Systems & Control Transactions

Original Research Article
Peer Reviewed Conference Proceeding
ESCAPE 36 - European Symposium on Computer Aided Process Engineering
Sheffield, UK. 21-24 June 2026

PSE PRESS

Data Reconciliation for Inventory Monitoring in a Petrol Refinery

Jakub Gaborčík[a]*, Karol Ľubušký[b], and Radoslav Paulen[a]

[a] Faculty of Chemical and Food Technology, Slovak University of Technology in Bratislava, Bratislava, Slovakia
[b] Slovnaft, a.s., Bratislava, Slovakia
* Corresponding Authors: xgaborcik@stuba.sk

ABSTRACT

We study a data reconciliation problem in a petrol refinery. The problem is to reconcile inventory and flow measurements to estimate true values of measured and unmeasured flows respecting the mass conservation. The problem is formulated as a mixed-integer quadratic program (MIQP). Upon successful problem resolution, a neural network (NN) is trained to mimic the MIQP solver to study potential improvements in CPU time without compromising the solution quality. The results show a significant improvement in refinery monitoring and feasibility of NN-based reconciliation.

Keywords: data reconciliation, optimization, neural networks, oil refinery

INTRODUCTION

A typical inventory infrastructure in a refinery consists of multitude interconnected technological blocks with an extensive network of liquid storage tanks. Although measurements of input and output flows to some of these blocks and tank levels might be available, the precise flows to, from, and between individual storage tanks are not necessarily directly metered. The goal of this work is to design a data reconciliation system coupled to mass balance to estimate the incoming, interconnecting and outflowing streams for product storage tanks in the Slovnaft refinery, an industrial partner.

Effective data reconciliation [1] is a vital tool in process industry to support operational and economic decisions [2]. Over the past years, data reconciliation has been successfully applied in petrol industry [3, 4, 5]. The problem is usually resolved by formulating an optimization problem in a static or dynamic fashion [6, 7]. The optimization problem rectifies the values of measured variables by enforcing the mass and energy balance conservation laws as problem constraints.

In this work, we leverage historical operational data, technical data sheets, and fundamental physical laws to derive the necessary relationships and introduce an optimization problem to design a data reconciliation framework. The data reconciliation needs to assign network topology (pipeline connections) and thus the problem boils down to a mixed-integer quadratic program (MIQP). To study whether a simplified workaround can be used for data reconciliation, we develop a neural network model that mimics the rigorous MIQP solution.

PROBLEM DEFINITION

Figure 1 shows a simplified scheme of the product storage block that we study in this work. The measured quantities (flows, inventory levels) are depicted in green and unknowns are marked with red question marks. The technological block consists of $m = 7$ tanks with measured hold-ups and there are $n = 17$ inflows into the block. The inflows into tanks are assigned based on operational requirements. The assignment is done manually and is unknown to the plant database. It follows the rule that each input stream leads into a single tank while multiple input streams can be assigned into the same tank. There are $o = 3$ outflowing streams from the block and, analogically, the outflow from a tank leads to a single output flow. Moreover, there are connections between the tanks, whose assignment is also unknown but a connection can be established only between two tanks at a time.

Data Reconciliation

Data reconciliation is the process of comparing, matching, and balancing data from two or more independent sources to ensure accuracy and consistency across systems [7]. The use of reconciliation in this work can be represented by a flowchart in Figure 2, showing

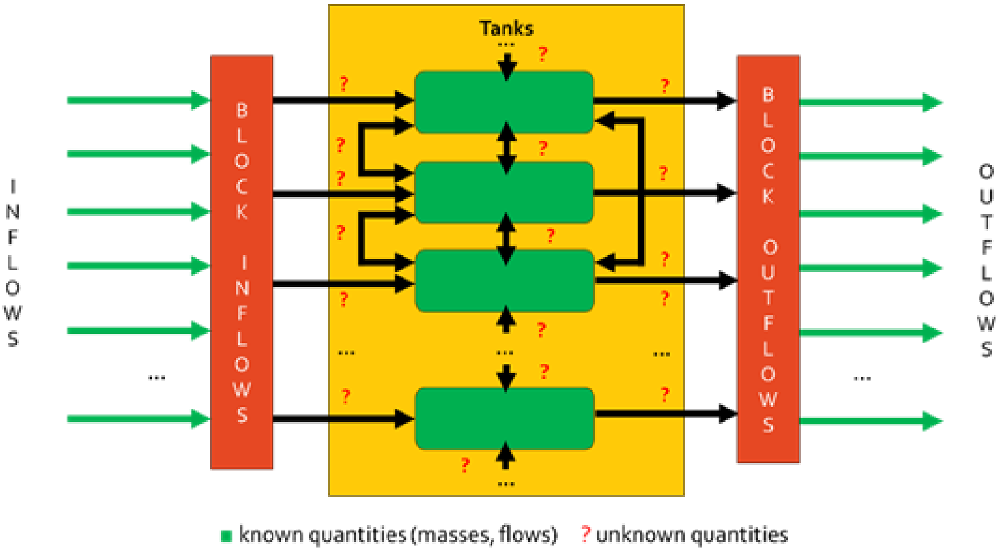

Figure 1. Schematic representation of a studied inventory problem, showing measured and unknown quantities.

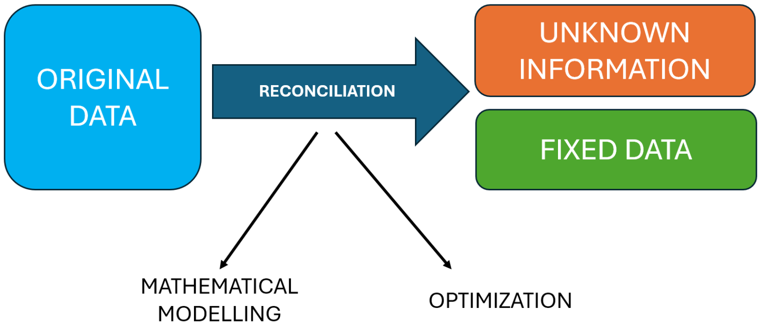

Figure 2. Schematic representation of data reconciliation methodology.

that we attempt to estimate unmeasured quantities from the measured quantities. At the same time, we also aim to rectify the value of measured variables so they would follow physical laws such as mass balance. All this is achieved using mathematical modelling and optimization.

The data reconciliation problem is formulated upon the following conditions and assumptions:

- The measured variables are denoted as y^{meas}
- The reconciled variables are denoted as y
- The dynamic mass balance model assumes the knowledge of actual state (hold-up) of the tanks and the state in the next sampling point
- According to the problem setup, we complement inflow and outflow of each tank with a binary variable in relation to block streams – a stream can either enter the tank or not and the outflow from the tank can enter only one outflow block stream
- The representation of interactions between tanks is realized with and m x m anti-symmetric matrix, representing all the possible inflows/outflows between each tank. The matrix diagonals are zero since they represent flow from/to the tank itself.

The described rules are transformed into the following mathematical relations. The vector

$$y^{meas} = \left\{ \begin{matrix} \dot{m}_{in}^{meas} \in \mathbb{R}^{n} \\ m_{tank,k}^{meas} \in \mathbb{R}^{m} \\ m_{tank,k+1}^{meas} \in \mathbb{R}^{m} \\ \dot{m}_{out}^{meas} \in \mathbb{R}^{o} \end{matrix} \right\} \tag{1}$$

involves measurements, i.e., masses m and mass flows $\dot{m}$. Subscripts $in, out, tank, k,$ and $tank, k+1$ represent inflows, outflows, and current and future tank hold-ups.

Similarly, the vector

$$y = \left\{ \begin{matrix} \dot{m}_{in} \in \mathbb{R}^{n} \\ m_{tank,k} \in \mathbb{R}^{m} \\ m_{tank,k+1} \in \mathbb{R}^{m} \\ \dot{m}_{out} \in \mathbb{R}^{o} \\ \dot{m}_{tank} \in \mathbb{R}^{m \times m} \end{matrix} \right\} \tag{2}$$

describes the reconciled quantities together with $\dot{m}_{tank}$ signifying interaction flows between the tanks. Further,

$$z = \left\{ \begin{matrix} Z_{in} \in \{0,1\}^{n\times m} \\ Z_{tank} \in \{0,1\}^{m\times m} \\ Z_{out} \in \{0,1\}^{o\times m} \end{matrix} \right\} \quad (3)$$

involves binary variables that represent the network connections. This specifically means the optimization problem for each time step involves a total of 272 variables: 189 binary (119 for inflows, 21 for outflows, and 49 for inter-tank connections) and 83 continuous (representing all mass flows and inventory levels at time k and $k+1$).

The objective of data reconciliation can be formulated as:

$$\min_{y,z\in\{0,1\}} \left(y^{meas} - y\right)^{\top} Q\left(y^{meas} - y\right) \quad (4)$$

where

$$Q_{ii} \coloneqq \sigma_y^{-2},\ Q_{ij} \coloneqq 0 \quad (5)$$

with σ_y representing the vector of standard deviations of measurement errors.

The objective function is subject to the constraints:

$$m_{tank,k+1} = m_{tank,k} + \Delta m \quad (6)$$

$$\Delta m = Z_{in}^{\top}\dot{m}_{in} + Z_{tank}\dot{m}_{tank}\mathbf{1}_{\boldsymbol{m}\times 1} - Z_{out}^{\top}\dot{m}_{out} \quad (7)$$

$$m_{tank,k+1},\ m_{tank,k},\ \dot{m}_{in},\ \dot{m}_{out} \geq 0 \quad (8)$$

$$\forall i \in \{1,\dots,n\}:\quad \textstyle\sum_{j=1}^{m} Z_{in}(i,j) = 1 \quad (9)$$

$$\forall j \in \{1,\dots,m\}:\quad \textstyle\sum_{l=1}^{o} Z_{out}(l,j) = 1 \quad (10)$$

$$\forall j \in \{1,\dots,m\}:\quad Z_{tank}(j,j) = 0 \quad (11)$$

$$\forall j \in \{1,\dots,m\}:\quad \textstyle\sum_{k=1}^{m} Z_{tank}(j,k) \leq 1 \quad (12)$$

$$\forall j \in \{1,\dots,m\}:\quad \textstyle\sum_{i=1}^{m} Z_{tank}(i,j) \leq 1 \quad (13)$$

$$Z_{\mathrm{tank}}(i,j) + Z_{\mathrm{tank}}(j,i) \leq 1 \quad (14)$$

that represent mass balance (Eqs. (6) and (7)), non-negativity constraints (Eq. (8)), assignment constraints in Eqs. (9) and (10), tank to tank constraints in Eqs. (11) – (13) and binary anti-symmetry in Eq. (14). Note that in Eqs. (6) and (7), we assume aligning the temporal difference Δm unit with the flow units.

Reconciled data can be potentially used for neural network training with original data as inputs and reconciliated data as outputs, giving us another tool beside optimization solver. Neural networks could improve real-time reconciliation should the MIQP solver turn out slow.

IMPLEMENTATION

We use MATLAB for framework implementation. The workflow contains the following steps:

1. Processing time series data with half-hour sampling time, covering several months of refinery operation (13, 726 half-hourly data points in total). Outliers are removed. Missing or removed values are linearly interpolated.
2. Unit and quantity conversion, where volumes and volumetric flows are transformed into masses and mass flows using measured densities.
3. Setting up the problem (4) – (14) using YALMIP [9] and solving it using GUROBI [10].
4. Training a neural network using *feedforwardnet* function in MATLAB (data split training / validation / testing = 70% / 15% / 15%). A feedforward network is created with three hidden layers sized 250, 150, and 80. The training algorithm is set to resilient backpropagation (*trainrp*). Each hidden layer's transfer function is set to *poslin*, which is ReLU. We exclude binary variables and use only continuous ones, because while binary variables are essential in mixed-integer problem definition, it is the set of continuous variables that carries the physical importance of the system. This step makes the problem simpler for neural network training, too. However, this approach makes the NN model topology-specific. If the physical network connections change (e.g., a new pipeline is installed or a tank is rerouted), the NN cannot dynamically adapt its binary logic and the surrogate model must be retrained using a new dataset generated by the updated MIQP formulation. Another specific settings, mostly for early stopping, are specified in Table 1.

Table 1: Feedforward neural network settings.

Parameter	Value
Maximum no. of epochs	2, 000
Sufficient performance	1E-05
Sufficient gradient	1E-06
Maximum validations check	15

RESULTS

Data Reconciliation using Optimization Solver

We demonstrate results of reconciliation via optimization in Figure 3. Figure 3.a) shows that the MIQP solver was able to find optimal solution in around 98% of the instances, with remaining 2% representing failed attempts, mostly due to numerical problems (98%) and few because of reached time/iteration limit of 60 s (2%).

Figure 3.b) represents the empirical cumulative distribution function of mass balance residuals between reconciled and measured variables. It should be noted that in the MIQP formulation, the reconciled flow is defined by the product of a binary existence variable and a continuous flow rate. While the optimization ensures that

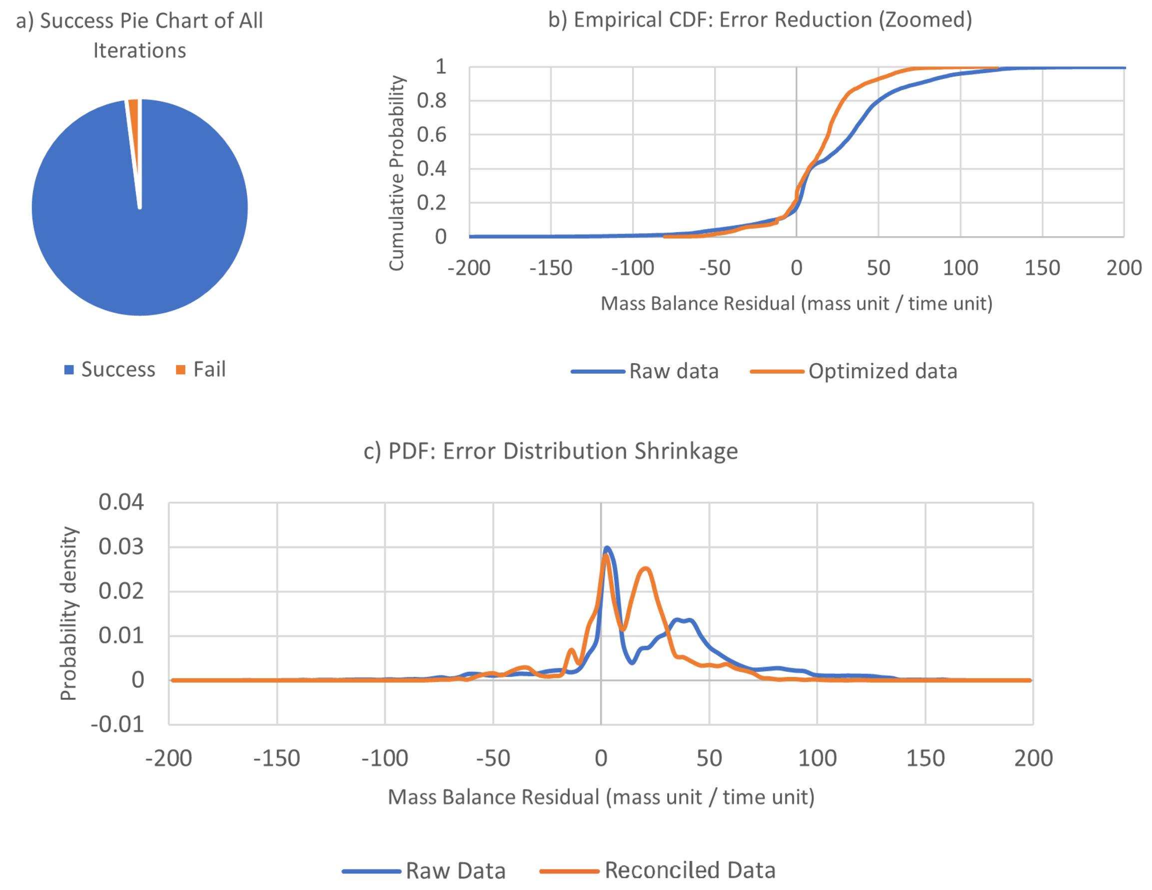

Figure 3. Performance plots of data reconciliation.

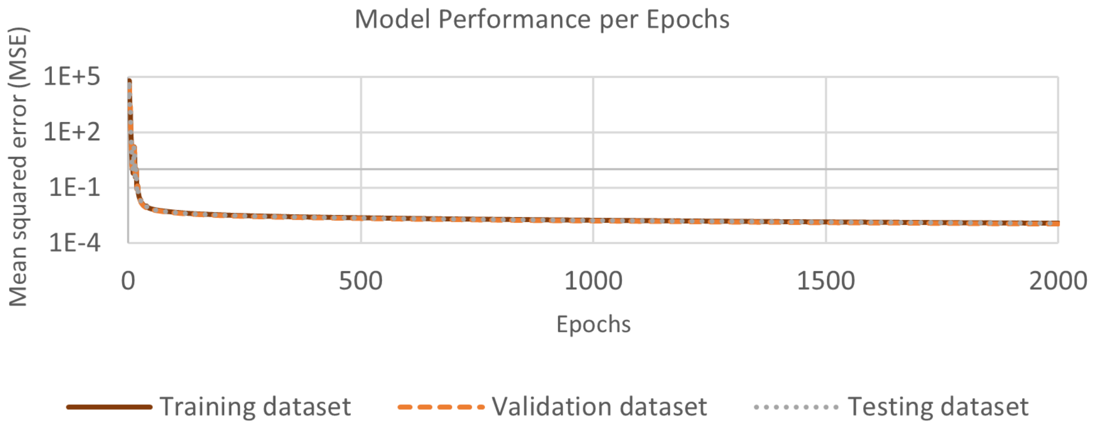

Figure 4. NN model performance over the training epochs.

mass conservation is strictly satisfied for these reconciled values, the residuals plotted here represent the deviation of raw plant measurements from this physically consistent logic. The steep slope around zero confirms high reconciliation precision for most observations. Approximately 70% of residuals are positive, indicating that the reconciliation process systematically compensates for under-measured inflows or unmeasured gains to satisfy mass conservation constraints. This indicates systematic (gross) errors and provides plant personnel with valuable monitoring information.

Complementarily, Figure 3.c) illustrates the probability density function (PDF), which showcases a significant reduction of long tails within the reconciliation process. While the raw measurements exhibit a broad and dispersed noise profile, the optimized residuals are concentrated into two dominant peaks near zero, confirming the effective suppression of random measurement noise. The presence of the secondary peak in the positive region of the horizontal axis further validates the model's handling of the binary-continuous products; it demonstrates the aforementioned systematic inflow errors, indicating that the model successfully identifies and compensates for mass balance inconsistencies within the monitored system by reconciling the raw data toward the enforced physical structure.

Data Reconciliation using Neural Network

We demonstrate the neural network training and validation performance in Figure 4. The plot illustrates the evolution of the Mean Squared Error (MSE) over the maximum 2, 000 epochs, since requirements for early stopping were not satisfied. The logarithmic scale of the graph highlights the rapid initial convergence and subsequent fine-tuning of the model. The training process successfully reduced the performance metric from an initial value of 8.24×10^3 to a final termination value of 1.15×10^{-3} as indicated by the best validation performance achieved at the final epoch.

The high fidelity of the resulting model is further quantified by the regression analysis (not reproduced here), where the network reached a correlation coefficient of $R = 0.9982$ on the independent test dataset. This corresponds to a coefficient of determination $R^2 \approx 0.9964$, signifying that the neural network accounts for approximately 99.64% of the variance in the optimization data. The close tracking of the training, validation, and test curves throughout the 2, 000 epochs confirms a robust learning process with no signs of overfitting, demonstrating that the network has effectively internalized the underlying physical constraints and mass balance logic of the MIQP solver.

Table 2 shows comparison of CPU time as a comparison between optimization solver execution time and neural network inference time, representing potential of neural network for real-time optimization since it is, based on mean values, around 300-times faster. Both minimum and maximum CPU times are exponentially smaller for neural network than for optimization solver.

Table 2: CPU time comparison for optimization solver vs neural network.

CPU time (s)	Optimization solver	Neural network
Minimum	0.059903	0.007845
Mean	2.787424	0.009257
Maximum	60.480676	0.025399

CONCLUSION

We studied inventory data reconciliation in an oil refinery. The presented approach led to a significant increase in the accuracy of tank inventory monitoring and allowed for the quantification of previously unknown flows within the entire refinery system. Consequently, the resulting model provides a valuable tool for the early detection of anomalies, improvement of mass balance calculations, and real-time optimization of liquid inventory management. The future goal would encompass a use of an SCIP solver [11] that can be licensed freely not only for academic but also for commercial purposes.

ACKNOWLEDGEMENT

This work is funded by the Slovak Research and Development Agency under the project APVV-24-0007 and by the Scientific Grant Agency of the Slovak Republic under the grant 1/0263/25, and by the European Union under the grant scheme NextGenerationEU projects no. 09I01-03-V05-00002 and no. 09I01-03-V04-00024.

AUTHOR IDENTIFIERS

Author ORCIDs:
Paulen R: 0000-0002-1599-2634

REFERENCES

1. Leibman MJ, Edgar TF, Lasdon LS. Efficient data reconciliation and estimation for dynamic processes using nonlinear programming techniques. Computers & Chemical Engineering 16:963-986 (1992). https://doi.org/10.1016/0098-1354(92)80030-d
2. Galan A, De Prada C, Gutierrez G, Sarabia D, Gonzalez R. Real-time reconciled simulation as decision support tool for process operation. Journal of Process Control 100:41-64 (2021). https://doi.org/10.1016/j.jprocont.2021.02.003
3. Plácido J, Campos AA, Monteiro DF. Data reconciliation practice at a petroleum refinery company in Brazil. Comput Aided Chem Eng 27:777-782 (2009) https://doi.org/10.1016/S1570-7946(09)70350-5
4. de Oliveira EC, Lourenço FR. Data reconciliation applied to the conformity assessment of fuel products. Fuel 300:120936 (2021). https://doi.org/10.1016/j.fuel.2021.120936
5. Lid T, Skogestad S. Data reconciliation and optimal operation of a catalytic naphtha reformer. Journal of Process Control 18:320-331 (2008). https://doi.org/10.1016/j.jprocont.2007.09.002
6. Taylor JH, del Pilar Moreno R. Nonlinear dynamic

data reconciliation: in-depth case study. 2013 IEEE International Conference on Control Applications (CCA) :746-753 (2013). https://doi.org/10.1109/cca.2013.6662839

7. Bai S, Thibault J, McLean DD. Dynamic data reconciliation: alternative to kalman filter. Journal of Process Control 16:485-498 (2006). https://doi.org/10.1016/j.jprocont.2005.08.002
8. DAMA International. DAMA-DMBOK: Guide to the Data Management Body of Knowledge. Technics Publications (2017).
9. Löfberg J. YALMIP: A Toolbox for Modeling and Optimization in MATLAB. In: Proceedings of the CACSD Conference. Taipei, Taiwan (2004). https://yalmip.github.io/
10. Gurobi Optimization LLC. Gurobi Optimizer Reference Manual. Gurobi Optimization LLC (2025). https://www.gurobi.com
11. Bestuzheva Z. et al. The SCIP Optimization Suite 9.0. Technical Report, Optimization Online (2023). https://www.scipopt.org