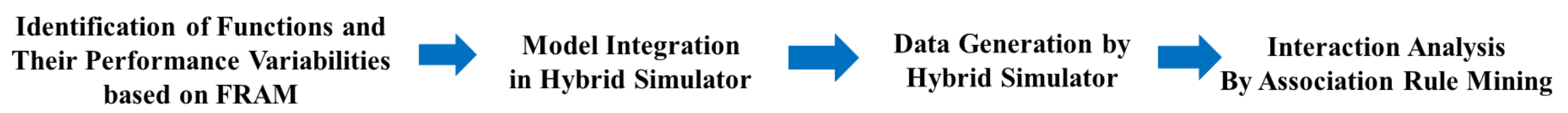
Generation of Synthetic Data using a FRAM-based Model and Identification of Hazards of Complex System using Association Rule Mining

Mengyi Yu[a,b], Noor Quddus[a], Costas Kravaris[b], and M. Sam Mannan[a,b],
[a] Mary Kay O'Connor Process Safety Center,
[b] Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX 77843-3122, USA

Background

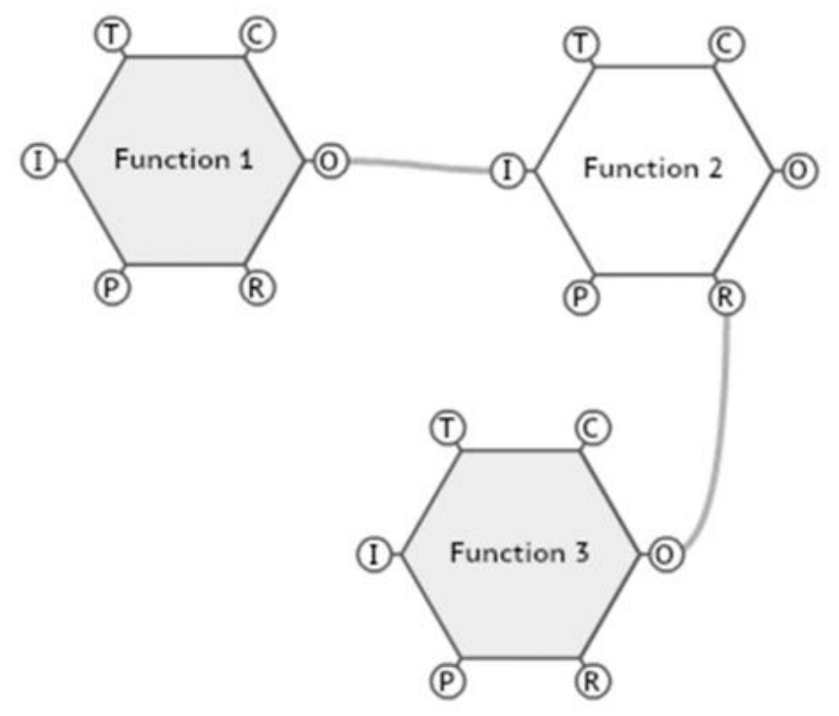
- Incidents still occur globally in the high-risk process industries.
- Process industries are complex system involving many functions.
- Operations in process industries are with high-degree of complexity and tight coupling.
- Interactions among different functions may aggregate into nonlinear and circular relationships, and lead to process incidents as emergent failures.
- Systematic examinations are needed for the proactive identification of potential hazards in a complex system.
- In a complex system, there are millions of interactions among the system variables and it is very hard to identify the critical ones.
- Industry data is not publicly available and we have minimal understanding of what data to look into to identify critical interactions.
- The Functional Resonance Analysis Method (FRAM) is getting attention in recent years to understand hazards emerging from function couplings in complex systems.
- However, FRAM has limitations to be applied in process industries since the function couplings are estimated based on qualitative analysis.
- In the study, a framework based on the FRAM was developed to build a simulator to simulate function interactions in a chemical plant.
- Associational rule mining was applied on the synthetic data to quantify function couplings and to understand the couplings leading to potential hazard scenarios.

Methodology


Identification of Functions and Their Performance Variabilities based on FRAM → **Model Integration in Hybrid Simulator** → **Data Generation by Hybrid Simulator** → **Interaction Analysis By Association Rule Mining**

STEP 1:

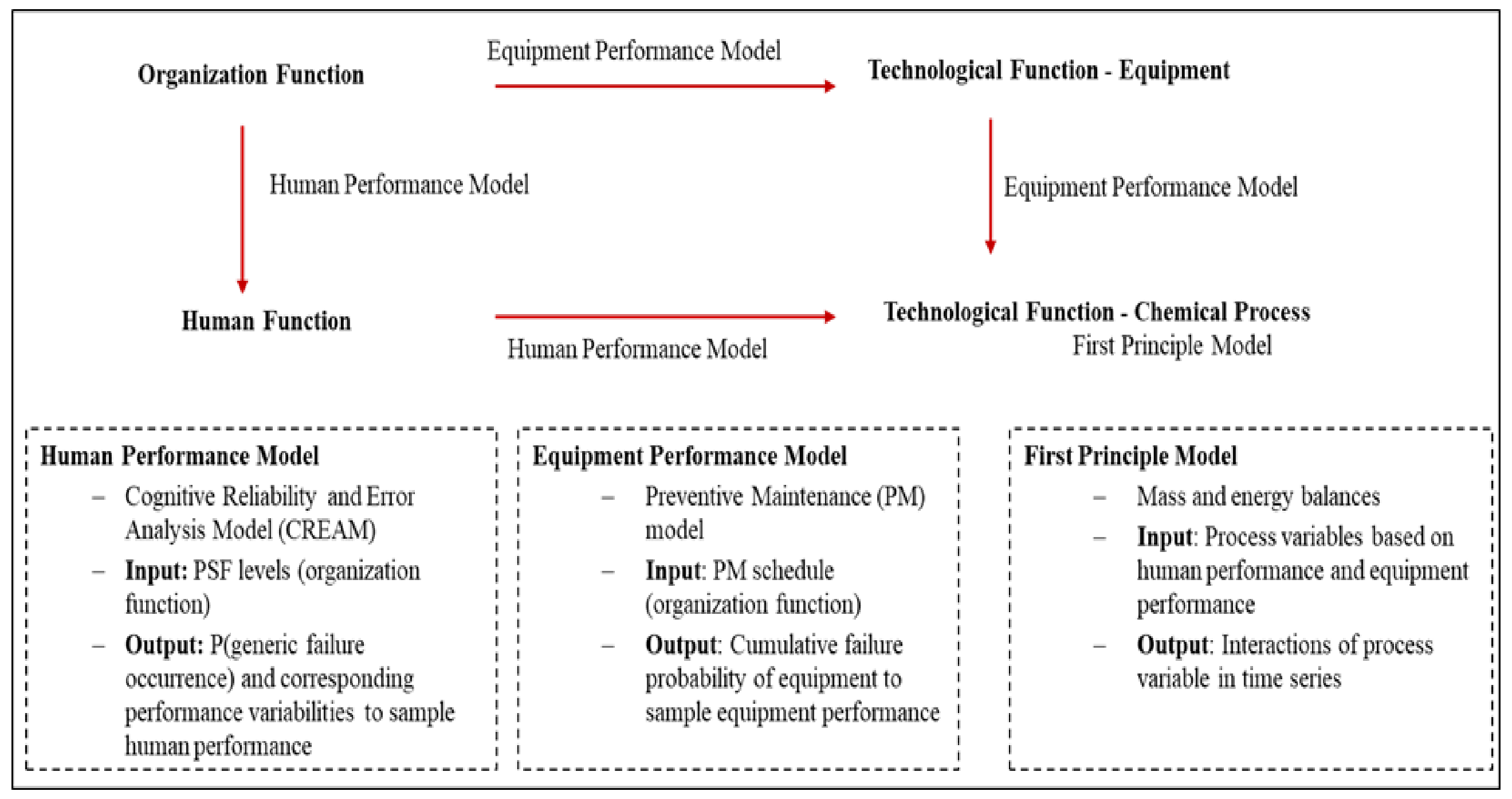
- In the study, FRAM was used to describe the selected system. Functions and their performance variabilities were identified.
- FRAM describes each function by six aspects (input, precondition, resource, control, time, and output).
- Couplings between functions are described by connecting appropriate aspects of upstream-downstream functions.



- Input (I): Starts the function
- Precondition (P): Exists before carrying out
- Resource (R) : Needed/consumed by the function
- Control (C) : Monitors/controls the function
- Time (T) : Temporal constraints
- Output (O): Result of the function

STEP 2:

- To overcome the limitations of the FRAM, the study integrated a human performance model, an equipment performance model, and a first-principle chemical process model into a hybrid simulator based on the FRAM. The hybrid simulator was built in MATLAB using Simulink and Stateflow modules.



Human Performance Model
- Cognitive Reliability and Error Analysis Model (CREAM)
- **Input:** PSF levels (organization function)
- **Output:** P(generic failure occurrence) and corresponding performance variabilities to sample human performance

Equipment Performance Model
- Preventive Maintenance (PM) model
- **Input**: PM schedule (organization function)
- **Output**: Cumulative failure probability of equipment to sample equipment performance

First Principle Model
- Mass and energy balances
- **Input**: Process variables based on human performance and equipment performance
- **Output**: Interactions of process variable in time series

STEP 3:

- Possible function interactions were simulated by sampling strategies using the hybrid simulator.
- Synthetic data was generated for interaction analysis.

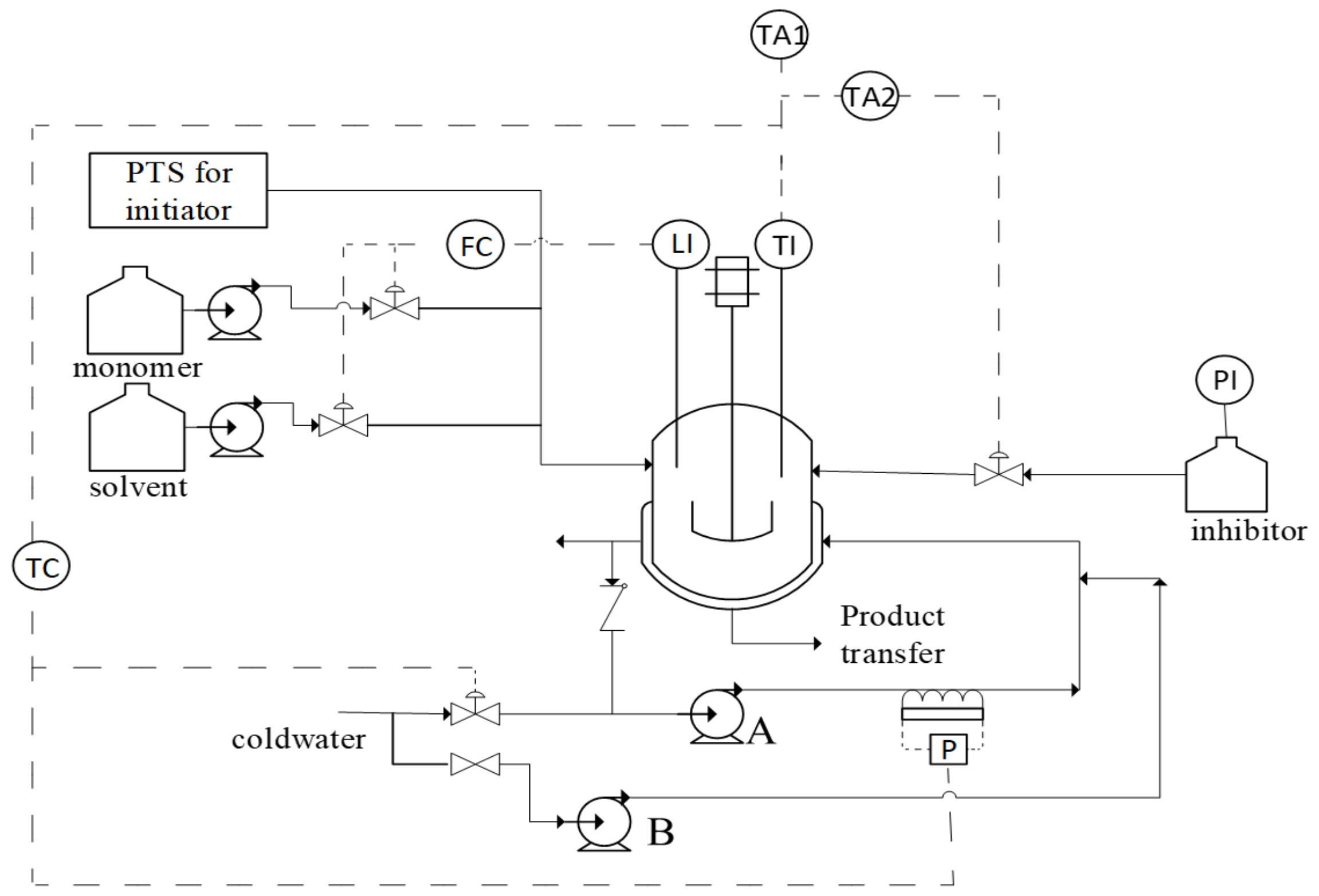
STEP 4:

- Association rule mining was applied on simulation data to quantify couplings between directly connected upstream and downstream functions.
- Association rule mining is widely applied to identify patterns from historical data.
- Association rules are in the format: Antecedents → Consequents
- To extract couplings between upstream and downstream functions, rules whose antecedents are upstream function performance and consequents are downstream function performance were extracted.
- Lift value of association rule was used as the quantitative metric to describe the coupling.

$$\text{Lift } (A \rightarrow C) = \text{Probability } (A \cap C) / (\text{Probability}(A) \text{ Probability } (C))$$

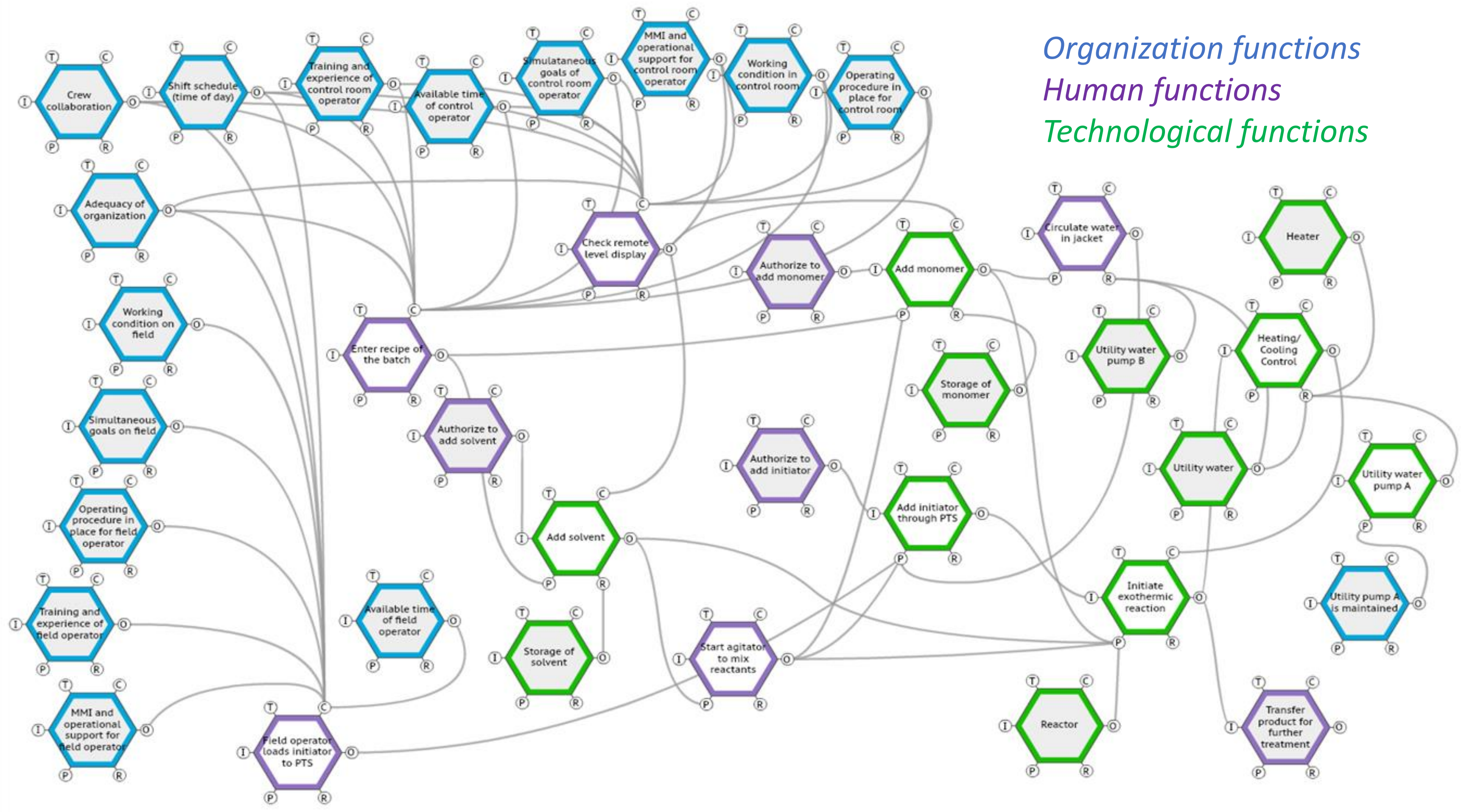
where A represents antecedents and C represents consequents.

Case Study: Batch Polymerization Process

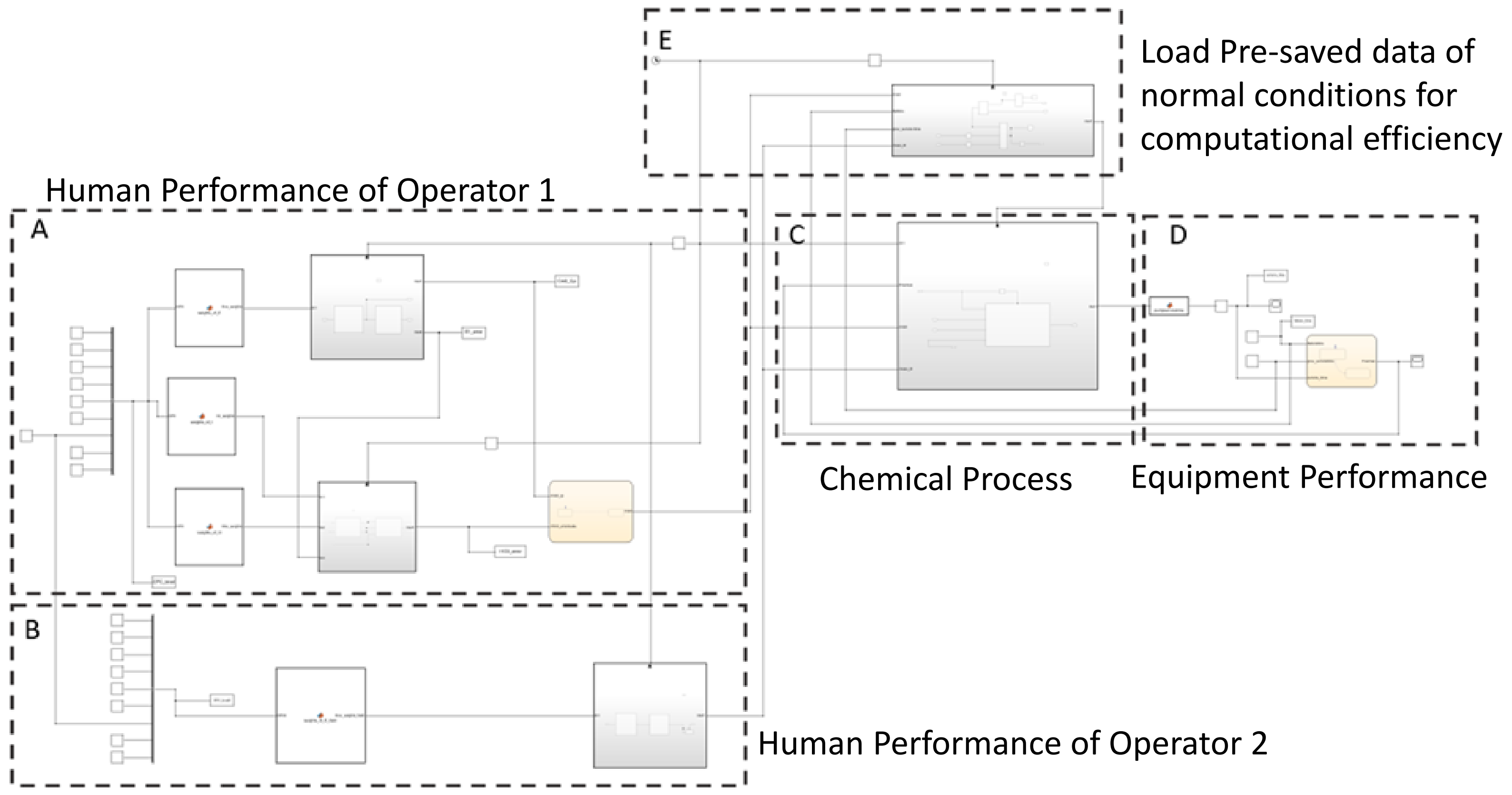


The polymerization process is a batch operation involving one control room operator (CRO) and one field operator (FO). The. During normal operations, FO is responsible to load 742-kg initiator into the power transfer system (PTS), and the CRO is responsible for charging reactants into the reactor to start an 8-hr batch production. On control panel CRO enters the batch recipe which requires manually feeding the initial monomer concentration (3.66 kgmol/m3) and the corresponding solvent concentration is automatically calculated based on the pre-defined batch size. After the recipe is entered, CRO authorizes the automatic solvent charging process, starts the agitator, then authorizes the automatic monomer charging process. After the solvent and monomer are charged into the reactor, CRO needs to check the remote level displays to ensure the appropriate amount of raw materials is charged. Otherwise, the amount of reactants needs to be adjusted accordingly. CRO circulates utility cooling water to the reactor jacket and authorizes releasing the initiator from PTS to the reactor to start the reaction. The heating and cooling control system controls the reactant temperature to make the reaction follow the pre-defined temperature profile. The reaction mass is transferred for further treatment after the batch runs for 8 hours.

FRAM of the Batch Polymerization Process



Hybrid-simulator to Simulate Function Interactions in FRAM



Simulation using High Performance Research Computing Clusters

- Cluster: Ada Cluster provided by High Performance Research Computing at Texas A&M University, College Station, TX.
- Software: MATLAB, Simulink and Stateflow modules
- Job submission: a job was submitted to simulate 3650 batches in series. Each batch was 8-hour in MATLAB simulation time, and data was exported by seconds.
- # Cores requested: 8
- Average memory used: 15,000 MB
- # run time: 10,000s ~ 30,000s. Run time varied since simulations were only executed for batches when initial conditions were different with normal condition. When initial conditions of a batch was the same as those of normal condition, the simulation of the batch was skipped and a pre-saved data was loaded to save computational time. Initial conditions were sampled stochastically thus number of batches which were actually simulated varied in each job submission.
- Data Generated: 7,300 batch operations in current phase

Quantify and Identify Function Couplings Leading to Potential Hazards

- **Example: Couplings between the downstream function of "Enter recipe of the batch" and its upstream functions**

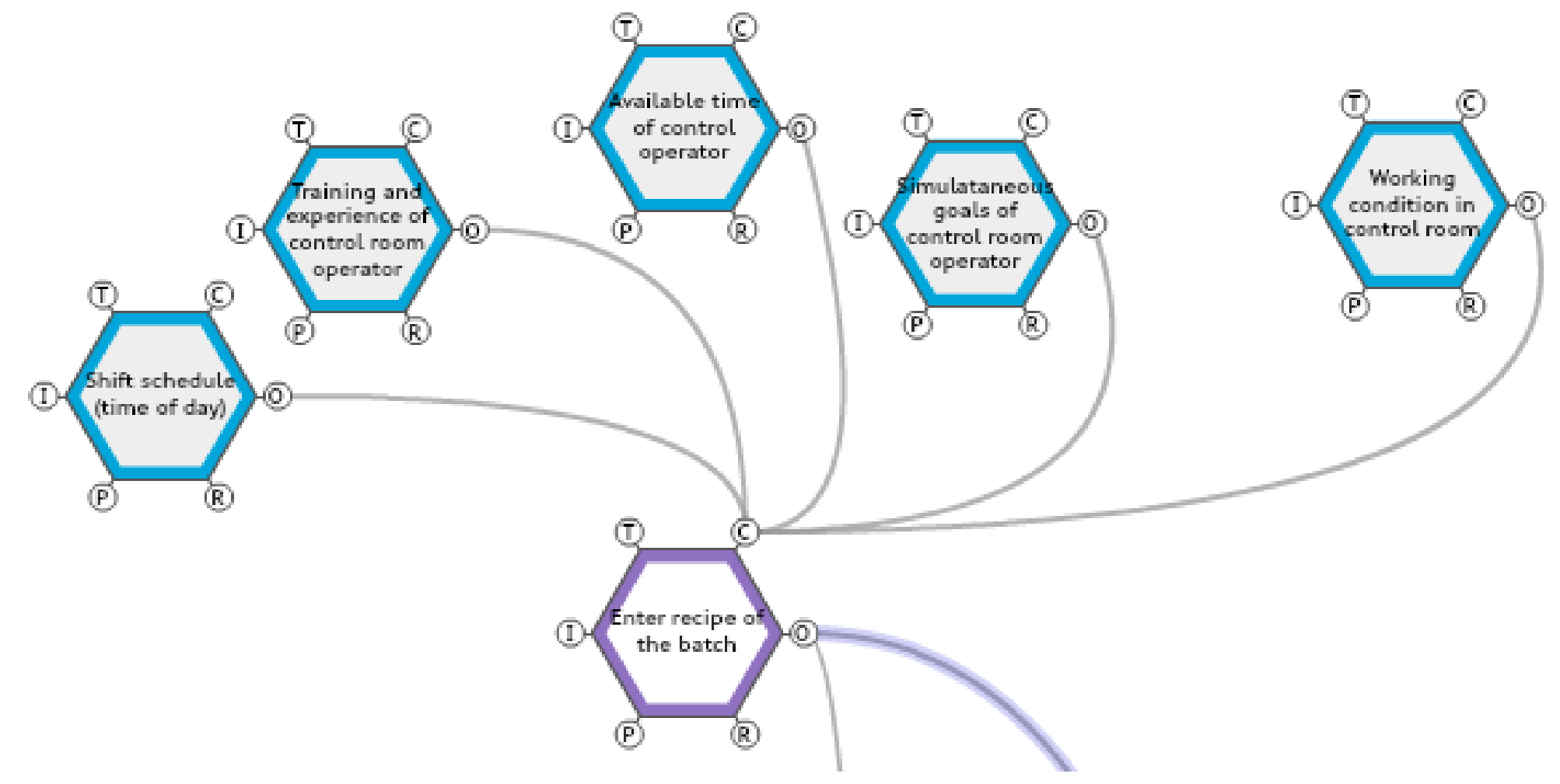


Table: Association rules between organization functions and the function "Enter recipe of the batch".

No.	Antecedents	Consequents	Lift
1	adequacy of training = 3, available time = 3, number of simultaneous goals = 3, time of day = 2, working conditions = 3	C_{m0}_cp = "incorrect"	10.58
2	adequacy of training = 2, available time = 3, number of simultaneous goals = 3, time of day = 2, working conditions = 3	C_{m0}_cp = "incorrect"	8.86
3	adequacy of training = 3, available time = 3, number of simultaneous goals = 3, time of day = 1, working conditions = 3	C_{m0}_cp = "incorrect"	6.75
4	adequacy of training = 3, available time = 3, number of simultaneous goals = 2, time of day = 2, working conditions = 3	C_{m0}_cp = "incorrect"	6.51
5	adequacy of training = 3, available time = 3, number of simultaneous goals = 1, time of day = 1, working conditions = 3	C_{m0}_cp = "incorrect"	6.14

- Antecedents in the table are performances of upstream functions. The performances are described as categorical variables. 1 represents the most competent levels, while 3 represents the lest competent levels.
- Consequences represents the human error when control room operator entered an incorrect amount of monomer to be charged into reactor, which could lead to potential temperature excursion in the process.
- The table shows the couplings between upstream organizational functions and the downstream human error.
- As the lift value of rule 1 shows, when all the upstream functions perform at the least competent levels, it has strongest couplings with the downstream function performance.
- When some upstream functions perform at the most competent levels, while others performance at the least competent levels, the couplings is difficult to be evaluated, if without quantification, As rule 5 shows, number of simultaneous goals =1 and time of day = 1 decrease the occurrence likelihood of the human error, while the other three functions are at the levels which increase the occurrence likelihood of the human error. The association rule indicates that the combination positively associates with the occurrence of the human error.
- Next stage of the study is to expand the analysis to entire system to identify critical paths showing how functions interact leading to potential temperature excursions.

Reference

Yu, M., Quddus, N., Kravaris, C.,Mannan, M. S. 2020. Development of a FRAM-based framework to identify hazards in a complex system. *Journal of Loss Prevention in the Process Industries, 63*, 103994. doi:https://doi.org/10.1016/j.jlp.2019.103994