

# Species Related Gas Tracking in Distribution Grids

Athanasios Alexiou

Munich University of Applied Sciences  
Lothstraße 34, 80335 Munich, Germany  
Email: athanasios.alexiou@hm.edu

Joachim Schenk

Munich University of Applied Sciences  
Lothstraße 34, 80335 Munich  
Email: joachim.schenk@hm.edu

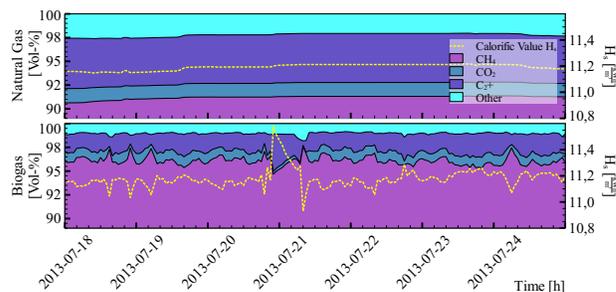
**Abstract**—Due to a wider diversification of gas sources, today tracking gas in distribution grids is of great interest for gas grid operators to provide fair invoicing of gas customers. Substitute natural gas (SNG), e.g. derived from raw biogas, injected concurrently into natural gas grids may differ in its calorific value  $H_s$ , compared to fossil natural gas in the grid. This is manifesting in deviating chemical compositions of injected grid gases. Remarkably, the chemical fractions of SNGs fluctuate significantly over time exhibiting time-dependent signatures. Sampling over relevant features of injected gases, e.g. the chemical species concentrations at standard temperature and pressure, by means of calibrated sensors, provides time-dependent signals which can be taken for gas tracking purposes. To that end, we present an accurate technique to estimate the transit times of gas between nodes, e.g. from an entry to an exit point. As a result, calorific value extrapolation from one gas grid node to a downstream node, with an accuracy sufficient for gas customer invoicing, is feasible. In an experimental section we show a normalized root-mean-square deviation (NRMSD)  $< 0.3\%$  with respect to calorific value estimation.

**Index Terms** — Viterbi algorithm, dynamic time warping, gas tracking, calorific value tracking, distribution grids

## I. INTRODUCTION

Powerful state of the art software for gas tracking, e.g. E.ON SmartSim [1, 2], utilize mainly numerical methods based on computational fluid dynamics (CFD). All recently applied methods implemented in software produce reliable and precise results, provided that grid topology features (i.e. the exact length, diameter and surface roughness of gas pipes, time dependent injection gas flows, temperatures and pressures) are submitted with sufficient accuracy by the gas grid operator and are correctly transferred into the simulation model. Any substantial deviations between model parameters and physical reality, typically lead to incorrect gas tracking and calorific value (CV) estimation. Challenging grid structures, e.g. loop topologies with multiple forks, can additionally result in ambiguous outcomes, due to insufficient model accuracy.

In this paper, we show that feasible gas and CV tracking can be implemented based on signal processing methods utilizing calibrated sensors along distribution grids. Our new gas tracking method is based on the Viterbi algorithm [3, 4]. It is capable of CV tracking by finding an optimal mapping between two signals that have been sampled at distinct locations (nodes) along a distribution grid. The following section II describes the procedure from data sampling to transit times and calorific value estimation. It is performed in three steps, outlined in the following sections II-A, II-B, II-C. Then, we explain our novel approach for gas tracking in-depth. In an experimental section III we demonstrate gas tracking in a real-



**Fig. 1:** By process gas chromatography (PGC) hourly measured fractions of the chemical components in natural gas (top) and upgraded biomethane (lower diagram). The dotted lines represent the corresponding calorific values. (Data collected at gas-to-grid plant BZBiogas Lüchow in Lüchow Germany. Timescale: 7 days, between 2013-07-18 00:00 and 2013-08-25 00:00)

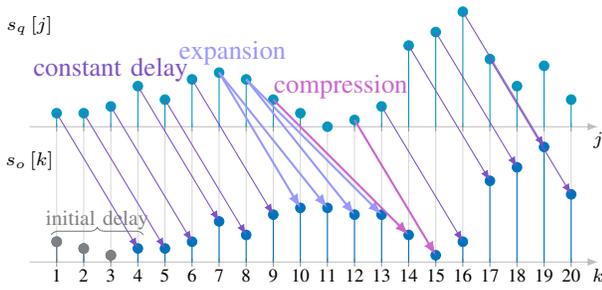
world gas grid. Finally we give a conclusion and an outlook in section IV.

## II. GAS TRACKING WITH SENSOR SIGNALS

Figure 1 shows the chemical compositions of Natural Gas (top) and upgraded biomethane (bottom), both concurrently injected into the distribution grid in Luechow, Germany. Remarkably, as can be seen in Figure 1, the chemical fractions in the substitute natural Gas (bottom) are exhibiting intrinsic fluctuations. This phenomenon implies that the compositions of SNGs vary over time and additional, partly nondeterministic, hidden factors impose time dependent signatures, i.e. fingerprints, onto the biogas compositions, that can be tracked along the piping over time.

### A. Sequential data recording by infrared sensor sampling

For our novel approach we use time series obtained continuously at equidistant points in time by means of calibrated sensors, e.g. by process gas chromatography (PGC), at distant locations along the piping network of natural gas grids. The minimum sampling rate results from required precision for gas and CV tracking. Currently, sampling gas with a rate of  $f_s = 0.28$  mHz, corresponding to one sample every  $\Delta t_s = 3600$  seconds, i.e. one sample every hour, proves to be sufficient for hourly tracking gas from one source  $q$  to an output node  $o$ . Our approach is limited to a pair of nodes  $(q, o)$ , for which the downstream node  $o$ , e.g. an exit node, is fed by one source  $q$  only. Then, accurate monthly gas customer invoicing on the node  $o$  becomes feasible with our approach. By sampling multiple time-varying gas features on relevant nodes simultaneously (e.g. the concentration of one



**Fig. 2:** Gas-signals  $s_q[j]$  and  $s_o[k]$  sampled at different locations, e.g. at an inlet  $q$  and an exit node  $o$ , appear delayed, additionally either compressed or expanded, i.e. nonlinearly warped along the time axis. Gas that has been sampled at an inlet is consumed along its way through the grid and flow velocities change rapidly over time. Natural gas distribution grids therefore appear as time-variant systems.

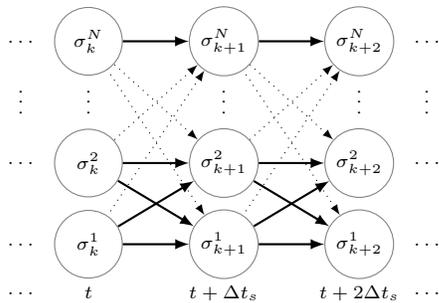
or more chemical species, the gas density and so forth) at the same thermodynamic condition (e.g. standard temperature and pressure (STP) [5]), redundant information is obtained and can afterwards be used for error-tolerant data processing, as will be shown in section III.

### B. Time series analysis utilizing the Viterbi algorithm

Two time series are recorded at two different grid nodes, an entry node  $q$  and an exit node  $o$ , simultaneously and continuously at equidistant points in time. Gas transit times  $\tau[k]$ , with  $k := [1 : M]$ , need to be estimated for gas tracking purposes by finding an optimal alignment between temporally overlapping time series  $s_q := (s_q[1], s_q[2], \dots, s_q[N])$  and  $s_o := (s_o[1], s_o[2], \dots, s_o[M])$  of length  $N, M \in \mathbb{N}$  meeting the following condition:

$$\sum_k (s_o[k] - s_q[k - \tau[k]])^2 \rightarrow \min \quad (1)$$

Sequence  $s_o$  is thereby the delayed signal sampled on a downstream node  $o$ , in the favored flow direction of gas, remote from the source node  $q$ . With a naive approach for minimization in 1, non-causal assignments will occur with transit times  $\tau < 0$ , which are physically impossible. At the same time, gas is a compressible medium and volumetric flow rates and consumption conditions on outlet nodes may change rapidly over time while gas travels the distance between two

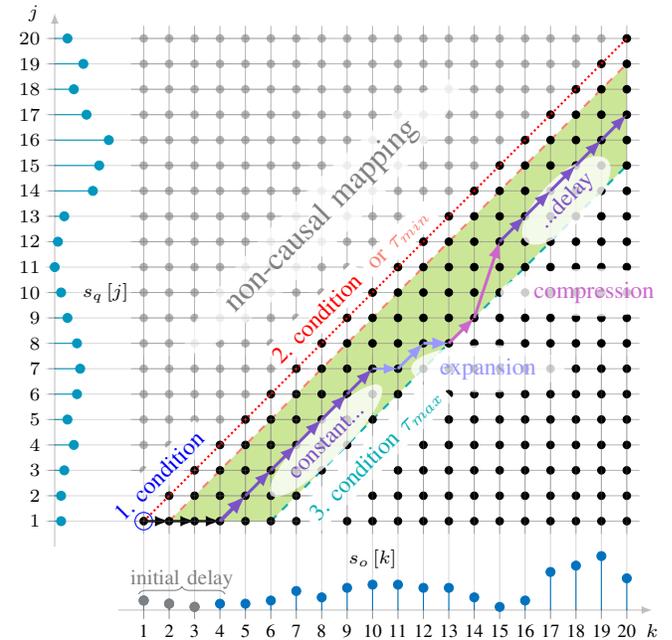


**Fig. 3:** The state  $\sigma_k^j$  thereby means “ $s_q[j]$  is aligned to  $s_o[k]$ ”. State transitions put two sequential alignments into relation.

sensors (e.g. gas may stand still in the pipes if no gas is taken out, the flow-direction may turn if, e.g., a valve is opened, etc.). Therefore, aligned sequences do not only appear constantly delayed, but additionally expanded or compressed along the time domain (see Figure 2).

In order to respect thermodynamic restrictions on the one hand, and to cope for varying delay times on the other hand, the alignment is modeled as a state machine (see Figure 3). The state  $\sigma_k^j$  thereby means “ $s_q[j]$  is aligned to  $s_o[k]$ ”. Every alignment between  $s_q$  and  $s_o$  is represented by one path  $\kappa$  through the trellis. For instance, the transition  $\sigma_k^{j_1} \rightarrow \sigma_{k+1}^{j_2}$  means “ $s_q[j_1]$  is aligned to  $s_o[k]$  and  $s_q[j_2]$  is aligned to  $s_o[k+1]$ ” resulting in successive entries in the path vector  $\kappa[k] = j_1$  and  $\kappa[k+1] = j_2$ .

Rolling out the state-machine over time leads to a trellis as shown in Figure 4. Hence, the optimal path through the trellis leads to an optimal alignment. Physical constraints translate naturally to path conditions:



**Fig. 4:** Rolling out the state transition diagram in Figure 3 over time, for the alignment of the time series  $s_q$  to  $s_o$  of figure 2, yields a trellis as shown in this figure. Any path  $\kappa_q$  through the trellis renders a possible alignment between  $s_q$  and  $s_o$ . Mapping of single features from  $s_q$  to multiple features of  $s_o$  (expansion – Figure 2), corresponds to horizontal steps along the path in the trellis. Vice versa is not feasible. Multiple features from sequence  $s_q$  can be omitted instead (compression – Figure 2) since counterparts in the sequence  $s_o$  are missing. For every omission the path  $\kappa[k]$  is shifted by one vertical state upwards. All paths are restricted to the band (light green window) enclosed by the transit times boundaries  $\tau_{min} \leq \tau[k] \leq \tau_{max}$ , i.e. condition 2. and 3.

1. Alignment starts with the index tuple  $(1, 1)$ .
2. A lower bound transit time must be considered for causality, i.e.  $\tau[k] \geq \tau_{min}$ , with  $\tau_{min} \geq 0$ . The red dotted line in Figure 4 represents  $\tau_{min} = 0$  and the light red dashed line  $\tau_{min} > 0$ .

3. An upper bound  $\tau_{max}$  of gas transit times can be expected to exist as long as gas is regularly consumed on the downstream node, so that  $\tau[k] \leq \tau_{max}$ .

The squared distance

$$d(s_q[j], s_o[k]) = (s_q[j] - s_o[k])^2 \quad (2)$$

with  $j = k - \tau_q[k]$ , represents a measure for the quality of alignment between the  $j$ -th element of  $s_q$  and the  $k$ -th of  $s_o$ , corresponding to the state  $\sigma_k^j$ . Indexes  $j$  of all alignments to the corresponding index  $k$  are stored in a path matrix, denoted  $\Psi$ . For all tuples  $(k, n)$  within a path  $\kappa_q[k] = n$  of transitions, the overall cost must be minimized for optimal alignment. By defining an accumulated cost or distance matrix  $\delta_q[n, k] \in \mathbb{R}_{\geq 0}^{N \times M}$ , an optimal path  $\kappa_q[k]$  of alignment between  $s_q$  and  $s_o$ , within the band defined by the constraints  $\xi$  in Eq. 9 and  $\gamma$  in Eq. 10 (light green area in figure 4), can be found utilizing the Viterbi algorithm following the steps 1 to 4:

- 1) Initialization

$$\delta[j, k] = \infty \quad (3)$$

$$\text{for } 1 \leq j \leq N \text{ and } 1 \leq k \leq M \quad (4)$$

$$\delta[1, 1] = d(s_q[1], s_o[1]) \quad (5)$$

$$\psi[1, 1] = 0 \quad (6)$$

- 2) Induction

$$\delta[j, k] = \min_{\xi \leq \lambda \leq \gamma} \delta[\lambda, k-1] + d(s_q[j], s_o[k]) \quad (7)$$

$$\psi[j, k] = \arg \min_{\xi \leq \lambda \leq \gamma} \delta[\lambda, k-1] \quad (8)$$

(Only paths within the band constrained by  $\xi$  in Eq. 9 and  $\gamma$  in Eq. 10 are allowed)

$$\xi = \min(N, \max(1, [k - \tau_{max} + 1])) \quad (9)$$

$$\gamma = \min(N, \max(1, [k - \tau_{min} + 1])) \quad (10)$$

$$\text{for } \xi \leq j \leq \gamma \text{ and } 2 \leq k \leq M$$

- 3) Termination

$$\kappa[M] = \arg \min_{\xi \leq \lambda \leq \gamma} \delta[\lambda, M] \quad (11)$$

$$\text{for } k = M$$

- 4) Path Backtracking

$$\kappa[k] = \psi[\kappa[k+1], [k+1]] \quad (12)$$

$$\text{for } k = M-1, M-2, M-3, \dots, 1$$

Conditions 2, i.e.  $\tau[k] \geq \tau_{min}$ , and 3, i.e.  $\tau[k] \leq \tau_{max}$ , displayed in Figure 4, limit the number of cells that are evaluated in the accumulated cost matrix  $\delta$ . That speeds up the Viterbi algorithm through complexity reduction. Similar constraints are widely used in dynamic time warping. Two of the most commonly used bounds are the Sakoe-Chiba Band [6] and the Itakura Parallelogram [7].

In the induction step Eq. 7 the overall costs for each possible

path of transitions beginning in the initialization (5) to the actual tuple  $(j, k)$  are stored in the accumulated cost matrix  $\delta[j, k]$ . In Eq. 8 only those indexes of  $j$  for  $k-1$  are stored in the path matrix  $\psi[j, k] \in \mathbb{N}_0^{N \times M}$  which exhibit minimum overall cost within  $\delta[\xi \leq j \leq \gamma, k-1]$ . By locating the minimal overall cost in the  $M$ th column of accumulated cost matrix  $\delta$  (termination 11) the row index  $\kappa[M]$  for starting backtracking in 12 is being set.

After backtracking transit times  $\tau[k]$  for the source  $q$  to the node  $o$  can be derived:

$$\tau[k] = k - \kappa[k], \quad 1 \leq k \leq M \quad (13)$$

### C. Gas tracking and calorific value determination

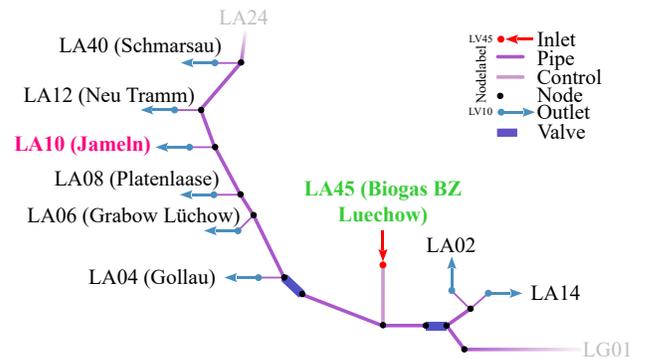
Once the transit times vector  $\tau$  for a pair of nodes  $(q, o)$  has been estimated, gas and therefore calorific value tracking can be performed. At gas grid entry points calorific values  $H_{s,q}$  are hourly measured by calibrated PGC. By utilization of  $\tau$  the calorific values in  $H_{s,q}$  sampled on the node  $q$  can be extrapolated with

$$H_{s,o}[i] = H_{s,q}[i - \tau[i]] \quad (14)$$

to succeeding sensor locations  $o$ , with  $i := [1 : M]$ , to derive the vector  $H_{s,o}$ . Then, CV tracking for customer invoicing is feasible. This is presented in the next section III.

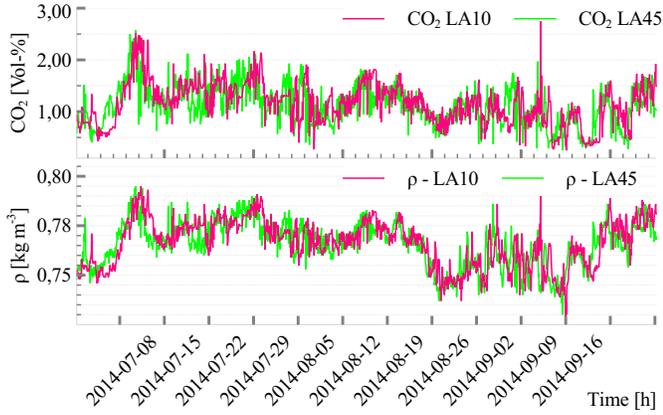
## III. RESULTS

Gas features, the CO<sub>2</sub> content in Vol.-% and the gas density (denoted  $\rho$ ) in  $\frac{kg}{m^3}$  taken for transit times estimation (see Figure 6), as well as the calorific values taken for validation and extrapolation, had been sampled at the same thermodynamic conditions on two source nodes (denoted LA45 and LG01) and one exit node (denoted LA10), in the natural gas grid Lüchow (Germany) operated by e.on Avacon AG (see Figure 5). The data had been recorded in the period from 2014-07-01 06:00 to 2014-09-30 05:00. The evaluation of our approach,

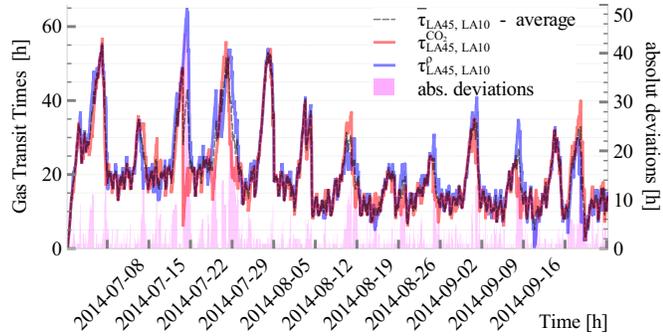


**Fig. 5:** Signals based on sampled gas features obtained in the gas grid topology of Lüchow (Germany), have been successfully tracked from the entry point Biogas BZ (LA45) (green bold characters) to the exit Jameln (LA10) (pink bold characters), corresponding to a distance of about 11 kilometers, utilizing our method

presented in this section, is performed by consideration of one dominant source  $q$  ( $= LA45$ ) only. Partial contributions of any other sources in the grid (e.g.  $q = LG01$ ) to the sampled gas on the exit node  $o$  ( $= LA10$ ), are neglected. Software based gas



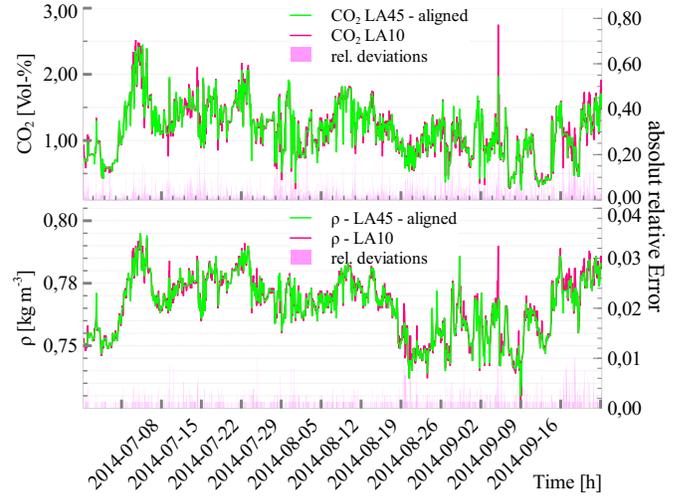
**Fig. 6:** Gas features, the CO<sub>2</sub> content in Vol.-% (top) and the gas density in  $\frac{kg}{m^3}$  denoted  $\rho$  (bottom), sampled at the gas-to-grid plant Biogas BZ Luchow (LA45) and the exit Jameln (LA10) (see Figure 5). Timescale: 91 days, 2014-07-01 06:00 to 2014-09-30 05:00.



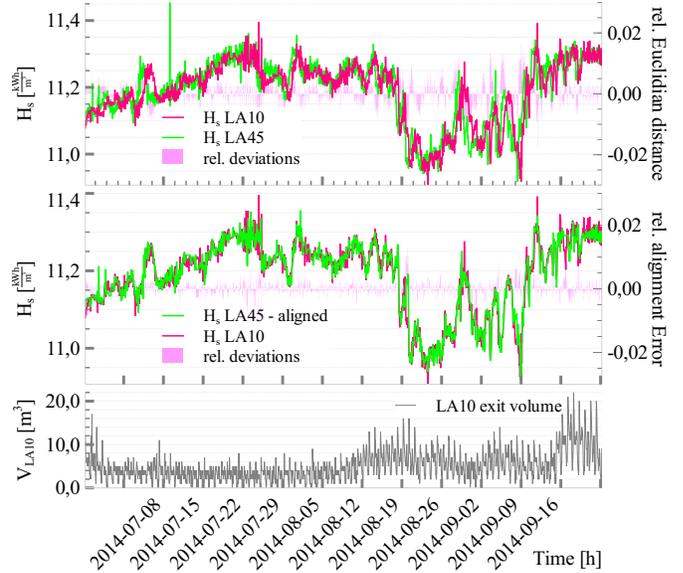
**Fig. 7:** By aligning signals of the same feature category from the source node LA45 onto signals sampled on the exit LA10, transit times  $\tau_{LA45, LA10}^{CO_2}$  and  $\tau_{LA45, LA10}^{\rho}$  can be estimated. As can be seen, results for both feature categories are slightly deviating. Thereby obtained transit times can afterwards be averaged to one single transit times vector  $\bar{\tau}_{LA45, LA10}$ , needed for extrapolation of calorific values  $H_{s, LA45}$  from the source  $q = LA45$  to the exit node  $o = LA10$ .

tracking (SmartSim[1]) provided valid transit time boundaries  $\tau_{min}$  and  $\tau_{max}$  (see Section II-B), as well as the assurance that the exit node  $o$  has almost exclusively been fed by the biogas inlet  $q$  within mentioned period. In order to estimate transit times of gas between  $q$  and  $o$ , signals of the same gas feature category, i.e. either the sampled CO<sub>2</sub> content or the sampled gas density  $\rho$ , are optimally aligned onto each other utilizing the Viterbi algorithm as mentioned in Section II-B. Resulting transit times are presented in the Figure 7. With obtained transit times vectors, the source signals can be warped along the time axis for error determination between the warped source and sampled signals of the exit node. The warped source signals with rel. deviations are presented in Figure 8. Derived transit times vectors  $\tau_{LA45, LA10}^{CO_2}$  and  $\tau_{LA45, LA10}^{\rho}$ , since they are deviating, are averaged into one single transit times vector  $\bar{\tau}_{LA45, LA10}$  needed for extrapolation of calorific values from the source to the exit, the gas customer, node with Eq. 14, so that

$$H_{s, LA10} [i] = H_{s, LA45} [i - \bar{\tau}_{LA45, LA10} [i]] \quad (15)$$



**Fig. 8:** As a measure for the quality of alignments between source and exit signals (CO<sub>2</sub> (top) and  $\rho$  (bottom)), and the transit times estimation, in this figure the extrapolated signals from the source (LA45) are presented on top of the signals sampled on the exit node (LA10) by calibrated measurements. Additionally the relative deviations (errors) between aligned and sampled exit node signals are shown.



**Fig. 9:** With estimated average transit times  $\bar{\tau}_{LA45, LA10}$ , derived through our proposed approach, the calorific values from the source  $H_{s, LA45}$  have been extrapolated to the downstream node (LA10) to obtain  $H_{s, LA10}$ , the estimated calorific values on the exit node. In the top diagram, both, the source as well as the exit node calorific values sampled by calibrated PGC, are presented together with the relative errors (euclidian distance, divided by the reference signal  $H_{s, o, PGC}$ ). In the diagram below, the extrapolated source (LA45), aligned with  $\bar{\tau}_{LA45, LA10}$ , as well as the measured calorific values  $H_{s, o, PGC}$  from the exit (LA10), and the deviations (errors), are presented. Additionally in the bottom diagram, the volumes of ejected gas on the exit node in the same period, is presented. The volumes are taken for the determination of monthly volume weighted deviations with Eq. 18.

for  $i := [1 : M]$  (see Figure 9). For validating (see Tab. I) our approach, as a measure for the quality of alignments between source and extrapolated exit node signals, for the CO<sub>2</sub> signal

and the normalized density  $\rho$  as well as for the calorific values  $H_s$ , we provide the Root Mean Square Deviation.

$$\text{RMSD} = \mathbf{E} \left[ (H_{s,o} - H_{s,o,\text{PGC}})^2 \right]^{\frac{1}{2}} \quad (16)$$

In Eq. 16 the extrapolated CVs from the source (LA45) are denoted  $H_{s,o}$  and the calibrated measurements from the exit node (LA10) are denoted  $H_{s,o,\text{PGC}}$  (see Eq. 15). The same equation 15 taken for CV extrapolation, can be taken for extrapolation of the  $\text{CO}_2$  and  $\rho$  source signals. In addition, for all sampled signals we give the Normalized RMSD yielding

$$\text{NRMSD} = \frac{\text{RMSD}}{\mathbf{E}[H_{s,o,\text{PGC}}]} \quad (17)$$

And for the calorific values only, we provide the monthly volume weighted deviations between extrapolated and measured CVs, which are of great relevance with respect to gas customers invoicing.

$$\bar{H}_V = \frac{\sum_{i=1}^M (H_{s,o,\text{PGC}}[i] - H_{s,o}[i]) \cdot V_{\text{LA10}}[i]}{\sum_{i=1}^M H_{s,o,\text{PGC}}[i] \cdot V_{\text{LA10}}[i]} \quad (18)$$

**TABLE I:** For all sampled gas feature signals the RMS deviation and the normalized RMSD in % between the extrapolated source and calibrated exit node signals are provided. For the calorific values the monthly volume weighted deviations in % between extrapolated CVs and the reference CVs measured by PGC demonstrate the in average high accuracy of the gas tracking approach presented in this paper.

Measure	Month	$\text{CO}_2$		
		July	August	September
<b>RMSD</b> [Vol-%]		0.114	0.09	0.165
<b>NRMSD</b> [%]		8.501	8.426	17.31
Measure	Month	Normalized gas density - $\rho$		
		July	August	September
<b>RMSD</b> $\left[ \frac{\text{kg}}{\text{m}^3} \right]$		0.002	0.003	0.004
<b>NRMSD</b> [%]		0.266	0.324	0.494
Measure	Month	Calorific values - $H_{s,\text{LA10}}$		
		July	August	September
<b>RMSD</b> $\left[ \frac{\text{kWh}}{\text{m}^3} \right]$		0.020	0.021	0.027
<b>NRMSD</b> [%]		0.180	0.192	0.221
<b>Vol-Weighted Dev.</b> [%]		0.016	0.023	0.003

#### IV. CONCLUSION

As the experiments in sec. III show, our method, alternatively to gas flow simulations, enables gas and CV tracking with high accuracy. In a future approach, additionally to what we have already presented in this publication, a source separation step will be introduced for enabling the decomposition of superimposed sources, sampled in gas mixtures on exit nodes, therefore generally improving our method over complex topologies. Additionally, the signal transmitting properties of gas piping will be investigated and discussed.

#### ACKNOWLEDGMENT

This work is supported by grand Ka. 15 49 TG 78 StMWi, Bavaria.

#### REFERENCES

- [1] J. Schenk, P. Schley *et al.*, "A new method for gas quality tracking in distribution grids," *Gas for Energy*, vol. 3, pp. 34–42, 2012.
- [2] S. Rickelt, P. Schley *et al.*, "Gas quality tracking with smartsim energy billing in gas grids with multipoint injection," *Gas for Energy*, vol. 2, pp. 24–32, 2016.
- [3] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inf. Theor.*, vol. 13, no. 2, pp. 260–269, Sep. 2006.
- [4] G. D. Forney, "The viterbi algorithm," *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, Mar. 1973.
- [5] A. D. McNaught and A. Wilkinson, "Compendium of chemical terminology," International Union of Pure and Applied Chemistry (IUPAC), Tech. Rep., 1997.
- [6] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 43–49, Feb 1978.
- [7] F. Itakura, "Minimum prediction residual principle applied to speech recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 23, no. 1, pp. 67–72, Feb 1975.