \in M$ for each day $d \in D$ and each hour $h \in H$. Let us define $M_d \subseteq M$ as a subset of the measurements before day d. The features $i \in F_h = \{1, \ldots, n_h\}$ are defined as arbitrary functions of historical flow values, $f_{h,i}(d) : D \to M_d, i \leq p_h \leq n_h$ and exogenous variables $f_{h,i}(d), i \in \{p_h + 1, \ldots, n_h\}$. We can approximate gas flow with weighted sum of features

$$p_{d,h} = \sum_{i \in F_h} w_{h,i} f_{h,i}(d) \tag{9}$$

where $p_{d,h}$ is the flow value which is approximated, and $w_{h,i}$ define the weights.
The approximation error is defined as

$$e_{d,h} = p_{d,h} - m_{d,h}$$

and the optimal weights are calculated by minimizing the sum of absolute errors for each day d and hour h

$$\min \sum_{d \in D, h \in H} |e_{d,h}|$$

This problem is not a LP because of the nonlinear absolute value in the objective function but it can be transformed into a LP. We can rewrite the error $e_{d,h}$ as the difference of two non-negative variables:

$$e_{d,h} = e^+_{d,h} - e^-_{d,h}$$

Then the transformed objective function is

$$\min \sum_{d \in D, h \in H} |e^+_{d,h} + e^-_{d,h}|$$

For a solution to be optimal regarding to the objective $e^+_{d,h} \cdot e^-_{d,h} = 0$ (ref) must be true, so we can write

$$|e^+_{d,h} - e^-_{d,h}| = |e_{d,h}| + |e^-_{d,h}| = e^+_{d,h} + e^-_{d,h}$$

and consequently the final LP problem becomes

$$\min \sum_{d \in D, h \in H} (e^+_{d,h} + e^-_{d,h})$$

$$\text{subject to} \sum_{i \in F} f_{h,i}(d) \cdot w_{h,i} - m_{d,h} = e^+_{d,h} - e^-_{d,h} \text{ for all } \in D, h \in H$$

$$e^+_{d,h}, e^-_{d,h} \geq 0$$

$$w_{h,i} \in \mathbb{R}$$

Furthermore, we can force our model to be unbiased by requiring

$$\sum_{d \in D, h \in H} (e^+_{d,h} - e^-_{d,h}) = 0$$

and setting bounds for the weights $l \leq w_{h,i} \leq u$ to limit the influence of a single specific feature.

For each day in the test set (out of sample days that we want to forecast) the forecasted flow values are computed by first computing the weights via an LP with 16 weeks of historical data and then using the weighted sum of features (9) for each hour to forecast the flow values. The lower and upper bounds for the weights are set to $l = -2$ and $u = 2$, respectively. For the computation of the forecasted flow values, it might be that also forecasted flow values of prior hours are used as input values for calculating the features, if the corresponding hours do not lie in the past.

Training: Feature Selection

In the training procedure of this method, a slightly different model is used which automatically chooses for each hour the B features which are most important, to limit over-fitting in the LP. Therefore, we add additional binary variables $x_{h,i}$ to the problem, which determine whether feature i is chosen for hour h, i.e., whether the weight of feature i and hour h is not equal to zero. Then, we link these variables to the weight variables

$$x_{h,i} \cdot l \leq w^{+}_{h,i} \leq x_{h,i} \cdot u$$

and limit the number of chosen features by B

$$\sum_{i \in F} x_{h,i} \leq B$$

The solution of the resulting MILP leads for each hour h to one feature set F_h of at most B features which are most important for this hour.

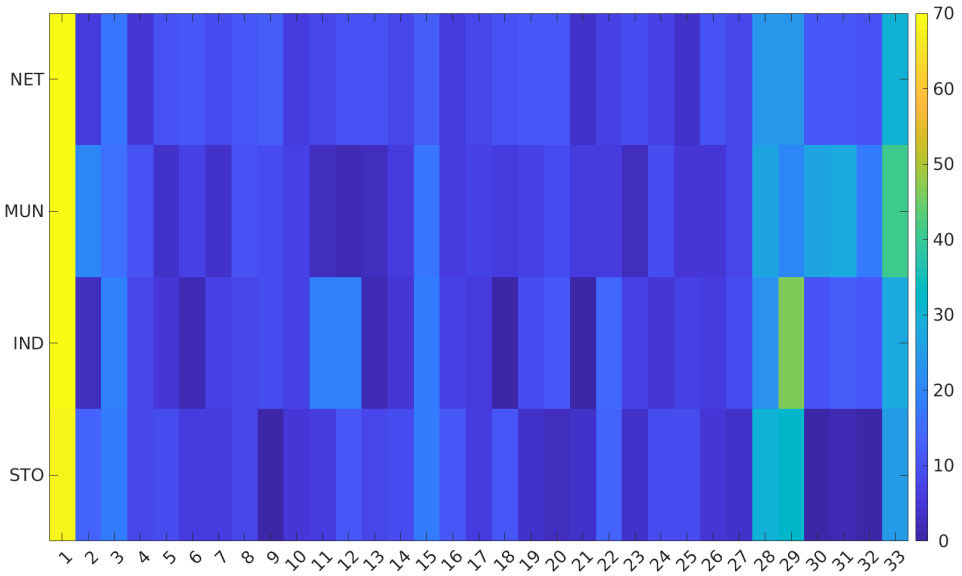
Table 4: List of features

Feature	Description
$f_1(d,h) = \begin{cases} m(d,h-1), & \text{if } h > 0 \\ m(d-1,23) & \text{otherwise} \end{cases}$	Prior hour
$f_2(d,h) = m(d-1,0)$	First hour yesterday
$f_3(d,h) = m(d-1,23)$	Last hour yesterday
$f_{4:10}(d,h) = m(d-(1,2,..,7),h)$	The same hour 1,2,..,7 days ago
$f_{11:12}(d,h) = m(d-1,0(h))/m(d-2,0(h))$	Ratio first(same) hour yesterday first(same) hour 2 days ago
$f_{13:14}(d,h) = m(d-1,0(h)) - m(d-2,0(h))$	Difference first(same) hour yesterday, first(same) hour 2 days ago
$f_{15:21}(d,h) = 1/24(\sum_{h\in H} m(d-(1,2,...,7),h))$	Mean flow 1,2,..,7 days ago
$f_{22}(d,h) = f_{15}(d,h)/f_{16}(d,h)$	Ratio mean flow yesterday, 2 days ago
$f_{23}(d,h) = f_{15}(d,h)/f_{21}(d,h)$	Ratio mean flow yesterday, 7 days ago
$f_{24}(d,h) = f_{15}(d,h)/(1/24(\sum_{h\in H} m(d-8,h)))$	Ratio mean flow yesterday, 8 days ago
$f_{25}(d,h) = f_{15}(d,h) - f_{16}(d,h)$	Difference mean flow yesterday, 2 days ago
$f_{26}(d,h) = f_{15}(d,h) - f_{21}(d,h)$	Difference mean flow yesterday, 7 days ago
$f_{27}(d,h) = f_{15}(d,h) - (1/24(\sum_{h\in H} m(d-8,h)))$	Difference mean flow yesterday, 8 days ago
$f_{28}(d,h) = \begin{cases} 0, & \text{if } h = 0 \\ m(d,0) & \text{otherwise} \end{cases}$	First hour today
$f_{29}(d,h) = \begin{cases} 0, & \text{if } h = 0 \\ 1/h \sum_{i=0}^{h-1}(d,i) & \text{otherwise} \end{cases}$	Mean flow today
$f_{30}(d,h) = t_d - t_{d-1}$	Difference mean temperature today and yesterday
$f_{31}(d,h) = \begin{cases} 1, & \text{if } day \in \{Saturday, Sunday\} \\ 0 & \text{otherwise} \end{cases}$	Weekend
$f_{32}(d,h) = \begin{cases} 1, & \text{if } day \in \{Friday, Saturday\} \\ 0 & \text{otherwise} \end{cases}$	Evening
$f_{33}(d,h) = 1$	Offset

The list of features used in this study is presented in Table 4. The whole set of features F we used consists of 29 different features based on historical flow values, one temperature feature and two different features describing position of the predicted gas day in the week and the offset feature. Using sensitivity analysis the number of chosen features was limited to six. One year of

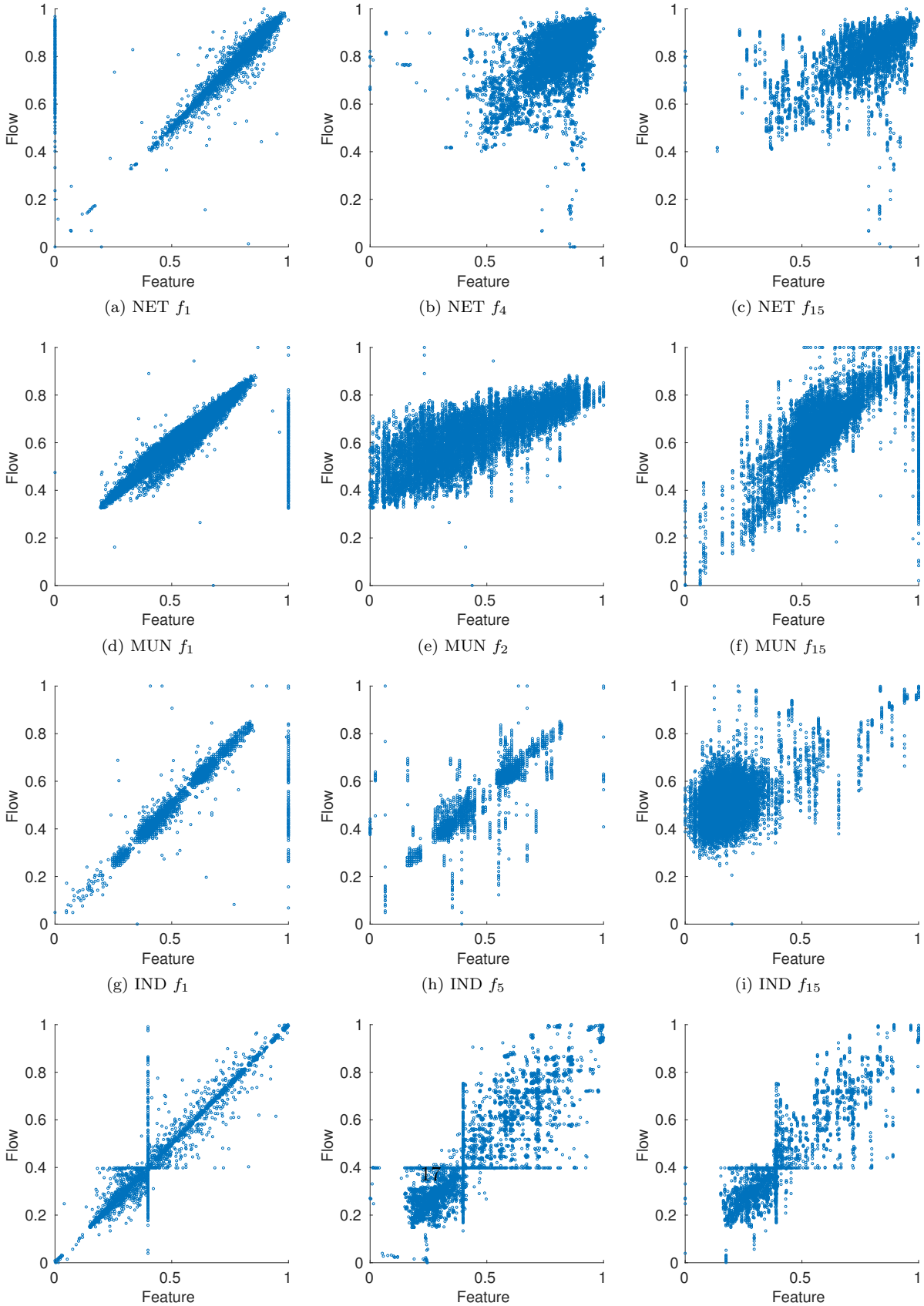
historical measurements was used for training and selecting optimal set of features for every node and hour. The Figure 3 is showing the heatmap of selected features for MP model for each group of nodes summed up for 24 hours.

Figure 3: Heatmap of selected features for different node group



For all nodes the feature representing the flow of the previous hour (f_1) is the mostly used feature. For all hours except the first predicted gas hour this feature value is calculated based on the the forecasted flow of previous hour when the final forecast is calculated. The same hour yesterday(f_4) is also widely selected among all groups. As it was expected Weekend (f_{31}), Evening (f_{32}) as well as Mean temperature difference feature (f_{30}) are usually chosen only for the Municipal utilities since the behaviour of those nodes shows strong daily, weekly and seasonal patterns. In the case of Industry nodes the features of Mean flow of the same and previous day (f_{29}, f_{15}) together with the Ratio features f_{11}, f_{12} are the most frequently chosen features. For Transfer Points and Storages features of mean flow of the same and previous day (f_{29}, f_{15}) are also dominating ones except for first gas hour where this pattern is not present. For all observed nodes the Offset feature (representing the bias in the model) is selected very frequently.

Figure 4 shows a scatter plot of flow amount versus the three most frequently chosen features among different types of nodes for all hours of the day. It can be seen that the flow depends linearly on f_4 while other features are showing a nonlinear dependency.



(a) NET f_1 (b) NET f_4 (c) NET f_{15}

(d) MUN f_1 (e) MUN f_2 (f) MUN f_{15}

(g) IND f_1 (h) IND f_5 (i) IND f_{15}

(j) STO f_1 (k) STO f_4 (l) STO f_{15}

Figure 4: Scatter plot of some frequently chosen features vs flow

3.3.1. Hybrid model

The main advantage of the mathematical programming method (MP) proposed in this paper is flexibility in the sense that adding new features when they are available in order to improve the forecast is very simple.

In this section we propose a hybrid model, combining mathematical programming (MP) with the two other proposed methods by adding the outputs from LSTM and FAR model as an exogenous inputs to the MP model. The optimal sets of features chosen for every node and hour were kept from previous MP training and extended with the forecasts from the LSTM and FAR model as additional features. The final forecast is calculated as weighted sum of all features from the extended features set:

$$p_{d,h} = p_{d,h} = \sum_{i \in F} f_{h,i}(d) \cdot w_{h,i} + \text{LSTM}(d,h) \cdot w_{h,\text{LSTM}} + \text{FAR}(d,h) \cdot w_{h,\text{FAR}}$$

New optimal weights are calculated by running an LP:

$$\min \sum_{d \in D, h \in H} (e^{+}_{d,h} + e^{-}_{d,h})$$

$$\text{subject to } p_{d,h} - m_{d,h} = e^{+}_{d,h} - e^{-}_{d,h} \text{ for all } d \in D, h \in H$$

$$p_{d,h} = \sum_{i \in F} f_{h,i}(d) \cdot w_{h,i} + \text{LSTM}_h(d) \cdot w_{h,\text{LSTM}} + \text{FAR}_h(d) \cdot w_{h,\text{FAR}}$$

$$e^{+}_{d,h}, e^{-}_{d,h} \geq 0$$

$$w_{h,i}, w_{h,\text{LSTM}}, w_{h,\text{FAR}} \in \mathbb{R}$$

where $\text{LSTM}_h(d)$ and $\text{FAR}_h(d)$ represent the forecasted values obtained from the LSTM and FAR models, respectively.

4. Testing and Results

4.1. Influence of temperature

The temperature is one of the most important factors that influence gas consumption. When the natural gas is consumed for heating such as in residential areas, the temperature usually has an inverse relationship with gas consumption which is also dependent on other environmental data such as the time of day, the day of week, the season, etc. Furthermore, temperature and time of day are the factors that mostly impact the forecast error ([39]).

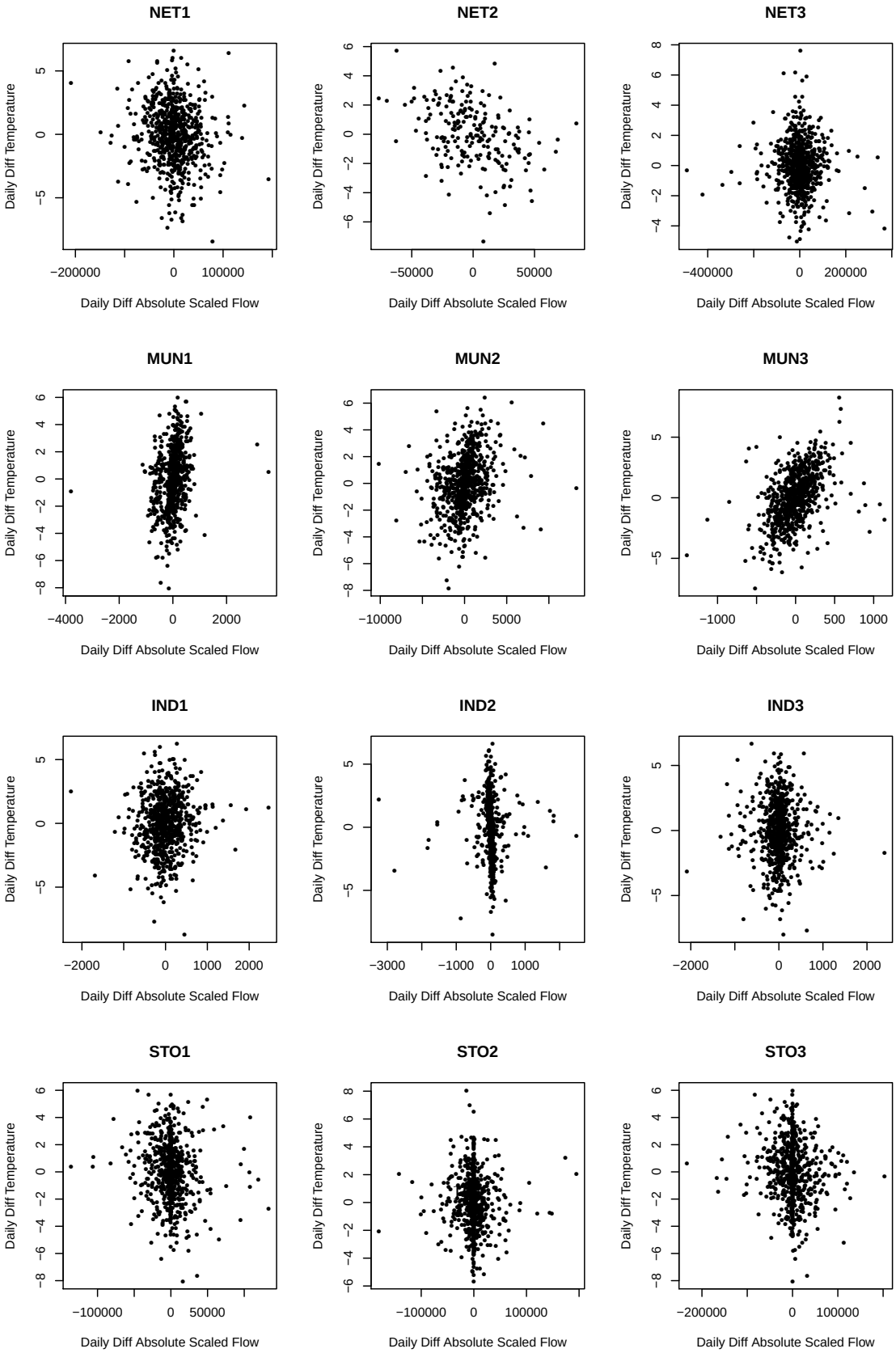
The majority of models presented in the literature are focused on residential and small commercial consumer at individual or aggregate level.

Several authors have considered the temperature (i.e. [26], [27], [12]) or meteorological data ([40]) in their models to forecast gas flow and reduce the prediction error.

In [44], the authors pointed out that the nonlinear characteristics of temperature has been assessed long time ago and gas consumption is proportional to the Heating Degree Day (HDD) ([4]). This proportionality is evidenced when plotting the average daily temperature versus the gas consumption.

The scatter plots of daily changes of temperature versus daily changes of gas flow of the nodes considered in this work are shown in Figure 5. Storage nodes have a high percentage of zero flows,

Figure 5: Scatter plot of gas consumption vs temperature



those are the nodes that better approximate the non linear relationship expressed by the HDD. As expected, the Municipal nodes present a positive correlation between the flow and the temperature. Among the network nodes, one of them presents a negative relationship with the temperature. The remaining nodes appear to be independent from the temperature.

4.2. Results

4.2.1. Objectives, setup and evaluation metrics

In this paper we observed a data set of hourly gas flow time series from twelve nodes with 17.520 observations (two years).

Our goal is to predict values $p_{d,0}$ to $p_{d,23}$ for a given $d \in D$. Of course only data for days $d-1$ and earlier can be used. All proposed methods are tested on the last 60 days of the data set.

Our basis comparison are the mean absolute deviation (MAD) between the hourly forecast and the measured flow during one day (24h) ahead forecast, defined as

$$MAD_d := 1/h \sum_h |p_h - m_h| \tag{10}$$

and mean absolute percentage error (MAPE) defined as

$$MAPE_d := 1/h \sum_h |(p_h - m_h)/m_h| \tag{11}$$

All results are also compared to a baseline (BAS) forecast defined as

$$\hat{p}_{h,d} := m_{h,d-1} \tag{12}$$

4.2.2. Forecasting Results

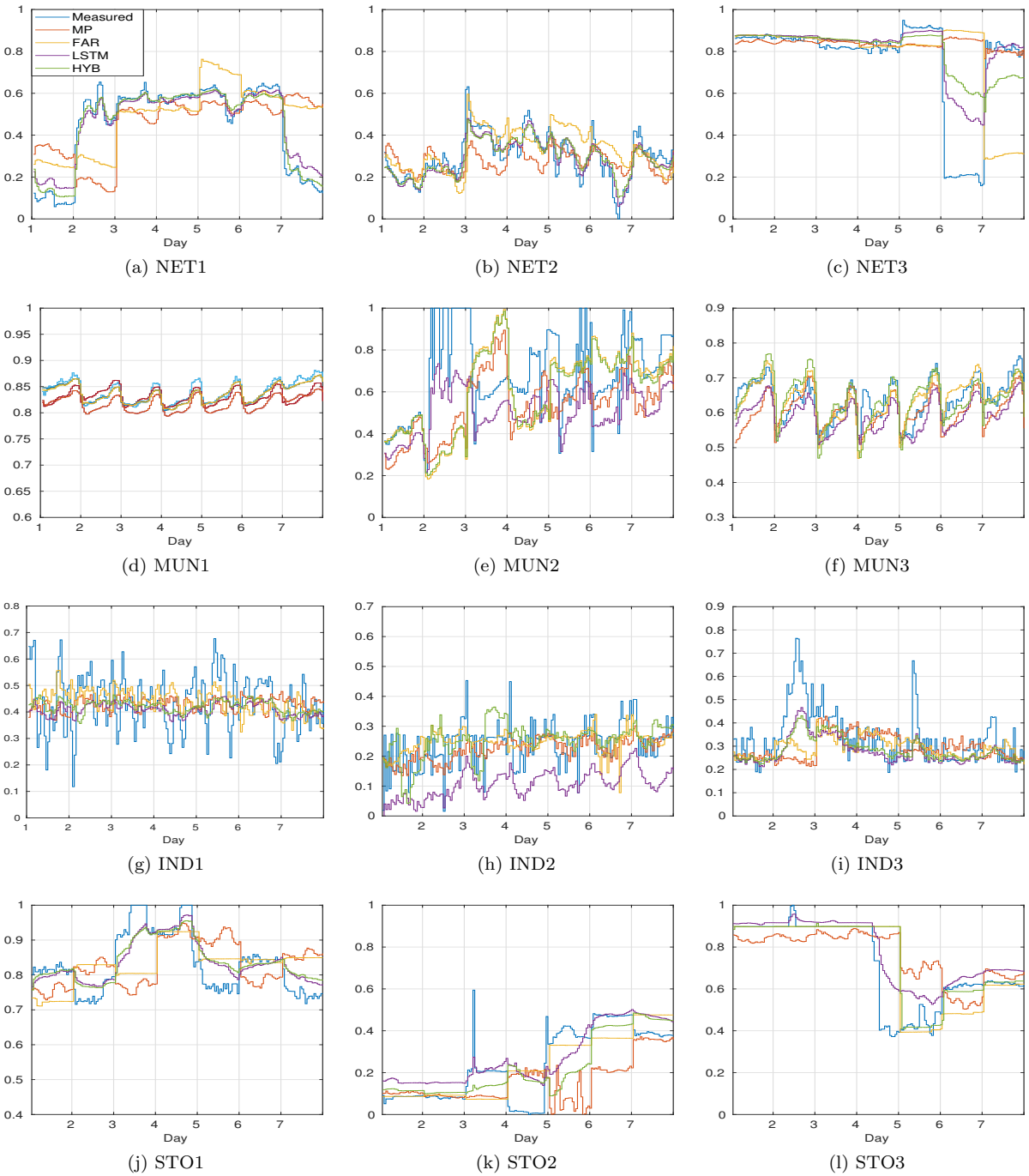
Mean MAD and MAPE (over 60 days in the test set) achieved by proposed models are presented in the Table 5.

Table 5: Comparison of mean daily performance

	MAPE					MAD				
	MP	FAR	LSTM	HYB	BAS	MP	FAR	LSTM	HYB	BAS
NET1	0.128	0.131	0.035	**0.033**	0.133	1269	1350	357	**331**	1366
NET2	0.176	0.195	**0.075**	0.088	0.211	727	828	**328**	378	901
NET3	0.084	0.096	0.056	**0.055**	0.098	1855	2245	1439	**1393**	2168
MUN1	0.032	0.101	0.09	**0.030**	0.069	6.36	19.03	29.02	**5.97**	11.47
MUN2	0.327	0.370	0.588	**0.324**	0.355	65.67	72.90	101.18	**64.8**	71.1
MUN3	0.064	0.088	0.071	**0.064**	0.089	5.36	7.37	**5.16**	5.61	7.41
IND1	0.131	0.137	**0.109**	0.114	0.171	19.37	19.99	**15.64**	16.66	25.5
IND2	0.035	**0.028**	0.057	0.030	0.034	3.85	**3.19**	6.24	3.71	3.76
IND3	0.103	0.135	0.087	**0.086**	0.142	9.78	12.32	7.48	**7.45**	13.44
STO1	0.550	0.689	0.313	**0.291**	0.764	609.28	753.42	313.79	**294.09**	681.31
STO2	0.298	0.343	0.226	**0.205**	0.411	428.24	560.78	384.33	**329.54**	624.38
STO3	0.895	0.970	**0.475**	0.502	0.997	895	1155	909	**816**	1442

It can be seen that between individual forecasting methods the LSTM model is the most robust one and obtained the best results for nodes from all 4 behavior groups. The MP model achieved the best results for all Municipal utilities, and the performance is especially good for MUN2 node where all models had a particularly high MAPE. FAR model outperformed others for the Industry node IND2. Even though FAR errors are slightly higher than other two proposed models (MP and LSTM) it can be seen that for most of the nodes the performance is very similar. For Storage nodes with intermittent behaviour none of the proposed individual methods has demonstrated adequate accuracy. The Hybrid model showed an improvement for all 4 groups nodes. The improvement is especially significant for Storage nodes, where the average MAPE is lower for more than 2(%) for nodes STO1 and STO2. For node STO3 the LSTM model has the lowest MAPE but the lowest MAD is achieved by the HYB model.

Figure 6 shows calculated 24 hours ahead forecast and the measured flow of all proposed models for a one week period.



(a) NET1 (b) NET2 (c) NET3

(d) MUN1 (e) MUN2 (f) MUN3

(g) IND1 (h) IND2 (i) IND3

(j) STO1 (k) STO2 (l) STO3

Figure 6: 24 hours ahead forecast for one week test set

5. Conclusions

In this paper we proposed a robust and powerful hybrid forecast model combining Mathematical Programming with Functional AutoRegressive and Long Short Term Neural Network model for forecasting gas flows at the boundary nodes of gas transport network. Our experiments are based on real world data from one of Germany's largest transmission system operators, Open Grid Europe. We showed that the proposed method is appropriate for choosing optimal set of features and forecasting various behaviours from different nodes groups in the complex gas transmission network. From obtained results it is clear that even though in some specific cases single forecast models outperform the Hybrid model, the proposed method can achieve stable accuracy close to the best individual model and in some cases brings a significant improvement to the forecast quality.

6. Acknowledgments

This work has been supported by the Research Campus Modal Mathematical Optimization and Data Analysis Laboratories funded by the Federal Ministry of Education and Research (fund number 05M14ZAM).

[1] Alfares, H. and Nazeeruddin, M. (2002). Electric load forecasting: Literature survey and classification of methods. *International Journal of Systems Science*, 33(1):23–34.

[2] Antoniadis, A. and Sapatinas, T. (2003). Wavelet methods for continuous-time prediction using hilbert-valued autoregressive processes. *Journal of Multivariate Analysis*, 87:133–158.

[3] Azadeh, A., Asadzadeh, S., and Ghanbari, A. (2010). An adaptive network-based fuzzy inference system for short-term natural gas demand estimation: Uncertain and complex environments. *Energy Policy*, 38(3):1529 – 1536. Security, Prosperity and Community Towards a Common European Energy Policy? Special Section with Regular Papers.

[4] Baker, D. (1975). Effect of observation time on mean temperature estimation. *Journal of Applied Meteorology*, 14(4):471–476.

[5] Baldacci, L., Golfarelli, M., Lombardi, D., and Sami, F. (2016). Natural gas consumption forecasting for anomaly detection. *Expert Systems with Applications*, 62:190 – 201.

[6] Besse, P., Cardot, H., and Stephenson, D. (2000). Autoregressive forecasting of some climatic variations. *Scandinavian Journal of Statistics*, 27:673–687.

[7] Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications.* Springer-Verlag, New York.

[8] Brabec, M., Konr, O., Pelikn, E., and Mal, M. (2008). A nonlinear mixed effects model for the prediction of natural gas consumption by individual customers. *International Journal of Forecasting*, 24(4):659 – 678. Energy Forecasting.

[9] Degiannakis, S. (2018). Multiple days ahead realized volatility forecasting: Single, combined and average forecasts. *Global Finance Journal*, 36:41 – 61.

[10] Dombayc, O. (2010). The prediction of heating energy consumption in a model house by using artificial neural networks in denizliturkey. *Advances in Engineering Software*, 41(2):141 – 147.

[11] Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14.

[12] Forouzanfar, M., Doustmohammadi, A., Menhaj, M., and Hasanzadeh, S. (2010). Modeling and estimation of the natural gas consumption for residential and commercial sectors in iran. *Applied Energy*, 87(1):268 – 274.

[13] Guillas, S. (2001). Rates of convergence of autocorrelation estimates for autoregressive hilbertian processes. *Statistics & Probability Letters*, 55:281–291.

[14] Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8):1735–1780.

[15] Hribar, R., Potonik, P., ilc, J., and Papa, G. (2019). A comparison of models for forecasting the residential natural gas demand of an urban area. *Energy*, 167:511 – 522.

[16] Izadyar, N., Ong, H. C., Shamshirband, S., Ghadamian, H., and Tong, C. W. (2015). Intelligent forecasting of residential heating demand for the district heating system based on the monthly overall natural gas consumption. *Energy and Buildings*, 104:208 – 214.

[17] Karadede, Y., Ozdemir, G., and Aydemir, E. (2017). Breeder hybrid algorithm approach for natural gas demand forecasting model. *Energy*, 141:1269 – 1284.

[18] Khotanzad, A., Elragal, H., and Lu, T. (2000). Combination of artificial neural-network forecasters for prediction of natural gas consumption. *IEEE Transactions on Neural Networks*, 11(2):464–473.

[19] Kokoszka, P. and Zhang, X. (2010). Improved estimation of the kernel of the functional autoregressive process. Technical report, Utah State University.

[20] Liang, Z., Liang, J., Wang, C., Dong, X., and Miao, X. (2016). Short-term wind power combined forecasting based on error forecast correction. *Energy Conversion and Management*, 119:215 – 226.

[21] Liu, C., Jin, Z., Gu, J., and Qiu, C. (2017). Short-term load forecasting using a long short-term memory network. In *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, pages 1–6.

[22] Mas, A. and Pumo, B. (2011). Linear processes for functional data. In *In: F. Ferraty and Y. Romain (Ed.), The Oxford Handbook of Functional Data*, pages 47–71. Springer-Verlag.

[23] Mirasgedis, S., Sarafidis, Y., Georgopoulou, E., LalasD.P., Moschovits, M., Karagiannis, F., and Papakonstantinou, D. (2006). Models for mid-term electricity demand forecasting incorporating weather influences. *Energy*, 31(2):208 – 227.

[24] Mokhtari, F. and Mourid, T. (2003). Prediction of continuous time autoregressive processes via the reproducing kernel spaces. *Statistical Inference for Stochastic Processes*, 6:247–266.

[25] Nguyen, H. T. and Nabney, I. T. (2010). Short-term electricity demand and gas price forecasts using wavelet transforms and adaptive models. *Energy*, 35(9):3674 – 3685.

[26] Özmen, A., Ylmaz, Y., and Weber, G.-W. (2018). Natural gas consumption forecast with mars and cmars models for residential users. *Energy Economics*, 70:357 – 381.

[27] Panapakidis, I. and Dagoumas, A. (2017). Day-ahead natural gas demand forecasting based on the combination of wavelet transform and anfis/genetic algorithm/neural network model. *Energy*, 118:231 – 245.

[28] Pascanu, R., Mikolov, T., and Bengio, Y. (2013). On the difficulty of training recurrent neural network. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28.

[29] Plunkett, K. and Elman, J. (1997). *Exercises in rethinking innateness : a handbook for connectionist simulations*. MIT Press.

[30] Potočnik, P., Soldo, B., Šimunović, G., Šari, T., Jeromen, A., and Govekar, E. (2014). Comparison of static and adaptive models for short-term residential natural gas forecasting in croatia. *Applied Energy*, 129:94 – 103.

[31] Pumo, B. (1998). Prediction of continuous time processes by $c_{[0,1]}$-valued autoregressive process. *Statistical Inference for Stochastic Processes*, 1:297–309.

[32] Ravnik, J. and Hriberek, M. (2019). A method for natural gas forecasting and preliminary allocation based on unique standard natural gas consumption profiles. *Energy*, 180:149 – 162.

[33] Rumelhart, D., Hinton, G., and Williams, R. (1986). Learning representations by backpropagating errors. *Nature*, 323(6088):533–536.

[34] Sen, D., Erdem Gnay, M., and Tun, K. M. (2019). Forecasting annual natural gas consumption using socio-economic indicators for making future policies. *Energy*, 173:1106 – 1118.

[35] Shaikh, F., Ji, Q., Shaikh, P., Mirjat, N., and Uqaili, M. (2017). Forecasting chinas natural gas demand based on optimised nonlinear grey models. *Energy*, 140:941 – 951.

[36] Soldo, B. (2012). Forecasting natural gas consumption. *Applied Energy*, 92:26 – 37.

[37] Soldo, B., Potočnik, P., Šimunović, G., Šarić, T., and Govekar, E. (2014). Improving the residential natural gas consumption forecasting models by using solar radiation. *Energy and Buildings*, 69:498 – 506.

[38] Su, H., Zio, E., Zhang, J., Xu, M., Li, X., and Zhang, Z. (2019). A hybrid hourly natural gas demand forecasting method based on the integration of wavelet transform and enhanced deep-rnn model. *Energy*, 178:585 – 597.

[39] Szoplik, J. (2015). Forecasting of natural gas consumption with artificial neural networks. *Energy*, 85:208 – 220.

[40] Taşpnar, F., Ƴelebi, N., and Tutkun, N. (2013). Forecasting of daily natural gas consumption on regional basis in turkey using various computational methods. *Energy and Buildings*, 56:23 – 31.

[41] Taylor, J. (2003). Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society*, 54(8):799–805.

[42] Thaler, M., Grabec, I., and Poredoş, A. (2005). Prediction of energy consumption and risk of excess demand in a distribution system. *Physica A: Statistical Mechanics and its Applications*, 355(1):46 – 53. Market Dynamics and Quantitative Economics.

[43] Tonković, Z., Zekić-Sušac, M., and Somolanji, M. (2009). Predicting natural gas consumption by neural networks. *Tehniki vjesnik*, 16(3):51–6.

[44] Vitullo, S. R., Brown, R., Corliss, G. F., and Marx, B. (2009). Mathematical models for natural gas forecasting. *The Canadian Applied Mathematics Quarterly*, 17(7):807827.

[45] Šebalj, D., Mesarić, J., and Dujak, D. (2017). Predicting natural gas consumption a literature review. In *Proceedings of the Central European Conference on Information and Intelligent Systems*.

[46] Wang, Q., Li, S., Li, R., and Ma, M. (2018). Forecasting u.s. shale gas monthly production using a hybrid arima and metabolic nonlinear grey model. *Energy*, 160:378 – 387.

[47] Yu, F. and Xu, X. (2014). A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved bp neural network. *Applied Energy*, 134:102 – 113.

[48] Zhou, H., Su, G., and Li, G. (2011). Forecasting daily gas load with oihf-elman neural network. *Procedia Computer Science*, 5:754 – 758. The 2nd International Conference on Ambient Systems, Networks and Technologies (ANT-2011) / The 8th International Conference on Mobile Web Information Systems (MobiWIS 2011).

[49] Zhu, L., Li, M., Wu, Q., and Jiang, L. (2015). Short-term natural gas demand prediction based on support vector regression with false neighbours filtered. *Energy*, 80:428 – 436.