5. System identification design

In this chapter, the Multi-inlet, Multi-WT model, which has been derived based on first principles in Chapter 4: System Modelling, is reformulated such that it is suitable for system identification. First, a model structure is selected, then a Neural Network(NN) based Radial Basis Function(RBF) model is presented. In the first case, the identification is carried out on a simple example, then on the Randers WSS EPANET model. In the second case, the identification is carried out based on measurements from the real-world network.

5.1 Model structure of the Multi-inlet, Multi-WT system

In Chapter *4: System Modelling*, the model of a multi-inlet WSS with the extension of multiple WTs has been derived. The presented model has been given by a non-linear SS representation, which consists of the equations describing the dynamics, the outputs and the constraints on the network. As the model derivation is based on first principles, some insight into the structure and the input-output relation is available.

5.1.1 Output equation

The model is going to be utilized for identification purposes, therefore let us recall the governing equations. The output vector $\bar{p}_{\mathcal{K}}$ of the inlet pressures is given in discrete form such that

$$\bar{p}_{\mathcal{K},k} = K^T \bar{H}_{\mathcal{T}}^{-T} f_{\mathcal{T}}(A_2 q_{\mathcal{C},k} + A_3 K \bar{d}_{\mathcal{K},k} - A_3 D v_{\mathcal{D}} \sigma_k) - K^T \bar{H}_{\mathcal{T}}^{-T} \hat{H}_{\mathcal{T}}^T (\hat{p}_k + \hat{h}) - K^T \bar{h}, \quad (5.1)$$

where

$$A_1 = \hat{H}_{\mathcal{C}}^T - \bar{H}_{\mathcal{C}}^T \bar{H}_{\mathcal{T}}^{-T} \hat{H}_{\mathcal{T}}^T,$$
$$A_2 = -\bar{H}_{\mathcal{T}}^{-1} \bar{H}_{\mathcal{C}},$$
$$A_3 = \bar{H}_{\mathcal{T}}^{-1}.$$

Furthermore, let us recall the constraint on $q_{\mathcal{C}}$, and rewrite it in discrete-time form such that

$$f_{\mathcal{C}}(q_{\mathcal{C},k}) - A_1(\hat{p}_k + \hat{h}) + A_2^T f_{\mathcal{T}}(A_2 q_{\mathcal{C},k} + A_3 K \bar{d}_{\mathcal{K},k} - A_3 D v_{\mathcal{D}} \sigma_k) = 0, \quad (5.2)$$

with $\bar{d}_{\mathcal{K},k}$ inlet flows, σ_k total demand, $v_{\mathcal{D}}$ distribution parameter, $q_{\mathcal{C},k}$ flows in set $\mathcal{C}$, $\hat{p}_k$ pressures in the WTs, $\hat{h}$ elevations of the WTs and $K^T \bar{h} = \bar{h}_{\mathcal{K}}$ the elevations of the pumping stations. The corresponding values of the pressures and flows in the network are evaluated at each time step k.

It is important to point out that the model in Chapter *4: System Modelling* has been derived in a general manner, taking into account that $v_{\mathcal{D}}$ is a time-varying distribution parameter of the demands. However, in the further description, let us restrict ourselves and assume that $v_{\mathcal{D}}$ is constant. This assumption is beneficial from the system point of view, as the total consumption σ_k can be represented simply by the sum of the hourly demand variations $1^T \bar{d}_{\mathcal{D}}$. From the technical point of view, the identification becomes less complex, as in this case $v_{\mathcal{D}}$ is a linear constant parameter. In *Equation: (5.2)* and *Equation: (5.1)* this assumption is already taken into account, therefore $v_{\mathcal{D}}$ does not have any time index.

The constraint on $q_{\mathcal{C}}$ in *Equation: (5.2)* is given by an implicit expression for which analytical soulution has not been derived. Therefore, the constraint cannot be substituted explicitly into *Equation: (5.1)*, but can be given in the following ~~following~~ implicit form in *Equation: (5.3)*.

$$q_{\mathcal{C},k} = q_{\mathcal{C}}((\hat{p}_k + \hat{h}), \bar{d}_{\mathcal{K},k}, \sigma_k). \tag{5.3}$$

It is shown in *Equation: (5.3)* that the $q_{\mathcal{C}}$ flows in the network depend on the same physical measures, i.e. the same variables as the outputs $\bar{p}_{\mathcal{K}}$. By substituting *Equation: (5.3)* into *Equation: (5.1)*, we get the following output equation

$$\begin{aligned} \bar{p}_{\mathcal{K},k} + \bar{h}_{\mathcal{K}} =& K^T \bar{H}_{\mathcal{T}}^{-T} f_{\mathcal{T}}[A_2 q_{\mathcal{C}}((\hat{p}_k + \hat{h}), \bar{d}_{\mathcal{K},k}, \sigma_k) + A_3 K \bar{d}_{\mathcal{K},k} - A_3 D v_{\mathcal{D}} \sigma_k] \\ & - K^T \bar{H}_{\mathcal{T}}^{-T} \hat{H}_{\mathcal{T}}^T (\hat{p}_k + \hat{h}). \end{aligned} \tag{5.4}$$

In *Equation: (5.4)*, the total head in the pumping stations $(\bar{p}_{\mathcal{K},k} + \bar{h}_{\mathcal{K}})$ is given by the expression on the right-hand side. Let us write *Equation: (5.4)* in a form where the non-linear expression on the right-hand side is replaced with a non-linear function $\tilde{f}_1(\cdot)$,which has an unknown structure but has the same variables in the argument. Thus, a reformulated output equation can be given such that

$$\tilde{y}_k = \bar{p}_{\mathcal{K},k} + \bar{h}_{\mathcal{K}} = \tilde{f}_1((\hat{p}_k + \hat{h}), \bar{d}_{\mathcal{K},k}, \sigma_k) + \tilde{a}_1(\hat{p}_k + \hat{h}), \tag{5.5}$$

The static model described in *Equation: (5.5)* is a mapping defined by the non-linear function $\tilde{f}_1$ and the linear term, which maps the input set, $u = \{(\hat{p}_k + \hat{h}), \bar{d}_{\mathcal{K}}, \sigma\}$ to the outputs $\tilde{y} = \{\bar{p}_{\mathcal{K},k} + \bar{h}_{\mathcal{K}}\}$. In the input set, the total consumption can be calculated according to the mass-balance in the whole network such that

$$\sigma_k = 1^T \hat{d}_k + 1^T \bar{d}_{\mathcal{K},k}. \tag{5.6}$$

In *Equation: (5.6)*, we assume that the flows in the WTs are measured.

5.1.2 State equation

The state equation is a first-order system of ODEs, which has been formulated on the pressures $\hat{p}$ in the WTs. In order to give a discretization for the approximate solution of the ODEs, Euler-method is applied. The Euler-method is the simplest Runge-Kutte method, which provides an acceptable precision for our problem[28]. Thus, the discretized state equation yields as follows

$$\Lambda \frac{1}{T_s}(\hat{p}_{k+1} - \hat{p}_k) = -(\hat{H}_{\mathcal{C}} - \hat{H}_{\mathcal{T}} \bar{H}_{\mathcal{T}}^{-1} \bar{H}_{\mathcal{C}}) q_{\mathcal{C},k} - \hat{H}_{\mathcal{T}} \bar{H}_{\mathcal{T}}^{-1} K \bar{d}_{\mathcal{K},k} + \hat{H}_{\mathcal{T}} \bar{H}_{\mathcal{T}}^{-1} D v_{\mathcal{D}} \sigma_k. \tag{5.7}$$

Substituting the constraint on $q_{\mathcal{C}}$ into *Equation: (5.7)*, and expressing the approximation of the derivative term on the left-hand side, the following yields

$$\begin{aligned} \hat{p}_{k+1} - \hat{p}_k = T_s \Lambda^{-1} \big[& -(\hat{H}_{\mathcal{C}} - \hat{H}_{\mathcal{T}} \bar{H}_{\mathcal{T}}^{-1} \bar{H}_{\mathcal{C}}) q_{\mathcal{C},k}((\hat{p}_k + \hat{h}), \bar{d}_{\mathcal{K},k}, \sigma_k) \\ & - \hat{H}_{\mathcal{T}} \bar{H}_{\mathcal{T}}^{-1} K \bar{d}_{\mathcal{K},k} + \hat{H}_{\mathcal{T}} \bar{H}_{\mathcal{T}}^{-1} D v_{\mathcal{D}} \sigma_k \big]. \end{aligned} \tag{5.8}$$

where

T_s is the sampling time. [h]

The state equation in *Equation: (5.7)* describes the linear combination of the the flows $q_{\mathcal{C}}$, the total head in the WTs $(\hat{p}_k + \hat{h})$ and the total consumption σ. However, by substituting the $q_{\mathcal{C}}$ flows with their implicit non-linear expression, the structure of the state equation is not a linear combination of the corresponding signals anymore. Therefore, let us write *Equation: (5.8)* in a form where the non-linear and linear terms are described by a non-linear $\tilde{f}_2(\cdot)$ function with unknown structure and linear terms with parameters, respectively. Thus, a reformulated state equation can be given such that

$$\hat{p}_{k+1} - \hat{p}_k = \tilde{f}_2((\hat{p}_k + \hat{h}), \bar{d}_{\mathcal{K},k}, \sigma_k) - \tilde{a}_2 \bar{d}_{\mathcal{K},k} + \tilde{a}_3 \sigma_k, \tag{5.9}$$

where $\tilde{f}_2$ is a non-linear function, $\tilde{a}_2$ and $\tilde{a}_3$ are parameters of the inlet flows $\bar{d}_{\mathcal{K},k}$ and total consumption σ_k.

5.2 RBFNN model of the Multi-inlet,Multi-WT system

As a result of substituting the constraints on the flows $q_{\mathcal{C},k}$, the system description has been reduced to a non-linear SS problem with state equation given in *Equation: (5.9)* and output equation in *Equation: (5.5)*. The complete identification model is summarized in *Equation: (5.10)*.

$$\begin{cases} \hat{p}_{k+1} - \hat{p}_k = \tilde{f}_2((\hat{p}_k + \hat{h}), \bar{d}_{\mathcal{K},k}, \sigma_k) - \tilde{a}_2 \bar{d}_{\mathcal{K},k} + \tilde{a}_3 \sigma_k, \\ \tilde{y}_k = \tilde{f}_1((\hat{p}_k + \hat{h}), \bar{d}_{\mathcal{K},k}, \sigma_k) + \tilde{a}_1(\hat{p}_k + \hat{h}). \end{cases} \tag{5.10}$$

The main goal of the system identification is to find a realization of the functions $\tilde{f}_1(\cdot)$ and $\tilde{f}_2(\cdot)$, furthermore to find the parameters $\tilde{a}_1$, $\tilde{a}_2$ and $\tilde{a}_3$. Therefore, the parameters need to be identified, such that the model is able to reproduce the approximate of the state derivatives $(\hat{p}_{k+1} - \hat{p}_k)$ and the outputs $\tilde{y}_k$ from any input set $u = \{(\hat{p}_k + \hat{h}), \bar{d}_{\mathcal{K}}, \sigma\}$ within the operating regions where we are mapping from.

The identification model shown in *Equation: (5.10)* is an abstraction of the first principle model derived in Chapter *4: System Modelling*. For the constraint on $q_{\mathcal{C},k}$, existence has been implied, however exact structure has not been given. I.e. we know that the relationship exists but an analytical first principle solution for $q_{\mathcal{C},k}$ has not been derived. Therefore, by substituting the implicit expression of the constraint into the state and output equations, some of the insights on the structure of the model are lost. Thus, it is crucial to put a structure on the non-linear functions $\tilde{f}_1(\cdot)$ and $\tilde{f}_2(\cdot)$ in *Equation: (5.10)*.

From a technical point of view, it is beneficial to describe the system by a linear-in-the-parameters model. By restricting ourselves such that the structure of both functions $\tilde{f}_1(\cdot)$ and $\tilde{f}_2(\cdot)$ are linear in the parameters, the parameters of the model can be estimated by simple linear optimization methods, such as Least Squares(LS). For any linear optimization, only the parameters have to enter linearly, as the inputs can depend on any non-linear way on the input data sets.

By making the restriction on $\tilde{f}_1(\cdot)$ and $\tilde{f}_2(\cdot)$, the two non-linear terms will be approximated by some non-linear functions in both the state and output equations of the model. The tools for carrying out such identification procedure leads to the discussion of basis functions and neural networks, which is discussed in *Appendix: E* in detail.