Joshua Pulsipher | Research Statement

pulsipher@cmu.edu • pulsipher.info

Overview

My research in **process systems engineering develops computational methods in data-science and decision-making under uncertainty to tackle environmental, societal, and sustainability challenges in engineering and science** (see Figure 1). These challenges exhibit interactions over a **hierarchy of space-time scales** that experience uncertainty stemming from environmental factors, fitted parameters, and/or molecular length-scales; characterizing uncertainty allows me to capture real-world behavior and quantify the **risk/reward** of candidate decisions. Furthermore, I operate at the intersection of chemical engineering, mathematics/statistics, and computer science in open-source **software-accelerated research** that expedites scientific discoveries and makes them broadly accessible. To date, this vision has yielded **13 completed scholarly publications**, **1 pending patent**, 5 software products, contributions to **proposals totaling $698K**, 33+ presentations/seminars (e.g., **AICHE 2022 plenary speaker**), and my network of 27+ collaborators in academia, national laboratories, and industry. My future research group will continue this vision forward by designing **sustainable REE-CM supply chains**, creating **data-driven wildfire mitigation** strategies, advancing **computer vision aided control**, and pursuing advancements in **random field optimization** and **neural operator surrogates**.

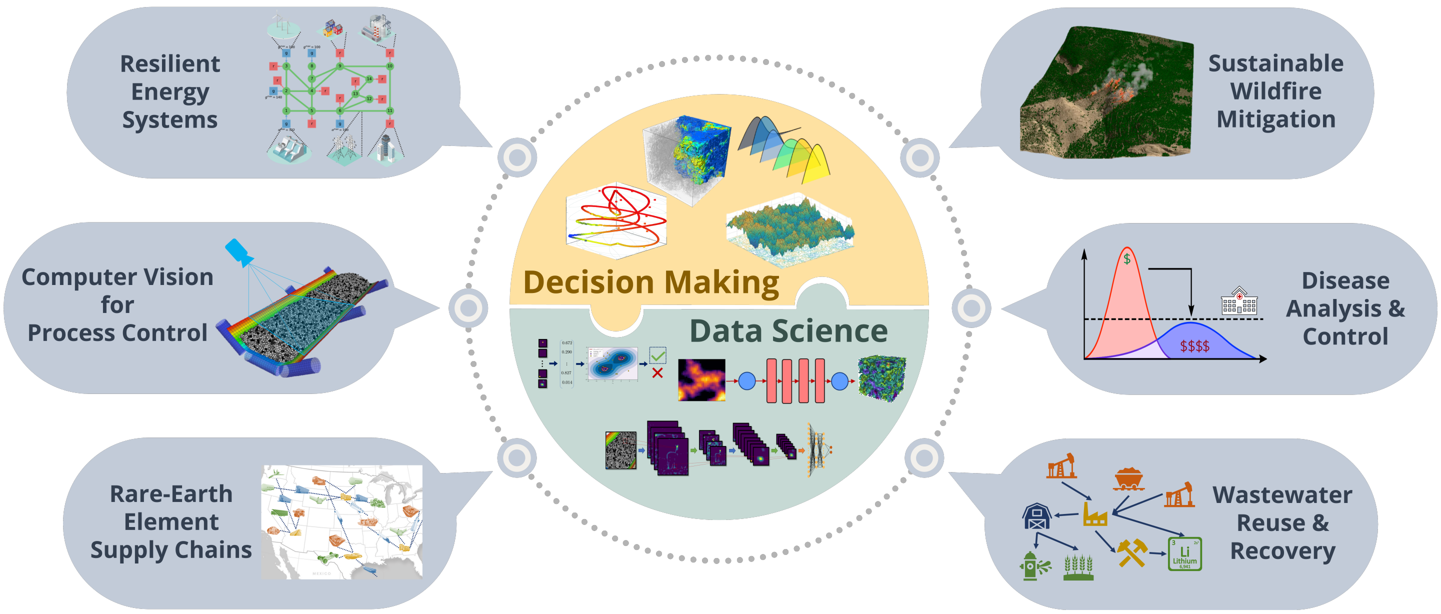

Figure 1: Summary of highlighted research interests.

Research Experience

This section highlights portions of my research experience related to robust energy infrastructure [1, 2, 3, 4, 5], computer vision aided process control [6, 7], REE-CM recovery networks, and software-accelerated scientific discovery [8].

Robust Energy Infrastructure Assessment and Design

Rare, high-impact events (e.g., extreme weather, international conflicts) can inflict severe disruption/damage to energy infrastructure. In early 2021, extreme winter weather caused a power crisis in Texas (due to inadequate weatherproofing) killed 200+ people and cost $200 billion in damages [9, 10]. The Department of Energy (DoE) project Multifaceted Mathematics for Rare, High-Impact Events in Complex Energy and Environment Systems (MACSER) sought mathematical strategies to address this challenge. Under MACSER, I first proposed

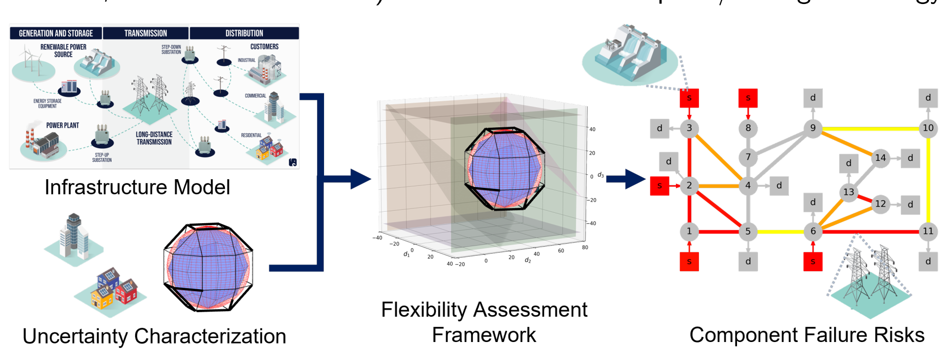

Figure 2: Flexibility assessment framework for energy grids.

a framework to quantify the **flexibility** (the ability to feasibly operate with random fluctuations) of **complex energy infrastructure** (illustrated in Figure 2) which I implemented in `FlexibilityAnalysis.jl` [1, 2]. I also developed a flexible energy system design strategy that is **orders of magnitude faster** than previous techniques, enabling us to design real-world sized systems [3]. With these insights, I also developed a framework for assessing and designing the **reliability** (the ability to operate with random equipment failure) of energy infrastructure networks, critical to avoid disruptions like the Texas power crisis [4]. Finally, I created a **unifying mathematical abstraction** for decision-making over infinite spaces (e.g., time, space, uncertainty) that facilitates theoretical crossover and collaboration (see Figure 3).

Software-Accelerated Scientific Discovery

I create **open-source software modeling frameworks** that accelerate computational discovery and ensure that such discoveries are **accessible and reproducible** across diverse disciplines. For instance, my unifying abstraction for infinite-dimensional optimization problems is implemented in `InfiniteOpt.jl`. This enables computational experimentation for **rapid theoretical crossover** that facilitates scientific discoveries (see Figure 3) such as highly accurate dynamic model identification [8], new risk metrics for decision-making [11], characterizing uncertainty via random field theory [12], the use of neural operators in decision-making [13], and event-constrained optimization [5]. Notably, I am invited to give a **plenary talk on event-constrained optimization** (which makes less conservative decisions via system specific logic) at AICHE 2022. I plan to refine the use of **neural operator surrogates** and **random field uncertainty** characterizations to enable the proposed applications in my future research.

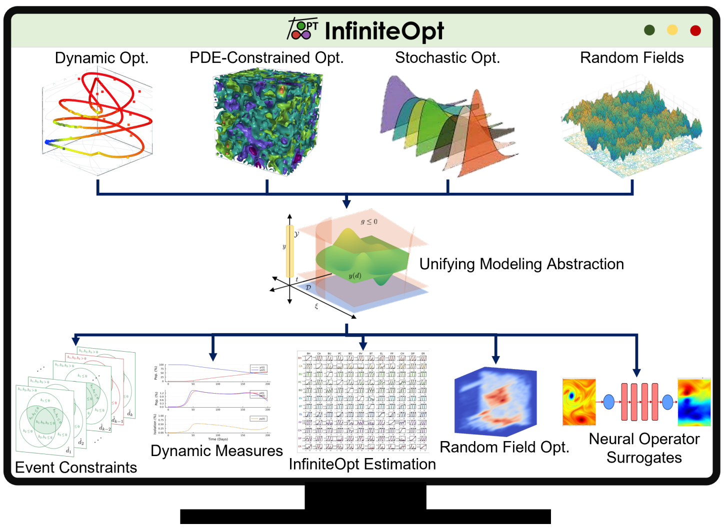

Figure 3: Scientific discovery via `InfiniteOpt.jl`.

Computer Vision for Process Control

Computer vision sensors rapidly measure states from video data and are increasingly deployed in process control. For instance, to combat climate change incurred by flare stack emissions, companies have started using computer vision sensors (which use convolutional neural networks) to automatically monitor emissions. A vulnerability these sensors introduce is **erroneous measurement** when subjected to visual disturbances, which can incur severe safety/profitability consequences. Working with ExxonMobil, I developed a **tailored novelty detection framework** to assess sensor prediction quality in real-time, which has a **lower overhead and higher accuracy** than off-the-shelf methods (**patent pending**). Recently, to generalize this work I proposed the Sensor Activated Feature Extraction One-Class Classification (SAFE-OCC) novelty detection framework (see Figure 4) [7].

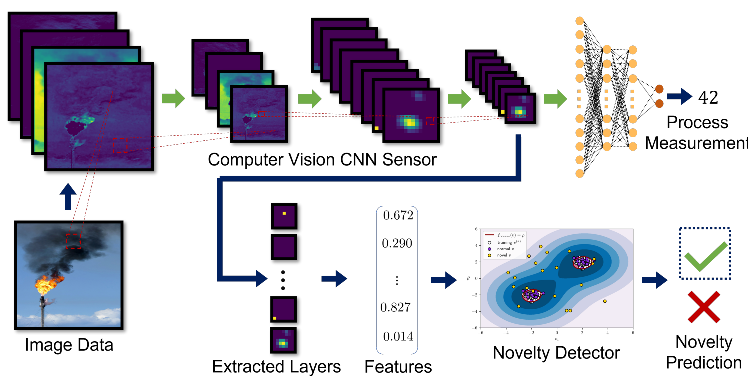

Figure 4: The SAFE-OCC novelty detection framework.

Rare Earth Element & Critical Material Recovery Network Design and Operation

Rare earth elements (e.g., yttrium) and critical materials (e.g., lithium) are crucial for manufacturing electronics, pharmaceuticals, renewable energy generators, and more. Intriguingly, certain **wastewater streams** can contain REE-CMs, which may incentivize costly wastewater treatment to mitigate its impact. Hence, Prof. Carl Laird and I proposed to develop a **decision-making framework** to explore recovering REE-CM from wastewater and how this decreases the cost of treatment. This framework will consider

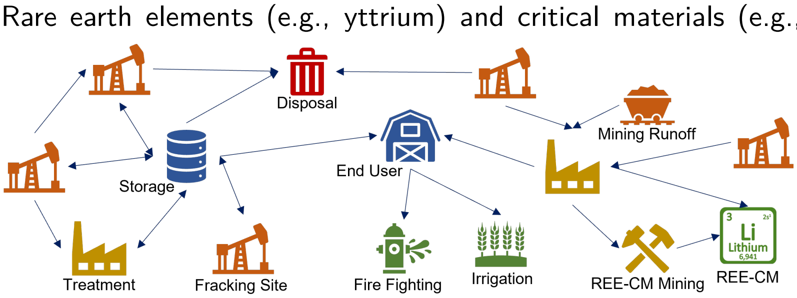

Figure 5: Candidate multi-enterprise water network configuration.

the design and operation of a **multi-enterprise water network** (see Figure 5) to treat/transport/reuse water from diverse sources. This proposal was recently approved by the DoE for ∼**$550k in funding**. We will collaborate with the Produced Water Application for Beneficial Reuse, Environmental Impact, and Treatment Optimization (PARETO) project teams at National Energy Technology Laboratory (NETL) and Lawrence Berkeley National Laboratory which work closely with a board of industrial stakeholders.

Future Research

My research group will develop **tailored computational frameworks** in data-science and decision-making under uncertainty for high-impact engineering challenges in **energy systems**, **the environment**, **sustainability**, and other pertinent areas. Here, my **software-accelerated research** paradigm will give my group a competitive edge. This vision will not simply extend my current research, but will also expand into other problem areas (e.g., wildfire mitigation) and actively exploit **state-of-the-art computational strategies** from other research domains (e.g., neural operators). Below, I propose four short to medium term projects that will guide my group in fulfilling this long-term research vision.

Project 1: Methods Toward Sustainable REE-CM Extraction, Use, & Recovery

Establishing a **sustainable and economically viable domestic REE-CM supply chain** is a high priority to industry and the U.S. government, and is essential to avoid disruptions in manufacturing key technological products. Hence, the DoE recently awarded $19 million to 13 REE-CM related projects at universities throughout the U.S. under the direction of NETL. These efforts seek to better identify sources of REE-CMs from coal and coal by-products (e.g., coal ash) in U.S. coal basins and further develop separation/refinement technologies. However, determining how to design a viable REE-CM supply chain remains a largely unaddressed challenge. Here, key decisions pertain to source selection, technology placement, raw-material transportation, and required government policies (if any), which are highly coupled and **multi-scale** in nature. Moreover, these considerations are subject to **time-dependent uncertainties** such as REE-CM recovery yields and material demand.

Hence, my group will develop a framework to design REE-CM supply chains that enables us to **model candidate infrastructure** configurations (e.g., Figure 6), optimally choose components (i.e., sources, technologies, transportation, end-uses), and facilitate **techno-economic analysis** under transient uncertainty (e.g., recovery yields, demand). We can also investigate economic opportunities for cooperation across this multi-enterprise network. We will leverage state-of-art approaches such as **neural operator surrogates** and decision-making with **random field uncertainty**. My current collaboration with NETL on REE-CM projects, coupled with my research expertise, uniquely equips my group to tackle this engineering challenge. We will collaborate with subject matter experts in my existing network at NETL, CMU, and in industry.

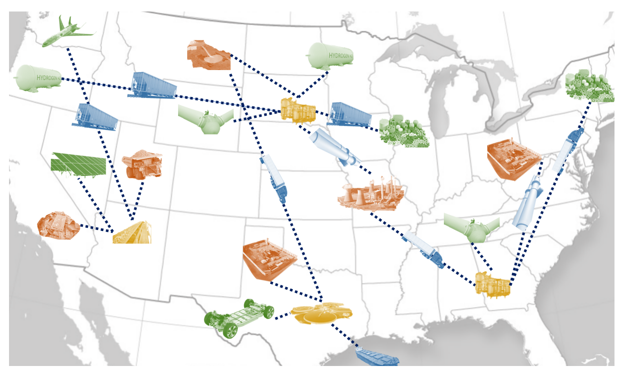

Figure 6: *Candidate REE-CM supply chain with sources (orange), refinement technologies (gold), transportation routes (blue), and end-uses (green).*

Potential Funding Sources: DOE (extending my research with NETL), SNL (proposed REE-CM projects), AIMS (proposed REE-CM projects), DoD (REE funding), and NSF (NSF 22-060, CAREER, NSF 20-570, & PD 22-8084).

Project 2: Data-Driven Modeling for Sustainable Wildfire Mitigation

Climate change is **increasing wildfire extent and severity**. The 2018 Camp Fire destroyed ∼19,000 buildings (costing $16.5 billion), killed 85 people, and caused hundreds of excess deaths from poor air quality [14]. Consequently, the U.S. spends $1+ billion annually to combat wildfires. Increases in fire activity, wildland urban interfaces, suppression costs, and controlled burn hazards suggest that current **fire management practices are unsustainable** and a shift to proactive responses is needed, which requires more advanced modeling and decision-making tools. However, wildfires entail highly complex characteristics such as intricate combustion dynamics and external space-time uncertainties (e.g., weather, vegetation composition). Research is underway for **high-fidelity wildfire models** that accurately capture behavior and widespread impacts (e.g., air pollution). Challenges include modeling large terrains, procuring adequate data, and embedding these models in decision-making strategies.

To address these challenges, my group will engineer **scalable surrogate wildfire models** that guide data collection (monitoring) and readily embed into decision-making tools (see Figure 7). Inspired by the success of neural operators for chaotic Navier-Stokes flow systems, we plan to develop **physics-informed neural operators** to create high-fidelity models that rapidly predict wildfire behavior. With these we can conduct **uncertainty propagation** to guide data collection (monitoring). Moreover, the computational efficiency of these surrogates facilitates the integration of high-fidelity models (e.g., fire and weather simulators) into decision-making tools, promoting **better informed fire-management strategies**. For instance, we can consider how controlled burn smoke impacts surrounding population areas. We will collaborate with experts in wildfire research such as Profs. Coty Jen and Ryan Sullivan at CMU who are developing high-fidelity models on the interplay between wildfires, weather, and air quality.

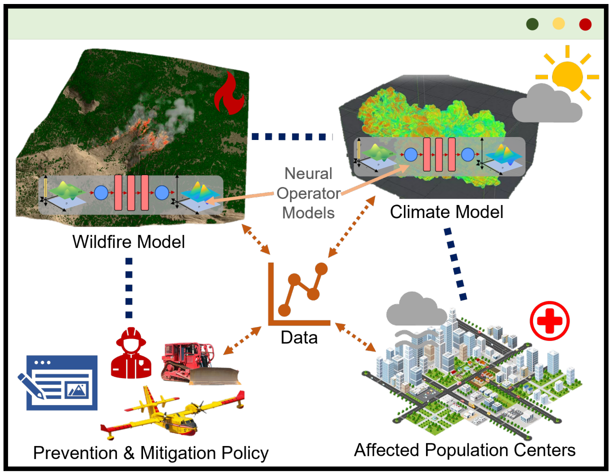

Figure 7: Proposed wildfire mitigation framework.

Potential Funding Sources: The Joint Fire Science Program (annual FOAs), EPA (Air RFAs), NIST (Fire Research Grants program), FEMA (under the Fire Prevention and Safety Grant program), SFPE Foundation (student research grant), RMRS (Fire, Fuel, and Smoke program), DOE (wildfire mitigation for power grid funding), and NSF (NSF 21-124, PD 22-1525, CAREER, & PD 22-8084).

Project 3: Advanced Computer Vision Methods for Process Control

Computer vision sensors **rapidly interpret video data** signals to inform automated decision-making/control workflows (e.g., optimizing production and detecting anomalies) which promote the **increased safety, efficiency, sustainability, and reliability** of process systems. Illustrative applications include flare stack control, conveyor belt monitoring, and temperature distribution control. Key challenges that limit industrial adoption include **validating** control architecture robustness, designing computer vision strategies with **limited data**, and **assessing sensor health** in real-time.

Therefore, my group will develop: (1) a computer vision process simulator (see Figure 8), (2) **optimization-based validation** using (1), and (3) a sensor novelty detection framework with **uncertainty quantification**. Establishing (1) will enable us to generate large synthetic datasets, test sensor configuration designs, access a wide scenario envelope in (2), and **accelerate future research**. Moreover, (2) and (3) will enable us to assess computer vision **sensor robustness** both offline and online. My previous experience in this area coupled with my expertise puts my group in a unique position to tackle (1)–(3). We will collaborate with industry (e.g., ExxonMobil, Dow Chemical) and leverage open-source simulation tools used by my existing collaborators like `Chrono` (3D environments) and `IDAES` (chemical processes).

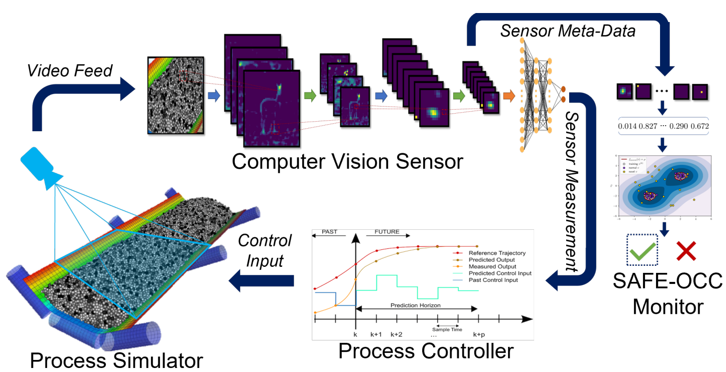

Figure 8: Computer vision aided process control simulator.

Potential Funding Sources: ExxonMobil (contact Dr. Tyler Soderstrom), Dow Chemical (contact Dr. Ranjeet Kumar), DoE (Advanced Scientific Computing Research initiative), NIH (AI, ML, & Deep Learning initiative), and NSF (NSF 20-1403, NSF 21-616, NSF 22-567, PD 18-7564, NSF 20-570, CAREER, & PD 22-8084).

Project 4: Decision-Making with Random Fields & Neural Operators

Many engineering systems (e.g., REE-CM supply chains, wildfire simulations, process systems, microbial communities, complex fluid flows, molecular dynamics) are modeled using variables that are indexed over **continuous domains** (e.g., space-time), making them infinite-dimensional. My unifying abstraction (InfiniteOpt) enables simultaneous innovation across these areas. Here, **random phenomena** (e.g., wind, porosity, random particle trajectories) is rarely accounted for due to

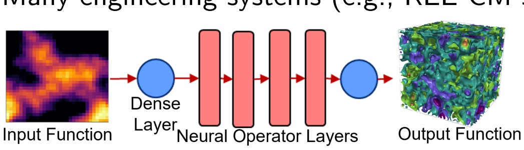

Figure 9: Neural operator surrogate.

the difficulty in representing/solving such problems. Moreover, some complex systems cannot be modeled from 1st principles and/or incur **high computational cost**. My research, which uses a unifying abstraction (InfiniteOpt) to simultaneous innovate across all these systems, demonstrates that **random field theory** (a powerful mathematical framework for characterizing random phenomena over space-time) and **neural operator surrogates** (state-of-the-art machine learning models for PDEs, see Figure 9) have great potential to address these challenges [12, 13].

Building upon my random field optimization framework [12], we will develop tractable **high-fidelity solution** approaches, potentially by adapting PDE-constrained optimization decomposition techniques (see Figure 10). To facilitate InfiniteOpt surrogates, we will also engineer **constraint representations of neural operators** that embed in InfiniteOpt formulations. Taking inspiration from physics-informed neural network based control, we will also create a **physics-informed decision-making framework that uses neural operators**. We will implement our scientific discoveries in `InfiniteOpt.jl` to promote accessibility and accelerate research in a wide breadth of applications. Potential expert collaborators include Dr. Alexander Smith (random field theory) at the University of Minnesota, Dr. Sungho Shin (PDE decomposition) at ANL, and Prof. Fani Boukouvala (data-driven optimization) at Georgia Tech, and Prof. Ruth Misener (decision-making surrogates) at Imperial College London.

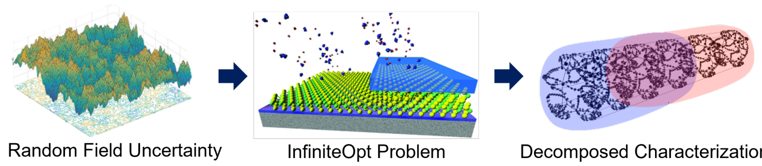

Figure 10: *Decomposition for random field optimization.*

Potential Funding Sources: NSF (PD 20-1403, PD 18-7607, PD 18-1269, PD 19-072Y, NSF 22-023, CAREER, & PD 22-8084), DoE (contact DoE lab colleagues), Alfred P. Sloan Foundation (Data & Computational Research initiative), Project 1 (random phenomena in REE-CM models), and Project 2 (climate and wildfire surrogates).

Resources

My research group will provide interested students and post-doctoral associates to work on the above and related topics in a **collaborative and inclusive environment** following the philosophy outlined in my diversity statement. To support these efforts, my group will require a moderate computer cluster with high-performance CPU and GPU hardware for rapid prototyping and establishing a collaborative hub. Finally, we will require personal computing equipment (e.g., laptops, keyboards, monitors) for each student and myself.

Existing and Potential Collaborations

My research is **highly collaborative**, giving me the privilege to work with 26+ talented researchers in diverse disciplines as detailed in my curriculum vitae. Above, I highlighted my collaborations with researchers at UW-Madison (Profs. Victor Zavala and Jim Luedtke), ANL (Drs. Mihai Anitescu and Sungho Shin), SNL (Drs. Bill Hart and Michael Bynum), PNNL (Drs. Henry Huang, David Barajas-Solano, and Jing Li), CMU (Profs. Carl Laird, Ignacio Grossmann, and David Bernal), NETL (Drs. Miguel Zamarripa Perez and Markus Drouven), and ExxonMobil (Drs. Tyler Soderstrom, Kevin Furman, and Merve Merakli) over a scope of research projects. Additionally, I am starting up research at CMU on **surrogate model decision-making** approaches with Imperial College London (Prof. Ruth Misener, Prof. Calvin Tsay, and Dr. Francesco Ceccon), SNL (Drs. Micheal Bynum, Bill Hart, and Emma Johnson), and CMU (Prof. Carl Laird).

References

[1] **J. Pulsipher** et al. "A mixed-integer conic programming formulation for computing the flexibility index under multivariate gaussian uncertainty". In: *CompChemEng* (2018).

[2] **J. Pulsipher** et al. "A computational framework for quantifying and analyzing system flexibility". In: *CompChemEng* (2019).

[3] **J. Pulsipher** et al. "A scalable stochastic programming approach for the design of flexible systems". In: *CompChemEng* (2019).

[4] **J. Pulsipher** et al. "Measuring and optimizing system reliability: a stochastic programming approach". In: *TOP* (2020).

[5] **J. Pulsipher** et al. "Characterizing event constraints with generalized disjunctive programming". In: (Under Review).

[6] J. V. Kadam et al. *Systems and methods of monitoring and controlling an industrial process.* Patent App. 17/126151, 2020.

[7] **J. Pulsipher** et al. "SAFE-OCC: A novelty detection framework for convolutional neural network sensors and its application in process control". In: *Journal of Process Control* (2022).

[8] **J. Pulsipher** et al. "A unifying modeling abstraction for infinite-dimensional optimization". In: *CompChemEng* (2022).

[9] J. Hellerstedt. *February 2021 winter storm-related deaths–texas.* Texas Health and Human Services. 2021.

[10] C. W. King et al. *The timeline and events of the feb 2021 texas electric grid blackouts.* UT-Austin Energy Institute. 2021.

[11] **J. Pulsipher** et al. "New measures for shaping trajectories in dynamic optimization". In: *IFAC PapersOnLine* (In Press).

[12] **J. Pulsipher** et al. "Random field optimization". In: *CompChemEng* (2022).

[13] **J. Pulsipher** et al. "Data-driven surrogates for infinite-dimensional optimization problems". In: *CORS/INFORMS.* 2022.

[14] *Camp fire incident.* 2021. URL: `www.fire.ca.gov/incidents/2018/11/8/camp-fire/#incident-damages-losses`.