Workshop Report

Building AI-Driven Digital Twins for the Process Industry

Author: Dr. Kamal Al-Malah
Focus Case Study: C3 Splitter Optimization and Fault Detection

---

1. Introduction: Closing the “Reality Gap” in Modern Engineering

Classical chemical engineering education relies heavily on *clean*, idealized simulations using tools such as Aspen Plus and Aspen HYSYS. While these simulators are indispensable for design and analysis, they do not reflect the true operating environment of industrial plants, where sensor drift, equipment fouling, process disturbances, and noisy data are unavoidable.

This workshop introduces a **hybrid digital-twin philosophy**, combining first-principles process models with artificial intelligence (AI) techniques. The objective is to bridge the gap between theoretical rigor and plant-floor reality, enabling engineers to extract reliable insight from imperfect industrial data.

Key Skills Acquired

At the end of the workshop, participants will be able to:

1. **Synthetic Data Engineering:** How to create industrial datasets when real plant data is proprietary or unavailable.
2. **Soft-Sensor Development:** Building a virtual instrument that "cleans" noisy physical sensors in real-time.
3. **Predictive Maintenance:** Identifying the "Residual" (the difference between the AI prediction and the noisy sensor) to detect reboiler fouling before it causes a shutdown.
4. **MATLAB Fluency:** Transitioning from simple scripts to advanced AI objects (fitrgp).

Final message: The modern engineer does not merely read plant data; the *hybrid engineer* designs digital twins that transform noisy measurements into actionable insight.

This is a sample of the full Workshop Report. The full version includes the complete MATLAB source code and training datasets.

2. Step 1 – Establishing the Ground Truth (Physics-Based Baseline)

2.1 Interactive Exploration: C3 Splitter Mini-Simulator

Reference File: `C3_Shortcut_Column_App.m`

The workshop begins with an interactive MATLAB App designed to explore the sensitivity of a C3 splitter using the classical **Fenske–Underwood–Gilliland (FUG)** framework. Trainees can manually adjust key operating variables, such as feed propylene mole fraction and column pressure and immediately observe their impact on:

- Theoretical number of stages (N)
- Minimum number of stages (Nmin)
- Reflux ratio (R)
- Minimum reflux ratio (Rmin)

Figure 1 illustrates the graphical interface, which provides immediate feedback and builds physical intuition.

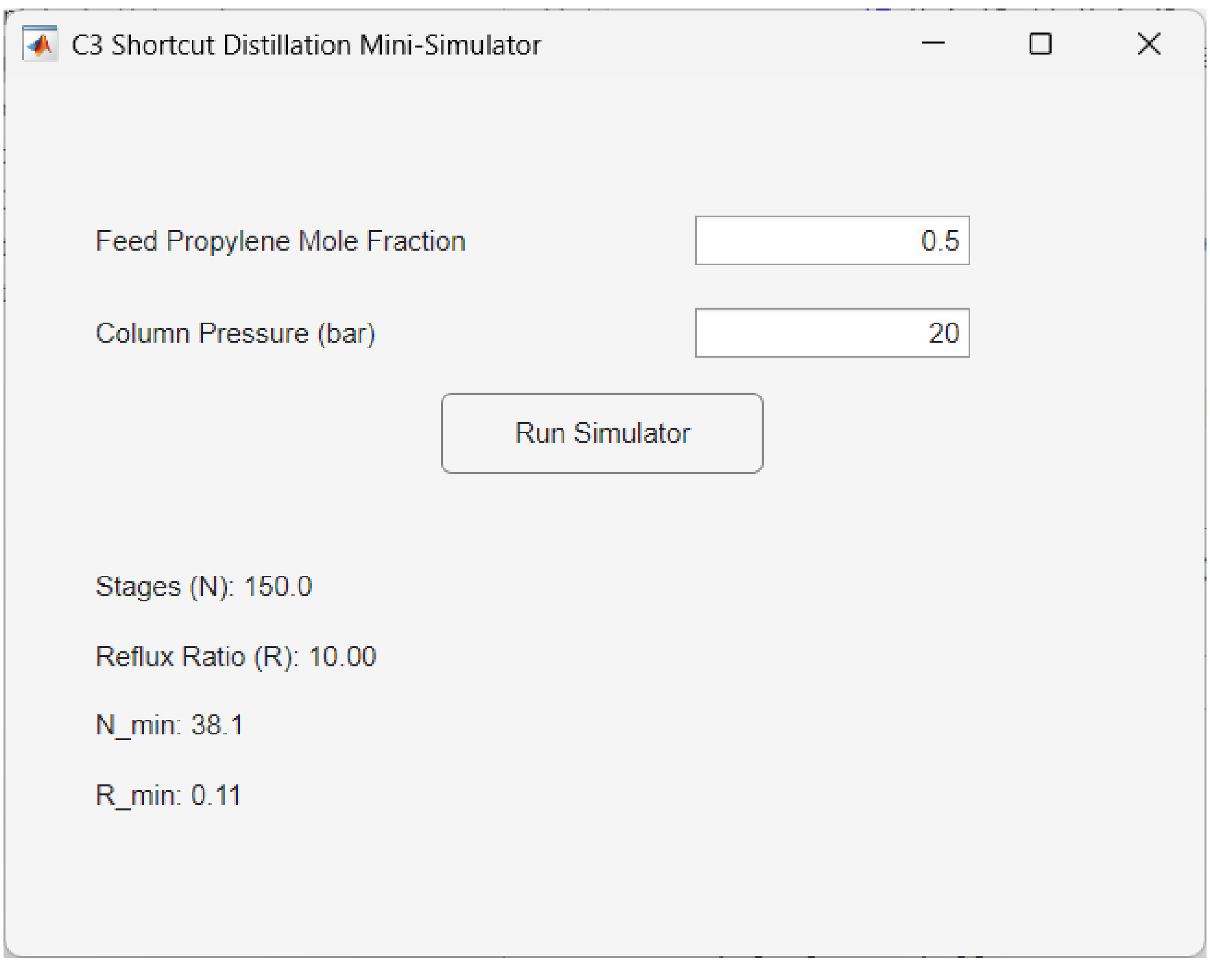

Figure 1. Interactive GUI for adjusting feed propylene mole fraction and column pressure to observe the resulting stages and reflux ratio using the Fenske–Underwood–Gilliland method.

This is a sample of the full Workshop Report. The full version includes the complete MATLAB source code and training datasets.

In traditional engineering, simulations often represent pristine, steady-state conditions that do not reflect industrial reality. To bridge this "**Reality Gap**", controlled data decay is deliberately introduced into clean datasets to simulate the three primary enemies of a process engineer: instrument drift, equipment fouling, and stochastic process shocks. This transition utilizes high-fidelity failure signatures, such as the Kern-Seaton model for reboiler fouling and linear calibration drift for sensors, to move students from theoretical modeling toward managing a "**Digital Twin**" of a failing plant. By injecting disturbances like ±5°C feed temperature spikes, the environment tests the AI's capacity to distinguish systemic failures from random noise.

The solution to these industrial inaccuracies is the "AI Bridge", which employs Gaussian Process Regression (GPR) to reconstruct true process signals from corrupted measurements. While a human operator might only see a confusing "cloud" of noisy data points, the GPR model is trained using clean historical data as the "known truth" to filter out noise while preserving underlying physics. This "magic" effectively generates an AI-corrected signal that ignores electrical jitter and sensor vibration, tracking the theoretical truth perfectly. Ultimately, this allows Digital Twin to restore operational clarity and provides a foundation for predictive maintenance before equipment failures cause a shutdown. Figure 2 shows that:

- The AI-driven "soft-sensor" effectively bridges the gap between industrial reality and theoretical physics by using a Gaussian Process Regression (GPR) model to reconstruct the true process signal from a cloud of noisy measurements.
- While the human eye might see only a chaotic "cloud" of red dots caused by sensor noise and vibration, the AI-corrected green line ignores these inaccuracies to track the blue physics-based truth perfectly.

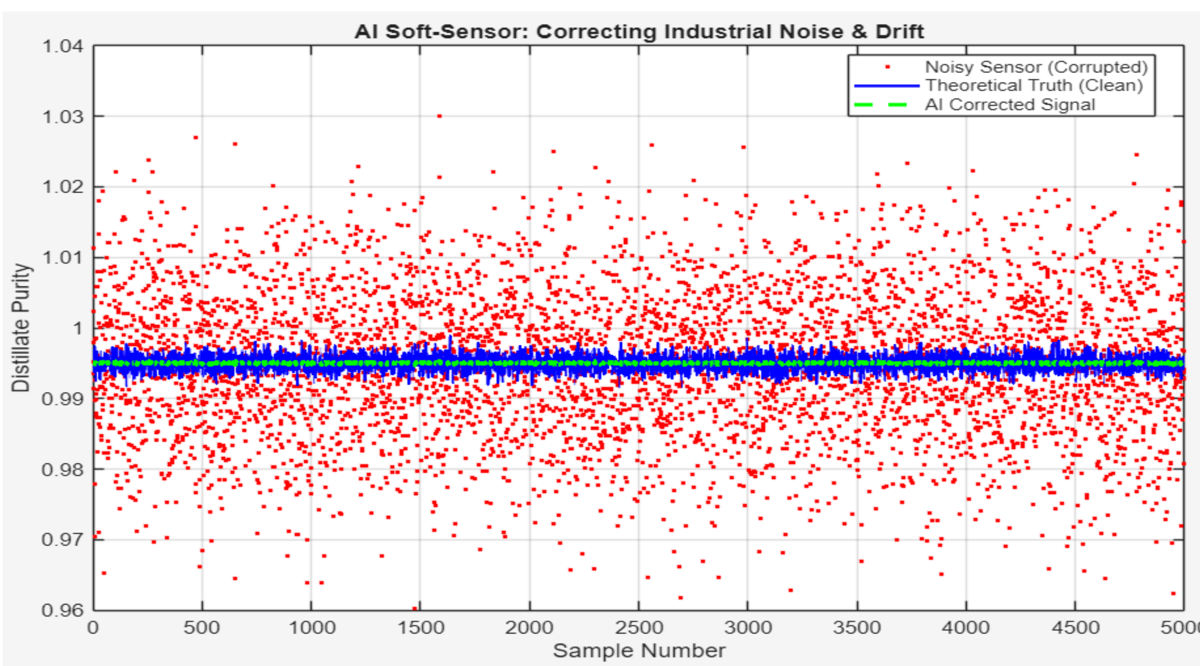

Figure 2: The plot of distillate purity using the clean data (blue vibrating curve), the noisy sensor (red dots), and the AI-corrected signal (green dotted line).

This is a sample of the full Workshop Report. The full version includes the complete MATLAB source code and training datasets.