

# Simultaneous calibration of relative flow metering systems within a hydraulic network

Johannes Lanzersdorfer

Andritz AG, Graz, Austria

E-mail (corresponding author): johannes.lanzersdorfer@andritz.com

---

## Abstract

The waterway of a hydroelectric power plant from intake to tailrace is nothing more than a flow network. The estimation of edge flows (e.g., fluid flow rates in conduits) based on noisy measurements in a network is an ongoing research topic in computer science with numerous applications in diverse engineering disciplines. This paper presents a numerical approach which provides edge flow estimates with respect to a reference flow. Each edge flow estimate is described by a flow function which includes initially unknown coefficients to account for the underlying measurement principle of the flow metering device. These functions can be linear or non-linear in their unknowns but shall not include an intercept. Solving a system of equations which must fulfill the continuity law at each inner vertex provides the coefficient values. A search algorithm and model evaluation criteria are introduced to find the most suitable combination of flow functions. Finally, it provides a relative in-situ calibration of all edge flow metering devices simultaneously. Flow monitoring and optimization of plant output are suitable applications of it.

---

## 1. Introduction

It is a reliable and cost-efficient approach to perform index tests on hydraulic turbomachine units either to verify power vs. total head guarantees or to determine its optimum operating range. The use of a secondary flow metering technique (e.g., Winter-Kennedy differential pressure, acoustic clamp-on flow meter and others), which at least features proper short-period reproducibility, is sufficient for this purpose. The output of these flow meters is a measurand which requires some kind of scaling to provide a flow rate quantity with values of appropriate magnitude. The scaling is commonly based on model test data or on any reference operating condition. Since index tests at a multi-units powerplant will be executed successively unit by unit, the flow metering system on each unit will be scaled independently from other units. This proceeding gives the required information as listed above, but it cannot provide reliably the relative fluid portions flowing through the individual units, which is needed to operate all units under optimum power conditions.

The numerical method described in this paper overcomes this drawback and provides flow estimations of each branch within the hydraulic network based on simultaneous indicative measurements of branch flows. The flow estimations are relative rather than absolute. The method consists of the next steps.

First, we must assign *flow functions*, which reflect the characteristics of the branch flow metering device, to each branch. This also includes some unknown coefficients, which need to be determined later. One of

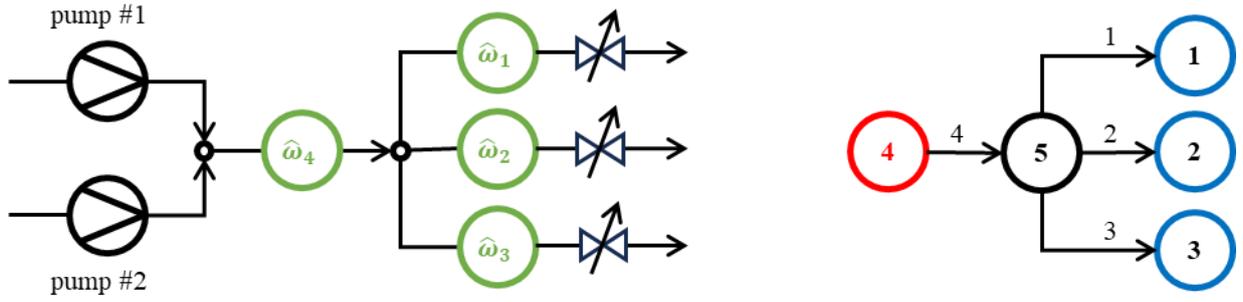
these flow meters serves as *reference*, which all calibration results of the other metering devices are related to. The reference flow is also described by a flow function but without any unknown parameters.

The flow functions are then combined to meet the network's structure and to fulfill the *continuity law* at each *inner vertex*, i.e., a vertex within a hydraulic network, which is neither source nor sink, and which is directly connected to at least two other vertices. This *model* provides an *over-determined system of equations*, which, for our purposes in doing hydraulic measurements, is mostly linear in the unknown coefficients. Solving the system of equations by the least squares method leads to unbiased estimates of the formerly unknown coefficients aside other statistical parameters.

In cases where the mathematical relationship of flow functions is unknown, finding a proper model can be a tedious task. Since the number of possible combinations of flow functions grows *exponentially* with the maximum number of unknowns per flow function, a suitable strategy is needed to find the best model. A simple *backward elimination method*, which only requires a few models to evaluate, is introduced. The best model among the evaluated ones is then chosen by a suitable *information criterion*.

Finally, the favored model undergoes a *visual and statistical examination* which includes hypothesis testing on the calculated coefficients, residual diagnostics regarding unusual pattern of observations and to identify possible outliers, and curvature inspection of the branch flow estimates and their confidence intervals.

The implementation of this method is shown next with real measurement data. Since a typical hydraulic network



**Figure 1: Hydraulic scheme (left image) with edge flow metering devices (green circles) and its translation into a flow network (right image) with source (red circle) and sinks (blue circles)**

for hydropower or pumping applications can be reduced to a single feeding line which is connected to several turbomachine units, we focus only on structures with one inner vertex. Details concerning the application on complex networks can be found in [1].

## 2. Method

The individual steps of the method should be described by using former measurement data, which have been recorded during an index test campaign on two parallel operated main cooling pumps [2]. The left image of Figure 1 provides the hydraulic scheme of this site. Two pumps in parallel arrangement deliver water into a collector pipe, which is split downstream into three edges (i.e., branches) with individually adjustable valves by a trifurcation. Upstream the valves the three edges were equipped with clamp-on flow meters providing the measured edge flow rates  $\hat{\omega}_i$  with  $i \in \{1,2,3\}$ . The collector flow rate  $\hat{\omega}_4$  could not be measured directly because the whole length of this conduit was not accessible.

### 2.1 Flow network

The translation of the hydraulic scheme into a flow network (Figure 1) reveals that we have with vertex #5 a single inner vertex ( $p = 1$ ). Furthermore, the network consists of  $n_v = 5$  vertices, i.e., one source (red circle) and three sinks (blue circles), and  $n_e = 4$  edges. First, we need to derive the incidence matrix  $\tilde{\mathbf{B}} \in \mathbb{Z}^{n_v \times n_e}$  whose elements are defined by

$$[\tilde{\mathbf{B}}]_{ij} = \begin{cases} +1, & \text{if edge } j \text{ enters vertex } i, \\ -1, & \text{if edge } j \text{ leaves vertex } i, \text{ and} \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

This gives for the network depicted on the right side of Figure 1

$$\tilde{\mathbf{B}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & -1 & -1 & 1 \end{pmatrix}. \quad (2)$$

We only need the rows belonging to inner vertices. Therefore, let us now infer the reduced incidence matrix  $\mathbf{B} \in \mathbb{Z}^{p \times n_e}$  from Equation (2) by removing all rows which contains only one non-zero element yielding

$$\mathbf{B} = \mathbf{b}_1^T = (-1, -1, -1, 1) \quad (3)$$

In this example  $\mathbf{B}$  consists of only one row which simplifies the next calculations.

### 2.2 Flow functions and reference edge flow

The reason for defining a flow function  $f_i(\hat{\boldsymbol{\omega}}^T; \boldsymbol{\beta}) \in \mathbb{R}$ , which is assigned to edge  $1 \leq i \leq n_e$ , is to provide a better approximation to the true relative edge flow rate  $q_i$  than the measured  $\hat{\omega}_i$  can do. The function arguments above, i.e.,  $\hat{\boldsymbol{\omega}}^T = (\hat{\omega}_1, \dots, \hat{\omega}_{n_e})^T$  and  $\boldsymbol{\beta} = (\beta_1 \dots \beta_m)^T$ , denote the vector of measured edge flow rates  $\hat{\omega}_i$  and the vector of unknown coefficients, respectively. Table 1 contains the parameter values of the measured edge flow rates for  $n = 9$  measuring points. The measuring points have been recorded under constant pump head but with different openings of the three downstream valves. The constant pump head also ensured that  $\hat{\omega}_4$  kept constant. Consequently, it was not needed to measure this edge flow rate, and we can set it equal to an arbitrary value, for instance,  $\hat{\omega}_4 = \hat{\omega}_{sp}$ .

**Table 1: Measured means of relative edge flow rates (source: [2])**

# (-)	$\frac{\hat{\omega}_1}{\hat{\omega}_{sp}} (-)$	$\frac{\hat{\omega}_2}{\hat{\omega}_{sp}} (-)$	$\frac{\hat{\omega}_3}{\hat{\omega}_{sp}} (-)$	$\frac{\hat{\omega}_4}{\hat{\omega}_{sp}} (-)$
1	0.54583	0.55630	-0.00024	1.00000
2	0.48615	0.60867	-0.00006	1.00000
3	0.42719	0.65810	-0.00010	1.00000
4	0.61598	0.48504	-0.00049	1.00000
5	0.69592	0.42131	-0.00002	1.00000
6	0.50881	0.52906	0.05453	1.00000
7	0.59396	0.46033	0.05448	1.00000
8	0.52974	0.46139	0.10939	1.00000
9	0.47430	0.51102	0.11070	1.00000

Although we expect very good proportionality between the true values  $q_i$  and the measured ones  $\hat{\omega}_i$  due to the measurement principle in use, we design our flow functions in the following manner:

$$\begin{aligned} f_1(\hat{\boldsymbol{\omega}}^T; \boldsymbol{\beta}) &= \beta_1 \hat{\omega}_1 + \beta_2 \hat{\omega}_1^2, \\ f_2(\hat{\boldsymbol{\omega}}^T; \boldsymbol{\beta}) &= \beta_3 \hat{\omega}_2 + \beta_4 \hat{\omega}_2^2, \\ f_3(\hat{\boldsymbol{\omega}}^T; \boldsymbol{\beta}) &= \beta_5 \hat{\omega}_3 + \beta_6 \hat{\omega}_3^2, \text{ and} \\ f_4(\hat{\boldsymbol{\omega}}^T; \boldsymbol{\beta}) &= \hat{\omega}_4. \end{aligned} \quad (4)$$

That is, we assign a parabolic polynomial without intercept to the first three flow functions. The flow function of the collector pipe  $f_4$  does not contain any unknown coefficient and serves as reference edge flow ( $i^* = 4$ ). From a mathematical point of view, the definition of a non-zero reference edge flow is required to obtain an unambiguous estimate for  $\boldsymbol{\beta}$ . Again, it must be pointed out that all flow functions provide edge flow estimates in relation to the reference flow. Here we dispense with the treatment of non-linear flow functions, e.g., including a variable Winter-Kennedy exponent. For details on this matter see Reference [1].

### 2.3 Ordinary least squares estimates

A flow network with  $p$  inner vertices can be described by the linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (5)$$

with the response vector  $\mathbf{y} \in \mathbb{R}^{p \cdot n}$ , the design matrix  $\mathbf{X} \in \mathbb{R}^{p \cdot n \times m}$  and the error vector  $\boldsymbol{\epsilon} \in \mathbb{R}^{p \cdot n}$ . We may divide Equation (5) into  $p$  sub-models yielding

$$\begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_p \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_p \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \boldsymbol{\epsilon}_1 \\ \vdots \\ \boldsymbol{\epsilon}_p \end{pmatrix} \quad (6)$$

with  $\mathbf{y}_i \in \mathbb{R}^n$ ,  $\mathbf{X}_i \in \mathbb{R}^{n \times m}$  and  $\boldsymbol{\epsilon}_i \in \mathbb{R}^n$ . If sub-model  $i$  includes the reference edge flow  $i^*$  then  $\mathbf{y}_i = \mathbf{f}_{i^*}$ , and  $\mathbf{y}_i = \mathbf{0}$ , otherwise.

Let us keep the calculation scheme here simple by using ordinary least squares (OLS). The derivation with weighted least squares (WLS) can be found in [1]. OLS implies that  $\text{var}(\mathbf{y}_i) = \text{var}(\boldsymbol{\epsilon}_i) = \sigma^2 \mathbf{E}_n$ , where the parameters  $\sigma^2$  and  $\mathbf{E}_n \in \mathbb{R}^{n \times n}$  are the true model variance and the identity matrix, respectively. This also enables us to define the sum of squared errors, i.e., our cost function which must be minimized, by

$$SSE(\boldsymbol{\beta}) = \sum_{i=1}^p (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^T (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}). \quad (7)$$

Setting the gradient of Equation (7) with respect to  $\boldsymbol{\beta}$  equal to the zero vector provides the least squares estimate (LSE) of the solution vector yielding

$$\hat{\boldsymbol{\beta}} = \left( \sum_{i=1}^p \mathbf{X}_i^T \mathbf{X}_i \right)^{-1} \cdot \left( \sum_{i=1}^p \mathbf{X}_i^T \mathbf{y}_i \right). \quad (8)$$

Furthermore, inserting Equation (8) into Equation (7) and division by  $d_f = p \cdot n - m$  denoting the statistical degrees of freedom provides the LSE of the model variance

$$\hat{\sigma}^2 = \frac{1}{d_f} \sum_{i=1}^p (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})^T (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}) \quad (9)$$

and the computation of other related parameters. Setting up the  $\mathbf{y}_i$ 's and the  $\mathbf{X}_i$ 's is not trivial. A possible approach using the row vectors of the reduced incidence matrix in Equation (3) is given by

$$\mathbf{y}_i = \mathbf{Z} \cdot \text{diag}(\mathbf{C}_e \mathbf{b}_i) \cdot \mathbf{c}_y \quad (10)$$

and

$$\mathbf{X}_i = -\mathbf{Z} \cdot \text{diag}(\mathbf{C}_e \mathbf{b}_i) \cdot \mathbf{C}_x \quad (11)$$

with  $\mathbf{Z} \in \mathbb{R}^{n \times (m+1)}$ ,  $\mathbf{C}_e \in \mathbb{Z}^{(m+1) \times n_e}$ ,  $\mathbf{c}_y \in \mathbb{Z}^{(m+1)}$  and  $\mathbf{C}_x \in \mathbb{Z}^{(m+1) \times m}$ . That is, these four parameters must be set up only once, and then it provides the responses and design matrices of all sub-models. The matrix  $\mathbf{Z}$  holds the predictor variables  $\mathbf{z}_{i,j} = \partial \mathbf{f}_i / \partial \beta_{\delta(i,j)}$  ( $= m$  columns) and the response variable  $\mathbf{z}_{i^*,1} = \mathbf{f}_{i^*}$  ( $= 1$  column) of  $n$  measuring points ( $=$  rows) and it yields

$$\mathbf{Z} = (\mathbf{z}_{1,1}, \mathbf{z}_{1,2} \dots \mathbf{z}_{n_e, m_{n_e}}) \quad (12)$$

with  $\sum_{i=1}^{n_e} m_i = m + 1$  and the index  $\delta(i, j)$ , which gives  $\forall i, j \in \{\mathbb{N} | (i \leq n_e) \wedge (i \neq i^*), j \leq m_i\}$ :

$$\delta(i, j) = \begin{cases} \sum_{k=1}^{i-1} m_k + j, & \text{if } i < i^*, \text{ and} \\ \sum_{k=1}^{i-1} m_k + j - 1, & \text{if } i > i^*. \end{cases} \quad (13)$$

A case in point: The flow function design in Equation (4) gives

$$\mathbf{Z} = (\hat{\boldsymbol{\omega}}_1, \hat{\boldsymbol{\omega}}_1^2, \hat{\boldsymbol{\omega}}_2, \hat{\boldsymbol{\omega}}_2^2, \hat{\boldsymbol{\omega}}_3, \hat{\boldsymbol{\omega}}_3^2, \hat{\boldsymbol{\omega}}_4) \quad (14)$$

where  $\hat{\boldsymbol{\omega}}_i \in \mathbb{R}^n$  denotes the vector of  $n$  measured flow estimates belonging to edge  $i$ . The representation  $\hat{\boldsymbol{\omega}}_i^y$  is only symbolical and it stands for

$$\hat{\boldsymbol{\omega}}_i^y \equiv \begin{pmatrix} \hat{\omega}_{i,1}^y \\ \vdots \\ \hat{\omega}_{i,n}^y \end{pmatrix}. \quad (15)$$

The elements of the remaining parameters  $\mathbf{C}_e$ ,  $\mathbf{c}_y$  and  $\mathbf{C}_x$  are either 0's or 1's. The diagonal matrix  $\text{diag}(\mathbf{C}_e \mathbf{b}_i) \in \mathbb{Z}^{(m+1) \times (m+1)}$  assigns the factors  $\in \{-1, 0, +1\}$  to the columns of  $\mathbf{Z}$ . The matrix  $\mathbf{C}_x$  and the vector  $\mathbf{c}_y$  assign the columns of  $\mathbf{Z}$  to the contributing flow functions. A case in point: The flow function design in Equation (4) yields

$$\mathbf{C}_e = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \mathbf{c}_y = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \text{ and} \quad (16)$$

$$\mathbf{C}_x = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Continuing our example with one single inner vertex by inserting Equations (3), (14) and (16) into Equations (10) and (11), we obtain

$$\mathbf{y} = \mathbf{y}_i = \hat{\boldsymbol{\omega}}_4 \quad (17)$$

and

$$\mathbf{X} = \mathbf{X}_i = (\hat{\boldsymbol{\omega}}_1, \hat{\boldsymbol{\omega}}_1^2, \hat{\boldsymbol{\omega}}_2, \hat{\boldsymbol{\omega}}_2^2, \hat{\boldsymbol{\omega}}_3, \hat{\boldsymbol{\omega}}_3^2). \quad (18)$$

Finally, let us provide a last auxiliary parameter definition denoted by master matrix which yields

$$\mathbf{M} = \sum_{i=1}^p \text{diag}(\mathbf{C}_e \mathbf{b}_i) \cdot \mathbf{Z}^T \mathbf{Z} \cdot \text{diag}(\mathbf{C}_e \mathbf{b}_i) \quad (19)$$

$$\mathbf{M} \in \mathbb{R}^{(m+1) \times (m+1)}$$

and which provides simplified representations for Equations (8), (9) and the covariance matrix

$$\hat{\boldsymbol{\beta}} = (\mathbf{C}_X^T \mathbf{M} \mathbf{C}_X)^{-1} \cdot (-\mathbf{C}_X^T \mathbf{M} \mathbf{c}_y) \in \mathbb{R}^m, \quad (20)$$

$$\hat{\sigma}^2 = \frac{1}{d_f} (\mathbf{c}_y + \mathbf{C}_X \hat{\boldsymbol{\beta}})^T \mathbf{M} (\mathbf{c}_y + \mathbf{C}_X \hat{\boldsymbol{\beta}}) \in \mathbb{R}, \quad (21)$$

$$\text{cov}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2 (\mathbf{C}_X^T \mathbf{M} \mathbf{C}_X)^{-1} \in \mathbb{R}^{m \times m}. \quad (22)$$

#### 2.4 Automated variable selection

The number of possible flow function combinations grows exponentially in  $n_e$  and yields

$$n_{comb} = \prod_{i=1, i \neq i^*}^{n_e} (2^{\max(m_i)} - 1). \quad (23)$$

Consequently, the search for the *best* model needs a strategy to find it and a quantitative measure to evaluate it. For the latter, the corrected Akaike information criterion, e.g., [3], given by

$$AIC_c = pn \cdot \ln \left( \frac{SSE(\hat{\boldsymbol{\beta}})}{pn} \right) + 2m + 2 \frac{m^2 + m}{pn - m - 1} \quad (24)$$

is expected to suit well also for a low number of observations  $pn$ . The lower the value in the equation above, the better the model. The first term in Equation (24) accounts for the accuracy of the model to avoid *underfitting*, and the remaining two terms penalize too high complexity, which is intended to avoid *overfitting*. There exist numerous strategies in the scientific literature to find an acceptable good model, e.g., [4]. Here, we make use of a simple *backward elimination algorithm* which is expected to find an acceptable model after  $n_m \leq (\max(m) - n_e + 2)$  evaluations. This search algorithm works in the following manner:

1. Start with a regression model holding the maximum number of predictor variables  $M_j = \max(m)$  with  $j = 1$  (e.g.,  $M_j = 6$  in our example).
2. Compute  $AIC_c^{(j)}$  with Equation (24).
3. Compute all percentiles  $\forall i \leq M_j$ :

$$t_i = \frac{\hat{\beta}_i}{se(\hat{\beta}_i)} \quad (25)$$

with the standard error

$$se(\hat{\beta}_i) = \sqrt{[\text{cov}(\hat{\boldsymbol{\beta}})]_{ii}} \quad (26)$$

4. Eliminate the predictor variable  $\mathbf{z}$  in Equation (14) whose absolute value of the corresponding percentile is the smallest. Remove it only if the flow function has more than one predictor variable. Otherwise remove one with the next smallest percentage from another flow function. Assuming the  $i$ -th predictor variable must be eliminated, proceed as follows:
  - a. Remove the  $i$ -th column in  $\mathbf{Z}$  if  $i < i^*$ . Otherwise remove column  $(i + 1)$ .
  - b. Remove the  $i$ -th column in  $\mathbf{C}_X$ .
  - c. Remove the  $i$ -th row in  $\mathbf{C}_e$ ,  $\mathbf{C}_X$  and  $\mathbf{c}_y$  if  $i < i^*$ . Remove row  $(i + 1)$  otherwise.
5. Increase the index  $j \rightarrow j + 1$ .
6. Compute the new regression model having  $M_j = M_{j-1} - 1$  predictor variables.
7. Compute  $AIC_c^{(j)}$  with Equation (24).
8. Continue to step 3, if  $AIC_c^{(j)} \leq AIC_c^{(j-1)}$  and if  $M_j \geq n_e$ .
9. Stop the algorithm and apply diagnostic tools onto the most suitable model, which has been found.

Back to our example where we have  $n_{comb} = 27$  possible flow function combinations. The backward elimination algorithm requires only  $n_m \leq 4$  models to evaluate. The results of this search procedure are listed in Table 2. The columns from left to right denote the index  $j$ , the number of regressors  $m$ , the degrees of freedom  $d_f$ , the coefficient of determination  $R^2$ , the information criterion  $AIC_c$  from Equation (24) and the model term with the lowest percentile and which is to be eliminated in the model which follows. The complexity and the accuracy of the evaluated models decrease from top to bottom. It is typical that  $R^2$  favors the most complex model ( $j = 1$ ) because its definition does not involve any penalty term. The adjusted coefficient of determination  $R_{adj}^2$  accounts slightly for the complexity. However, both tend to overfit given data [4].

**Table 2: Results of the backward elimination algorithm starting with the flow functions given in (4)**

$j$ (-)	$m$ (-)	$d_f$ (-)	$R^2$ (-)	$AIC_c$ (-)	eliminate
1	6	3	.999993	-94.6	$\beta_6 \hat{\omega}_3^2$
2	5	4	.999993	-96.5	$\beta_2 \hat{\omega}_1^2$
3	4	5	.999993	-98.4	$\beta_4 \hat{\omega}_2^2$
4	3	6	.999991	-99.1	-

The parameter  $AIC_c$  favors the simplest model ( $j = 4$ ), which uses the flow function definition

$$\begin{aligned} f_1(\hat{\boldsymbol{\omega}}^T; \boldsymbol{\beta}) &= \beta_1 \hat{\omega}_1, \\ f_2(\hat{\boldsymbol{\omega}}^T; \boldsymbol{\beta}) &= \beta_2 \hat{\omega}_2, \\ f_3(\hat{\boldsymbol{\omega}}^T; \boldsymbol{\beta}) &= \beta_3 \hat{\omega}_3, \text{ and} \\ f_4(\hat{\boldsymbol{\omega}}^T; \boldsymbol{\beta}) &= \hat{\omega}_4. \end{aligned} \quad (27)$$

The model above coincides exactly with the model choice in [2].

### 2.5 Hypothesis testing and residual diagnostic

After a possible candidate for a suitable model has been found by the search algorithm in the previous section, we need to perform obligatory checks with respect to hypothesis tests on the LSEs, non-randomized residual pattern and outlier detection, for instance, by using deletion studentized residuals, e.g., [4], [1].

Table 3 summarizes the result of  $t$ -tests on the individual LSEs  $\hat{\beta}_i$ . The columns from left to right reveal the index  $i$ , the values of  $\hat{\beta}_i$ , their standard error  $se(\hat{\beta}_i)$ , the percentile using Equation (25) and the probability of rejection  $pr \in (0,1)$ . We commonly choose a significance level of  $\alpha = 0.05$  and recognize, that for all LSEs it yields  $pr \ll \alpha$  revealing high significance of the chosen regression model.

**Table 3: Results of the  $t$ -test with model in Equation (27)**

$i$ (-)	$\hat{\beta}$ (-)	$se(\hat{\beta})$ (-)	$t$ (-)	$pr$ (-)
1	0.8572	0.0084	102.14	5.9E-11
2	0.9621	0.0086	111.24	3.6E-11
3	0.9288	0.0257	36.20	3.0E-08

The residual vector is defined by the difference between response vector and fitted vector and it gives

$$\hat{\epsilon} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \in \mathbb{R}^{pn}. \quad (28)$$

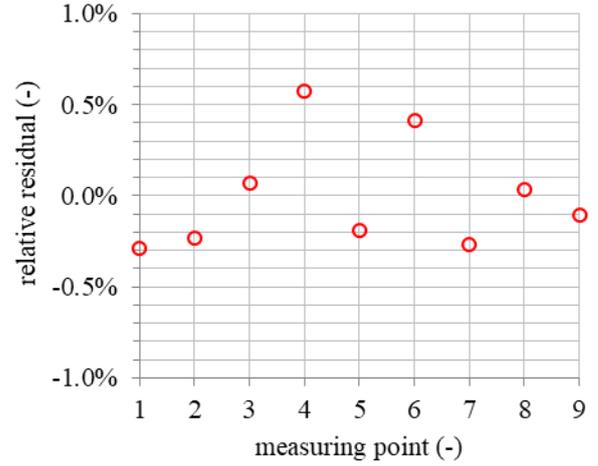
Figure 2 provides a diagram showing the relative residuals

$$\frac{\hat{\epsilon}_i}{y_i} = \frac{[\hat{\epsilon}]_i}{[\mathbf{y}]_i} \in \mathbb{R} \quad (29)$$

on the primary vertical axis versus the measuring point number. The deviations from the horizontal axis appear randomly distributed and there is no clear indication of a possible outlier. This statement can also be confirmed by considering deletion studentized residuals, which provide a more sophisticated outlier detection method, but which has been omitted here. Due to the low number of measuring points all observations are kept in. Finally, we can state that the model in Equation (27) describes well the available network data with a regression standard deviation from Equation (21) yielding  $\hat{\sigma} = 0.0036$  for  $d_f = 6$  degrees of freedom.

### 2.6 Mean and confidence interval of edge flow rates

Now we can deliver the most important output of the calibration method. First, a mean edge flow rate  $\hat{q}_i$  can be calculated in using Equation (27) and the results



**Figure 2: Residual plot of the model in Equation (27)**

given in Table 3. It yields

$$\begin{aligned} \hat{q}_1(\hat{\omega}_1) &= 0.8572(84) \cdot \hat{\omega}_1, \\ \hat{q}_2(\hat{\omega}_2) &= 0.9621(86) \cdot \hat{\omega}_2, \\ \hat{q}_3(\hat{\omega}_3) &= 0.929(26) \cdot \hat{\omega}_3, \text{ and} \\ \hat{q}_4(\hat{\boldsymbol{\omega}}^T) &= \hat{q}_1(\hat{\omega}_1) + \hat{q}_2(\hat{\omega}_2) + \hat{q}_3(\hat{\omega}_3). \end{aligned} \quad (30)$$

The computation of the confidence interval requires the variance of the mean flow rate, which is obtained with Equation (22) providing

$$se^2(\hat{q}_i) = \sum_{j=1}^m \sum_{k=1}^m \left( \frac{\partial \hat{q}_i}{\partial \beta_j} \right) \left( \frac{\partial \hat{q}_i}{\partial \beta_k} \right) [cov(\hat{\boldsymbol{\beta}})]_{jk} \quad (31)$$

and the  $100\% \cdot (1 - \alpha/2)$ -th percentile of the  $t_{d_f}$ -statistics yielding

$$t = t\left(1 - \frac{\alpha}{2}, pn - m\right) \quad (32)$$

Consequently, with the positive square root of Equation (33) and with Equation (32) the  $100\% \cdot (1 - \alpha)$  confidence interval of an edge flow rate  $q_i$  is given by

$$\hat{q}_i - se(\hat{q}_i) \cdot t \leq q_i \leq \hat{q}_i + se(\hat{q}_i) \cdot t. \quad (33)$$

For our problem with  $t = 2.447$  and Table 3 we obtain

$$\begin{aligned} 0.8367 \cdot \hat{\omega}_1 &\leq q_1 \leq 0.8778 \cdot \hat{\omega}_1 \\ 0.9410 \cdot \hat{\omega}_2 &\leq q_2 \leq 0.9833 \cdot \hat{\omega}_2 \\ 0.8660 \cdot \hat{\omega}_3 &\leq q_3 \leq 0.9916 \cdot \hat{\omega}_3 \end{aligned} \quad (34)$$

A diagrammatic comparison between the measured edge flow rates  $\hat{\omega}_i$  and the mean flow rates  $\hat{q}_i$  can be found in Figure 3. It clearly shows the outcome of Table 3, that the change of edge flow 1 is higher than for the others to fulfill the continuity law with the chosen reference flow in Table 1. In Figure 4 The curvature (i.e., here just a straight line) of the one-sided confidence intervals  $se(\hat{q}_i) \cdot t$  for mean edge flow 1 (red line) and 2 (blue line) is very similar. This is not surprising because their edge flows have been measured with the same type of

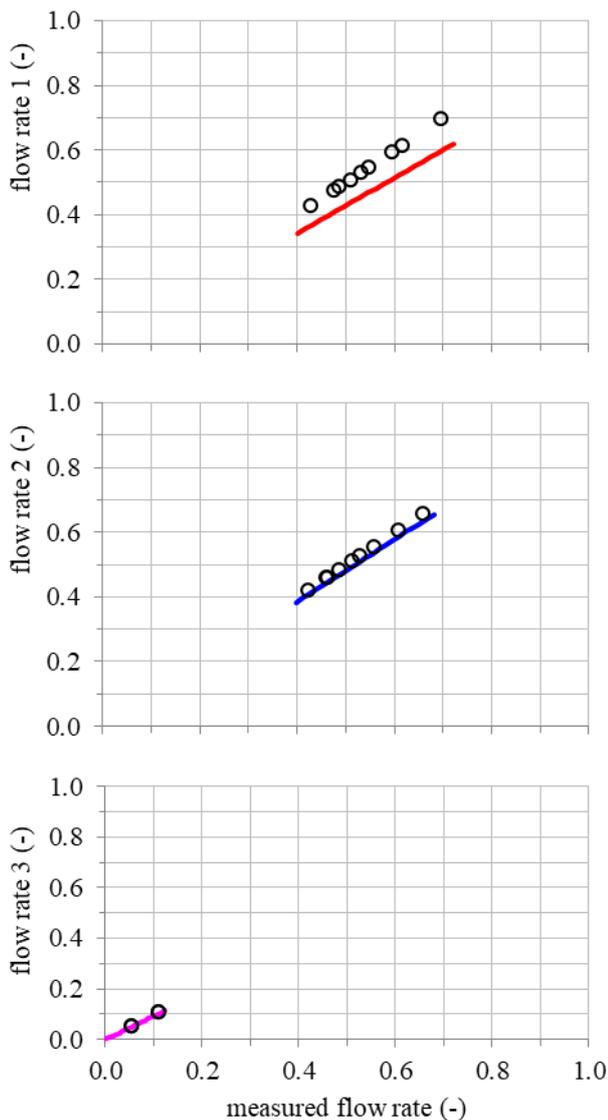


Figure 3: Comparison of measured (circles) and mean (solid line) flow rates for edge 1 (top), 2 (middle) and 3 (bottom) versus measured edge flow rates

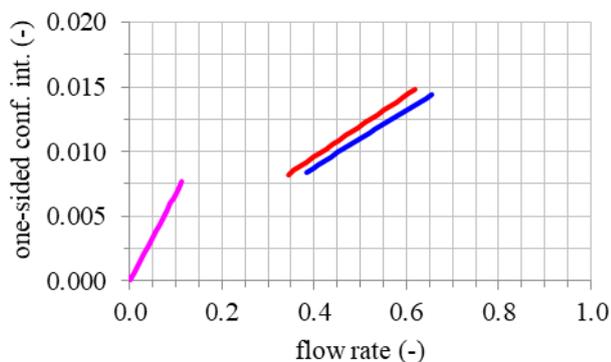


Figure 4: One-sided confidence intervals  $se(\hat{q}_i) \cdot t$  for edge 1 (red), 2 (blue) and 3 (magenta) versus mean flow rates  $\hat{q}_i$

metering system on pipe sections with identical dimension and they are using the same type of flow function. The slope of the one-sided confidence interval

for mean edge flow 3 (magenta line) is steeper than the others, which bases on the lower absolute values. However, the drawback of this method becomes evident by examining the curvatures in Figure 4: all one-sided confidence intervals are zero at zero mean edge flow. This is because the flow functions in (27) have been designed without intercept. On the one side, if two or more flow functions, which contribute to the same inner vertex, kept an intercept, the design matrix  $X$  in Equation (5) would not have full column rank anymore and the system of equations could not be solved. On the other side, if an intercept was assigned to a single flow function, the system of equations could be solved, but the regressor of this intercept would compensate any irregularities of all contributing flow functions to meet the continuity law, and, therefore, the mean edge flows would become meaningless.

### Conclusion

A numerical method, which enables the simultaneous calibration of edge flow metering systems within a hydraulic system, has been introduced. The output of one metering system serves as reference, which the calibration of the remaining metering systems are related to. It is cost-efficient and advantageous in reducing the measurement uncertainty, that the necessary flow measurements can be conducted in-situ and, usually, under normal operating conditions. The search for a suitable regression model, which describes the portions of each edge flow rate best, is done by combining a backward elimination algorithm with the corrected Akaike information criterion. The steps in performing the computations of this method have been given in appropriate detailedness for a hydraulic network with one inner vertex, which represents the case a practitioner faces most often on hydroelectric sites. The treatment of more complex systems, e.g., fresh water supply systems with multiple inner vertices, and the use of non-linear flow functions is described in detail in the provided references.

### References

- [1] J. Lanzersdorfer, *An in-situ calibration method for metering devices in flow networks with applications in hydraulic systems*, Master Thesis, Graz: Technical University of Graz, 2024.
- [2] J. Lanzersdorfer, M. Schmidt and J. Pichler, "Flow rate evaluation in parallel pump arrangements: Two case studies," in *11th Conference of the International Group for Hydraulic Efficiency Measurement*, Linz, Austria, 2016.
- [3] B. Chen, Y. Zhu, J. Hu and J. C. Principe, *System Parameter Identification: Information Criteria and Algorithms*, Elsevier, 2013.
- [4] S. Weisberg, *Applied linear regression*, 3rd ed., Wiley series in probability and statistics, 2005.