Data Fusion and State Estimation Using Belief Propagation in Gas Distribution Networks

Goekhan Demirel†*, Steven de Jongh†, Felicitas Mueller† and Thomas Leibfried†
†Institute of Electrical Energy Systems and High Voltage Technology (IEH),
Karlsruhe Institute of Technology, Karlsruhe, Germany
Email: {steven.dejongh, felicitas.mueller, thomas.leibfried}@kit.edu
*Institute for Automation and Applied Informatics (IAI),
Karlsruhe Institute of Technology, Eggenstein-Leopoldshafen, Germany
Email: {goekhan.demirel}@kit.edu

***Abstract*—This paper proposes a solution to the state estimation problem in gas networks using the distributed belief propagation (BP) algorithm. Power system identification applications require precise and robust state estimatiors as well as various sensor information. Compared to augmenting the power system with a very large number of sensors, a limited number of sensors and probabilistic graphical models can be used to infer the system state and reduce hardware investments. A novel BP algorithm propagates the pressure quantities at nodes in the gas network based on pressure manometer signals and applies a correction based on the information of neighboring nodes in the fusion step by using additional supporting sensors. Finally, the data fusion algorithm is demonstrated for a 14-node gas distribution network based on real data. This paper presents a novel algorithm aimed at tackling the traditional weighted least squares method to validate the developed novel approach in order to highlight the advantage of the distributed inference algorithm over traditional methods.**

***Index Terms*—Belief propagation (BP), data fusion, distribution network, gas, probabilistic graphical model (PGM), state estimation (SE).**

I. INTRODUCTION

Almost all approaches to state estimation (SE), a data fusion problem, still rely heavily on observability of the distribution network. With the rise of distributed renewable energy generation and more complex heterogeneous sensor networks, there has been a surge of interest in data fusion and processing strategies for monitoring global system behavior. This increased interest in the new distributed network component is triggering a significant change in the overall energy infrastructure by introducing new uncertainties and new varying operating conditions. In distribution networks, especially in medium- and low-pressure networks, measurement equipment is almost non-existent. Current solution techniques, such as the Newton method, require many measurement equipments at each node in the network to ensure safe operation and observability. Moreover, the alternative weighted least squares (WLS) method, which is a traditional solution approach for the SE problem, leads to instability for very sparsely distributed measurement configurations and requires the regularized weighted least squares (RWLS) approach for robust SE. A solution to the aforementioned problems is provided by belief propagation (BP) algorithm and is represented by probabilistic graphical models (PGMs). Due to its robustness, the BP algorithm allows intelligent operation with low investments and can be applied in networks with low number of measurements. The present paper applies factor graphs in the SE problem, which are realized by the distributed BP algorithm. The proposed BP algorithm allows to use different types of measurements and shows high robustness even for ill-conditioned systems.

The SE problem is generally solved by the maximum *a posteriori* method using a distributed algorithm (sum-product or max-product algorithms) in a BP approach, as demonstrated in [1], [2] and [3]. There are few studies that solve the SE problem using probabilistic inference approaches in an electric power system. In the pioneering study [2], the BP algorithm was, for the first time, applied to the distribution networks to estimate the system state as a stochastic variable. To bridge the gap between the traditional energy management system (EMS) and modern microgrids that include distributed energy sources, a real-time state estimator based on smart and flexible solutions was employed in [3]. Here, historical data and loads were implemented into the model. In particular, [2] and [3] exhibited the use of factor graphs and linearization techniques to realize the BP through a linear approximation of the nonlinear functions. BP was applied to real-time SE for the DC model in [4], while in [5], an extended DC model that contributes to the establishment of BP in decentralized electrical networks was analyzed. The potential of the BP framework in energy networks was further expanded in [6]. It guarantees the same accuracy as the centralized algorithm while retaining the advantages of the BP algorithm. Fusion of heterogeneous measurement data was performed in [7], and associated measurement models based on the BP were applied.

In the present work, BP methods are used to solve the SE problem of highly nonlinear gas distribution networks. In a gas distribution system, state variables represent the pressure variables at the network nodes. In contrast to previous applications, where BPs were considered for SE problems in

This paper was written when authors were at the Institute of Electrical Energy Systems and High Voltage Technology, Karlsruhe Institute of Technology, Germany.

G. Demirel is now with the Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, Germany.

electrical networks only, this paper applies data fusion and SE using BP in PGM-based gas distribution networks. To verify the performance on a case study, the resulting SEs are applied to a small test system based on real-world data, and the accuracies of the estimation are compared.

II. SE ALGORITHM

This Section II has been organized as follows: In the first Section II-A, the novel BP algorithm and general characteristics of the WLS method are explained; the Section II-B presents a discussion on the structure of the factor graphs and the required information as well as the calculations of the novel Gaussian belief propagation (GaBP) algorihtm; the concluding Section II-C details the derivation of the pressure model, the application of the pressure model to the measurement functions along the pipelines and the associated assumptions required to solve the SE problem.

A. SE in Gas Networks

Bayesian fusion [1] is the basic theory used for the consistent fusion of many uncertain data sources:

$$\underbrace{p(x|z)}_{\text{Posterior}} = \underbrace{\frac{\overbrace{p(z|x)}^{\text{Likelihood}} \cdot \overbrace{p(x)}^{\text{Prior}}}{p(z)}}_{\text{Normalization constant}} \tag{1}$$

It consists of the *a posteriori* probability, which describes an unknown state x, given measurements z, likelihood function, *a priori* probability $p(x)$ and normalization constant $p(z)$. The normalization constant for the accumulated conditional probability of the *a posteriori* probability over all outcomes of z normalizes to one.

The measurement model of a sensor describes the relationship of state variables x with measurements z by a nonlinear mapping:

$$\mathbf{z} \ = h(\mathbf{x}) + \ \epsilon \tag{2}$$

where vector $\mathbf{x}$ takes n system states as system state vector $\mathbf{x} = [x_1, \ldots, x_n]^T$; the measurement function is a general nonlinear known function $h(\mathbf{x}) = [h_1(\mathbf{x}), \ldots, h_k(\mathbf{x})]^T$, which is used to denote the measurement vector $\mathbf{z} = [z_1, \ldots, z_k]^T$ for k measurements; and the zero-mean uncorrelated Gaussian measurement error vector is represented by $\epsilon = [\epsilon_1, \ldots, \epsilon_k]^T$ in the SE problem. In general, the SE problem is an overdetermined system, leading to $k > n$ [8]. The goal of the SE algorithm is to find the system state vector $\mathbf{x}$ by using the network topology, network parameters and measurements $\mathbf{z}$ to determine the best estimate for the unknown state variable $\mathbf{x}$.

Assume that a measurement dataset $\mathcal{M}$ of different measurements $\mathbf{z} = (z_1, \ldots, z_k)^T$ or $\mathcal{M}_i \in \mathcal{M}$ is given in the form of a measurement vector $\mathbf{z}$, which is obtained from the measurement function $h_i(\mathbf{x})$ and the error ϵ_i for i-th measurement $i \in k = 1, \ldots, k$. All system variables are initially defined by their mean μ and variance σ^2 values. For variables that are not part of the measurement set $\mathcal{M}$, this variance is high, since no information about their actual state is present. A solution to the WLS problem is given for the case when the diagonal weighting matrix $\mathbf{W}$ is equivalent to the inverse covariance matrix of the sensors, i.e., $\mathbf{W} = \text{diag}\left(\frac{1}{\sigma^2_{z_1}}, \ \ldots, \frac{1}{\sigma^2_{z_k}}\right) \overset{!}{=} \mathbf{\Sigma}^{-1}$ to the WLS method. WLS is equal to the maximum likelihood estimation (MLE) and can be expressed as follows:

$$\hat{\mathbf{x}} = \underbrace{\max_{\mathbf{x}} \prod_{i=1}^{k} \mathcal{N}\left(z_i|\mathbf{x}, \sigma^2_{z_i}\right)}_{\text{MLE}} = \underbrace{\min_{\mathbf{x}} \sum_{i=1}^{k} \frac{[z_i - h_i(\mathbf{x})]^2}{\sigma^2_{z_i}}}_{\text{WLS}} \tag{3}$$

B. Factor Graphs and BP Algorithm

The SE problem can be described analogously by using factor graphs, and it can be calculated using the BP algorithm. The goal of the BP algorithm is to efficiently compute the marginal values in a system of random variables $\mathbf{x} = [x_1, \ \ldots, x_n]^T$ described by the composite probabilities $p(\mathbf{x}|\mathbf{z})$. Assuming that the function $p(\mathbf{x}|\mathbf{z})$ can be factorized to a product of local functions

$$p(\mathbf{x}|\mathbf{z}) = \prod_{i=1}^{k} \psi_i(\mathcal{V}_i) \tag{4}$$

with $\mathcal{V}_i \subseteq \{x_1, \ \ldots, x_n\}$, the estimation problem can be solved efficiently with the BP algorithm. A factor graph is formed in the first step, which describes a bipartite graph. The structure of the factor graph comprises the set of factor nodes $\mathcal{F} = \{f_1, \ \ldots, f_k\}$, where each factor node f_i represents a local function $\psi_i(\mathcal{V}_i)$, and the set of variable nodes $\mathcal{V} = \{x_1, \ \ldots, x_n\}$. The factor node f_i connects to the variable node x_s only if each $x \in \mathcal{V}_i$ [9]. Hence, a factor graph represents a graphical model that allows graph-based representation of probability density functions.

1) Factor graph construction: For the purpose of clarity, the important components of a factor graph model are explained subsequently. The system state variable of the gas networks is given by $\mathbf{x} = \mathbf{p}^T$. The set of factor nodes $\mathcal{F} = \{f_1, \ldots, f_k\}$ is defined by the set of measurements $\mathcal{M}$, which includes the measurement functions of the gas networks. Each factor node $f_i \in \mathcal{F}$ is associated with a subset of the variable nodes $x_s \in \mathcal{V}$ if and only if the state variable x_s has a probabilistic dependence on the corresponding measurement functions $h_s(x_s)$. The number of variable nodes is described by state variables of gas networks $\mathcal{V} = \{p_1, \ldots, p_n\} \equiv \{x_1, \ldots, x_n\}$, the values of which are to be estimated but are not always directly available or observable. In this process, the subset of variable nodes associated with a given factor node f_i is denoted by $\mathcal{V}_\text{i} \subseteq \mathcal{V}$ [9]. Bipartite linkage patterns of a factor graph decomposes the factorizability structure of the whole probabilistic model, so that all factor nodes $f_i \in \mathcal{F}$ are independent of each other.

The i-th factor nodes or local functions $\psi_i(\mathcal{V}_i)$ are defined as follows:

$$\psi_i(\mathcal{V}_i) \ = \ Z \cdot \exp -\frac{1}{2} \cdot [(\mathbf{z} - \mathbf{h}(\mathcal{V}_i))^T \mathbf{\Sigma}(\mathbf{z} - \mathbf{h}(\mathcal{V}_i)] \tag{5}$$

This expression represents the probability of obtaining a measurement vector $\mathbf{z}$ from the distributed sensor. The factor node f_i is a function of the set of variables involved $\mathcal{V}_i$ and, thus, a subset of the total state variables $\mathcal{V}$. The function is in the form of a quadratic exponential function with a normalization factor Z to normalize the probability distribution, the value of which need not be calculated. The measurement z_i defines local likelihood functions $\mathcal{N}(z_i|\mathbf{x}, \sigma_{z_i}^2)$ resp. $\mathcal{L}(z_i|\mathbf{x})$, which in turn are equal to the local functions $\psi_i(\mathcal{V}_i)$ associated with factor nodes f_i. The number of variable nodes is defined by the state variables of the gas network $\mathbf{x} = \mathbf{p}$. Consequently, the multivariate normal distribution over all variable nodes is the product of all factor nodes in (4).

In summary, the following three information properties are necessary for the construction of a Gaussian factor graph: a) the measurement functions $\mathbf{h}(\mathcal{V}_i)$ that depend on the local state variables $\mathcal{V}_i$, b) the measurements $\mathbf{z}$ from the distributed sensors and c) the symmetric covariance matrix $\mathbf{\Sigma}$ of the measurements $\mathbf{z}$.

2) BP algorithm: The BP algorithm efficiently computes the marginal distribution of the state variables by passing two types of message operations along the factor graph edges: i) message type $\mu_{x_s \to f_i(x_s)}$ from a variable node $x_s \in \mathcal{V}$ to a factor node $f_i \in \mathcal{F}$ and ii) message type $\mu_{f_i \to x_s(x_s)}$ from a factor node $f_i \in \mathcal{F}$ to a variable node $x_s \in \mathcal{V}$. Under the assumption of Gaussian message types (consisting of mean and variance), the algorithm is called GaBP.

Message type $\mu_{x_s \to f_i}(x_s)$ from a variable node to a factor node can be written as:

$$\mu_{x_s \to f_i}(x_s) \;=\; \prod_{f_a \in \mathcal{F} \setminus f_i} \mu_{f_a \to x_s}(x_s) \tag{6}$$

and is equivalent to the product of all incoming messages from factor nodes to variable nodes that arrive at all other adjacent edges. It should be noted that $\mathcal{F}_s \setminus f_i$ describes the set of all adjacent factor nodes of variable node x_s, where f_i is excluded from the product. Moreover, each message $\mu_{f_a \to x_s}(x_s)$ describes a function of the variable x_s.
The message in (6) is proportional to a multivariate normal distribution:

$$\mu_{x_s \to f_i(x_s)} \;\propto\; \mathcal{N}(x_s | z_{x_s \to f_i},\; \sigma_{x_s \to f_i}^2) \tag{7}$$

with mean $z_{x_s \to f_i}$ and variance $\sigma_{x_s \to f_i}^2$ calculated as:

$$z_{x_s \to f_i} = \left(\sum_{f_a \in \mathcal{F}_s \setminus f_i} \frac{z_{f_a \to x_s}}{\sigma_{f_a \to x_s}^2} \right) \sigma_{x_s \to f_i}^2 \tag{8a}$$

$$\frac{1}{\sigma_{x_s \to f_i}^2} \;=\; \sum_{f_a \in \mathcal{F}_s \setminus f_i} \frac{1}{\sigma_{f_a \to x_s}^2} \tag{8b}$$

After the variable node x_s receives the messages from all adjacent factor nodes from the set $\mathcal{F}_s \setminus f_i$, it evaluates the message $\mu_{x_s \to f_i}(x_s)$ according to (8b) and sends the parameters of the normal distribution to the factor node f_i.

Message type $\mu_{f_i \to x_s}(x_s)$ from a factor node f_i to a variable node x_s is defined as a product of all incoming messages from the variable node to factor node, arriving from other incident edges, multiplied by the function $\psi_i(\mathcal{V}_i)$ associated with the factor node f_i. Marginalization over the number of connected variable nodes L or the number of nodes connected to the incoming messages is required as well:

$$\mu_{f_i \to x_s}(x_s) \;=\; \underbrace{\int_{x_l} \ldots \int_{x_L}}_{\text{Marginalization}} \psi_i(\mathcal{X}_i) \prod_{x_b \in \mathcal{V}_i \setminus x_s} (\mu_{x_b \to f_i}(x_b) \cdot dx_b) \tag{9}$$

where $\mathcal{V}_i \setminus x_s$ represents the set of all adjoint variable nodes of x_s, and the variable node x_s is excluded from the product. Due to the linearity of the measurement functions, closed-form expressions for these messages are possible and follow a normal distribution form:

$$\mu_{f_i \to x_s}(x_s) \;\propto\; \mathcal{N}(x_s | z_{f_i \to x_s},\; \sigma_{f_i \to x_s}^2) \tag{10}$$

This message type can only be calculated if all incoming variable-to-factor messages, i.e., the node pressures, are known. That can be achieved by a synchronous scheduling of the messages. Assuming that there are Gaussian messages in the factor nodes, the corresponding Gaussian function of the factor node f_i is given as

$$\mathcal{N}(z_i | x_s,\; x_l,\; \ldots,\; x_L, \sigma_{z_i}^2) \propto \exp\left\{ \frac{[z_i - h_i(x_s,\; x_l,\; \ldots,\; x_L)]^2}{2 \cdot \sigma_{z_i}^2} \right\} \tag{11}$$

where $h_i(\mathbf{x})$ represents the linear measurement functions. Measurements are processed by linearizing the measurement function. The message types as well as the marginalization can be calculated according to (8b), (12b), and (15b).

Message type $\mu_{f_i \to x_s}(x_s)$ from the factor node f_i to the variable node x_s is represented by the normal distribution function in (10). The mean $z_{f_i \to x_s}$ and the variance $\sigma_{f_i \to x_s}^2$ of the message type are calculated as

$$z_{f_i \to x_s} = \frac{1}{C_{x_s}} \left(z_i - \sum_{x_b \in \mathcal{V}_i \setminus x_s} C_{x_b} z_{x_b \to f_i} \right) \tag{12a}$$

$$\sigma_{f_i \to x_s}^2 = \frac{1}{C_{x_s}^2} \left(\sigma_{z_i}^2 - \sum_{x_b \in \mathcal{V}_i \setminus x_s} C_{x_b}^2 \sigma_{x_b \to f_i}^2 \right) \tag{12b}$$

where coefficients C_{x_s} with $x_s \in \mathcal{V}_i$ describe the Jacobian elements of the measurement function associated with the factor node f_i:

$$C_{x_s} = \frac{\partial h_i(x_s,\; x_l,\; \ldots, x_L)}{\partial x_s} \tag{13}$$

After the factor node f_i receives the messages from all neighboring variable nodes from the set $\mathcal{V}_i \setminus x_s$, it evaluates the message $\mu_{f_i \to x_s}(x_s)$ according to (12b) and sends the parameters of the normal distribution to the variable node x_s.

Marginal inference of the state variable can be written as

$$p(x_s) \propto \; \mathcal{N}(x_s | \hat{x}_s, \hat{\sigma}_{x_s}^2) \tag{14}$$

where mean $\hat{x}_s$ and variance $\hat{\sigma}^2_{x_s}$ can be calculated using the following equations:

$$\hat{x}_s = \left(\sum_{f_c \in \mathcal{F}_s} \frac{z_{f_c \to x_s}}{\sigma^2_{f_c \to x_s}} \right) \sigma^2_{x_s} \tag{15a}$$

$$\frac{1}{\hat{\sigma}^2_{x_s}} = \sum_{f_c \in \mathcal{F}_s} \frac{1}{\sigma^2_{f_c \to x_s}} \tag{15b}$$

Finally, the mean $\hat{x}_s$ is expressed as the estimated value of the state variable x_s, and its uncertainty is expressed by the variance $\hat{\sigma}^2_{x_s}$.

C. Gas Network SE

A model for the pressure drop along a pipeline in gas networks is derived using the Darcy–Weisbach equation. Moreover, a measurement model based on the gas model is derived using the measurement functions and equations of the gas network state estimator. While the measurement function represents the physical gas laws of the pipe hydraulics, it also connects the measurement quantity and the state variable x.

A gas network with arbitrary topologies is given as an graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with a set of nodes $\mathcal{V} = \{1, \dots, n\}$ with $n \in \mathbb{N}^+$ and a set of edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ describes the edges $\mathcal{E} = \{1, \dots, m\}$ with $m \in \mathbb{N}^+$. The system state vector is given by

$$\mathbf{x} = [p_1, \dots, p_n]^T \tag{16}$$

The pressure drop $\Delta p_{i,j}$ along a pipeline from node i to node j can be calculated using the Darcy–Weisbach equation given by [10]:

$$|\Delta p_{ij}| = |p_i - p_j| = R_{ij}(\lambda, Z)\ \dot{V}^2_{ij} \tag{17}$$

where $\dot{V}_{ij}$ is the volume flow and the pipe resistance R_{ij} can be described by the dimensionless pipe friction coefficient λ and the compressibility factor Z. Given the initial volume flows $\dot{V}_{ij}$ and input data for gas networks, we have sufficient information to compute the pipe resistance [11]:

$$R_{ij}(\lambda, Z) = \frac{16 \cdot \lambda \cdot p_n \cdot \rho_n \cdot Z_m \cdot T_m}{\pi^2 \cdot d^5 \cdot T_n} \tag{18}$$

where d is the pipe diameter, ρ_n is the normalized fluid density, p_n denotes the normalized pressures, T_n is the normalized temperature, Z_m represents the average compressibility factor and T_m is the average temperature between nodes i and j. These calculated pipe resistances are fixed parameters for the SE problem. The pipe friction coefficient λ can be determined using the Colebrook–White equation described in [11] and [12]. Due to its implicit nature, it can only be solved iteratively based on (17). First, the gas volume flow rates can be solved using (17) as

$$\dot{V}_{ij} = \mathrm{sgn}(p_i - p_j) \sqrt{\frac{|p_i - p_j|}{R_{ij}}} \tag{19}$$

For $p_i \geq p_j$, the edge volume flow is defined as a positive value including zero. For $p_i < p_j$, the edge volume flow has negative values. Equation (19) provides the fundamental basis for the measurement function describing the measurements of pipe flows and edge volume flows in gas networks. The gas model includes a total of three measurement functions:

- Pressure measurement function at node i:

$$h_{p_i}(\cdot) = p_i \tag{20}$$

- Volume flow measurement function at edge from i to j:

$$h_{\dot{V}_{ij}}(\cdot) = \mathrm{sgn}(p_i - p_j) \cdot \sqrt{\frac{|p_i - p_j|}{R_{ij}}} \tag{21}$$

- Volume flow injection or extraction measurement function derived using Kirchhoff's first law at each node i:

$$h_{\dot{V}_i}(\cdot) = - \sum_{k \in N_{b(i)}} \mathrm{sgn}(p_i - p_k) \cdot \sqrt{\frac{|p_i - p_k|}{R_{ik}}} \tag{22}$$

where $N_{b(i)}$ is a set of adjacent nodes of node i, and k represents the adjacent nodes.

In conclusion, the measurement model presented in (21) and (22) behaves nonlinearly. According to [1] and [13], this leads to a restriction in the GaBP algorithm and, therefore, cannot be considered in the framework of linear Gaussian laws. In general, the nonlinear measurement function is developed in a Taylor series and terminated after the linear term:

$$\tilde{h}(\mathbf{x}) = h(\mathbf{x})\Big|_{\mathbf{x}=\mathbf{x}_{\mathrm{OP}}} + \frac{\partial h(\mathbf{x})}{\partial \mathbf{x}}\Big|_{\mathbf{x}=\mathbf{x}_{\mathrm{OP}}} (\mathbf{x} - \mathbf{x}_{\mathrm{OP}}) \tag{23}$$

It is assumed that the nonlinear measurement function $h(\mathbf{x})$ have a continuous character at least around the operating points $\mathbf{x}_{\mathrm{OP}}$ for suitable linearization [13]. Measurement data of the gas model $\mathcal{M} = \{M_{\dot{V}_{ij}}, M_{\dot{V}_i}, M_{p_i}\}$ contain the edge volume flow $M_{\dot{V}_{ij}}(i,j) \in \mathcal{E}$ and the node volume inflow or outflow $M_{\dot{V}_i}$ $i \in \mathcal{V}$ obtained from indirect pressure measurements such as flow meters. The node pressure M_{p_i} $i \in \mathcal{V}$ is obtained from direct pressure measurements such as manometers.

In the measurement model described in (2), an edge volume flow measurement denoted by $\mathbf{z}_{\dot{V}_{ij}}$ flows along the pipelines, $\mathbf{z}_{\dot{V}_i}$ is the node volume flow injection or extraction (feeding or loads) and $\mathbf{z}_{p_i}$ is the node pressure at the node. The measurement vectors can be represented to derive the linear measurement model as follows:

$$\underbrace{\begin{bmatrix} \mathbf{z}_{\dot{V}_{ij}} \\ \mathbf{z}_{\dot{V}_i} \\ \mathbf{z}_{p_i} \end{bmatrix}}_{\mathbf{z}:=} = \underbrace{\begin{bmatrix} \mathbf{H}_{\dot{V}_{ij}} \\ \mathbf{H}_{\dot{V}_i} \\ \mathbf{H}_{p_i} \end{bmatrix}}_{\mathbf{H}:=} \big[\mathbf{x}\big] + \underbrace{\begin{bmatrix} \epsilon_{\dot{V}_{ij}} \\ \epsilon_{\dot{V}_i} \\ \epsilon_{p_i} \end{bmatrix}}_{\epsilon:=} \tag{24}$$

where ϵ is the combined measurement error vector of the different measurement equations, and the measurement matrix $\mathbf{H}$ is obtained by calculating the Jacobian matrix from the measurement functions around the operating point at $\mathbf{x} = \mathbf{x}_{\mathrm{OP}}$. Due to the assumption of uncorrelated measurement errors, the covariance matrix of measurement uncertainty $\mathbf{\Sigma} \in \mathbb{R}^{k \times n}$ takes the form of a diagonal matrix in which only the values of the main diagonals are non-zero.

III. SIMULATION

This section presents the simulation results for the SE of an benchmark gas distribution network. First, the benchmark network and the available data are described. For the evaluation of the SE methods, three measurement configurations are presented for the benchmark network. Then, an error metric is presented to quantify the SE accuracy. Subsequently, the SE performances are analyzed and evaluated based on the measurement configurations.

A. Benchmark Network and Available Data

The performance of the proposed GaBP algorithm is evaluated based on a 14-node benchmark meshed gas distribution network. A ground truth is computed based on the network input data listed in [11]. Simulations of approved open-source network calculation software packages such as *pandapipes* [14] are used to provide the needed ground truth data. Since the measurements in reality mostly consist of noisy data, the synthetic measurement data from the simulation are perturbed with a simulated, normally distributed measurement error of the sensors. The results of the selected algorithm are then compared with the ground truth value. For each network node i, the ground truth value of the state variable is denoted by $x_{\text{i,true}}$. The system operates at low pressure (p_{N} = 3000 Pa).

B. Measurement Setup for Benchmark Network

For the evaluation of the estimation algorithm, sensor placement scenarios for the 14-node benchmark network are presented in Table I. The first two scenarios contain a number of measurements often proven in practice. This number is equal to twice the number of unknown target variables, so the measurement system has sufficient measurements to provide good estimates. The third configuration represents an ill-conditioned measurement configuration. The measurements in the three configurations are divided into two categories as shown in Fig. 1: direct measurement information at the node and indirect measurement information at the node and at the edges of the gas network. Direct measurements at the node are associated with the virtual, direct and slack pressure measurements at the node; these are shown with blue, green and yellow circles, respectively. The triangles denote indirect injection measurements at the corresponding node. The orange triangles represent an available indirect injection measurement, while the blue triangles represent a virtual injection measurement. Futhermore, squares show volume flow measurements at the edges as indirect measurements, with the blue squares representing the virtual measurements and the orange ones the available measurements.

TABLE I
OVERVIEW OF MEASUREMENT CONFIGURATION FOR SE EVALUATION

Measurement configuration	**1**	**2**	**3**
Direct measurements	14	7	1
Indirect measurements	14	21	16
Total	28	28	17

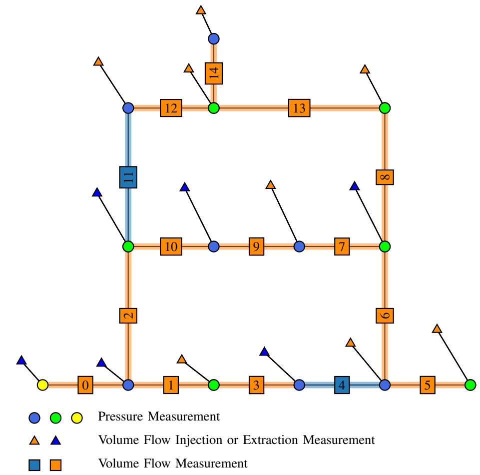

Fig. 1. 14-node network in measurement configuration 2.

C. Performance Evaluation

The following standard deviations refer to the ground truth values of the related variables. The standard deviations between the measurements at the node of the gas network are given as σ_1 for the pressure measurement and σ_2 for the volume flow measurement at the edges and nodes. In the gas network, the pressure at the slack node is controlled constantly by a pressure regulator. Hence, the pressure values are considered to be precise, leading to a very low standard deviation of σ_3. External networks represent the higher pressure network level connection and are modeled as slack nodes in both the SE and the pipe flow calculation. Since the amount of smart measurement points is limited, virtual measurements are introduced to all types of measurements, which use no prior information and take the nominal values for pressures. The values of these measurements from the nominal operating points at no *a priori* knowledge come with a large standard deviation σ_4. Randomly generated normally distributed measurement errors are estimated with the set of standard deviations $\sigma = [\sigma_1, \sigma_2, \sigma_3, \sigma_4] = \{0.01^2, 0.01, 0.01^3, 10.0^6\}$[p.u.] added to the raw data, resulting in noisy measurements. The choice of standard deviations illustrates that the true measurements are subject to only small uncertainties. To obtain a state estimate from the direct, indirect and virtual measurements, the loopy GaBP algorithm from Section II-B is applied to a linear gas model according to Section II-C. A loopy GaBP converges via a linear model to a local fixed point [15], which is a solution of an equivalent WLS problem according to (3) from section II-A. To obtain different accuracies, the SE algorithms are calculated using the mean relative estimation error per network node i:

$$\Delta \overline{x}_{\text{rel}} = \frac{1}{n} \sum_{i=1}^{n} \frac{|\Delta x_{\text{i,se}} - x_{\text{i,true}}|}{x_{\text{i,true}}} \tag{25}$$

where $x_{\text{i,se}}$ is the SE solution and $x_{\text{i,true}}$ represents the ground truth values.

D. State Estimation Results

The SE problem is solved using the GaBP algorithm in a PGM-based gas network. The data fusion algorithm is used to combine indirectly measured flow rates at the pipelines and the direct measured pressure values at the nodes. This helps obtain a pressure estimate for the gas network nodes. For the verification of the GaBP algorithm, the conventional WLS method is compared with the first and second measurement configurations.

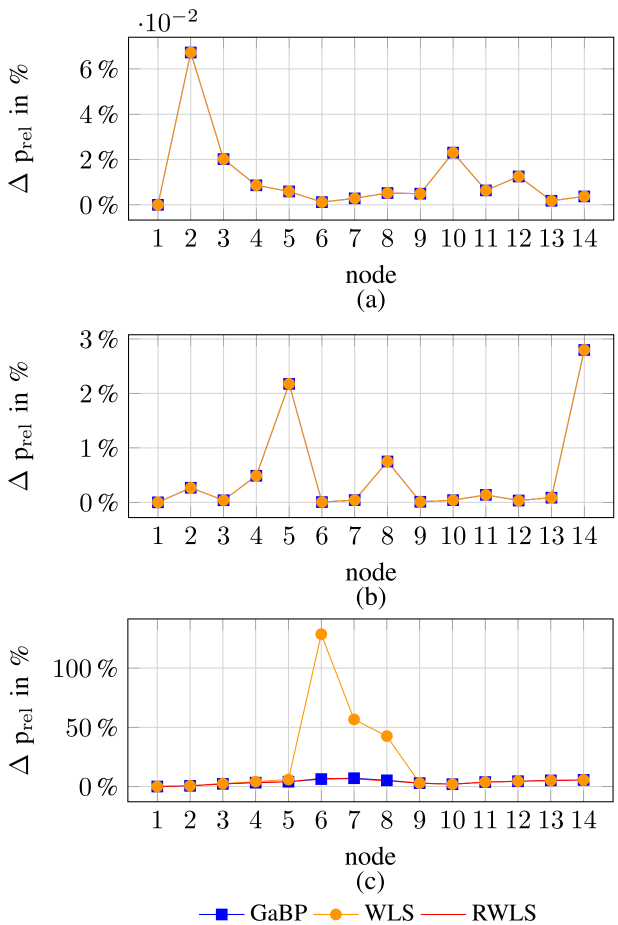

Fig. 2. Mean relative error estimated with GaBP (blue), WLS (orange) and RWLS (red) in measurement configurations 1, 2 and 3 (subfigures a, b and c).

As indicated in the legend, the blue line represents the GaBP estimated values, the orange ones the WLS estimated values, and the red ones the RWLS estimated values. The results show that both GaBP and WLS methods provide equivalent results for an identical measurement configuration, which are confirmed in theory by the common quality criteria, as evident in Fig. 2. The mean relative error per node is $\approx$ 0.01% of $x_{\text{i,true}}$. An accurate estimate is obtained with a sufficient number of measurements. The second measurement scenario also highlights that a good solution can be obtained if the number of measured node pressures is reduced by half. The mean relative error per node increases to $\approx$ 0.50% for the GaBP and WLS estimates. Since the number of direct measurements in Table I is reduced by half, from 14 to 7, the third measurement scenario also confirms the robustness of the pressure estimates of the BP algorithm implemented in this work. In ill-conditioned measurement matrix, the RWLS approach accepts information bias to benefit from a similar GaBP solution, while the conventional WLS method without regularization requires matrix inversion.

IV. Conclusion and Outlook

The paper addresses the SE problem in PGM-based gas networks using the BP algorithm. The proposed BP algorithm provides the same solution as the conventional WLS method; this is empirically shown by applying both methods on a benchmark network with a common performance criteria. Consequently, the application range and robustness of the SE solution using the BP framework can be increased. Therefore, the proposed SE solution is suitable for distribution networks. For future work, we plan to extend the proposed algorithm for multimodal energy networks. This may be of interest for the SE related to coupled networks that are largely equipped with smart meters.

References

[1] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, New York, NY: Springer, 2007.

[2] Y. Hu, A. Kuh, T. Yang and A. Kavcic, "A belief propagation based power distribution system state estimator," in *IEEE Comput. Intell. Mag.*, vol. 6, no. 3, pp. 36–46, Aug. 2011.

[3] Y. Hu, A. Kuh, A. Kavcic and D. Nakafuji, "Real-time state estimation on micro-grids," in *Int. Jt. Conf. Neural Netw.*, 2011, pp. 1378–1385.

[4] M. Cosovic and D. Vukobratovic, "Fast real-time dc state estimation in electric power systems using belief propagation," in *Proc. IEEE Smart Grid Commun.*, Oct. 2017, pp. 207–212.

[5] M. Cosovic and D. Vukobratovic, "State estimation in electric power systems using belief propagation: An extended DC model.," in *Proc. IEEE Int. Workshop on Signal Processing Adv. Wireless Commun.*, 2016, pp.1–5.

[6] M. Cosovic and D. Vukobratovic, "Distributed Gauss–Newton method for state estimation using belief propagation," in *Proc. IEEE Trans. Power Syst.*, vol. 34, 2019, pp. 648–658.

[7] F. Fusco, S. Tirupathi and R. Gormally, "Power systems data fusion based on belief propagation," in *Technol. Conf. Eur. (ISGT-Europe)*, Sep. 2017, pp. 1–6.

[8] A. Abur and A. G. Exposito, *Power System State Estimation: Theory and Implementation*. New York, NY: Marcel Dekker, 2004.

[9] F. R. Kschischang, B. J. Frey and H. -A. Loeliger, "Factor graphs and the sum-product algorithm," in *IEEE Trans. Inf. Theory*, vol. 47, pp. 498–519, Feb. 2001.

[10] Pfeifer, M. *et al.*, "Weighted least squares state estimation for coupled power and gas distribution networks," in *53rd IEEE Universities Power Engineering Conf.*, Glasgow, UK, 2018, pp. 1–6.

[11] G. Cerbe and B. Lend, *Fundamentals of Gas Technique [Grundlagen Der Gastechnik]*, 6th ed. Carl Hanser Verlag Munich, 2017.

[12] R. Eberhard und R. Hüning, *Handbook of Gas Supply Technology [Handbuch der Gasversorgungstechnik]*, 2nd ed. Oldenbourg, 1990.

[13] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.

[14] D. Lohmeier, D. Cronbach, S. R. Drauz, M. Braun, and T. M. Kneiske, "Pandapipes: An open-Source piping grid calculation package for multi-energy grid simulations," *Sustainability*, vol. 12, no. 23, p. 9899, Nov. 2020, doi: 10.3390/su12239899.

[15] Y. Weiss and W. T. Freeman, "Correctness of belief Propagation in Gaussian graphical models of arbitrary topology," *Neural Comput.*, vol. 13, no. 10, pp. 2173–2200, Oct. 2001.