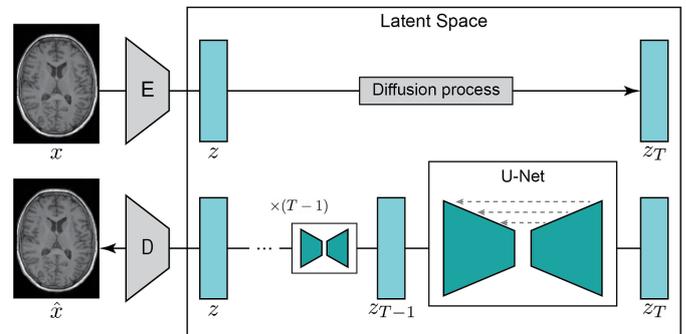


# Diffusion Generative Models and Gaussian Processes

## Introduction

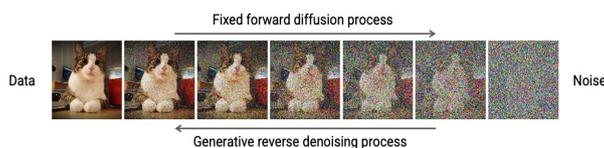
Diffusion Generative models are a class of generative model which aim to learn the backward denoising process for a predetermined forward Markov-chain data noising process. The forward process is designed such that the limit distribution as the noising steps approach infinity is indistinguishable from an isotropic Gaussian (i.e. our data becomes random noise). A neural network then learns the parameters of series of Gaussian distributions, representing the backwards denoising steps given the noisy data.



Latent diffusion model on medical imaging data.  
Image taken from TowardsDataScience.com

## Why Diffusion?

Learning to generate samples from the data distribution given a generic noise distribution allows for diverse data generation. By drawing random noise, the learned backward process has the ability to generate a data point from the data distribution that the model had never seen in the first place.



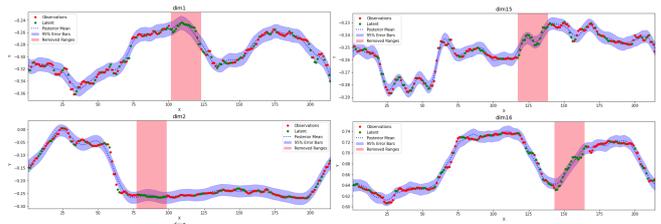
Fixed forward process and the learned generative process.  
Image from CVPR 2022.

## Classifier-free Guided Diffusion

Sampling from an LDM involves generating a latent  $T$  steps along the forward process, and performing  $T-1$  reverse steps to recover a latent from the data distribution. This latent is then decoded into data-space. However, guided diffusion involves feeding conditioning information into the denoising process at each reverse step. This can be generated from a partner model such as a classifier, or indeed an unsupervised encoder.

## Gaussian Process Guided Diffusion

We propose a Sparse Gaussian Process Variational autoencoder to generate the conditioning information for a video. Each latent noise sample is then guided by conditioning information for each frame of this video to produce a contiguous set of data-samples which could be drawn from the same video.



Latent trends for the first four dimensions of a 20 dimension latent vector for frames of a video. A Gaussian Process is fit to the data.

## Model Architecture

Our model uses the popular U-Net neural network architecture, with a top layer encoder-decoder pair to form a Latent Diffusion model (Rombach et al. 2022). We diffuse on latent representations of images, which are fed forward through multiple layers of residual connections and convolutions. We choose a small enough bottle-neck dimension such that we can incorporate **condition** information for guided sampling.

# Physically-Aware Autonomous Cyber Defence

Adam Neal

Supervised by Simon Maskell, Peter Green, Nick Hare, Ian Miles

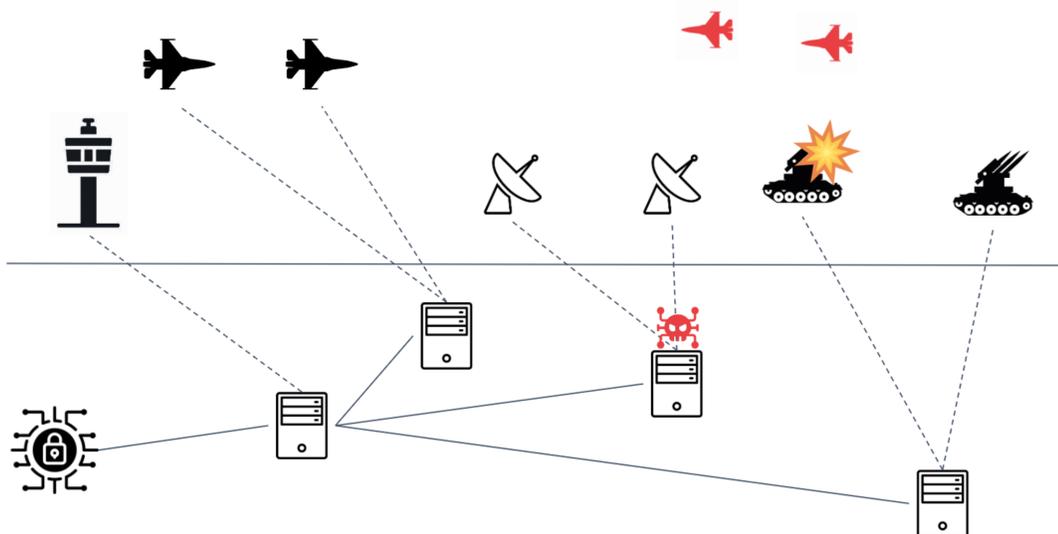


## 1. Introduction

Current state-of-the-art cyber defence strategies often deploy a network-centric approach, meaning they only consider service availability when making decisions in retaliation to a cyber attack.

As these approaches are only limited to information about the network it's protecting, it doesn't have the situational awareness to know how its actions affect us in the real-world.

By giving a system real-world context, like information about an enemy interceptor's location, we can better understand about what effect protecting the network will have on the real-world environment it sits in.

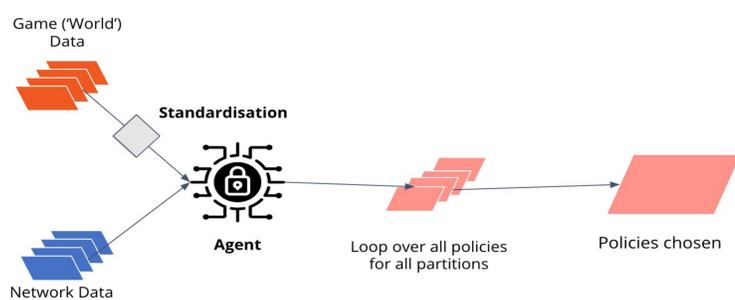


## 2. Project MIDGARD

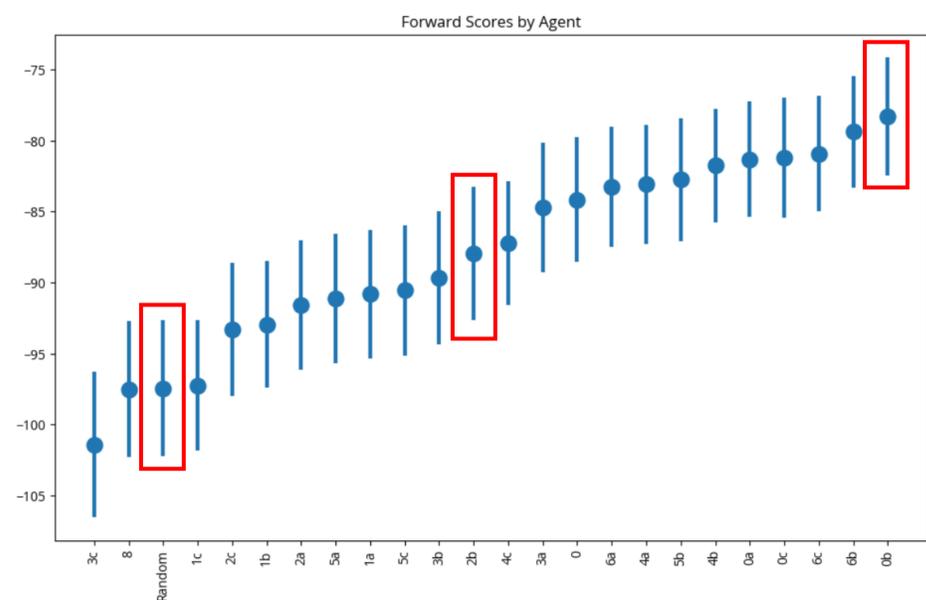
Project MIDGARD allows for real-world reasoning by implementing both a network simulation and real-world simulation of an air defence system.

A Gaussian process approach is used to learn about what actions are best to take to mitigate the cyber attack given the state of both the real-world and network simulation it's defending.

A Gaussian process is a probability distribution over possible functions that fit to a set of points. They provide a non-parametric way to infer a function from a set of points, capturing both the functions expected value and uncertainty at each point.



## 3. Results



Random: untrained random agent  
2b: trained on only network data  
0b: trained on network and real-world data

Utilising both the network and real-world data allows for a significant performance improvement over just a network-centric approach. An even bigger improvement can be seen when considering the GPs uncertainty measurement.

## 4. Future Work

- Scale up to a higher fidelity network simulation.
- Implement a Bayesian Reinforcement Learning approach to improve decision making.
- Find the balance between a real-world approach and a network-centric approach.

# Scheduling of Distributed Information Processing



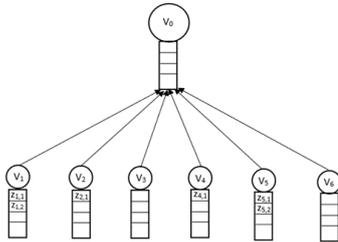
Alex Bird, Supervised by Prudence Wong, Jelena Milisavljevic, Alasdair Hunter

EPSRC Centre for Doctoral Training in Distributed Algorithms, University of Liverpool, Liverpool, UK

## Abstract

**Problem:** We need to be able to reason about what we might calculate elsewhere within a network both in present and in the future, without actually performing the calculation itself.

We consider a network of sensors connected to a central server. A scheduler needs to decide how to distribute the processing of sensed data to maintain various performance metric.



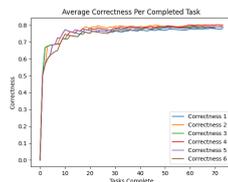
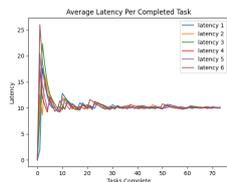
## Model

In order to investigate my research problem I am constructing a model and corresponding simulation to test different optimisation strategies within different network settings.

The current model consists of a set of sensor nodes and a central server. The sensor nodes randomly receive input which they must then choose to either compute locally or send to the central server for processing. The central server can compute the tasks with a higher degree of accuracy however this must be balanced with the speed and efficiency at which the tasks are completed. The overall goal of the model is to maximise the accuracy of the computed information whilst minimising the time taken to compute tasks.

## Simulation

In conjunction with this model we have created a Python based simulation to test out different optimisation techniques / parameters. The simulation utilises four different decision functions, based on two optimisation variables, to decided whether a task should be computed locally at a sensor or at the central server. We chose the Python programming language to create the simulation as it offers comprehensive libraries for a simple simulation implementation as well as extensive libraries for the future extensions such as statistical emulation and reinforcement learning.



## ILP

To investigate the model further we have been working to phrase it as an integer linear programming (ILP) problem. This allows us to demonstrate how we are optimising for our given objective function based on the input data. A generalised form of our ILP is as follows:

$$\begin{aligned}
 &J_i = \text{job at sensor } i && \text{Assumption: } s_0 > s_i \\
 &p_i = \text{processing time of job } J_i \\
 &s_i = \text{scaling factor for sensor } i \\
 &s_0 = \text{scaling factor for central server} \\
 &p_i * s_i = \text{processing time at sensor } i \\
 &p_i * s_0 = \text{processing time at server} \\
 &X_i * p_i * s_i + (1 - X_i) * p_i * s_0 = \text{Actual processing time} \\
 &\text{Completion time at sensor } i = X_i * p_i * s_i \\
 &\text{Completion time at server} = \sum_{i=1}^n (1 - X_i) * p_i * s_0 \\
 &C_{max} = \max \left\{ \sum_{i=1}^n (1 - X_i) * p_i * s_0, \sum_{i=1}^n X_i * p_i * s_i \right\} \\
 &acc_0 = \text{accuracy if job is completed at server} \\
 &acc = \text{accuracy if job is completed at sensor} \\
 &\text{Assumption: } acc_0 >> acc \\
 &A = \sum_{i=1}^n X_i * acc + \sum_{i=1}^n (1 - X_i) * acc_0 = \sum_{i=1}^n X_i * acc + (1 - X_i) * acc_0 \\
 &\max : \frac{1}{C_{max}} + A
 \end{aligned}$$

## Future Work

### Future Work

#### Refining the Model

The model and simulation are now capable of recording and optimising on multiple different optimisation criteria including energy consumption and task priority. The goal now is to consolidate the different versions and capabilities of the model to fully investigate my research problem.

#### Data Arrival

The first aspect we are considering is the data arrival times. The current model works with pre-determined arrival times, allowing us to account for factors such as sensor downtime. However, we want to be able to work with a stream of unknown incoming tasks, in a similar way to how an actual sensor network may perform.

#### Energy Consumption

Energy consumption currently plays a role in the model in dictating sensor uptime due to power usage and battery size. We are also investigating data transmission costs and different sensor types.

#### Priority Tasks

To further differentiate tasks, we have experimented with a task priority system. This allows us to give weight to urgent incoming tasks which may need to have a much shorter latency value.

#### Heterogenous Sensor Behaviour

The final expansion to the model we have discussed is the differentiating behaviours of sensors to incoming tasks.

## References

1. P. Tagade and H. Choi, "An emulator-based rapid source localization approach in informative sensor planning," *2012 American Control Conference (ACC)*, 2012, pp. 691-696
2. Pfeiffer, András & Kádár, Botond & Monostori, Laszlo. (2003). Evaluating and improving production control systems by using emulation.
3. J. Perazzone, M. Dwyer, K. Chan, C. Anderson and S. Brown, "Enabling Machine Learning on Resource-constrained Tactical Networks," MILCOM 2022 - 2022 IEEE Military Communications Conference (MILCOM), Rockville, MD, USA, 2022, pp. 932-937

# Deep Learning With Sequential Monte Carlo Samplers

Andrew Millard (a.mill212@liverpool.ac.uk),  
Prof. Simon Maskell (smaskell@liverpool.ac.uk),  
Dr. Simon Goodchild (simon.goodchild@stfc.ac.uk)



## GD Preconditioned SMC for Complex Probability Distributions

We propose an algorithm inspired by Pathfinder [1] which uses both gradient descent and SMC Samplers in conjunction with one another in order to be able to more effectively sample from complex probability distributions. We demonstrate this ability on a nine mode GMM model and a two-dimensional Rosenbrock PDF with the samples initialise by prior which is centred far from the target mass. In an upcoming paper, we demonstrate the effectiveness of this preconditioning compared to vanilla SMC Samplers and MCMC methods.

METHOD $N = 100$	AVERAGE FOUND MODES	$\sigma$
SMC NUTS	3.60	$\pm 0.97$
SMC NUTS w/SGD	4.55	$\pm 1.40$
METHOD $N = 1000$	AVERAGE FOUND MODES	$\sigma$
SMC NUTS	6.30	$\pm 0.95$
SMC NUTS w/SGD	8.60	$\pm 0.58$
METHOD $N = 4000$	AVERAGE FOUND MODES	$\sigma$
SMC NUTS	8.1	$\pm 0.77$
SMC NUTS w/SGD	9.0	$\pm 0.00$

Figure 1: Table of SGD SMC GMM Results

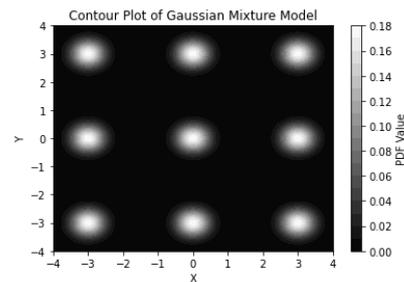


Figure 2: Contour of the Target GMM Distribution

## Proposing the parameters for a Bayesian Neural Network using SMC Samplers

We are currently using an SMC Sampler to propose the parameters for a Bayesian Neural Network. MCMC methods have been shown to converge to the true parameters of the distributions they navigate [2], while Variational Inference (VI) [3] gives Gaussian approximations in a quicker time than MCMC. We predict that SMC Samplers should be able to give better parameter estimates than VI while also being faster than MCMC. We have already demonstrated our ability to compete with SGD methods on raw accuracy and error rates on simple datasets and are now working on showing this ability on benchmark deep learning datasets such as MNIST and CIFAR-10.

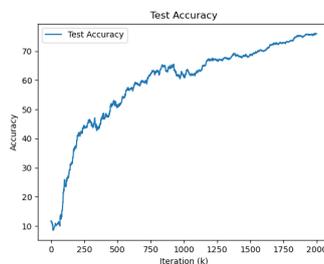


Figure 3: Test Accuracy from BNN trained by SMC Sampler on MNIST dataset

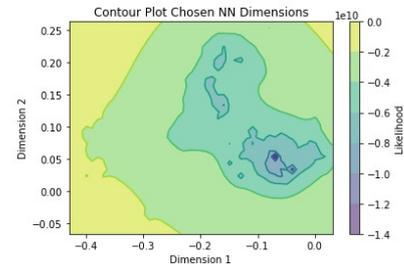


Figure 4: Estimated Contour from the samples generated in the SMC training process

## Combining SGD and SMC for Bayesian Neural Networks

This work follows on from the GD Preconditioned SMC Sampler. If we can use GD to find us a good initial starting point, we can speed up the convergence of the sampler in high dimensional and complex distributions, such as those experienced when navigating neural networks. So far, we have shown that using SMC Samplers, we can hold the accuracy that the SGD methods give us on complex datasets such as MNIST. The next stage is to show that we can successfully combine the two approaches to be successful on datasets and problems which require an uncertainty quantification.

## References

- [1] - Zhang, L., Carpenter, B., Gelman, A., & Vehtari, A. (2022). Pathfinder: Parallel quasi-Newton variational inference. *arXiv preprint arXiv:2108.03782*.
- [2] - Pavel Izmailov, Sharad Vikram, Matthew D Hoffman, and Andrew Gordon Wilson. What Are Bayesian Neural Network Posteriors Really Like? In ICML, 2021.
- [3] - Matthew D Hoffman, David M Blei, Chong Wang, and John Paisley. Stochastic variational inference. *The Journal of Machine Learning Research*, 14(1):1303–1347, 2013.



# Scheduling Surveillance of Space Objects

Benedict Oakes, Jason Ralph (UoL), Dominic Richards (STFC Hartree), Jordi Barr (Dstl)

sgboakes@liverpool.ac.uk, jfralph@liverpool.ac.uk  
dominic.richards@hartree.stfc.ac.uk  
jmbarr@dstl.gov.uk



## Abstract

As the number of satellites in space continues to increase, it becomes harder to keep track of them with limited ground-based sensing capability. This project aims to use reinforcement learning to improve non-myopic (long-term) space situational awareness by efficient scheduling of sensors.

## Background

- The UK is interested in becoming more self-reliant in the space domain, including space situational awareness (SSA) [1].
- Satellites have become essential for how we live our lives, used for banking, communication, GPS, and in many more areas.
- Low Earth orbit space is becoming increasingly congested [2], increasing the need for reliable tracking.
- The risk of collisions between satellites and producing debris becomes more likely, and needs to be mitigated against.
- Reinforcement Learning (RL) offers an opportunity to model high-dimensional problems easily.

## Reinforcement Learning

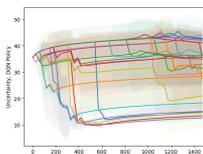
- RL: Intelligent agents making decisions to maximise a cumulative reward [3].
- We have implemented a range of RL algorithms on an environment of a ground-based telescope observing satellites.
- The implementations include Double Deep Q-Networks (DDQN), Soft Actor-Critic (SAC), and Proximal Policy Optimization (PPO).
- To cover a range of different scenarios, we have developed environments with different use cases.
- Some environments are designed for an agent to choose satellites from a given set, and others are designed for agents to find satellites in the field of regard in a search scenario.

## Fusion Paper

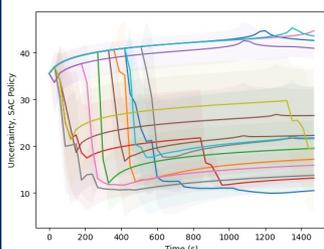


- We use a Double Deep Q Network (DDQN) on simulated satellites.
- Q-learning is a value iteration update on a Markov decision process. Q-values determine the reward gained by taking an action in a certain state.

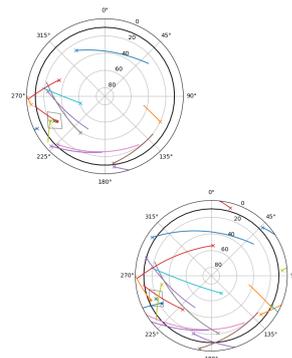
- Compared to a random baseline policy, we see drastically improved performance.
- We successfully track over half of the visible satellites with an extended Kalman filter and reduce uncertainties [4]



## RL Algorithms



- As well as the DDQN implementation in the fusion paper, we have begun generating results for the SAC and PPO algorithms.
- For the continuous action space agents (like SAC), we apply them to the environments in which a telescope must search the field of regard for satellites.
- For the discrete action space agents (like DDQN), we apply them to the environments in which a telescope must decide which target is the best to look at from a selection of possible targets.
- SAC has been successful in finding and measuring satellites, and reducing uncertainty
- PPO is beginning to find satellites (more results to follow)



## Future Work

- Producing meaningful, comparable results for different RL algorithms including SAC and PPO.
- Working on real satellite data in the form of Two-Line Element (TLE) sets, which are used with the Simplified General Perturbation (SGP) propagator. This model includes effects like solar radiation pressure, atmospheric drag, and more advanced gravity models.
- Developing environments to showcase significant non-myopic behaviour.

## References

- [1]: Ministry of Defence, National Security Strategy and Strategic Defence and Security Review, 2015, url: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/555607/2015\\_Strategic\\_Defence\\_and\\_Security\\_Review.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/555607/2015_Strategic_Defence_and_Security_Review.pdf)
- [2]: Carmen Pardini and Luciano Anselmo. "Evaluating the impact of space activities in low earth orbit". In: Acta Astronautica Volume 184 (2021), pp. 11–22.
- [3]: R.S. Sutton, A.G. Barto, "Reinforcement learning: An introduction", Malaysia: MIT press, 2018
- [4]: B. Oakes, D. Richards, J. Barr and J. Ralph, "Double Deep Q Networks for Sensor Management in Space Situational Awareness," 2022 25th International Conference on Information Fusion (FUSION), Linköping, Sweden, 2022, pp. 1-6

# Transfer Learning in Airborne Imagery



Benjamin Rise (benjamin.rise@liverpool.ac.uk), Murat Uney (M.Uney@liverpool.ac.uk), Xiaowei Huang (xiaowei.huang@liverpool.ac.uk)

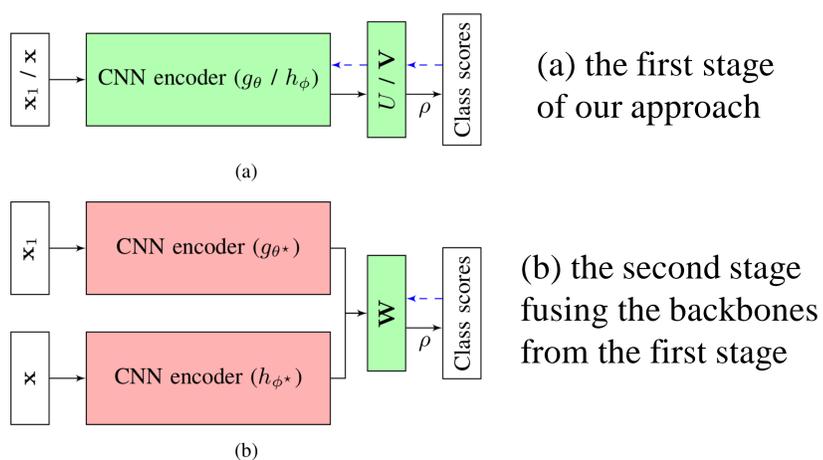
## Transfer Learning

In transfer learning we seek to use information from one or more source tasks, to improve performance or convergence speed in a target task [1]. This is typically achieved through fine tuning. Fine-tuning involves adjusting the weights of an already trained model on a new dataset, often resulting in more accurate predictions.

By utilising this information, we can potentially allow models to achieve good performance even in cases where we may have limited data and specialised tasks. This is an especially large problem in the domain of airborne imagery.

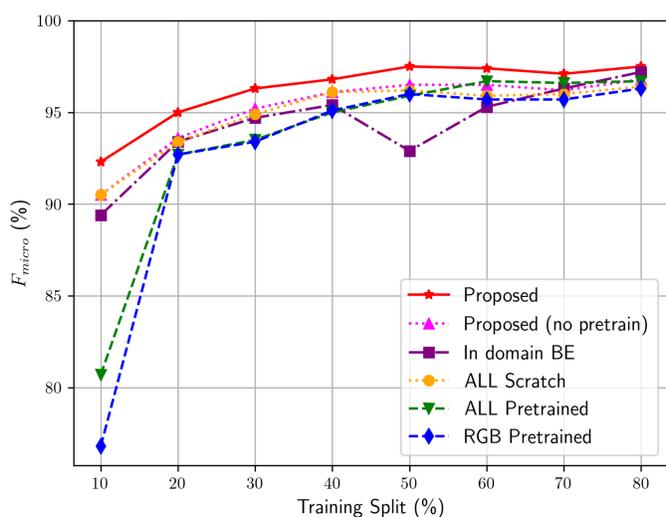
Much of our research so far has been focused on this problem in the context of hyperspectral airborne imagery. The rest of this poster outlines this.

## Schematic diagram of proposed methodology



## Results

Using our approach, we achieved enhanced performance on both evaluated datasets. Below are the results from the EuroSat dataset. The solid red line represents our proposed method, contrasted against various baselines.



## Datasets



(a) BigEarthNet



(b) EuroSAT

BigEarthNet [2] and EuroSAT [3] datasets, both sourced from the Sentinel-2 satellite. They are both *hyperspectral* datasets, with the RGB band versions shown to the left. The wavelengths include: RGB bands; bands capturing aerosols, water vapour, and cirrus detection; Near-Infrared bands; and Short-Wave Infrared bands.

## Methodology – Transfer learning in airborne hyperspectral imagery

Transfer learning has predominantly been explored in airborne hyperspectral imagery. We consider a scenario with access to pretrained sources for certain wavelengths like RGB but lack data for others, such as infrared. Our method involves a two-step approach:

Firstly, separate CNN models are trained on the RGB bands, initialized with ImageNet pretraining and fine-tuned, and another model on all bands without pretraining.

In the second phase, features from each model's final convolutional layer are merged. A new fully connected layer is introduced to integrate these features. All existing layers remain frozen, with only the newly introduced layer trained until convergence.

## Future work

- Additional Datasets: While the SEN12MS dataset is another Sentinel-2 option, testing other airborne hyperspectral datasets would be valuable.
- Multi-stage Coordination: Implementing more than two stages in our approach could lead to a further optimized model.
- Adversarial Training: Recent advancements in adversarial training suggest its potential in transfer learning, especially for enhancing model generalization in complex settings.

## References

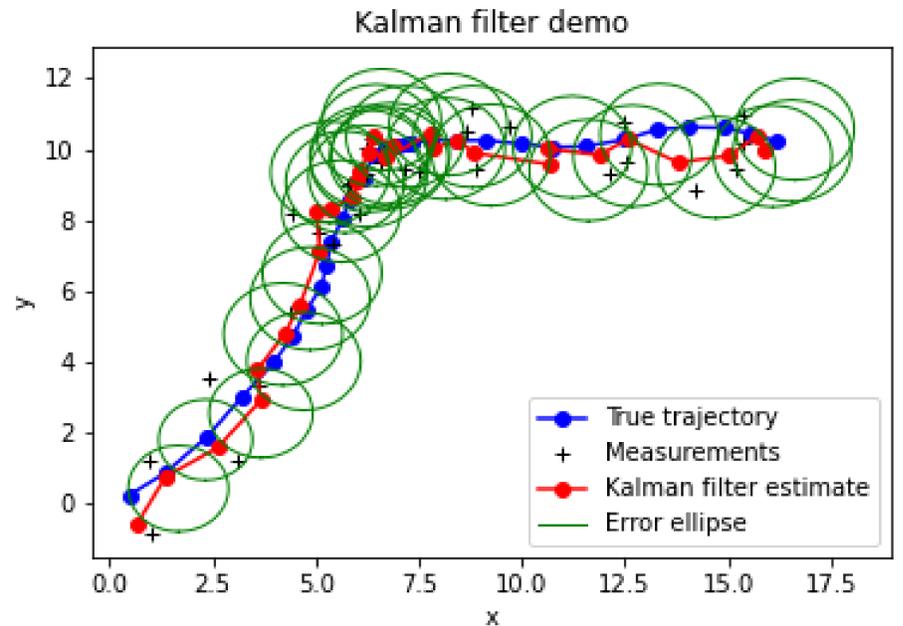
- [1] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," CoRR, vol. abs/1911.02685, 2019. [Online]. Available: <http://arxiv.org/abs/1911.02685>
- [2] G. Sumbul, M. Charfuehan, and V. Markl, "Bigearthnet: A large-scale benchmark archive for remote sensing image understanding," in IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, 2019, pp. 5901–5904.
- [3] P. Helber, B. Bischke, A. Dengel, and D. Borth, "Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 12, no. 7, pp. 2217–2226, 2019



# Learning transparent models from Data-driven Algorithms to Enhance Streaming Data Analysis

## Introduction

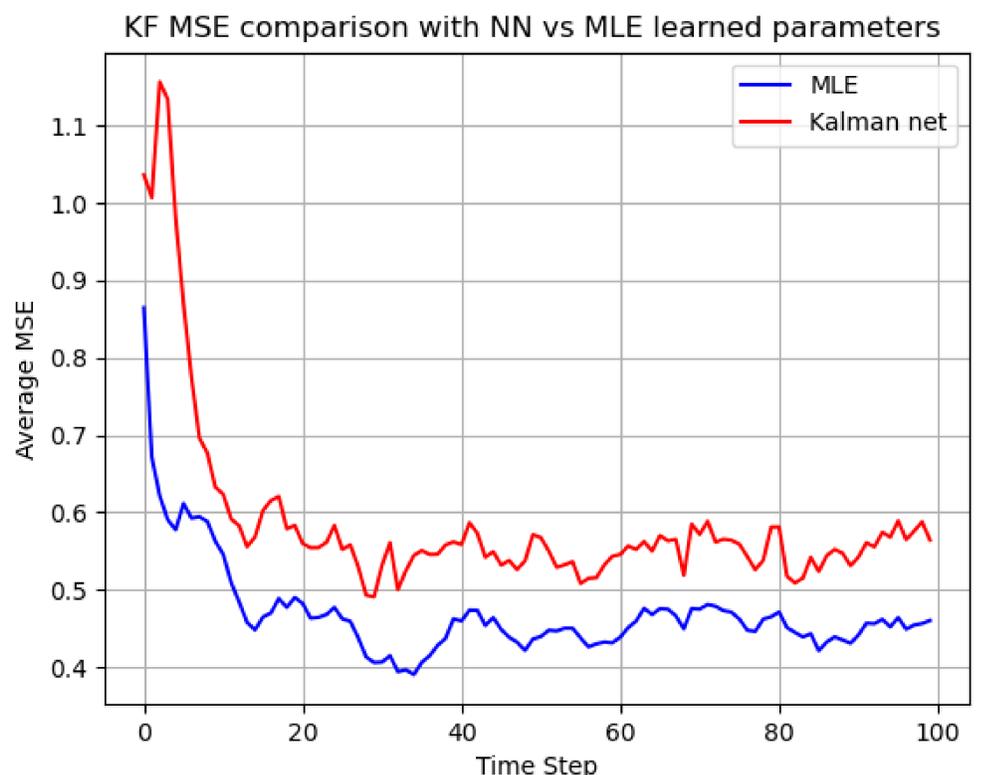
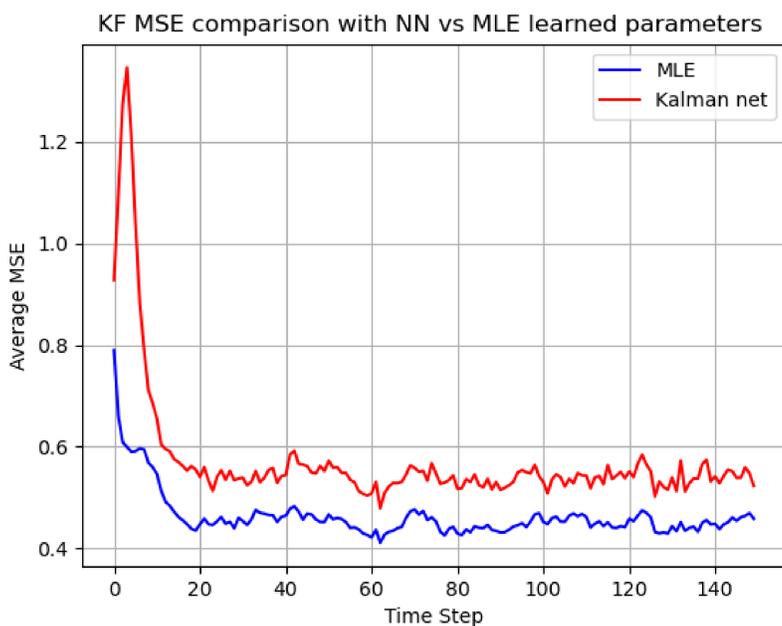
This project is focussed on the real time state estimation of dynamical systems with the aid of machine learning algorithms. Model based algorithms such as the Kalman Filter use fully transparent statistical models to address the state estimation problem. These are limited in that they rely on the linearity and accurate prior knowledge of the underlying state space model, which is not always available in practise. Furthermore, manual tweaking of the model parameters is required. Neural network algorithms can be used to overcome this, as they exploit previous data to train algorithms in order to be able to process future data, meaning that they can successfully operate even when the underlying SS model is unknown or only partially known. As a result, it is clear that a method to fuse these two approaches to create a particle filter model which works in conjunction with the algorithms to train neural networks is desirable, thus enhancing the Bayesian filter in the context of learning transparent models, which can be applied to fast streaming data.



Number of Steps	Noise parameter multiplier	MSE of MLE method	MSE of KalmanNet method
50	0.05	0.490	0.911
50	0.10	0.541	0.642
50	0.50	0.711	0.746
50	5.00	0.938	9.079
150	0.05	0.463	0.573

## Supervised learning in SS models

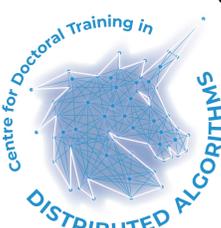
There are various ways to approach parameter estimation in Bayesian filtering, where the noise parameters are unknown. In the case of supervised learning, explored here, access to the ground truth and measurements is available. A comparison was run between a maximum likelihood method and a neural network based method of parameter estimation, in order to determine the efficiency of each method and directly compare their performances. The first method involved using maximum likelihood estimator (MLE) equations to determine values for the noise parameters Q and R, which represent the noise on the process model and measurement model respectively. The second method involved using a neural network (coined the KalmanNet) in order to learn a value for the Kalman gain, and then use this to determine the Kalman filter estimates. These methods were compared in terms of RMSE as a function of timestep, as well as comparing the computational time of each method to determine their efficiency. To successfully do this, a simulation of the ground truth was generated, and the measurements were then simulated by adding random noise to this. Both models were trained on the same training data and tested on the same Monte Carlo test sequences.



## Results and Conclusion

The results show clearly that the Maximum likelihood estimation method outperformed the Kalman net method. Throughout every timestep, the RMSE of the Kalman filter estimate compared to the ground truth was consistently lower for the MLE method. Whilst both methods follow a similar trend of having a larger MSE at lower timesteps, the MSE of the MLE method remained consistently lower throughout. Furthermore, the computational time taken to determine the Kalman filter estimate for the MLE method was consistently more than 10 times faster that of the KalmanNet. This is due to the long duration required to train the neural networks required to learn the Kalman Gain. In future research, these methods will be compared in the case where the process model is non-linear, as this may highlight potential limitations in the MLE method, meaning that there may be some situations in which the KalmanNet may be the more suitable method.

## Acknowledgements:



Bettina Hanlon  
sgbhanlo@liverpool.ac.uk  
Angel Garcia-Fernandez  
angel.garcia-fernandez@liverpool.ac.uk

# Using Machine Learning to train a Digital Test Pilot for missions in turbulent environments



Carole Liao (Carole.Liao@liverpool.ac.uk), Pr. Mark White (mdw@liverpool.ac.uk),  
Dr. Jony Castagna (jony.castagna@stfc.ac.uk), Dr. Sylvain Manso (sylvain.manso@novasystems.com)

## Problem

The safety limits of a helicopter operating to a ship depend on the aircraft's response to aerodynamic disturbances, the pilot's capabilities to counteract them, and the operating conditions.

They are traditionally defined on a 'test and declare' basis during at-sea flight trials when pilots fly repeated launch and recovery missions to a ship in many environmental conditions to provide as large as possible operational clearance envelope.

Due to the unpredictable nature of at-sea test conditions, determination of the Ship Helicopter Operational Limits (SHOL) can result in the definition of an incomplete and restricted operational envelope, achieved at very high financial cost with significant risk to the crew.

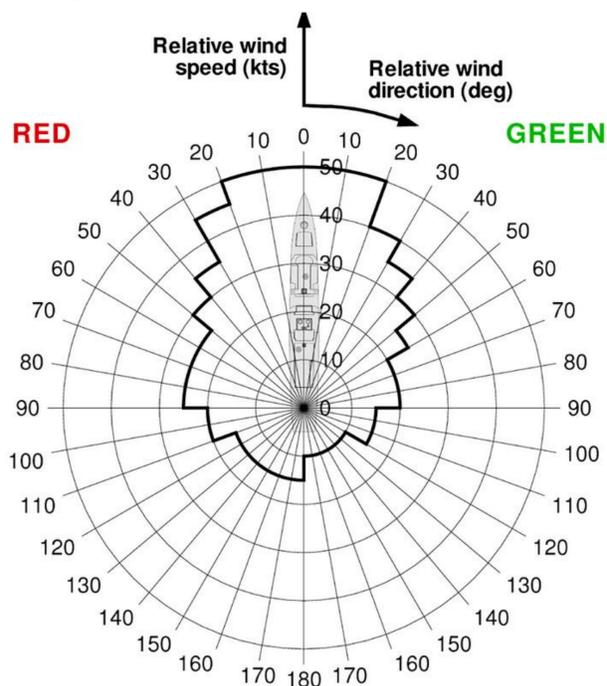


Figure 1: Ship Helicopter Operational limits (SHOL) [1]

## Objectives

Modelling and simulation can reduce cost and risk involved in real-world flight testing by examining a wide range of operational conditions.

The objective of this project is the development of a helicopter pilot model to support SHOL determination process.

From an early stage, the pilot model could conduct multiple virtual deck landings in a desktop simulation environment. The likely boundaries of a helicopter safe operational envelope could then be established. They would inform about the conditions where testing at sea should be conducted towards the edge of the operational envelope, improving the SHOL determination process efficiency and expanding operational clearance.

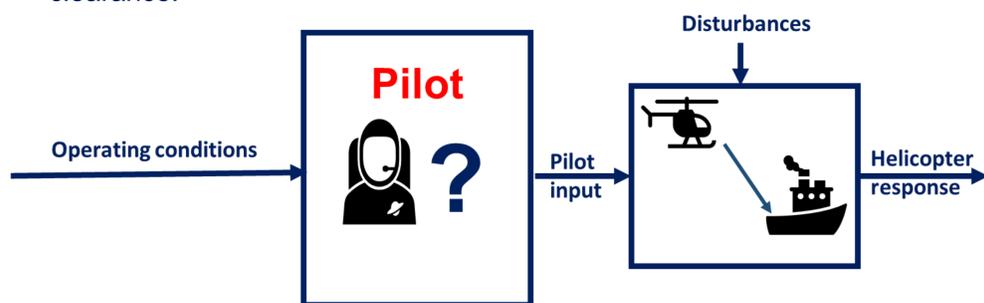


Figure 2: Helicopter pilot model

## Method

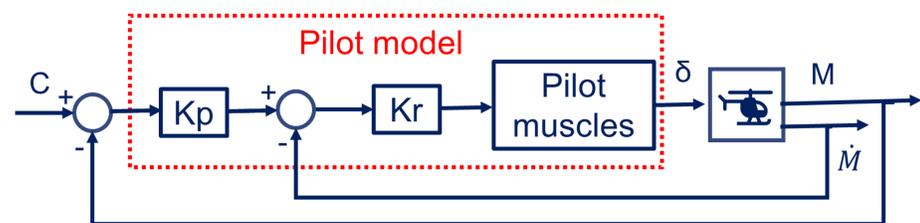


Figure 3 Helicopter pilot model cf. R. Hess [2]

The mathematical model of a human test pilot, the digital test pilot, is based on the control theoretic approach by R. Hess. It provides a framework for pilot modelling in pursuit tasks and is independent from the piloted vehicle dynamics.

For a single axis task, the digital test pilot consists in the tuning of gains  $K_p$  and  $K_r$ . The principle of this method can be extended to multi-axes tasks coherent to the demanding environment of deck landing operations.

The digital test pilot model will be trained using machine learning with data gathered during piloted simulator trials and real-world sea trials where possible.

## Data acquisition



Figure 4: Heliflight-R simulator [3]

Data from simulator trials is acquired via UoL Heliflight-R facility.

Test pilots are landing helicopter models of different handling quality levels in the simulator environment. The pilot inputs and helicopter model responses would be used to tune the digital test pilot model.

## References

- [1] Memon W., Development of High Fidelity Modelling and Simulation for the Helicopter Ship Dynamic Interface.
- [2] Hess R., Simplified Approach for Modelling Pilot Pursuit Control Behaviour in Multi-Loop Flight Control Tasks.
- [3] Hodge S., Dynamic Interface Modelling and Simulation Fidelity Criteria

# Data Driven Intelligence for Countering Crime



Efthymou Drousiotis (E.Drousiotis@liverpool.ac.uk),  
 Simon Maskell (SMaskell@liverpool.ac.uk), Paul  
 Spirakis (spirakis@liverpool.ac.uk)

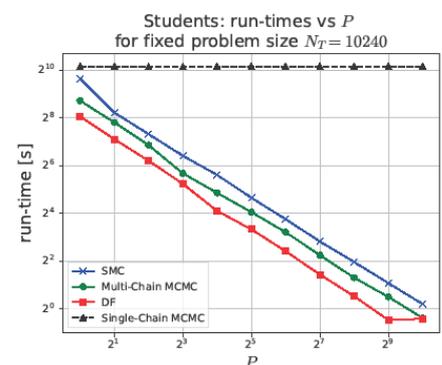
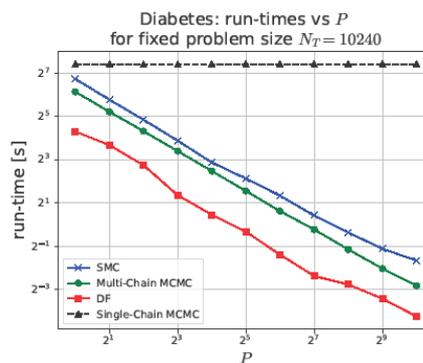
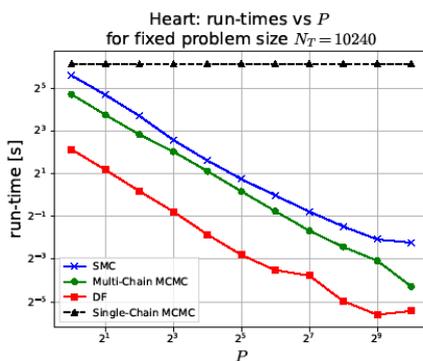
## Background

The main challenge when it comes to classification problems, is to find solutions that are both accurate and scalable. Decision Forest (DF) employs a collection of embarrassingly parallel Decision Trees (DTs), and is one of the most commonly used approaches when it comes to classification. Bayesian alternatives for DTs based on Markov Chain Monte Carlo (MCMC) methods have been shown to be more accurate but slower than DF. Also, the inherently sequential nature of MCMC makes it unsuitable to be performed in parallel unless the accuracy is sacrificed. This is particularly evident in Distributed Memory (DM) architectures, which normally provide access to large numbers of processors. Sequential Monte Carlo (SMC) samplers are a parallel alternative to MCMC, which do not trade off accuracy for parallelism. However, the performance of SMC samplers in the context of DTs is underexplored, and the parallelization is tricky due to the challenges in parallelizing its bottleneck, redistribution, especially on variable-size data types such as DTs

## Aims

- Construct the first SMC sampler for DTs
  - Operate in a Distributed environment
- Achieves asymptotically optimal  $O(\log N)$  time complexity
- Achieve same accuracy as MCMC DTs and greater accuracy compared to DF

## Results



Method	$P$	$N_T$	$N$	$K$	Hearts	Diabetes	Students
SMC	Any	-	1024	10	77.44%	73.27%	71.48%
Single-Chain MCMC	1	10240	-	-	77.01%	73.78%	71.64%
Multi-Chain MCMC	Any	-	1024	10	67.42%	66.76%	54.62%
DF	Any	10240	-	-	76.3%	71.0%	66.86%

## Future work

- Use a better proposal, such as HINTS to increase both accuracy and convergence time.
- Improve the prior which is going to improve the convergence time.
- Improve the backward kernel.
- Exploite alternative parallel computing solutions by mixing Distributed Memory programming models, such as MPI (which we have used here), with SMP models for CPUs or GPUs, such as CUDA or PyTorch
- Investigate better (either on shared or distributed memory) load-balancing strategies for redistribution, which is the run-time bottleneck in SMC.



# NON-MYOPIC APPROACHES TO SENSING AND SURVEYING

George Jones, Supervised by Ángel Garcia-Fernandez, Prudence Wong & Scott Page  
EPSRC Centre for Doctoral Training in Distributed Algorithms, University of Liverpool, Liverpool, UK

## 1. INTRODUCTION

Sensor management algorithms typically use Bayesian information theoretic approaches to evaluate the **value of different sensor combinations** and **platform actions**.

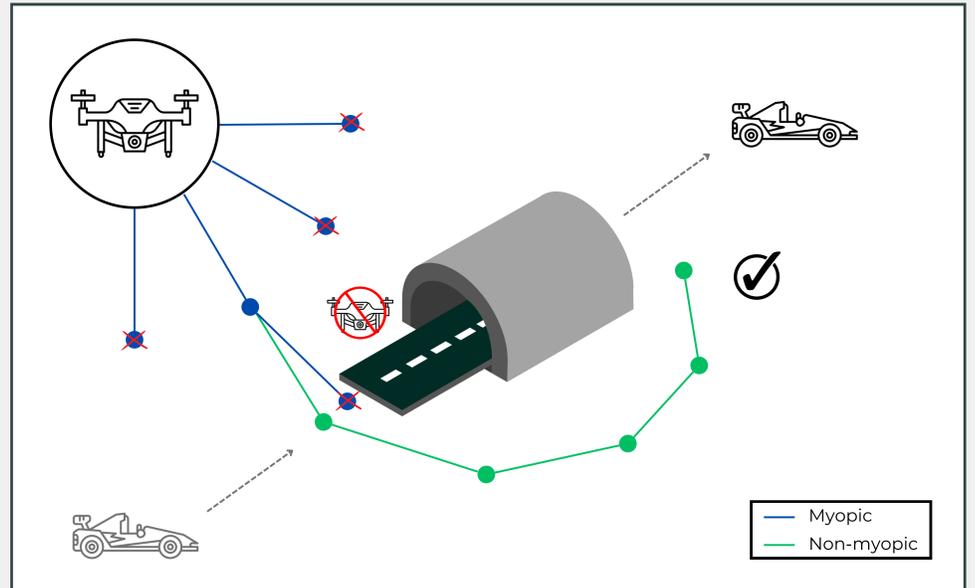
As we plan further into the future, the problem begins to suffer from **combinatorial explosion**.

Optimising over **multiple time-steps** (non-myopic) allows for **higher long-term gain** and enables mitigation of real-world challenges, such as obstacles to gathering information.

## 2. OBJECTIVES

The aim of the research project will be met by considering:

- Myopic algorithms for **Gaussian Bernoulli** ✓
- Non-myopic algorithms for Gaussian Bernoulli ✓
- Limitations of information theoretic approaches ✓
- Myopic algorithms for **Gaussian Multi-Bernoulli**
- Non-myopic algorithms for Gaussian Multi-Bernoulli



## 3. CURRENT SCENARIO

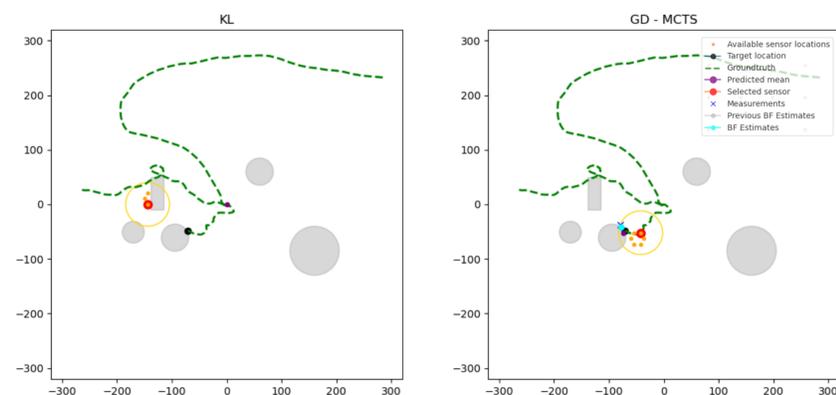
**Scenario** - only one object alive at any given time.

The ground-truth is a randomly generated path, born at the origin based on a transition matrix ( $F$ ) & some process noise ( $Q$ ).

Measurements are subject to localisation and cardinality errors. Meaning there are detections and clutter.

Obstacles in the surveillance region block sensor movement but not measurements or target movement.

Non-myopic approach driven by the GOSPA metric, benchmarked against three myopic algorithms (NS, GD & KL).

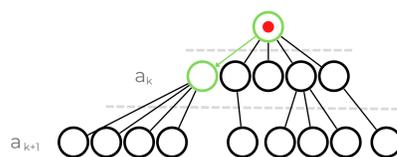


## 4. MONTE CARLO TREE SEARCH

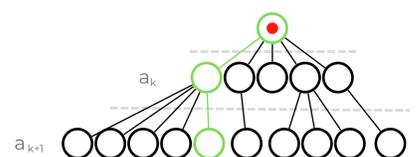
MCTS has four stages, as outlined below. It is a selective search algorithm that incrementally builds a search tree and decides where to explore next based on what it believes to be the most promising avenues.

The tree is continually grown until some predefined terminating condition is reached (such as number of iterations). Once this has been reached, the tree returns the best child of the root node, having considered a larger amount of actions within the planning horizon.

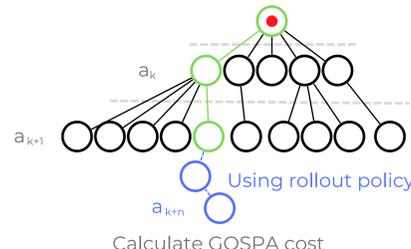
1. Selection



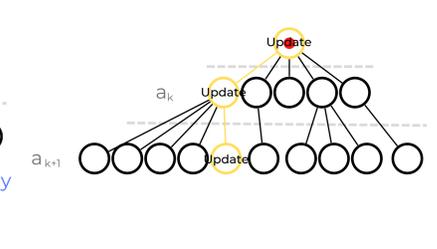
2. Expansion



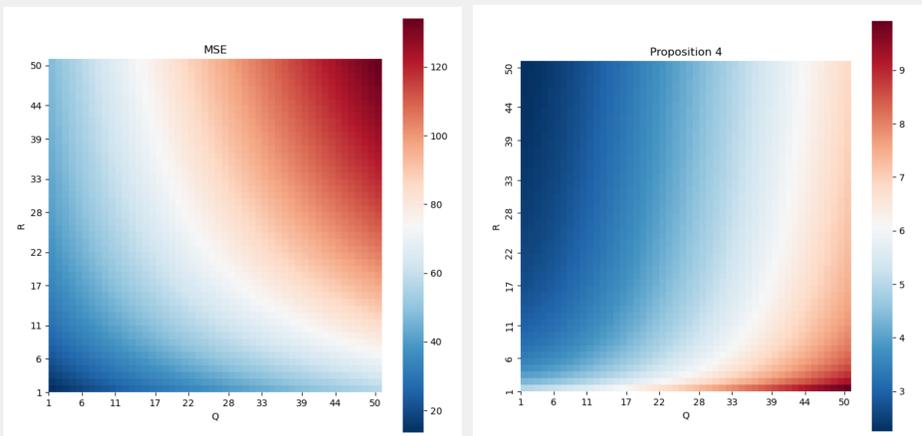
3. Simulation



4. Backpropagation



## 5. INFORMATION THEORETIC LIMITING CASE



If the actions that we select as a sensor manager, have an impact on the dynamics (either of the target or the measurements) then maximising the information theoretic divergence between two densities does not give desirable results.

The higher the values along the main diagonal of the  $Q$  and  $R$  matrices, the higher the value of the MSE, interestingly, also the higher the values of the KLD between the previous posterior and current posterior.

A similar pattern also shows in the Renyi Divergence of the same two densities and also when the two densities are the updated and predicted.

## 6. FUTURE WORK

Future work is to extend this line of non-myopic planning to multiple targets using a PMB filter.



Scan for more!



# Algorithms and Mechanisms in Distributed Settings

## Adversarial Equilibrium (AE): Contention Resolution (CR) – Combinatorial Group Testing (CGT)

**Model:**  $n$  stations,  $k < n$  of them are selected to be active and have to transmit a packet in shared channel. Communication happens in **synchronous** rounds. After the end of each round, the stations receive feedback from the channel.

**Problem 1:** Design distributed contention resolution protocols in which every station transmits with low latency.

**Problem 2:** Design such protocols that are robust under the selfish behaviour of stations.

**Adversarial Equilibrium:** An algorithm  $(s_1, s_2, \dots, s_n)$  for  $n$  players is called an  $(n, k)$ -Adversarial Equilibrium,  $(n, k)$ -AE, iff for any player  $i \in [n]$ , any unilateral deviation from strategy  $s_i$  to some other strategy  $s_i'$  results in either:

- no configurations  $K$  of  $k$  players where strategy  $s_i'$  improves the latency of station  $i$  **OR**
- if there exists a configuration  $K$  for which the strategy  $s_i'$  improves the latency of station  $i$ , then there must exist a configuration  $K'$  for which the strategy  $s_i'$  worsens the latency of station  $i$ .

### Results:

- Non-adaptive protocols that achieve *AE* with low latency for  $k = 2, 3$  [1]
- Adaptive protocols that achieve *AE* with approximately optimal latency [2]

## Maximal Extractable Value (MEV): Dynamics of Redistribution and Ordering Policies

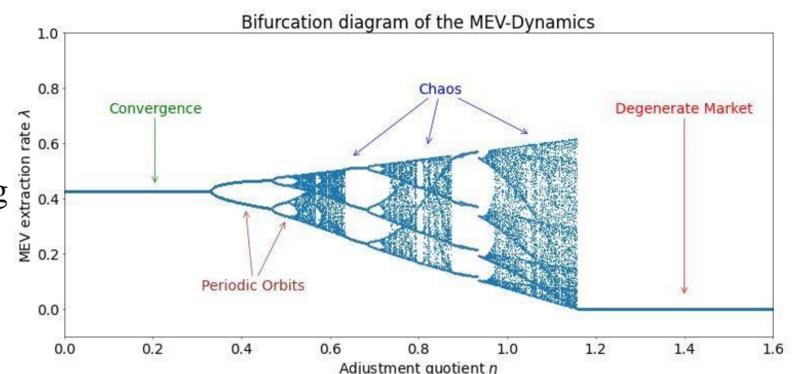
**MEV:** Any type of excess profit that a miner/validator can extract by adding, censoring and reordering users' transactions [3].

Two different approaches:

- Embrace MEV, but reduce its negative externalities
- Try to mitigate MEV by designing granular protocols (add communication and computational complexity overhead)

### Our Results:

- Study the evolution of a dynamic mechanism that redistributes part of the MEV extracted back to the users, with the goal of stabilizing at target value
- $\lambda_{t+1} = \lambda_t + \eta \lambda_t (1 - \lambda_t) (U \cdot \bar{F}(\lambda_t) - T \cdot M \cdot G(\lambda_t))$
- $\lambda_t$ : part of MEV that goes to the miners
- $F, G$ : tolerance distributions of users and miners resp. on MEV
- $\eta$ : adjustment quotient
- $T$ : target ratio between users and miners



## Future Work

- Study Contention Resolution by considering a less risk-averse behaviour of stations, such as a Bayesian model, where the type of each station follows a prior distribution.
- Improve the MEV Redistribution mechanism, by considering a more expressive target value.
- Analyse theoretically and empirically the incentive compatibility of such redistribution mechanisms.

## References

- [1] Georgios Chionas, Bogdan S. Chlebus, Dariusz R. Kowalski, and Piotr Krysta. Adversarial contention resolution games. In Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23
- [2] Georgios Chionas, Dariusz R. Kowalski, and Piotr Krysta. Combinatorial group testing with selfish agents. In Advances in Neural Information Processing Systems, volume 37. Curran Associates, Inc., 2023
- [3] P. Daian, S. Goldfeder, T. Kell, Yunqi Li, X. Zhao, I. Bentov, L. Breidenbach, and A. Juels. Flash boys 2.0: Frontrunning in decentralized exchanges, miner extractable value, and consensus instability. In 2020 IEEE Symposium on Security and Privacy (SP), 2020

# Developing Efficient Numerical Algorithms Using Fast Bayesian Random Forests

Harvinder Lehal (sma18hsl@liverpool.ac.uk),  
Prof. Simon Maskell (smaskell@liverpool.ac.uk),  
Dr. Lee Devlin (ljdevlin@liverpool.ac.uk)



## Motivation

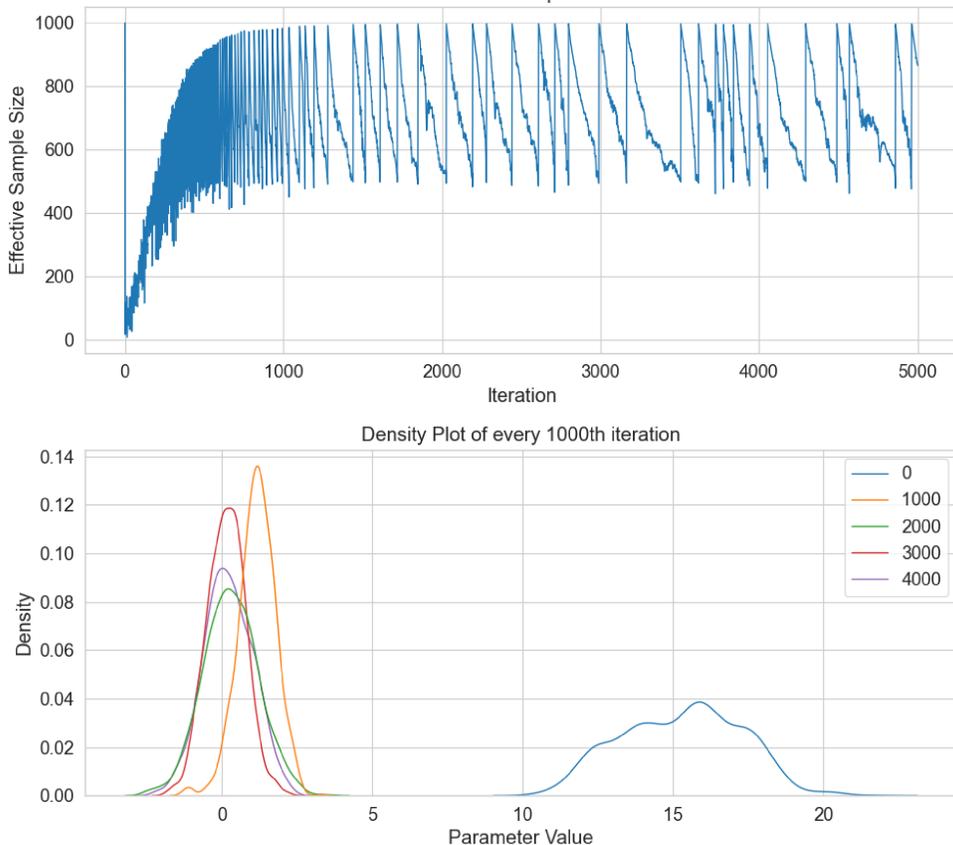
- Random Trees are inefficient and slow methods of classifying data that mostly rely on frequentist approaches that don't provide informed results with uncertainties.
- However, with Trees, the decision made at each node can be seen, unlike black-box algorithms such as Neural Networks.
- Unlike Trees, Random Forests are preferred as they can handle missingness, namely Missing completely at random (MCAR), Missing at random (MAR), and Missing not at random (MNAR).
- Such methods would be useful in areas where missingness is common such as opinion polls, surveys and databases [1,2,3].

## Proposal

- Use Bayesian approaches in the form of Sequential Monte Carlo (SMC) Algorithms to propose Random Forests [4].
- Look at novel proposals for SMC samplers such as Hierarchical Importance with Nested Training Samples (HINTS) [5].

## Sequential Monte Carlo (SMC)

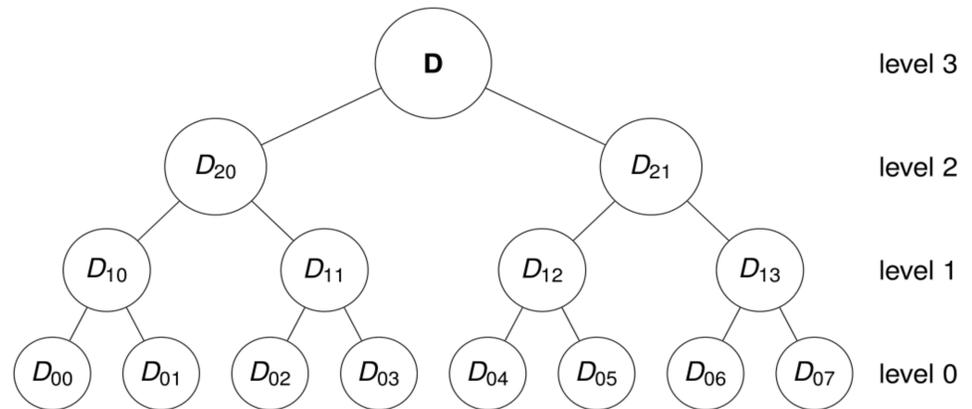
SMC Sampler of Gaussian with 1000 samples, 5000 iterations and stepsize 0.05



An example of SMC with a prior  $\mathcal{N}(15,3)$  target of  $\mathcal{N}(0,1)$ .

- By having proposal  $q(-)$ , backwards proposal  $L(-)$ , target  $\pi(-)$  and prior  $p_0(-)$  distributions, alongside a prior on the distribution being estimated, SMC is used to estimate the parameters of a target.
- Take a vector  $x^{(0)}$  of  $n \in 1, 2, \dots, N$  particles such that  $x_n^{(0)} \sim p_0(-)$ .
- For each iteration  $i \in 1, 2, \dots, M$  we calculate the new proposal for each particle  $x^{i+1} \sim q(x^i)$  and the normalised weights  $w_n^{(i)}$  for each particle.
- If the effective sample size  $N_{eff} = \frac{1}{\sum_{n=1}^N w_n^2}$  is below  $\frac{N}{2}$  we resample and re-set the weights to  $\frac{1}{N}$ , then continue as normal.

## Hierarchical Importance with Nested Training Samples (HINTS)

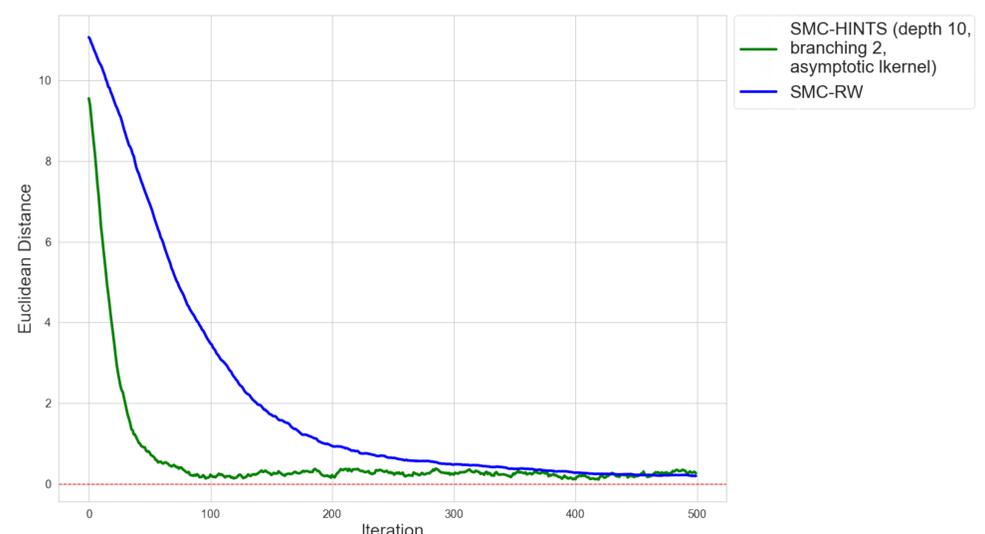


An example of the HINTS tree structure with depth 3 and branch factor 2.

HINTS is an advanced form of Markov Chain Monte Carlo (MCMC) that takes advantage of additive likelihood structures and datasets, denoted in the diagram above by  $\mathcal{D}$  and breaking it into subsets using a tree as shown and doing Metropolis-Hastings (M-H) steps on proposals using these subsets, going across the entire tree until it does a final M-H step on the entire set  $\mathcal{D}$ . This gives faster convergence to the target than standard MCMC, especially for higher dimensional data [5].

## Novel Idea: SMC-HINTS

Euclidean Distance from the Target Mean of an 11-D Gaussian using 256 particles, 500 iterations and 1024 data points



Preliminary results of SMC-HINTS compared to a random walk proposal.

By using a proposal  $q(-)$  as an iteration of HINTS in the SMC algorithm, we see a faster convergence to the target compared to a standard Random Walk proposal. So far it has worked on parameter estimation problems given a dataset, as shown above with a 11-D Gaussian Distribution with 1024 samples from the distribution.

## Moving Forward

- Use SMC-HINTS in discrete space problems where it can be used in place of Hamiltonian Monte Carlo (HMC) proposals.
- Apply SMC-HINTS to Trees and then Forests where it can take advantage of the additive structure of the likelihood function.

## References

- [1] Donald B. Rubin. *Inference and missing data*. 1976.
- [2] Roderick J. A. Little, Donald B. Rubin. *Statistical Analysis with Missing Data*. 2002.
- [3] J. L. Schafer. *Analysis of Incomplete Multivariate Data*. 1997.
- [4] Del Moral et al. *Sequential Monte Carlo Samplers*. 2006.
- [5] Malcolm J. A. Strens. *Efficient hierarchical MCMC for policy search*. 2004

# Computational Methods for *Real-Time* Subsampled Scanning (Transmission) Electron Microscopy



## ➤ Context and Motivation

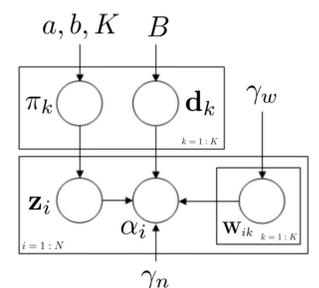
Despite their broad applicability to Electron Microscopy (EM) data, dictionary based inpainting algorithms have previously remained impractical for real-time usage with an Electron Microscope, as the reconstruction time for a single "frame" is orders of magnitude longer than the data acquisition time.

## ➤ Contribution

Development of efficient computational methods for  $N$ -dimensional batch signal generation (*unwrapping*) and recombination (*wrapping*), dictionary learning and sparse-coding for rapidly inpainting image data of arbitrary dimensions using a consumer GPU.

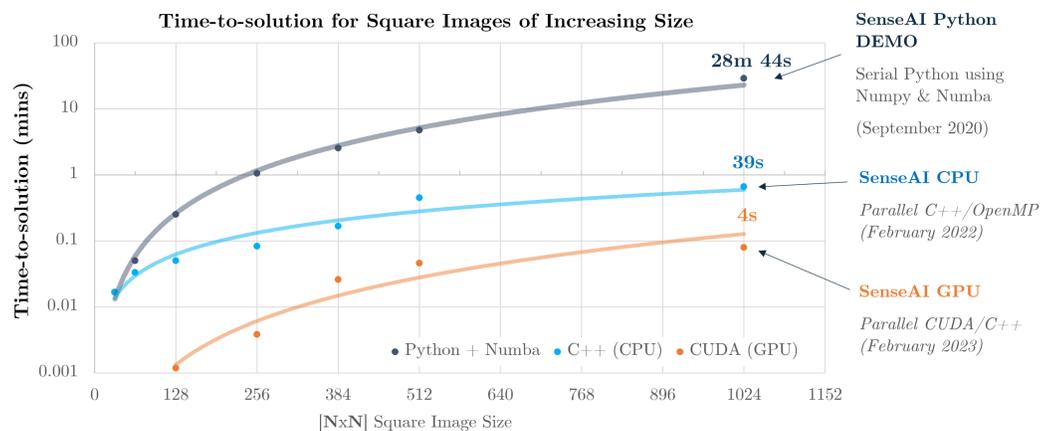
## ➤ The Algorithm

Beta-Process Factor Analysis (BPFA) is a hierarchical Bayesian model that uses the two-parameter beta process to define a sparse prior on an infinite dictionary space. Given a subsampled measurement  $Y$ , we first unwrap it into  $N_p$  overlapping patches  $\{y_i \in \mathbb{R}^{B^2}\}_{i=1}^{N_p}$ . We assume that each patch is sparse in some shared dictionary ( $x_i \approx D\alpha_i$ ), where  $D \in \mathbb{R}^{B^2 \times K}$  denotes a dictionary with  $K$  atoms and  $\alpha_i \in \mathbb{R}^K$  is a sparse vector of weights or coefficients.



## ➤ SenseAI: Real-Time Inpainting

SenseAI is a GPU-parallelised C++ library capable of efficient frame-by-frame reconstructions of a time-variable input feed, enabling the real-time subsampled operation and adjustment of an electron microscope. Over the last few years, iterative improvements to the method, such as the implementation of efficient N-dimensional wrapping and unwrapping of signals, algorithm parallelisation, and the use of a GPU has reduced standard megapixel reconstruction times from half an hour to just a few seconds (as shown in Figure 1).



**Figure 1:** Time-to-solution (2 epochs) for reconstructing images of increasing size. Each were performed on the same PC with an Intel Xeon E5-2678v3 (12 cores, 2.50 GHz) and an NVIDIA RTX 3060.

## ➤ Multi-Instance Reconstructions

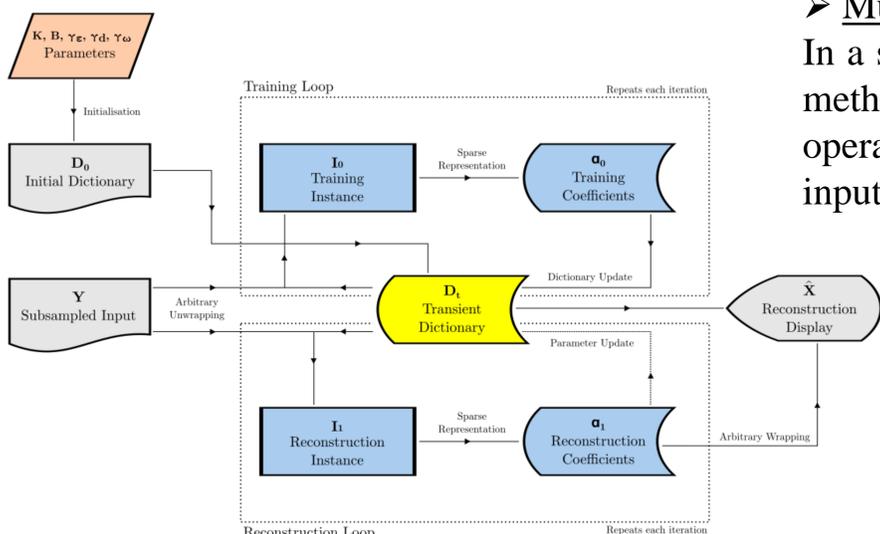
In a significant departure from previous methods, SenseAI introduces a method of multiple independent instances of the BPFA algorithm operating simultaneously on the same reconstruction of a time-variable input feed using a shared transient dictionary (each instance results in an independent estimate of  $\alpha$  for each batch but updates and performs sparse-coding using the shared dictionary parameters  $K, B, \pi, \gamma_\epsilon, \gamma_d, \gamma_w$ ).

## ➤ Software Capabilities

- ✓ N-Dimensional BPFA
- ✓ Arbitrary Patch Selection
- ✓ Time-Variable Input Feeds
- ✓ Live Dictionary Transfer
- ✓ STEM Hardware Support

## ➤ EM Applications

- ✓ STEM Simulations
- ✓ 2D STEM (BF/DF)
- ✓ 3D SEM (Cryo-FIB)
- ✓ 3D IPF Maps (EBSD)
- 4D STEM (Ptychography)



**Figure 2:** Flow-chart describing a multi-instance reconstruction of a time-variable subsampled input  $Y$ .

# Dynamic Semantic-based Graph Convolution Network for Skeleton-based Human Action Recognition



Jiayang Xie (Jiayang.Xie@liverpool.ac.uk), Yalin Zheng, Nguyen Anh, Xiaoyun Yang

## Background

- Spatial-Temporal GCNs have been the most popular for human action recognition since they can capture inherent interaction between body joints through node aggregation scheme.
- Previous works ignored the semantics of the skeleton. They simply assumed all joints/edges as the same type, making them insufficient to capture the semantic properties of actions.

## Contribution

- Encoding the dynamical semantic information of joints and edges in GCNs aggregation process implicitly.
- Generated a dynamic semantic-based graph neural network, and extensive experiments highlight that the proposed outperforms SOTA methods notably on published dataset

## Introduction

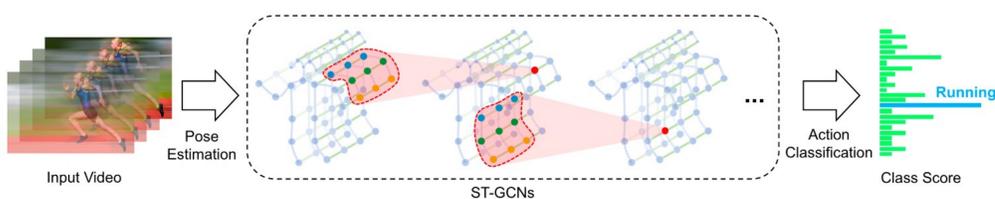


Figure 1. illustration of ST-GCNs for skeleton-based human action recognition

Figure 1 shows the basic framework of ST-GCN, where body joint and body link are considered as node and edge, and the skeleton graph generated based on body natural link, finally, ST-GCNs were applied for classification. However, the fixed graph limited the representation of GCN and is inefficient in capturing the changeable human movement. Thus, the dynamic graph generation methods were proposed as Figure 2 (a). Comparing with the previous work, our method aim to encode the semantic information of node into the graph, as shown in Figure 2 (b) and (c).

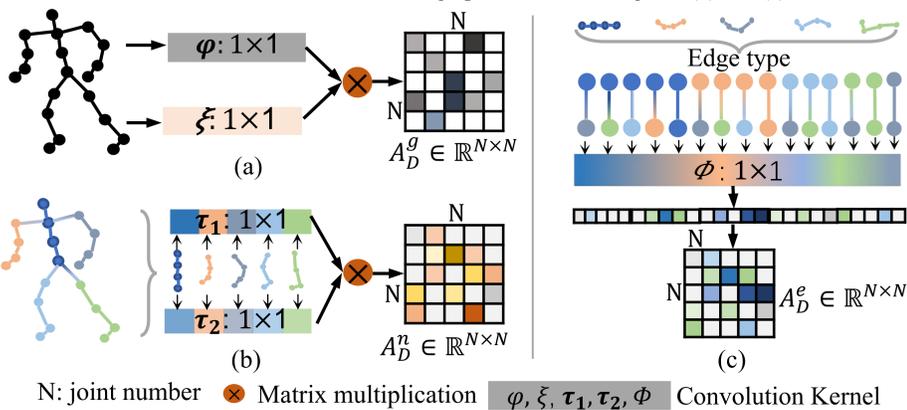


Figure 2. illustration of node and edge type-aware adaptive graph generation. (a) the previous adaptive graph generation. (b) node type-aware adaptive generation. (c) represent the edge type-aware adaptive graph generation.

## Method

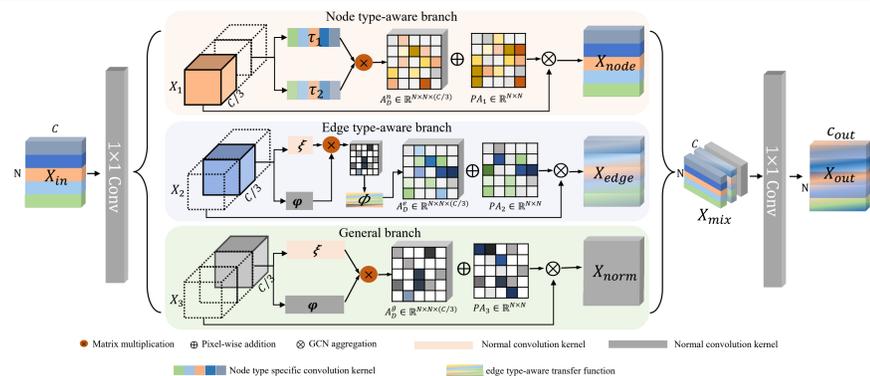


Figure 3. The framework of the proposed DS-GCN. The spatial graph convolution structure was decomposed into three branches, the node-type aware branch, the edge-type aware branch, and the general branch. In each branch, the corresponding semantic self-adaptive graph were applied to represent the skeleton.

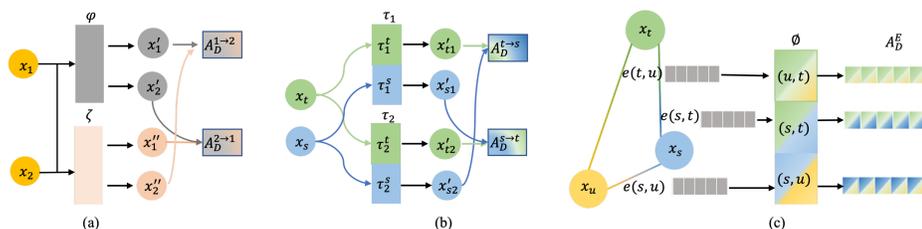


Figure 4. Illustration of the joint correlation calculation. (a) represents the standard non-local mechanism, for each transform function  $\varphi(\cdot)$  and  $\xi(\cdot)$ , the node features are updated by sharing the same parameters. (b) represents the node type-aware correction. In each transform function, the convolution kernels are divided into several parts, each of which corresponds to a specific node type, and then the node characteristics in different types were updated by their individual parameters set. The coloured circles denote different node types and the coloured squares denote different convolution kernels. (c) illustrates the edge type-aware correlation. For each type of edge, specific convolution kernels were designed and utilized for updating the edge feature. The coloured circles denote node types; mix-coloured squares denote corresponding edges with node pairs.

## Results

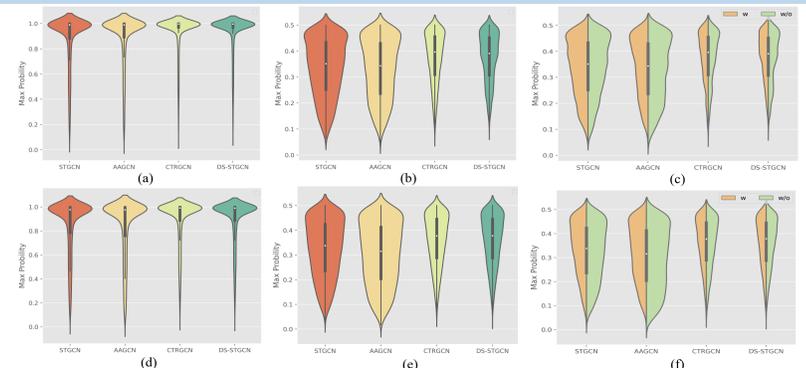


Figure 5. Analysis of the classification distribution for each backbone. (a-c) represent the classification result distribution for all samples in NTU60 XSub. (d-f) represent the classification result distribution for all samples in NTU120 Xset. The max probability of each sample was utilized as the final result. In (a) and (d), the samples with max probability ranging from 0 to 1 were analysed. In (b-c) and (e-f), the samples with a max probability lower than 0.5 were analysed. In (e) and (f), w means that backbone with semantic modules, and w/o means that backbone without semantic modules. The area for each violin map indicates the number of samples. Observing in (a) and (d), we can observe that the distribution generated by the proposed DS-STGCN is more compact, also in (b) and (e), it can be found that the numbers of samples with max probability lower than 0.5 are significantly reduced when comparing with STGCN and AAGCN. When looking at (e) and (f), we can see that for each backbone, the area for a model with semantic encoding is decreased significantly when compared with the model without semantic encoding, implying the proposed semantic modules can make the classification more accurate.

Method	NTU60-XSub	NTU120-XSet
ST-GCN (Si et al. 2018)	87.8	85.5
2s-GCN (Shi et al. 2019b)	89.5	86.0
CTR-GCN (Chen et al. 2021)	89.6	86.0
DS-GCN	<b>90.8</b>	<b>87.2</b>

Method	NTU60-XSub	NTU120-XSet
DS-GCN <sub>shared</sub>	90.1	86.8
DS-GCN <sub>B-wise</sub>	<b>90.8</b>	<b>87.2</b>

Table 3: Comparison DS-GCN in different learnable weight manners. DS-GCN<sub>shared</sub> represents the DS-GCN with shared  $\lambda$  for all the branches, DS-GCN<sub>B-wise</sub> represent the DS-GCN with individual  $\lambda$  for different branches.

Table 2: Effectiveness of DS-GCN. The proposed DS-GCN can achieve the best performance.

Method	NTU60-XSub
DS-GCN w/o N&E	90.0
DS-GCN w/o N	90.5
DS-GCN w/o E	90.4
DS-GCN	<b>90.8</b>

Module	Encode stage	NTU60-XSub
DS-GCN w/o N&E	-	90.0
DS-GCN <sub>ini</sub>	[1-4]	90.2
DS-GCN <sub>mid</sub>	[5-7]	90.7
DS-GCN <sub>end</sub>	[8-10]	90.5
DS-GCN	[1-10]	<b>90.8</b>

Table 4: Ablation on the edge/node type encoding.  $N$  represents the node type-aware encoding, and  $E$  represents the edge type-aware encoding.  $w/o$  means without, representing that the corresponding semantic encoding is replaced with the general branch.

Table 5: Exploration on the semantic encoding stage. DS-GCN w/o N&E represents no semantic module is utilized, DS-GCN<sub>ini</sub> represents just utilized DS-GCN in layer [1-4], DS-GCN<sub>mid</sub> represents just utilized DS-GCN in layer [5-7], DS-GCN<sub>end</sub> represents just utilized DS-GCN in layer [8-10], DS-GCN represents DS-GCN is utilized in all the layers.

Module	NTU60-Xsub	NTU60-Xview	NTU120-Xsub	NTU120-Xset	Kinetics-400
ST-GCN (Si et al. 2018)	81.5	88.3	70.7	73.2	30.7
SGN (Zhang et al. 2020)	86.6	93.4	-	-	-
AS-GCN (Li et al. 2019)	86.8	94.2	78.3	79.8	34.8
RA-GCN (Song et al. 2020)	87.3	93.6	78.3	79.8	34.8
2s-GCN (Shi et al. 2019b)	88.5	95.1	-	-	-
DGNN (Shi et al. 2019a)	89.9	96.1	-	-	-
FGCN (Yang et al. 2021)	90.2	96.3	85.4	87.4	-
ShiftGCN (Cheng et al. 2020b)	90.7	96.5	85.9	87.6	-
DSTA-Net (Shi et al. 2020a)	91.5	96.4	86.6	89.0	-
MS-G3D (Liu et al. 2020)	91.5	96.2	86.9	88.4	38.0
CTR-GCN (Chen et al. 2021)	92.4	96.8	88.9	90.6	-
ST-GCN++ (Duan et al. 2022b)	92.6	97.4	88.6	90.8	49.1
PoseConv3D (Duan et al. 2022c)*	<b>94.1</b>	<b>97.1</b>	86.9	90.3	47.7
DS-GCN	93.1	<b>97.5</b>	<b>89.2</b>	<b>91.1</b>	<b>50.6</b>

Table 6: Classification accuracy comparison against state-of-the-art methods.

## Conclusion

- This work proposed to implicitly encode the joints and edge types for skeleton-based human action recognition. Extensive experiments show that the semantic modules are generalizable and can be exploited in various backbones for boosting recognition accuracy
- The proposed DS-GCN surpasses the SOTAs on two challenging benchmarks, confirming its superior capability and effectiveness.

# AI for Fast Discovery of Novel Materials for Healthcare

Jinhao Gu supervised by Ángel F. García-Fernández, Robert E. Firth (STFC), Rasmita Raval, Joanne O'Keeffe (Unilever)

Email: jinngu@liverpool.ac.uk, angel.garcia-Fernandez@liverpool.ac.uk, robert.firth@stfc.ac.uk



## Background & Aim

Innovation in healthcare and personal care both depend on the creation of novel materials for public hygiene and infection treatment. In the molecule synthesis process, a chemist would prefer to know the characteristics of the molecules before deciding which ones to synthesise.

As machine learning has achieved great success in various areas, it can and has already been applied to chemistry. In this project, we target Gaussian Process models (Kernel design) for molecule property prediction for the discovery of novel materials.

## Ongoing Work

The molecular graph is a popular representation of molecules, which can contain all the information within a molecule.

We are now designing a new metric to measure the difference between graphs with different sizes.

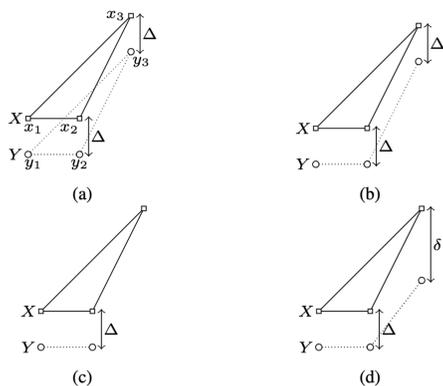


Fig. 1. Example to illustrate the node and edge mismatch costs for the same ground truth graph X, and different estimated graphs Y ( $\Delta \ll c$ ).

In each example, the mismatch between graph X and graph Y is:

- Three properly assigned nodes and no edge mismatch;
- Three properly assigned nodes and one missed edge;
- Two properly assigned nodes, one missed node and two half-edge mismatch penalties;
- Two properly assigned nodes and two unassigned nodes ( $\delta \gg c$ ), three half-edge mismatch penalties.

Our new graph GOSPA metric is composed of 4 separate types of errors, localisation error, false/miss node error, edge mismatch error and three hyperparameters. We can change the parameters to control the characteristics of the metric.

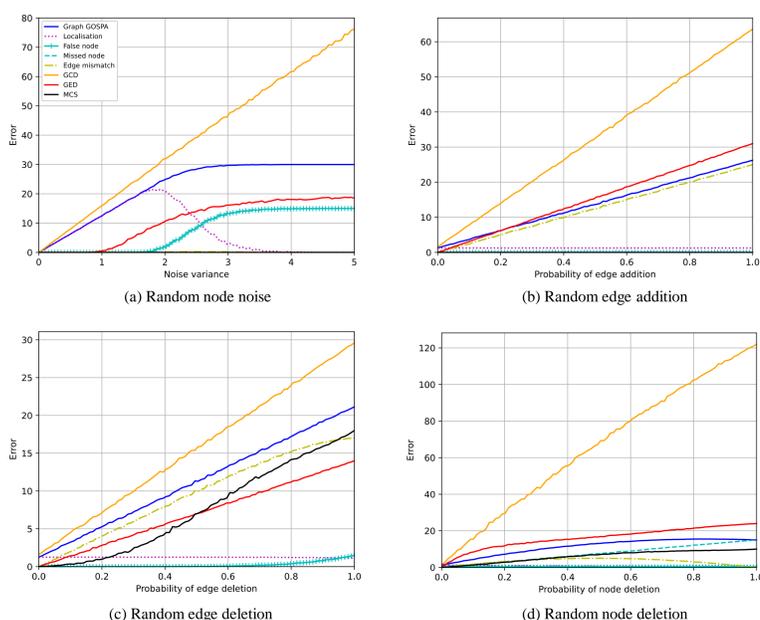


Fig. 2. Plots of average **graph GOSPA metric** (proposed), generalised chemical distance (GCD), graph edit distance (GED) and maximum common subgraph (MCS) distance comparing the ground truth graph of 10 nodes with (a) graphs with random Gaussian noise with increasing noise variance in node attributes; (b) graphs with random edge addition in adjacency matrix and Gaussian noise with  $\sigma^2 = 0.1$  in node attributes; (c) graphs with random edge removal in adjacency matrix and Gaussian noise with  $\sigma^2 = 0.1$  in node attributes; (d) graphs with different sizes and Gaussian noise with  $\sigma^2 = 0.1$  in node attributes.

## Experimental Results

We tested our metric and other existing graph distances (GCD, GED, MCS) over a molecule dataset, which contains 5 collections in total.

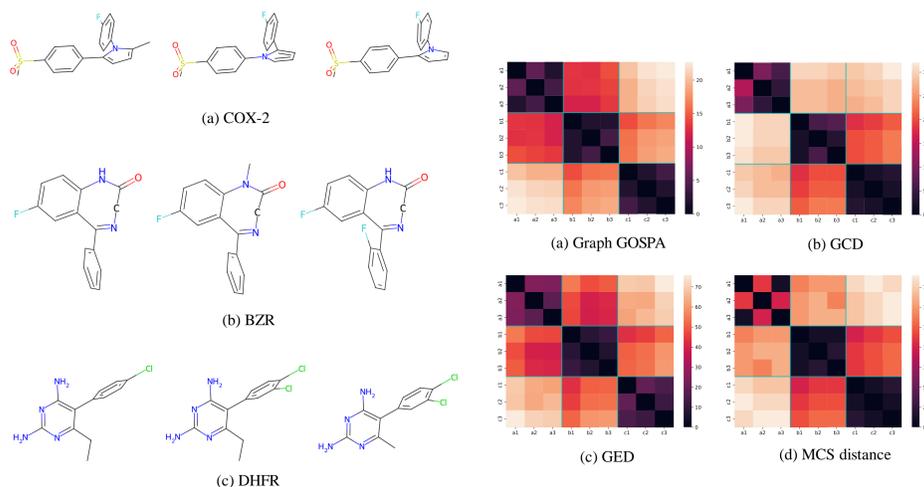


Fig. 3. Plots of three molecular graph samples from the following collections: (a) cyclooxygenase-2 (COX-2) inhibitors; (b) benzodiazepine receptor (BZR) ligands; (c) dihydrofolate reductase (DHFR) inhibitors

Fig. 4. Illustration of the distance matrices computed using (a) graph GOSPA, (b) GED, (c) GED and (d) GCD for molecules from three different collections.

Table 1. Average distance between molecule COX-2,1-1 and 100 molecules from each family.

	COX-2	BZR	DHFR
Graph GOSPA	11.78	19.15	23.49
GCD	51.58	52.17	70.27
GED	21.42	32.57	37.65
MCS	15.54	17.41	23.15

Table 2. Decomposition of the average graph GOSPA error between molecule COX-2,1-1 and 100 molecules from each family.

	COX-2	BZR	DHFR
Localisation	2.04	3.36	4.75
False node	4.94	3.45	6.74
Miss node	0.24	4.16	3.70
Edge mismatch	4.56	8.19	8.30

## Conclusion

This project is to discover novel materials for healthcare with the aid of machine learning tools.

Currently, we are trying to propose a graph GOSPA metric which can measure the discrepancy of graphs with different sizes. Unlike existing popular graph distances, it satisfies the metric properties (identity, symmetry, triangle inequality) for graphs with attributes and different sizes.

## References

- [1] A. S. Rahmathullah, Á. F. García-Fernández, and L. Svensson, "Generalized optimal sub-pattern assignment metric", in *2017 20th International Conference on Information Fusion (Fusion)*, Jul. 2017, pp. 1–8. doi: [10.23919/ICIF.2017.8009645](https://doi.org/10.23919/ICIF.2017.8009645).
- [2] J. Bento and S. Ioannidis, "A family of tractable graph metrics," *Applied Network Science*, vol. 4, no. 107, 2019.
- [3] A. A. Hagberg, D. A. Schult, and P. J. Swart, "Exploring network structure, dynamics, and function using NetworkX," in *Proceedings of the 7th Python in Science Conference*, G. Varoquaux, T. Vaught, and J. Millman, Eds., Pasadena, CA USA, 2008, pp. 11–15.
- [4] H. Bunke, "On a relation between graph edit distance and maximum common subgraph," *Pattern recognition letters*, vol. 18, no. 8, pp. 689–694, 1997.
- [5] J. J. Sutherland, L. A. O'Brien, and D. F. Weaver, "Spline-fitting with a genetic algorithm: A method for developing classification structure-activity relationships," *Journal of Chemical Information and Computer Sciences*, vol. 43, no. 6, pp. 1906–1915, 2003.

# Distributed Hypothesis Generation and Evaluation

Jordan Robinson (J.Robinson9@liverpool.ac.uk),  
 Katie Atkinson (katie@liverpool.ac.uk),  
 Simon Maskell (smaskell@liverpool.ac.uk),  
 Todd Robinson (Dstl), and  
 Chris Reed (University of Dundee)



## Background & Aim

Intelligence analysis is currently conducted by distributed teams of expert human agents who use their domain knowledge, combined with a variety of structured analytical techniques, to generate and evaluate sets of conflicting hypotheses to inform potential high-stake decision making. Analysis can be tedious but it requires the full attention of human agents as the context can be such that what would otherwise be a minor detail has a significant impact on the likelihood of a hypothesis, and so on the downstream decision making. This project aims to enhance the speed and scale of intelligence analyses through the development of decision-support tools which combine explainable artificial intelligence algorithms with human expertise, in the form of human-machine teams, to aid intelligence analysts in evaluating complex and competing hypotheses. The tools created will combine techniques found within the Natural Language Processing, Computational Argumentation, Probability and Information Theory literature and should enable analysts to focus their attention where it's needed most by assisting them throughout the analytical pipeline.

## Argument Data

The reasoning about the hypotheses in the ACH example was captured using Walton's argumentation schemes, specifically the *Argument from Cause to Effect* [2].

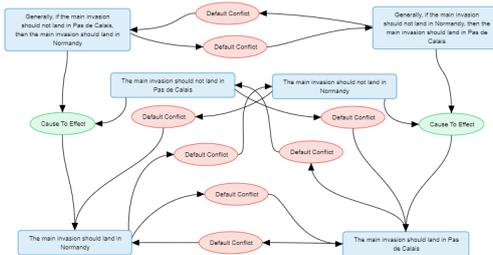


Figure 1: The Normandy and Calais hypotheses for the Operation Fortitude example.

The stereotypical patterns of reasoning about the evidence in the example was also captured using Walton's argumentation schemes [2].

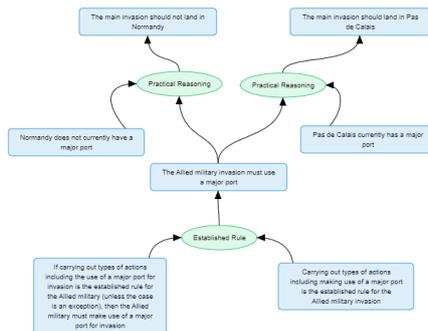


Figure 2: An example of the reasoning involved in the evidence, specifically the reasoning about an Allied invasion of a major port.

## Analysis of Competing Hypotheses

Heuer's Analysis of Competing Hypotheses (ACH) is a structured analytical technique where an analyst identifies a set of hypotheses, systematically evaluates evidence that is consistent (+) or inconsistent (-) with each hypothesis, and rejects hypotheses that contain too much inconsistent information [1]. The steps in the ACH are as follows:

1. Identify all possible hypotheses;
2. List all the significant information;
3. Create a *diagnosticity* matrix and use this to analyse the available information;
4. Refine the matrix;
5. Draw tentative conclusions about the relative likelihoods of the hypotheses;
6. Analyse the sensitivity of the tentatively-held conclusions.
7. Report the conclusions;
8. Identify indicators or milestones for future observation.

This research project has studied an example from World War 2, called Operation Fortitude. Dstl conducted the ACH on this example, instantiating a *diagnosticity* matrix, which this research has then taken and mapped to an argumentation framework (AF).

Table 1: An ACH diagnosticity matrix for Operation Fortitude.

	Normandy	Calais
PORT	-	+
STRONGPOINT	+	-
SHORTEST	-	+
EXPANSION	+	-
VSITES	-	+
FUSAG	+	+
FUSAGLOC	-	+
PATTON	-	+
21AG	+	+
21AGLOC	+	-
SPIES	-	+
SIGINT	-	+
RECON	-	+
AIRPLAN	-	+

## The Diagnostic Argument Identifier

Heuer refers to the sensitivity analysis conducted in step 6 of the ACH as a crucial task because it allows an analyst to understand how dependent the likelihoods of their hypotheses are on the items of evidence within their analysis. To do this, analysts sequentially remove each item of evidence from the ACH matrix and re-evaluate it to discover whether their initial, tentatively-held conclusions change.

The fundamental concept underpinning the Diagnostic Argument Identifier (DAI) is the notion of sensitivity analysis. The DAI combines Dung's abstract AFs from the computational argumentation literature [3], and Shannon's mutual information (I) between two random variables from the probability theory and information theory literature [4], to conduct sensitivity analysis, in accordance with step 6 of Heuer's ACH. The DAI sequentially removes arguments from AFs and statistically quantifies the change in the evaluation after an argument's removal. Automating this reasoning task should benefit intelligence analysts by providing them with an algorithm which quantifies the change that removing an item of evidence has on the likelihood of all the hypotheses included in the analysis.

$$\begin{aligned}
 I(x_1, x_{2:n}) &= \sum_{x_1} \sum_{x_{2:n}} p(x_{1:n}) \log_2 \frac{p(x_{1:n})}{p(x_1)p(x_{2:n})} \\
 &= \sum_{x_1} \sum_{x_{2:n}} p(x_{1:n}) \log_2 \frac{p(x_1)p(x_{2:n}|x_1)}{p(x_1)p(x_{2:n})} \\
 &= \sum_{x_1} \sum_{x_{2:n}} p(x_{1:n}) \log_2 \frac{p(x_{2:n}|x_1)}{p(x_{2:n})}
 \end{aligned}$$

## An Abstract Example

Imagine we have an initial AF  $\mathcal{G}$  with:

- Arguments:  $\{p, q, r, s\}$
- Relations:  $\{(q, p), (r, q), (r, s), (s, q), (s, r)\}$

We evaluate the AF  $\mathcal{G}$  using preferred semantics to find a set of labellings,  $\mathcal{L}$ .

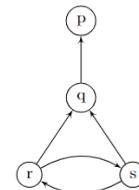


Figure 3: The initial AF  $\mathcal{G}$ .

Table 2: Preferred labellings of  $\mathcal{G}$ .

$\mathcal{L}$	Argument Labels			
	p	q	r	s
L1	IN	OUT	IN	OUT
L2	IN	OUT	OUT	IN

Let's now calculate the mutual information between the labels of each argument and the rest of the arguments within  $\mathcal{G}$ .

Table 3: Mutual information between each argument and the rest of the initial AF  $\mathcal{G}$ .

	Mutual Information (bits)
$I(p, rest(\mathcal{G}))$	0
$I(q, rest(\mathcal{G}))$	0
$I(r, rest(\mathcal{G}))$	1
$I(s, rest(\mathcal{G}))$	1

Now let's remove the argument  $r$  from the initial AF  $\mathcal{G}$  to create the sensitive AF  $\mathcal{G}^r$  with:

- Arguments:  $\{p, q, s\}$
- Relations:  $\{(q, p), (s, q)\}$

The sensitive AF  $\mathcal{G}^r$  is then evaluated using the same semantics as above to discover the set of sensitive labellings.

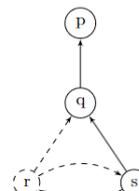


Figure 4: The sensitive AF  $\mathcal{G}^r$ .

Table 4: Preferred labellings of  $\mathcal{G}^r$ .

$\mathcal{L}$	Argument Labels		
	p	q	s
L1	IN	OUT	IN

We can now calculate mutual information between each argument in the sensitive AF  $\mathcal{G}^r$  and the rest of the network.

Table 5: Mutual information between each argument and the rest of the initial AF  $\mathcal{G}^r$ .

	Mutual Information (bits)
$I(p, rest(\mathcal{G}^r))$	0
$I(q, rest(\mathcal{G}^r))$	0
$I(s, rest(\mathcal{G}^r))$	0

Finally, if we wanted to understand the change in mutual information for, say, the argument  $s$  before and after the removal of the argument  $r$ , then we would compute the difference as follows:

$$\begin{aligned}
 \mathcal{D}(r | \mathcal{G}, \mathcal{G}^r) &= I(s, rest(\mathcal{G})) - I(s, rest(\mathcal{G}^r)) \\
 &= 1 \text{ bit}
 \end{aligned}$$

## References

1. Heuer, R. J. (1999). *Psychology of Intelligence Analysis*.
2. Walton, D., Reed, C., & Macagno, F. (2008). *Argumentation Schemes*.
3. Dung, P. M. (1995). *On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games*.
4. Shannon, C. E. (1948). *A Mathematical Theory of Communication*.

# Maximising Detection Using High-Performance Processing of Multi-Sensor Data

Joshua J. Wakefield (sgjwakef@liverpool.ac.uk)

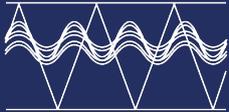
Prof. Jason F. Ralph (jfralph@liverpool.ac.uk), Dr Stewart Haslinger (sgh@liverpool.ac.uk),

Robert Taylor (R.Taylor@uk.thalesgroup.com)



This work utilises ray tracing techniques to simulate sound propagation in complex underwater environments, enabling the accurate prediction of acoustic signal paths. Conventional array beamforming techniques are then employed to process the received signals and extract directional information for target localisation. A particle filter tracking algorithm is implemented to estimate the trajectory of underwater objects based on bearing-only measurements.

## Ray Traced Acoustic Propagation Modelling



In underwater acoustic modelling, ray tracing is a fundamental technique used to simulate the transmission of sound waves through the ocean. Ray-based methods solve the equations governing propagation of sound waves by tracing the paths of individual sound rays, considering interactions with the seafloor and sea-surface.

The speed and direction of the rays are determined based on the local properties of the medium, such as sound speed ( $c$ ) and density ( $\rho$ ). These interactions are modelled using principles from geometrical and physical optics, incorporating mathematical expressions for the laws of reflection, refraction, and diffraction. This is achieved through the use of 3D vectorised Snell's Law where the speed of sound at either side of the boundary is calculated by an appropriate equation for deep or shallow water, e.g. the Munk or Mackensie equations.

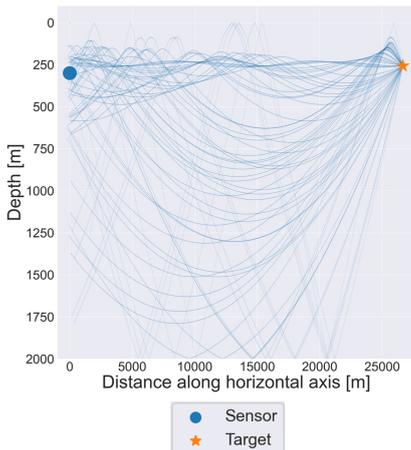


Fig. 1. Underwater acoustic propagation from source to sensor array, characterized by the Mackensie equation's sound speed profile.

## Conventional Beamforming



Conventional beamforming is a signal processing technique employed to enhance desired acoustic signals from a specific direction while minimizing noise and interference originating from other directions. By employing time delays and weightings of signals received by an array of hydrophones, conventional beamforming creates a spatial filter that focuses on a particular direction, effectively improving the signal-to-noise ratio.

The power spectral density (PSD) for a specific steering direction can be derived by computing the squared magnitude of the complex signal's Fourier transform, represented as:

$$PSD = |F(x)|^2.$$

The steering direction at which the PSD is maximum serves as the estimated direction of arrival and the measurement to be used in the update procedure of the particle filter.

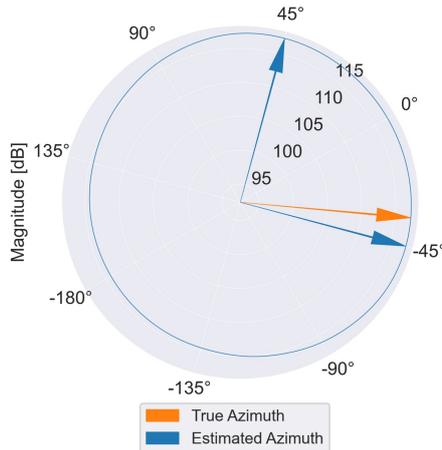
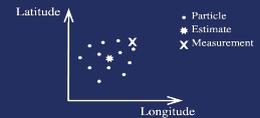


Fig. 2. Polar plot depicting bearings and directional power of a received signal at SNR of -15dB.

## Particle Filtering for Bearing-Only Measurements



A particle filter is a sequential Monte Carlo method used for state estimation in dynamic systems. By employing a set of weighted particles to represent possible target positions, the filter can effectively track the target's movement over time.

**Predict:** The particles are propagated using the system's dynamic model to predict their positions at the next time step.

$$x_k^i = Fx_{k-1}^i + v_{k-1}, \quad i = 1, \dots, N_s$$

**Update:** The weights of the particles are updated based on the likelihood of the measurements given the predicted positions.

$$w_k^i = e^{\frac{1}{2}d(x_{k-1}^i, z_k)} \cdot w_{k-1}^i, \quad i = 1, \dots, N_s$$

**Resample:** To prevent the degeneracy of the particle set, resampling is often performed. Particles with high weights are duplicated while particle with low weights are removed to ensure the particle set maintains diversity and represents the true posterior distribution.

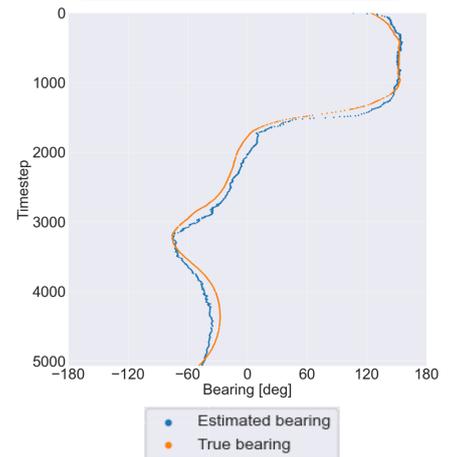


Fig. 3. Bearing-time record of estimated and true bearings over time

## Future Work

Track-before-detect techniques will be explored to improve the detection and tracking of low signal-to-noise ratio (SNR) targets. Additionally, the Probabilistic Multi-Bernoulli Mixture (PMBM) filter will be integrated to handle multiple target tracking scenarios, providing a robust solution for tracking multiple underwater objects simultaneously. Furthermore, the incorporation of Minimum Variance Distortionless Response (MVDR) adaptive beamforming will be investigated to enhance the spatial resolution and reduce the effects of ambient noise, thereby improving the accuracy and reliability of the target localisation process. These advancements are expected to significantly enhance the performance and versatility of the system, making it more adept at addressing complex underwater tracking and localization challenges.

# Towards Data Driven Aerodynamic Models

Mehdi Anhichem (m.anhichem@liverpool.ac.uk),  
Dr. Sebastian Timme (Sebastian.Timme@liverpool.ac.uk),  
Dr. Jony Castagna (jony.castagna@stfc.ac.uk),  
Dr. Andrew Peace (apeace@ara.co.uk),  
Moira Maina



## Background & Aims

The study of aerodynamics is crucial for estimating aircraft performance characteristics during the various design phases. Aerodynamic analysis and design rely heavily on three main sources of information; flight testing, wind tunnel testing and computational fluid dynamics (CFD). In wind tunnel experiments, a scaled model is manufactured to collect data about the air flow around the model through measurements. Wind tunnel testing can be expensive and is subject to multiple sources of uncertainty. Aerodynamic data can also be computed with deterministic numerical simulations such as using an incarnation of the Navier–Stokes equations. With the ever-increasing computing power, such numerical simulations have become a must in aerodynamics. However, when dealing with complex configurations and phenomena, particularly near the edge of the flight envelope, numerical simulations tend to lack the required accuracy and come with a significant cost burden.

**Research hypothesis:** leverage information from different aerodynamic data sources while controlling the definition and propagation of uncertainty towards the decision-making level.

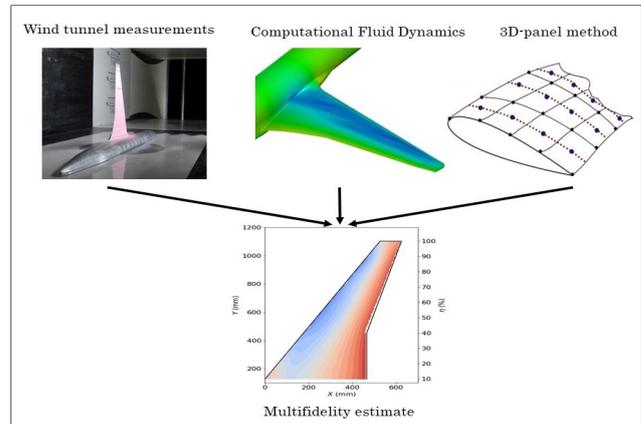


Fig 1: Schematic representation of the multifidelity data fusion.

## Bayesian surrogate models

A **Bayesian surrogate model** is a probabilistic approximation of a complex real-world system or function that provides predictions and quantifies uncertainty. Here, two different models (Bayesian neural network and Gaussian processes) are compared on an aerodynamic test case.

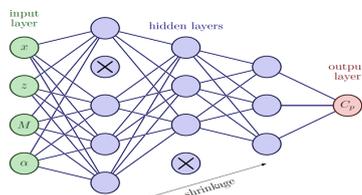


Fig 2: Bayesian neural network with Monte Carlo dropout architecture example.

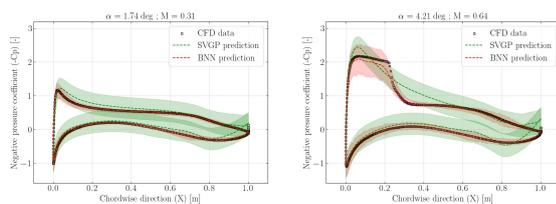


Fig 3: A comparison between the predictions made by BNN and SVGP models on points of the test set.

## Current & Future Work

- **Bayesian surrogate modelling:** Explore additional options for such models (based on both BNN and GP) and apply them to higher dimensional, practical aerospace problems with more complex geometry.
- **Multifidelity data fusion:** Continue the development of a detailed definition of the fidelity function such as considering variations in the reliability of an information source under different flow conditions. Compare approach to hierarchical data fusion approach from the literature.

## Non-hierarchical data fusion

**Multifidelity methodology:** Non-hierarchical multifidelity approach based on a combination of Gaussian process surrogate models.

The required amount of data to study the surface flow on a wing implies the use of a 'big data' extension of Gaussian process regression based on stochastic variational inference [1].

Incorporation of expert opinion into fidelity function.

The library adopted in the implementation of the data fusion framework is GPflow, a Python module based on TensorFlow.

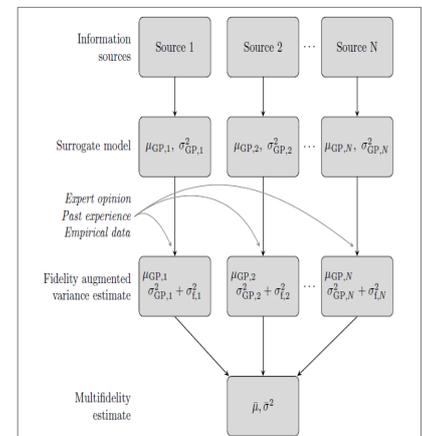


Fig 4: Combination of Gaussian processes with augmented variance. [2]

## Conclusion

This project analyses Bayesian surrogate models and a multifidelity data fusion framework applied to rich data describing large aircraft wing pressure distributions obtained from experimental and computational methods [3]. While non-Bayesian surrogate modelling has been widely employed in aerodynamics, this work presents a step change by adopting Bayesian surrogates, introducing a probabilistic framework that guarantees accuracy and captures uncertainties, enabling robust decision-making in the field of aerodynamic design and optimisation.

## References

- [1] Hensman et al. (2013), <https://arxiv.org/abs/1309.6835>
- [2] Feldstein et al. (2020), <https://doi.org/10.2514/1.J058388>
- [3] Anhichem et al. (2022) <https://doi.org/10.2514/6.2022-3526>

# Bayesian Block Sparse Spectral Unmixing

Oisín Boyle: ORBoyle@liverpool.ac.uk  
Xinping Yi: Xinping.Yi@Liverpool.ac.uk

Murat Uney: M.Uney@liverpool.ac.uk  
Joseph Brindley: Joseph.Brindley@Gencoa.com



## Background Theory

Spectral Unmixing [1] is the process of decomposing a spectrum to find the abundancies of each endmember. A standard way of doing this is to use a Linear Model.

$$\mathbf{y} = \mathbf{D}\mathbf{x} + \boldsymbol{\lambda}$$

As each of our endmembers is either an ionisation of an element or a series of a molecule, we want to use a block model:

$$\mathbf{x} = (x_{1_1}, \dots, x_{1_{d_1}}, \dots, x_{g_1}, \dots, x_{g_{d_g}})$$

We also assume that the majority of endmembers are not present, thus we consider a block sparse model.

A statistical inference method for solving such problems is the Block Sparse Bayesian Learning (BSBL) framework. Two algorithms of interest include BSBL-Bound Optimisation (BSBL-BO) [2] which maximises the posterior, and BSBL-Fast Marginalisation (BSBL-FM) [3] which maximises the marginal distribution of  $\mathbf{y}$ .

We look to exploit certain structures in our problem to improve accuracy and performance of the spectral unmixing.

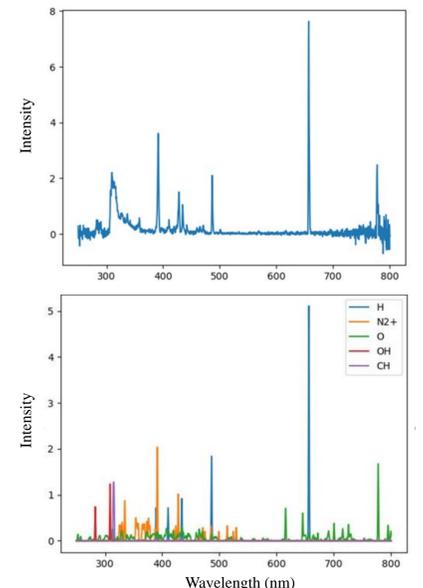


Figure 1: Top 5 endmembers contributions from spectral unmixing of air

## Non-Negative BSBL

In general, the abundancies of the endmembers are non-negative. For instance, we cannot have a negative volume of a certain gas. However, BSBL-BO and BSBL-FM return answers that are in the full real space. In industrial problems, the dictionary matrix is not a perfect representation of the spectral pattern of endmembers. This can lead to negative overfitting.

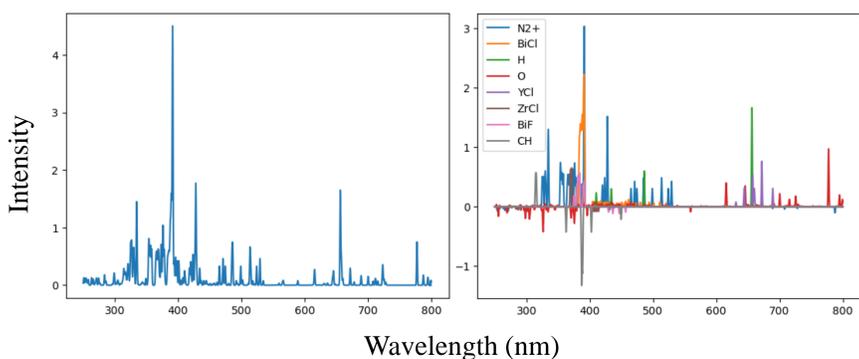


Figure 2: Left: Plasma Optical Emission Spectra of air. Right: Plot of the top 8 endmembers of air.

There are algorithms that solve for non-negative  $\mathbf{x}$ , such as Non-Negative Least Squares and NN-Block Orthogonal Matching Pursuit. However, these algorithms have certain drawbacks such as not considering a block structure or requiring a number of non-zero blocks being specified. We therefore adapt the BSBL framework to solve for non-negative  $\mathbf{x}$ .

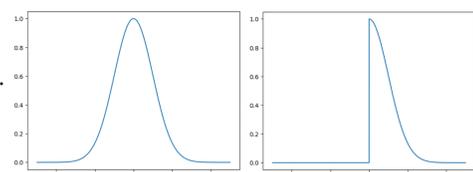


Figure 3: Gaussian distribution and rectified Gaussian Distribution with zero mean and unit covariance

We propose using Rectified Multivariate Gaussian Priors and Posteriors. The maximum likelihood and covariance of the posterior are calculated using the moment generating function. Our method gives more accurate results to BSBL-BO and BSBL-FM due to not overfitting the data. An example result is shown in figure 4.

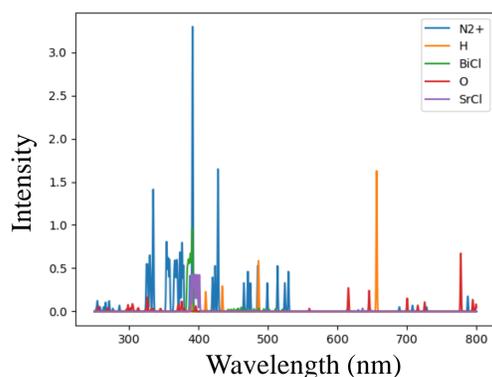


Figure 4: Contributions of the top 5 Endmembers of Air from NNBSBL.

## Graph Partitioned BSBL

We want to apply these statistical inference based spectral unmixing algorithms to real time data streams. These data streams update approximately every second. Both BSBL-FM and (especially) BSBL-BO are slower than desired for this practical application. We therefore need to speed them up (whilst minimising accuracy loss).

In many Spectral Unmixing applications, the matrix  $\mathbf{D}$  is sparse. Thus BSBL-BO and BSBL-FM are performing a lot of multiplications by zero. We wish to exploit this sparsity to speed up the algorithm. We therefore propose the Graph Partitioned-BSBL algorithm (GP-BSBL).

We use the method illustrated in Figure 5. This allows us to distribute the problem into  $w$  independent sub-problems. This guarantees speed up as the problem is solved in greater-than-linear time with respect to the size of the measurement.

The algorithm GP-BSBL-BO performs as fast as BSBL-FM with 4 or more partitions with  $M = 3000$ . Currently, this method is being applied to BSBL-FM, as this algorithm also has a greater than linear runtime with respect to the measurement vector size.

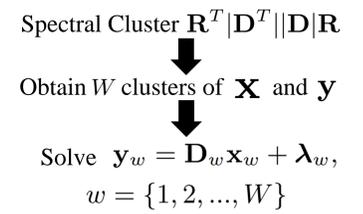


Figure 5: GP-BSBL pseudocode.

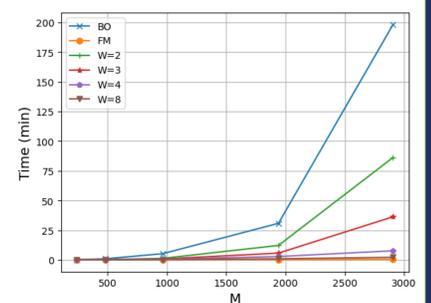


Figure 6: Time taken for BSBL-BO, BSBL-FM and GP-BSBL to converge.

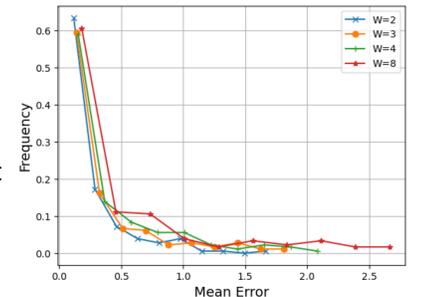


Figure 7: Mean Error of the GP-BSBL-BO algorithm.

## Conclusions & References

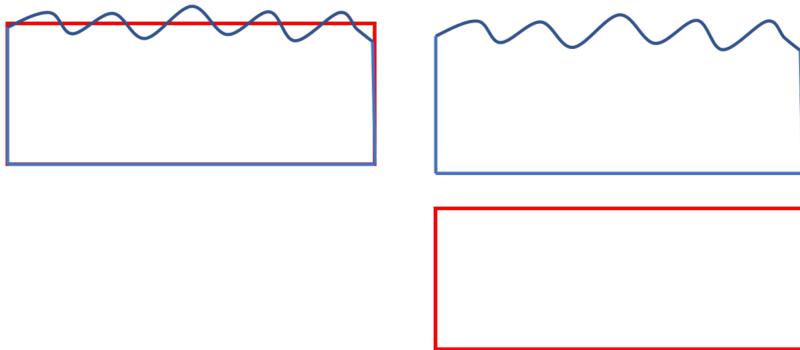
Here we present two methods for improving runtime and accuracy of Spectral Unmixing using a BSBL Framework. The first contribution, NNBSBL, prevents negative overfitting, whilst the second, GP-BSBL, reduces convergence time of the algorithm and allows for distributed implementation. The two algorithms can easily work together, though this has not been implemented yet.

1. Keshava, N. and Mustard, J.F., 2002. Spectral unmixing. *IEEE signal processing magazine*, 19(1), pp.44-57.
2. Zhang, Z. and Rao, B.D., 2013. Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation. *IEEE Transactions on Signal Processing*, 61(8), pp.2009-2015.
3. Liu, B., Zhang, Z., Fan, H. and Fu, Q., 2012. Fast marginalized block sparse bayesian learning algorithm. *arXiv preprint arXiv:1211.4909*.

## 1) Introduction

**Aim: continuously control the optimal input parameters to achieve a desired prespecified material thickness.**

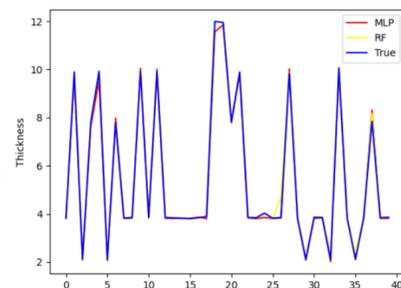
- Parameters determining the material thickness of certain industrial processes are often still adjusted by human experts.
- Unconsidered factors cause material thickness to vary, even when all controllable parameters are held constant.
- Different requirements of the customers result in various material thicknesses.



## (2) Surrogate Model

- Use real-world-data to train a surrogate model (MLP/ RF) to provide  $s'$

$$r_t = \frac{|MT_{previous} - MT_{target}| - |MT_{current} - MT_{target}|}{MT_{target}}$$



- At the end of each episode an additional reward of +1 is added if the MT is within the interval.
- Tolerance interval for target thickness is  $[MT_{target} - 0.2mm, MT_{target}]$ .

## (3) Model

$$s_t = (FH_t^j, FH_t^{j+1}, \dots, FH_t^J, MT_t^{current}) \in S$$

### Discrete

- DQN
- $a_t = \{\alpha_t^1, \alpha_t^2, \dots, \alpha_t^{J+J}\}$
- In each timestep, change one parameter by  $\zeta$

### Continuous

- PPO & DDPG
- $a_t = \{\alpha_t^1, \alpha_t^2, \dots, \alpha_t^J\}$
- In each timestep, change all parameters

### Note

$\zeta$  is a hyperparameter introduced in the discrete setting, determining the increase/decrease of the respective parameter.

## (4) Algorithm

### Algorithm 1 Material thickness control via PPO

```

Trained surrogate model S
Initialize replay memory D of capacity N
Initialize policy and value function parameters  $\theta_0, \phi_0$ , respectively
Desired target material thickness  $MT^{target}$  and lower and upper bound for acceptance interval I
Number of episodes N
Number of epochs K
1: for episode=1, ..., N do
2:   Initialize a random state  $s_0$ 
3:   for  $t = 0, 1, \dots, T$  do
4:     Sample action  $a_t \sim \pi(\cdot|s_t)$ 
5:     Execute action  $a_t$  in the environment
6:     Observe  $MT_t^{current}$  based on surrogate model S and consequently  $s_{t+1}$ 
7:     Observe  $r_t$  based on  $MT_t^{current}$ ,  $MT_{t-1}^{previous}$  and  $MT^{target}$ 
8:     Store trajectory  $\tau_t = \{s_t, a_t, r_t, s_{t+1}\}$  in D
9:     if  $MT_t^{current} \in I$  then
10:      break
11:   end if
12: end for
13: for epoch = 1, ..., K do
14:   for batch = 1, ..., BatchSize do
15:     Compute advantage estimates,  $\hat{A}_t$  based on the previous collected set of trajectories  $\{\tau_t\}$  and current value function  $V_{\phi_k}$ 
16:     Update policy parameters  $\theta_{k+1} \leftarrow \theta_k + \alpha \nabla_{\theta} \mathcal{L}$ 
17:     Update value function parameters  $\phi_{k+1} \leftarrow \phi_k - \alpha \nabla_{\phi} \mathcal{L}$ 
18:   end for
19: end for
20: end for

```

## (5) Validation

### Model validation

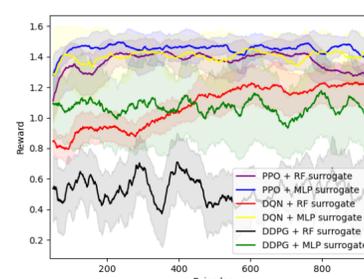
- Established an RL environment by training a surrogate model on real-world data to predict the material thickness given the input parameters.
- Train an agent by interacting with the surrogate model.
- Compare the model surrogate combinations based on the achieved cumulative reward.

### Surrogate model validation

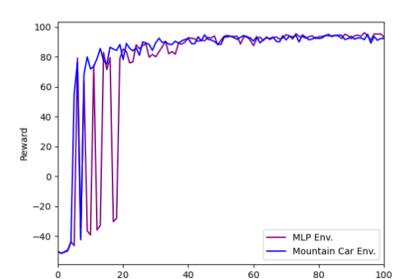
- Generate expert trajectories from a PPO agent acting in the mountain car environment
- Train a surrogate model based on expert trajectories.
- Train a second PPO agent by interacting with the surrogate model.
- Compare both PPO performance either trained in the real environment and the other train by interacting with the surrogate model.

## (6) Results

### Model validation



### Surrogate validation



# Data Driven Intelligence For Countering Crime

Panagiotis Pentaliotis (p.pentaliotis@liverpool.ac.uk)

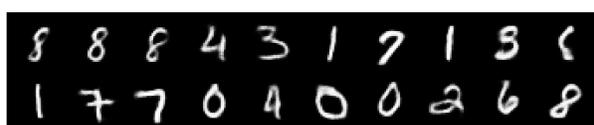
Prof Simon Maskell

Prof Paul Spirakis



## Random Projection

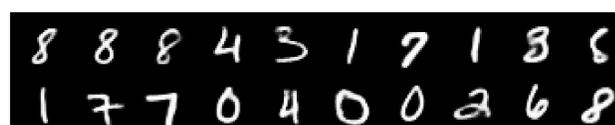
Random Projection is a dimensionality reduction technique. That more efficiently approximates a matrix than a diagonal. Especially when we require information from the off-diagonals. For a example a covariance's correlations.



Full rank covariance

## Random Projection Sampling

If we reverse the RP's intended application, we can project a smaller full-rank covariance to a larger dimensional space. Any stochastic algorithm that requires sampling from a Normal distribution can use RPS to sample from a full-rank covariance distribution.



Lower dimensional full rank covariance

## Auto-differentiable Ensemble Kalman Filters

Kalman filters assimilate data through Bayesian statistics and iterated observations. The AD-EnKF variant can capture non-linear relationships through Monte-Carlo simulations and optimise its prior from the data to produce estimations.

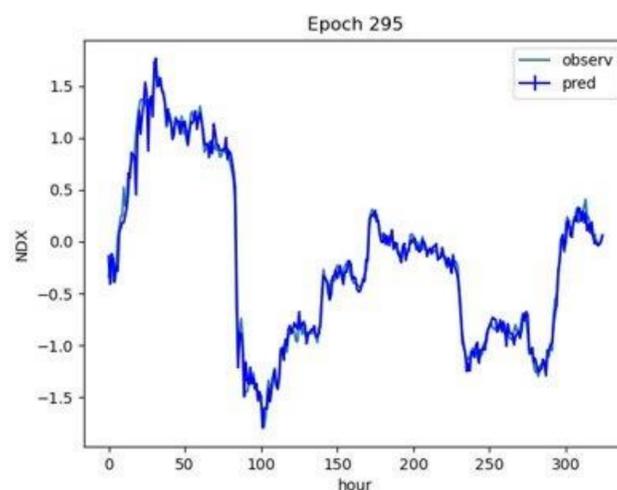
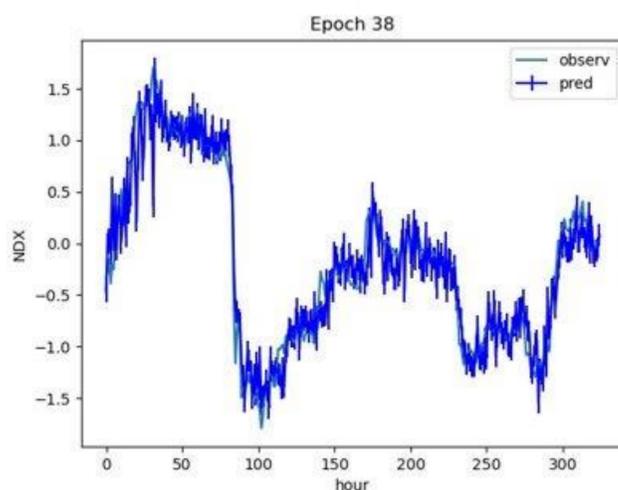
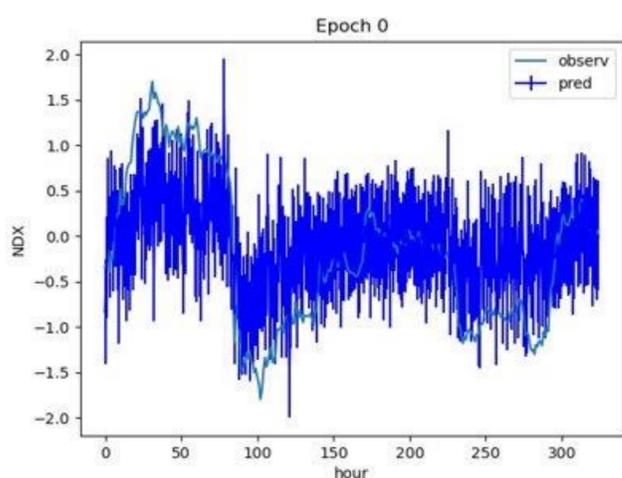
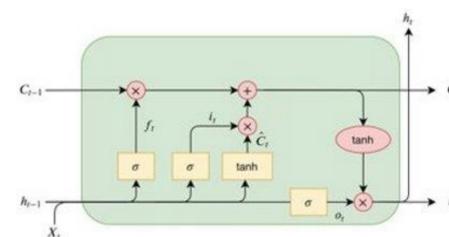
$$x_t = f(x_{t-1}) + w_t \quad w_t \sim N(0, Q)$$

$$y_t = h(x_t) + u_t \quad u_t \sim N(0, R)$$

## Auto-differentiable Ensemble Kalman Filters and Long-Short Term Memory

AD-EnKFs utilise LSTMs to propagate the Monte Carlo simulations through time and train the Neural Network model to assimilate the data better. Eliminating human bias given by the prior

$$x_t = f(x_{t-1}) + w_t$$



# Faster Uncertainty Quantification of Hydrocodes

Sarah M. Askevold ([S.M.Askevold@liverpool.ac.uk](mailto:S.M.Askevold@liverpool.ac.uk)),

Supervised by: Suzie Abbs ([scabbs@dstl.gov.uk](mailto:scabbs@dstl.gov.uk)), Veronica Bowman, ([vbowman@dstl.gov.uk](mailto:vbowman@dstl.gov.uk)),

Leszek Gasieniec ([L.A.Gasieniec@liverpool.ac.uk](mailto:L.A.Gasieniec@liverpool.ac.uk)), Brianna Heazlewood

([B.R.Heazlewood@liverpool.ac.uk](mailto:B.R.Heazlewood@liverpool.ac.uk)), Ubaid Qadri ([ubaid.qadri@stfc.ac.uk](mailto:ubaid.qadri@stfc.ac.uk)),



## Motivation and Summary

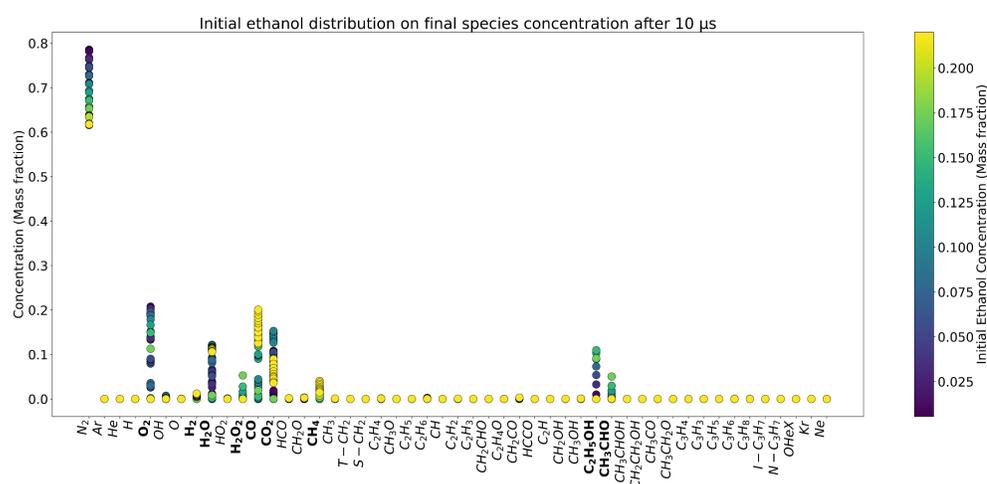
- To develop a single, integrated approach to analysing and speeding up a **hydrocode** that models blasts [1] and gain an understanding of the combined physics and chemistry processes involved, with a view to performing **uncertainty quantification**.
- Emulate a chemical reaction model using a **Gaussian Process** [2] or similar methods.
- Integrate new chemistry model with the **EDEN** hydrocode to increase simulation speed.
- Test **accuracy** and **parameter sensitivity** of hydrocode in simulation of **complex blasts**.

## Background

**EDEN** was developed by *Fluid Gravity Engineering Ltd* (FGE). It models impact scenarios with multiple materials and different phases, while also considering **afterburn**, as many explosives are rich in fuel and would potentially undergo secondary combustion. The hydrocode uses a **stoichiometric afterburn combustion model**, which is a type of model based on conservation of mass (so the reaction equations are balanced). Afterburning is significant, as it affects the temperature within the instigating fireball, which in turn affects the rate of change/ reactions in the scenario [3].

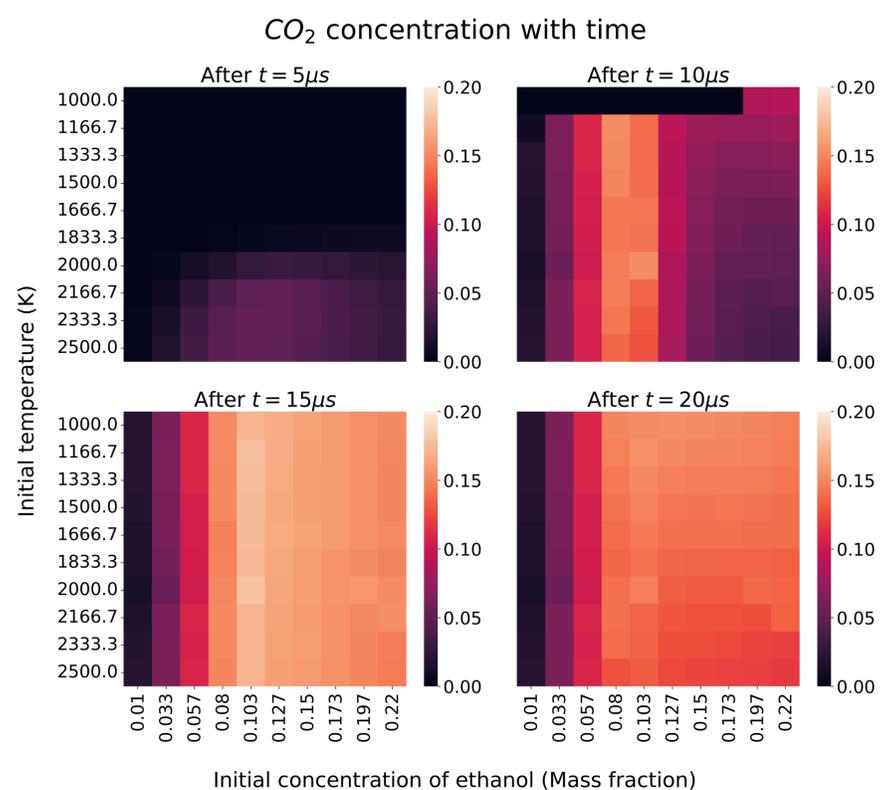
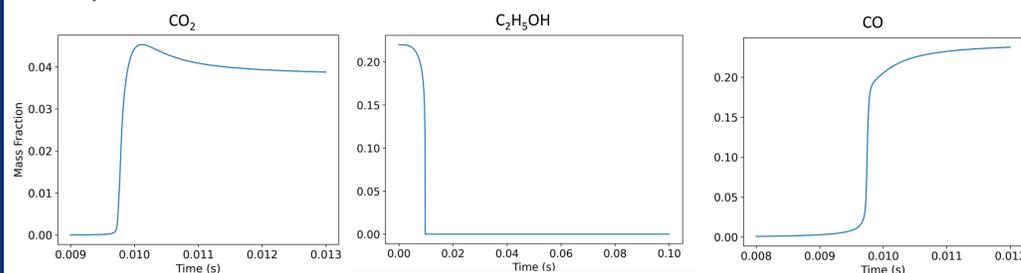
## Ongoing work

We are currently working on modelling Ethanol ( $C_2H_5OH$ ) reactions [4] in the python package *Cantera*, to identify relationships between initial conditions and output. These relationships could be used to emulate said package, and replace its function in the running of the hydrocode.



**Above:** Product distribution after  $10\mu s$ , based on 100 runs with varying initial conditions, grouped by initial concentration of ethanol.

**Below:** Concentration of  $CO_2$ ,  $C_2H_5OH$  and  $CO$  with time, based on an initial temperature of 1000 K and 22% initial concentration of ethanol.



We have generated and studied tables of species concentration with time, given initial concentration of ethanol =  $[0.01, 0.22]$  + air, and temperature =  $[1000 \text{ K}, 2500 \text{ K}]$ . The above mesh plots shows concentrations of  $CO_2$  after times  $[1, 4] \times 5 \times 10^{-6} \text{ s}$ . Other time steps were also studied (but are out of the scope of this poster).

## Future Work

- Apply a Gaussian Process to generated data to identify relationships between initial conditions and output.
- Expand parameters studied to include variations of initial pressure, as well as different time steps.
- Introduce mixing from the hydrocode to the data, and identify how that affects the chemistry modelling.

## Conclusion

This project will assess the sensitivity of the hydrocode model to input parameters and test if it is a good model for reality.

We will attempt to increase the speed of the code, without too much degradation in accuracy, so that it can be used in a more time-sensitive context.

Initial attempts at fitting a Gaussian process proved difficult due to the broad parameter space, but further exploration of the data could suggest a reduced training set on a more manageable sub-space.

## References

- [1]: David J Benson. "Computational methods in Lagrangian and Eulerian hydrocodes". en. In: *Comput. Methods Appl. Mech. Eng.* 99.2-3 (Sept. 1992), pp. 235–394.
- [2]: Robert B Gramacy and Herbert K H Lee. "Bayesian treed Gaussian process models with an application to computer modeling". In: *J. Am. Stat. Assoc.* 103.483 (Sept. 2008), pp. 1119–1130.
- [3]: Alec M Milne, Scott B Cargill, and Aaron W Longbottom, "Modelling of complex blast". In: *International Journal of Protective Structures* 7.3 (Aug. 2016), pp. 325–339. DOI: 10.1177/2041419616661431.
- [4]: Carsten Olm et al. "Development of an ethanol combustion mechanism based on a hierarchical optimization approach". en. In: *Int. J. Chem. Kinet.* 48.8 (Aug. 2016), pp. 423–441.



# Algorithms and Decision-Making Processes in Distributed Attacker-Defender Games

Tim Prokopenko, supervised by Igor Potapov, Jason Ralph, Gareth Rees

**Introduction:** The project is motivated by beyond-visual-range air combat scenarios and maps them into simpler/abstracted ‘canonical’ reachability problems in the language of multiplayer computational games in discrete and continuous space.

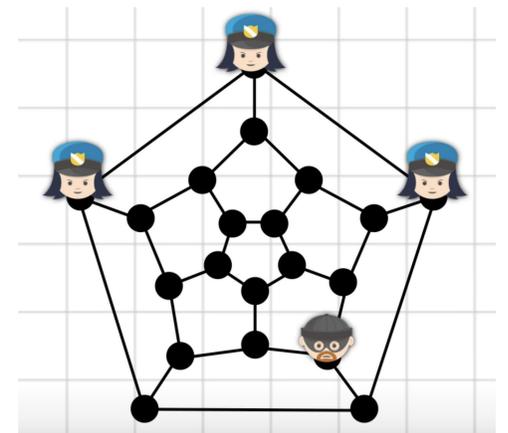
We consider several variants of **Cops and Robbers Games** for developing TCS techniques and solving strategic optimisation and combinatorial problems in different discrete arenas and geometric environments. The solution to these problems in a complex, *uncertain* and dynamic situation in real-time is a challenging computational task and may require developing approximation algorithms, applying machine learning techniques and distributed/decentralised high-performance computing methods.

## Our objectives are:

1. To map missions from the air combat domain into abstract computational games
2. To develop algorithms that can represent complex air-combat situations
3. To develop TCS techniques to overcome computational constraints

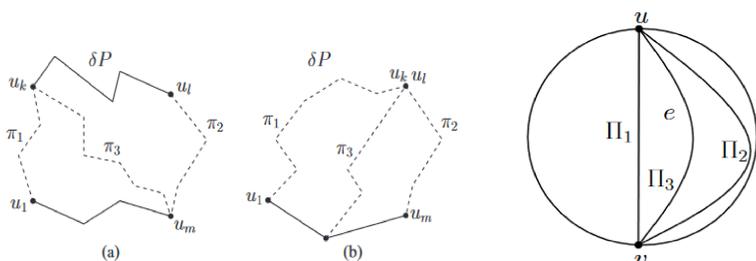
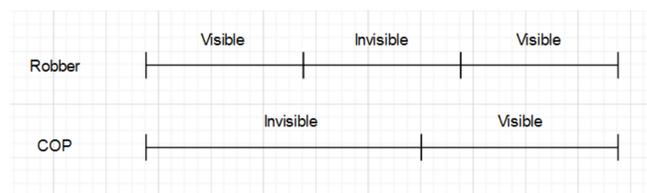
## COPS and ROBBERS game:

Two teams are involved in the game. A team of cops and a team of one or more robbers. At the initial stage, cops are placed in an arena (e.g. in the vertices of the graph  $G$ ) and robbers are placed in places (vertices) which are free of cops. Then the game is played alternately starting with the cop team. Each of the cops and the robber moves at a fixed distance in continuous space or along the edges of graph  $G$ . The cops win the game if for a finite number of steps one of the team can occupy the same vertex as the robber.



## Invisibility of Cops and Robbers

can change the complexity of the game and strategies[1]. At the initial stage we consider several solutions with full visibility and study how missing visibility can affect the solution of the problem.



## State-of-the-Art (zero vs full visibility)

We are interested in the interconnection between two parameters the cop number and the time of catching a robber in case of full or partial invisibility[1]. There are several effective solutions exist in case of full visibility[2]. Strategies for invisible robbers have been studied in very simple environments like trees and partial invisibility in more complex environments requires new solutions.

References:

1. Dereniowski, Dariusz, et al. "Zero-visibility cops and robber and the pathwidth of a graph." *Journal of Combinatorial Optimization* 29 (2015): 541-564.
2. Bhaduria, Deepak, et al. "Capturing an evader in polygonal environments with obstacles: The full visibility case." *The International Journal of Robotics Research* 31.10 (2012): 1176-1189.



Engineering and Physical Sciences Research Council



# Data Science and Artificial Intelligence for smart sustainable plastic packaging

William Jeffcott, Supervised by Prof. Vitaliy Kurlin (UoL), Dr. Tom McDonald (UoL), Dr. Sam Chong (UoL)

EPSRC Centre for Doctoral Training in Distributed Algorithms, University of Liverpool, Liverpool, UK

## Background & Aims

The project has started by looking at flexible packaging. Flexible packaging is made by building up layers of polyolefin films which are created by extrusion blow moulding. At the bottom a polyethylene (PE) layer is in contact with the product. There are two more PE layers above, followed by an adhesive layer and a print layer (see image below).

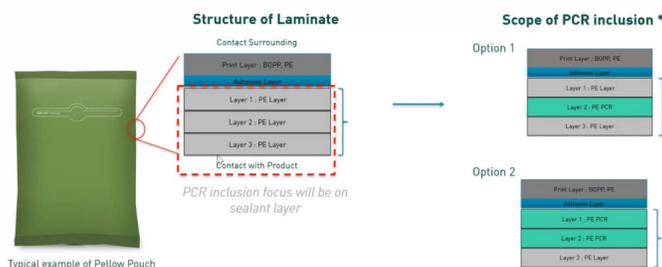


Fig 1: An example of a typical flexible pouch made by industrial partner Unilever.

Incorporation of 100% recycled plastic (PCR) is a challenge, because it can't be in contact with the product for safety reasons. We are therefore restricted to a maximum PCR content of ~80% (print layer + middle layers). At present most flexibles are ~30% PCR, so there is large room for improvement.

**So why can't we go 80% PCR straight away?** In general, PCR products contain contaminants from waste streams, become degraded by oxidation due to thermal processing and have a wide range of molecular weights present.

These factors can significantly lower the performance of the film for a given application and are normally indicated by the presence of gel spots.



Fig 2: Examples of some of the typical types of gel spots seen in PE films made from recycled plastics [1].

Many of these gel spots relate to specific contaminants. For example, fibre gels commonly indicate the presence of polypropylene (PP) or nylon contaminants. When stretched, the film is more likely to tear at these specific points.

### Overarching aims of the project:

- Characterise the type, size, shape and spatial distribution of gel spots for a range of different PE films with different PCR content/PCR feedstocks.
- Predictively link the gel distributions to performance properties of the films (sealability, likelihood to tear etc).
- Use the original PCR pellets to see whether the distribution of gel spots could be predicted from indirect structural information (e.g FTIR/DSC/XRD).

## Gel Identification and Classification

A batch of films with varying PCR content from different feedstocks has been provided by Unilever. The task is to accurately identify the **size, location and type of gels on a given film in a time efficient manner.**

We can get high quality images of the film using a simple transmission scanner. This allows us to identify thousands of gels for a given sample within minutes (traditional microscopy techniques are much slower), which leads to more concrete statistical conclusions.

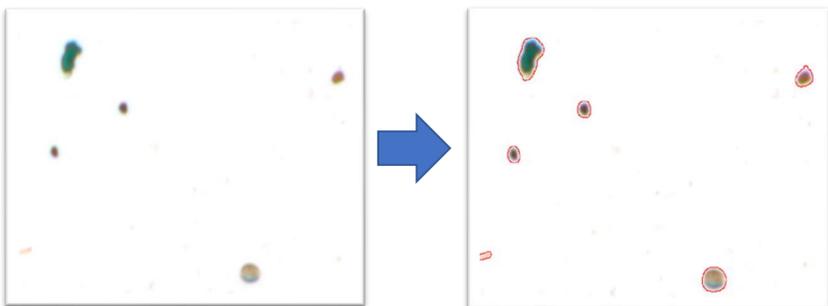


Fig 3: A small section of flexible polyethylene film imaged using a high resolution film scanner, the more prominent gels are circled in red (right image).

We identify the boundaries of the gels on a given film automatically using specifically adapted image thresholding techniques from computer vision. This allows us to see key statistics such as **gel frequency, size and area of film occupied.**



## Gel Distributions

By utilising features of the identified gel spots such as length to width ratio, colour differences, centre of mass position etc. we can classify the gels into one of three types which are guided by industry: **black point gels, transparent point gels and fibre gels.**

These gel types indirectly indicate the chemical contaminants present in the film, allowing manufacturers to liaise with recycling companies to have them removed from their waste stream.

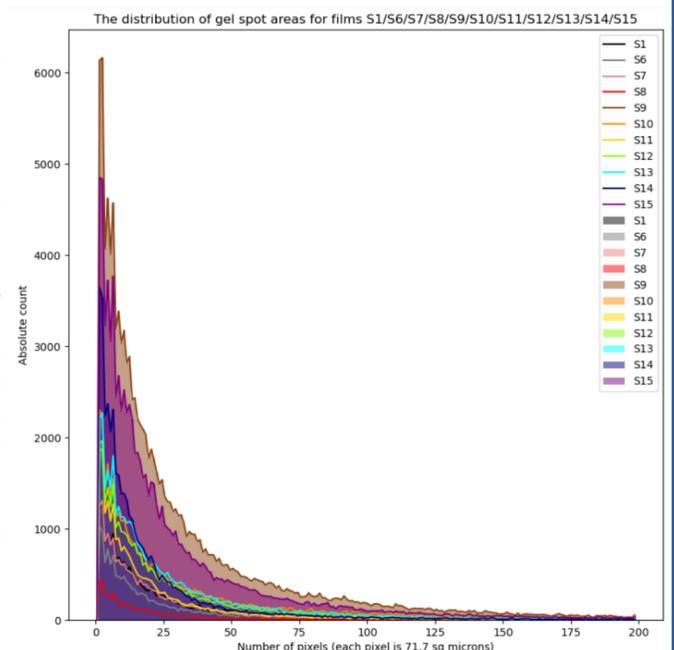


Fig 4 (above): A histogram showing the distribution of gel spot areas for different flexible films.

Fig 5 (left): Some examples of the different gel classifications.

## Future Work

- **Work on specific spot type identification:** Imaging allows us to classify the shape and distribution of the gels, but at present it does not allow us to understand the precise chemical cause of gels. By using a chemical technique like Raman microscopy, we can build a machine learning model to link visual information to chemical information.
- **Test the mechanical properties of the films:** A key part of this project is to understand the link between the presence of gel spots and the degradation in packaging performance. To do this we need to test the mechanical properties in a robust way.

## Conclusion

This project will explore whether we can accurately quantify the link between the inclusion of recycled materials and the resulting gel spot distributions. We will then attempt to link this gel spot distribution to both the performance properties of the film, and the structural properties of the original plastic pellet.

A good outcome would be to have the ability to decide whether a recycled plastic specimen will be appropriate in advance of making packaging from it, thus saving time/money. This will also reduce plastic waste in the environment, because choosing the ideal recycled pellets will allow us to maximise PCR inclusion in the product.

## References

[1] 'Gels in polyethylene films' information sheet - ExxonMobil

# Improving Passive SONAR Detection & Tracking using Machine Learning

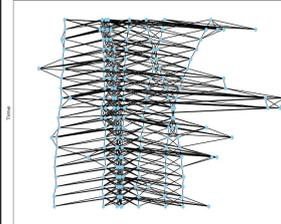


William Shaw (W.Shaw@liverpool.ac.uk), Dr Murat Uney (M.Uney@liverpool.ac.uk),  
 Dr Daniel Colquitt (mf0u60af@liverpool.ac.uk), Dr Cerys Jones (Cerys.Jones@ultra-  
 css.com)

## Introduction

In passive SONAR, tracking quiet transient targets becomes challenging due to non-linear target movements, fluctuating noise levels, and false alarms, especially amidst cluttered environments with multiple overlapping contacts. This PhD project harnesses the capabilities of advanced machine learning – including Bayesian and graph neural networks (GNN) – aiming to enhance detection and tracking algorithms in this demanding scenario.

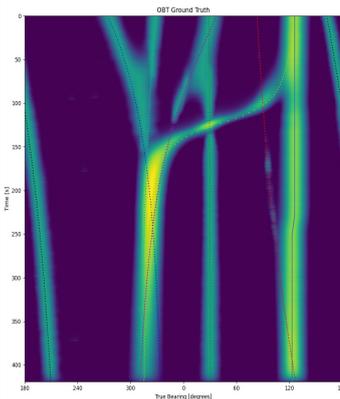
## Framing Tracking as a Graphical Problem



Nodes contain information such as bearing and time from the detections, while the edges represent the differences in that information.

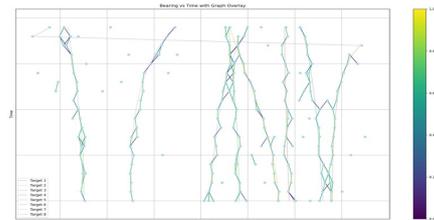
Detections are transformed into nodes and connected to adjacent detections to create a graph. A GNN then analyses this graph to discern the spatial and temporal patterns of the detections.

## Passive SONAR Data



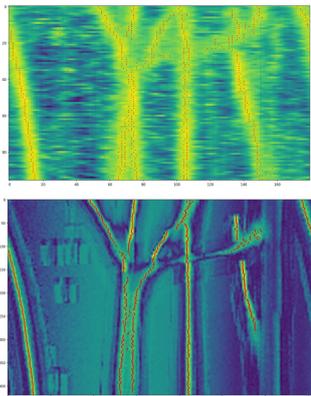
Waterfall plot of a passive SONAR simulation, highlighting targets with varying acoustic energy emissions and dynamics, the simulation also includes energy from the ownship.

## Graph Detection to Detection Data Association



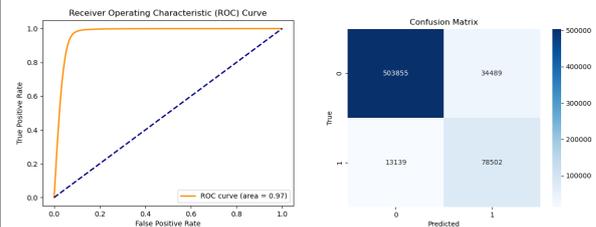
A GNN classifies edges between detections using higher-dimensional embeddings to differentiate between false alarms, non-same target, and same-target associations. However, it faces challenges with high bearing rate targets.

## Leveraging Uncertainty for Improved Detections



For (a) time-based detections and (b) sample-based detections, the model utilises probability and confidence from bearing/time bins to pinpoint areas with a high likelihood of target presence.

## GNN Classification Performance



The GNN excels with simulated detection data, though the dataset is imbalanced. Its performance declines when targets near the sensor exhibit high bearing rate tracks. Future efforts focus on iterative GNN enhancements and track formation.