Shahin Masoumi-Verki, PhD

Montreal, QC | +1 (438) 866 3810 | shahinmasoumi93@gmail.com | Google Scholar | ResearchGate | LinkedIn | GitHub

Summary

AI and data scientist with proven expertise in machine learning, applied mathematics, data engineering, data-driven control, dynamical systems, and IoT. Led the development and deployment of AI algorithms, reinforcement-learning-based control systems, and data models, while building robust data-engineering pipelines and driving data-governance initiatives for large-scale IoT projects. With an engineering background and strong cross-functional collaboration experience, I bridge mathematical intuition with practical, data-driven solutions that deliver measurable business and operational impact.

Education

Concordia University, Montreal, QC Sep. 2019 – Sep. 2022
PhD in Building Engineering

Iran University of Science and Technology, Tehran, Iran Sep. 2016 – Feb. 2019
MSc in Mechanical Engineering

Iran University of Science and Technology, Tehran, Iran Sep. 2012 – Aug. 2016
BSc in Mechanical Engineering

Experiences and Projects

Next Generation Cities Institute, Concordia University, Montreal, QC Jan. 2025 - present
Data Manager (Data Scientist and Data Engineer)

- Developed **advanced lightweight HVAC controller for office buildings** suitable for **edge devices**: (1) **SINDy-RL (SINDy: Sparse Identification of Nonlinear Dynamics)** hybrid controller blending **symbolic dynamics** with **Dyna-style model-based reinforcement learning**; (2) **Differentiable predictive control (DPC)** with offline policy optimization (*achieved almost **30% energy saving** through environment control*).
- Developed a **SINDy-based** model that discovers the sparse **CO_2 mass-balance equation** and **reconstructs real-time occupant** count, using CO_2 sensor measurements and desk occupancy sensors (*suitable for controlling indoor environments*).
- Developed and deployed a **scalable** and **secured IoT data pipeline** for hundreds of sensors, integrating **MQTT**, **Kafka**, **ontology-driven graph DB**, and **time-series storage**. Built preprocessing layers, **data warehouse**, **REST APIs**, and **Grafana** dashboards. Deployed the system on **Kubernetes** (*delivered in collaboration with **Desjardins** and **Schneider Electric***).
- Collaborating on **data governance** for a **smart-building** proof of concept at **Desjardins** (almost **500 IoT sensors** across two floors of **Complexe Desjardins, Montreal**), defining standards for **data quality**, **security**, **metadata**, and **lifecycle management**.
- Developed an **LLM-powered** pipeline for the Global Covenant of Mayors (**GCoM**) to extract **evidence** and **score Climate Action Plans (CAPs)** against **predefined indicators**, producing structured JSON outputs with page-level references.
- Developed **an end-to-end framework for full-field reconstruction from sparse sensor measurements,** leveraging **spatio-temporal models** (**spatio-temporal Transformers**, attention-based **LSTMs**), **sparse dimensionality reduction techniques** (e.g., **sparsity-promoting dynamic mode decomposition**). Designed to reconstruct **physical fields** (e.g., temperature, pollution, wind speed) efficiently with **minimal sensor measurements**, enabling timely and accurate estimations.
- Developed **a GPU-accelerated** tool for **Sparsity-Promoting Dynamic Mode Decomposition.**
- Developed **a scalable map-matching framework for noisy GPS data**, leveraging **PostGIS for efficient spatial data handling, segmentation, and tree-based querying**. Enhanced **kriging methods with covariance matrix reduction**, enabling user-defined strategies for **city-scale CO_2 emission estimation across Montreal's road network (~7M trips)** (*currently in use in **CityLayers' transportation mode***).
- **Lead data scientist and data engineer** in the **IoT showroom** project at the Next Generation Cities Institute to demonstrate smart building technologies in a live office environment (developed and deployed **IoT data pipeline**, **metadata** structuring using standard **ontologies**, **real-time data visualization**).
- Initiated and organized **knowledge-sharing sessions** within the team to foster collaboration and innovation.
- Designing coupled **indoor–outdoor physics-based digital twins** for **smart building** applications, aimed at enhancing occupant **comfort and sustainability**. Leveraging **differentiable programming** to integrate physical modeling with learning-based methods, enabling **real-time sensor data assimilation** and **continuous model calibration**. The system architecture will be optimized for **edge AI** inference and control.

Next Generation Cities Institute, Concordia University, Montreal, QC Jan. 2023 – Jan. 2025
Postdoctoral Fellow

- Developed various **dimensionality reduction** techniques: adversarial autoencoder (**AAE**) based on **Wasserstein** generative adversarial network (**WGAN**), variational autoencoder (**VAE**), convolutional autoencoder (**CAE**), sparsity-promoting dynamic mode decomposition (**sp-DMD**), proper orthogonal decomposition (**POD**).
- Developed **reduced-order model** (**ROM**) for **predicting** turbulent flow fields (**spatial-temporal** datasets) in urban areas, combining **AAE** and **bidirectional LSTM** (achieved **five** orders of magnitude **speed-up** on inference compared to conventional CFD simulations).
- Applied **transfer learning** (**TL**) to **ROMs** for predicting different flow field conditions to reduce the number of snapshots used for model training (**reduced** the number of the required snapshots for model training by **77%**, with **21%** of model performance improvement).
- Developed **physics-informed neural network** (**PINN**) using governing partial differential equations (**PDEs**) and **automatic differentiation** in TensorFlow to reconstruct velocity field from **sparse** and **limited** data points.
- **Reconstructed** the flow field around an isolated building from a **limited** number of data points using **convolutional neural networks** by taking advantage of **Voronoi** tessellation (captured **95%** of the variance of $60\times60\times2$ using **25** randomly distributed data points).
- Solved the **sparse sensor placement optimization** problem to find an optimal sparse representation of a 2D flow field in the wake of an isolated building using **PCA** (or **POD**) and **QR factorization** with **column pivoting** (captured **95%** of the variance of $60\times60\times2$ images using **71** sparse data points).
- Solved the **sparse sensor placement optimization** problem using **active (adaptive)** learning.
- Applied **DMD** to flow past a cylinder and airflow around an isolated building to identify the underlying physics.

- **Data assimilation** using Kalman filter to keep the digital twin model up to date (i.e., injecting data from sensors into trained models).
- Developed a **parameterized fast-response** model for predicting the flow past a cylinder using **POD** and **CAE**.
- Applied **decomposition** techniques (e.g., **Fourier transform**, **wavelet transform**, and singular value decomposition, **SVD**) on image datasets to compress them.
- Developed a **Transformer** model suitable for capturing both **spatial** and **temporal** dependencies among **sparse** data points.
- Developed a **sparse reconstruction** model, combining **sp-DMD**, various **time-series** algorithms (**bidirectional LSTM** with **temporal attention** mechanism, and **spatio-temporal Transformer**), and a **deep neural network**.
- Contributed to the **expansion** and **adaptation** of **ontologies** for integration with **urban-related** datasets, enhancing their suitability and applicability.

Concordia University, Montreal, QC

Sep. 2019 – Sep. 2022

Graduate Research Assistant (PhD Candidate)

- Developed **ROMs** for **predicting** flow fields in urban areas using various **convolutional autoencoders** (i.e., conventional, multi-scale, and self-attention-based) and parallel **LSTM**.
- Developed a **self-attention module** for image datasets, designed to capture **spatial dependencies** among data points, integrating the **physics** of turbulent flow fields into a fully data-driven model.
- Performed **POD** and **spectral analysis** on urban turbulent flow fields to understand the underlying physics of the problem.
- Conducted CFD simulations using embedded large eddy simulation (ELES) to reduce computational costs.

Skills

Programming & Computing
- **Languages:** Python, MATLAB
- **Big Data & Distributed Computing:** PySpark, Apache Kafka, MQTT

Machine Learning & Data Science
- **Frameworks & Platforms:** TensorFlow, Keras, PyTorch, AWS SageMaker, Scikit-learn, Pandas
- **Techniques:** Time-series analysis, Signal processing, Spatio-temporal modeling, Digital twin modeling, Large language models (LLM), Agentic AI, Retrieval-Augmented Generation (RAG)
- **Scientific / Hybrid ML:** Differentiable programming (PyTorch autograd), Physics-informed neural networks (PINN), Sparse sensing, Kalman data assimilation, Model-based reinforcement learning

Data Engineering & Databases
- **Databases:** PostgreSQL, TimescaleDB, PostGIS, Neo4j
- **Query Languages:** SQL, Cypher
- **Pipeline & Orchestration:** Apache Airflow, Docker, Kubernetes, AWS EC2
- **APIs & Backend Development:** FastAPI

Data Visualization & Monitoring
- **Visualization Tools:** Matplotlib, Seaborn, Tecplot
- **Real-time Monitoring & Dashboards:** Grafana

Mathematics & Theoretical Foundations
- **Mathematics & Modeling:** Linear Algebra, Probability & Statistics, Dynamical systems, Control theory
- **Physics & Engineering:** Fluid Dynamics, Turbulence, Convective Heat Transfer

Miscellaneous & Tools
- **Version Control & Scripting:** Git, GitHub, Bash Scripting
- **Productivity & Analytics:** MS Excel