ROBUST BAYESIAN DATA RECONCILIATION FOR COMPLEX COMPRESSION TRAINS

Matteo Barbanti, Lorenzo Ferrari
University of Pisa, Italy

Simone Bonacchi, Valerio Depau
Baker Hughes, Italy

ABSTRACT

In complex compression trains, available measurements often do not provide sufficiently accurate estimates of plant performance, primarily due to the presence of hard-to-measure quantities such as fluid leakages. Data reconciliation offers a powerful methodology for estimating unmeasured variables and improving the accuracy of existing measurements, thereby enabling robust performance assessment without requiring costly equipment upgrades. This work employs a novel Bayesian data reconciliation architecture that enables the definition of prior probability density functions for each variable, exploiting physical insight and historical knowledge of the plant under steady-state operating conditions. Results obtained from real industrial data are promising for an accurate estimation of the complete thermodynamic state of the plant, as well as for the capability of the algorithm to detect and isolate potential operational faults Results obtained from real industrial data are promising for an accurate estimation of the complete thermodynamic state of the plant, as well as for the capability of the algorithm to detect and isolate potential operational faults

INTRODUCTION

The core principle of data reconciliation is to use physical modeling as a filter to improve measurement accuracy. Traditionally, this task is formulated as an optimization problem [1,2]. In compression train systems, however, the data redundancy required to achieve significant accuracy improvements is rarely available with the existing sensor layout; consequently, conventional approaches typically provide only marginal benefits. Moreover, the statistical interpretation of classical data reconciliation remains rigorous primarily under linear constraints and Gaussian-distributed variables, conditions that are generally not satisfied in highly non-linear systems such as those of industrial turbomachinery applications.

The proposed architecture combines traditional optimization with a Bayesian model [3,4], introducing prior information to address low redundancy and recover a rigorous statistical interpretation of the reconciled variables. The available data are first processed using traditional Weighted Least Squares Reconciliation (WLSR) and are subsequently refined through Bayesian processing. The likelihood function is constructed directly from the physical model, while the resulting probabilistic space is explored using a Markov Chain Monte Carlo (MCMC) algorithm. The main implementation challenges addressed in this work include transforming the implicit physical model, based on conservation laws, into an explicit form suitable for likelihood evaluation, and mitigating the gross-error smearing effects that commonly affect traditional data reconciliation.

RESULTS AND DISCUSSION

In this work, a probabilistic data reconciliation framework was applied to a low-redundancy system, namely an industrial compression train. The case study includes 35 variables, 25 constraints, and only 10 available measurements, as shown in Figure 1 (left). By using deterministic WLSR to provide a highly informative initial guess, the Bayesian algorithm was able to explore the probabilistic space efficiently and reliably.

A key finding of this study concerns the behavior of the posterior distributions. Unlike traditional data reconciliation methods, which strictly rely on normality assumptions, namely Gaussian measurement noise and linearized constraints, the proposed Bayesian framework allows the algorithm to explore the probabilistic space without imposing such restrictive hypotheses. As a result, the posterior distributions are not artificially forced to be Gaussian and, as shown in Figure 1 (right), may exhibit skewed shapes. This skewness is a direct and valuable reflection of the strong non-linearities inherent in the thermodynamic equations governing gas compression. It demonstrates that the algorithm captures physical behaviors that would otherwise be masked by traditional linear approximations.

The proposed framework improved the accuracy of the measured variables, with uncertainty reductions of up to 40% for pressures and up to 53% for temperatures, as reported in Figure 2. For unmeasured variables, the framework effectively acts as a virtual sensor, exploiting the thermodynamic constraints of the system to substantially reduce the overall estimation uncertainty. The boxplot in Figure 2 shows the reduction in posterior standard deviations with respect to the corresponding priors for the pressure variables. Normalization was consistently applied to preserve the confidentiality of the industrial case study while maintaining rigorous comparability of the statistical dispersion.

The robustness of the estimation framework was further validated across multiple operating points and slightly different compression train configurations, showing consistent stability in all tested scenarios. Although Bayesian sampling approaches, such as MCMC methods, are commonly penalized by higher computational costs compared with deterministic solvers, the proposed framework achieved convergence within a few minutes. This computational efficiency was enabled by a "soft" implementation of the physical and thermodynamic constraints.

By aligning the sampling steps with the physical dependencies of the system, the algorithm significantly reduces rejection rates and captures the target distribution with high accuracy, even in high-dimensional and non-linear settings. This performance not only demonstrates the successful mitigation of the computational bottleneck, but also aligns well with the theoretical optimal acceptance range of approximately 23.4% identified by Gelman [4] for efficient exploration in high-dimensional spaces.

Looking ahead, this methodology provides a solid foundation for advanced plant diagnostics. In the presence of sensor drift, calibration errors, or incipient plant faults, the statistical distance between prior assumptions and posterior estimates can serve as a powerful diagnostic indicator. Future work will focus on integrating this probabilistic framework into a real-time Digital Twin environment, leveraging discrepancies between prior and posterior distributions to trigger automated anomaly detection and support predictive maintenance strategies.

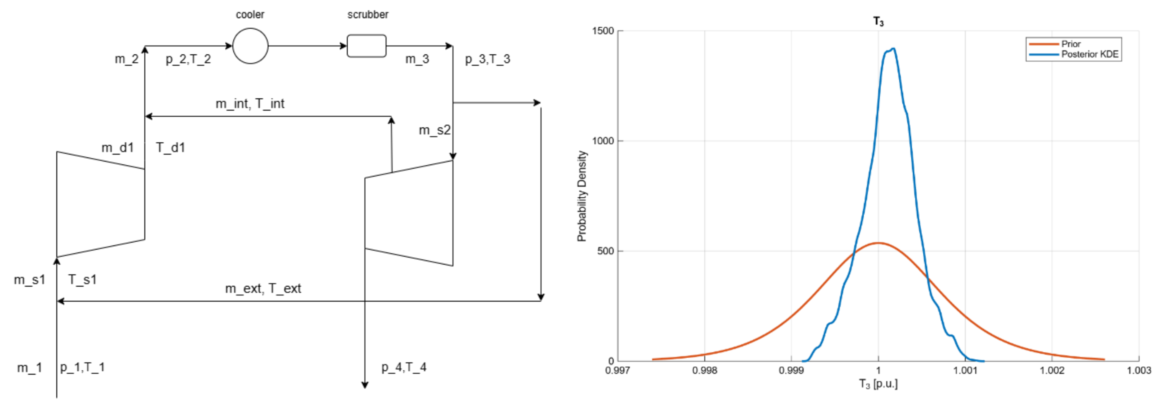

Figure 1: Plant scheme (left) and skewed shapes of a posterior distribution (right)

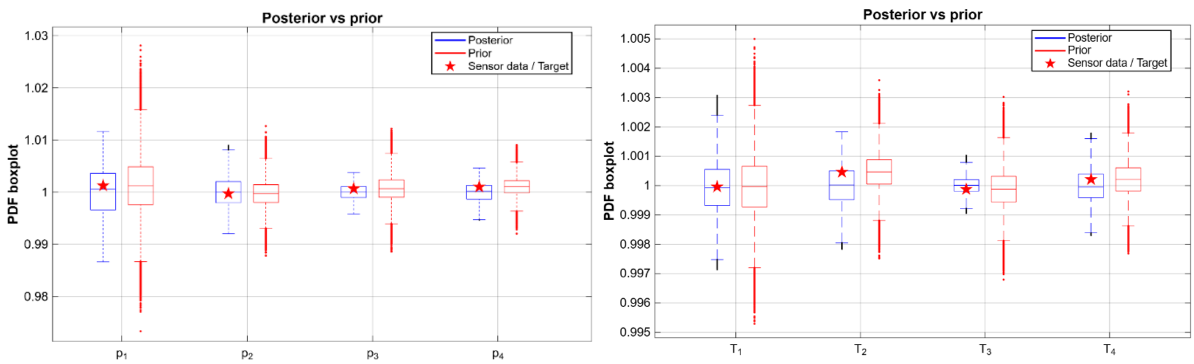

Figure 2: Priors and posteriors distributions for some pressures (left) and temperatures (right)

REFERENCES

[1] S. Narasimhan and C. Jordache, Data Reconciliation and Gross Error Detection: An Intelligent Approach. Houston, Texas: Gulf Professional Publishing, 1999.

[2] X. Jiang, P. Liu, and Z. Li, "Data reconciliation for steam turbine on-line performance monitoring," Applied Thermal Engineering, vol. 70, no. 1, pp. 122–130, 2014.

[3] O. Cencic and R. Fr¨uhwirth, "A general framework for data reconciliation—Part I: Linear constraints," Computers & Chemical Engineering, vol. 75, pp. 196–208, 2015.

[4] A. Gelman, J. B. Carlin, H. S. Stern et al., *Bayesian Data Analysis*, 3rd ed. Boca Raton, FL, USA: Chapman and Hall/CRC, 2013.