Reverse engineering atmospheric dust content from engine samples

A report for the ITT20 Rolls-Royce projects

(Students) **Amin Sabir**; Matt Evans, Daniel Hajnal, Bill Nunn
(Staff) Sergey Dolgov, Matt Nunes, Theresa Smith

---

1 Introduction

Rolls-Royce Holdings is a major aerospace and defence company that provides essential products and services to military and corporate aircraft companies worldwide. Not to be confused with Rolls-Royce Motor Cars, which is a subsidiary of BMW, Rolls-Royce Holdings (RR) also manufactures jet engines used in aeroplanes for over 300 commercial airlines across 80 different countries. Their Engine Environmental Protection (EEP) team deals with how these jet engines are affected by the environment over time, for instance through the weather and atmospheric dust conditions. The team has contracts with its customers for the maintenance and service of these engines.

During the 20$^{\text{th}}$ integrative think tank (ITT), RR proposed challenges about how environmental factors, in particular mineral dust (e.g. dirt, sand), have impacted their jet engines. The EEP team would like to charge maintenance costs to their customers more accurately, based on the use of their jet engines. Some of the challenges are outlined below:

1. **Reduce uncertainty in weather data affecting dust intake for engine performance -** the Copernicus Atmospheric Monitoring Service (CAMS) reanalysis dataset [Inness et al., 2019] uses the European Centre for Medium-Range Weather Forecasts (ECMWF) dataset to handle the atmospheric dust quantities such as chemicals and aerosols. These quantities are crucial for how much dust is affecting jet engines over time, for instance, the average dust intake for engines along a particular flight path. However, there is usually uncertainty across the quantities and it is hard to accurately predict how much on average a jet engine would intake.

2. **Reverse engineering dust and particulate content from engine samples -** over several flights, an engine will have visited numerous locations, including multiple origin and destination airports. Each airport has its local mineral dust composition (dust types), which will accumulate in an engine over time. Given the internal engine dust samples, can we approximate an airport's dust composition?

3. **Mathematical models for dust accretion and shedding in a jet engine -** during flights, dust particles are ingested into the engines [Ryder et al., 2023]. Depending on their mineral dust composition, these particles can melt and/or build up on interior surfaces such as the engine blades, affecting the engine's aerodynamics and if scattered across, can damage the engine components (see Figure 1). Can mathematical models be used to describe the dust accretion and shedding process?

In this report, we propose the use of numerical and statistical methods, involving the construction of a linear system of equations and solving this with the least-squares method and Bayesian statistics, to tackle challenge 2. First, we discuss the background of challenge 2 and how we consider the challenge mathematically (Sections 2 and 3). Next, we discuss the methods in detail and implement such methods on an engine dataset with preliminary results (Sections 3 and 4). We then conclude with future directions (Section 5) and how these methods can link to the challenges 1 and 3, alongside other proposals that were put forward during this ITT (Section 6).

2 Background and problem formulation

Before introducing our methodology and approaches, we start with the background and focus of challenge 2. This challenge aims to investigate how different types of dust accumulated across airports affect an engine over time. Note that dust is ingested also during the flight paths high above the ground, but we are mainly concerned with dust ingestion at the airports where the engines are travelling to and from [Ryder et al., 2023].

Throughout a jet engine's operating life it will encounter a range of airports, which have different amounts and types of dust levels (see Figure 1). The dust is ingested across these airports with some types of dust being more damaging than others [Clarkson, 2021]. At departing airports, when the aircraft is taxiing (i.e. moving along the runway) and taking off, the engines are often in a high-power phase and are more likely to ingest dust from the airport's surface than when the aircraft is stationary. For the landing airports, in general, less dust is ingested than at the take-off airport but its dust still contributes to the engine's overall condition.

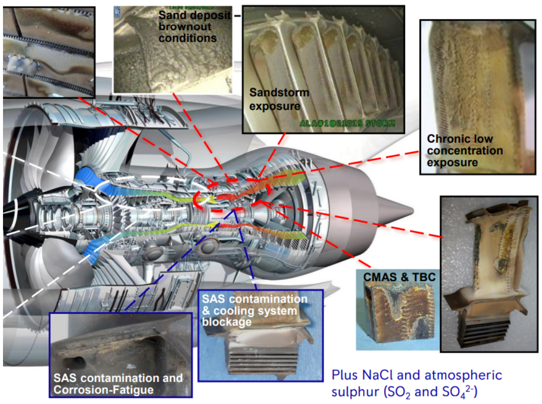

Figure 1: Inside of a jet engine filled with different dust that damage engine components over time [Clarkson, 2021]

Rolls-Royce wishes to construct a location-specific dust composition database for all airports their engines fly to and from. This way they can analyse how more damaging dust at certain airports affects their engines, charging more to airlines that frequently use these airports that have high concentrations of the damaging dust types.

Work has been started by their EEP team and the University of Manchester [Clarkson, 2021] to construct a database of atmospheric agent characteristics (DAAC) which starts to tackle the challenge. The goal here is to highlight the dust composition at a ground surface level but further extensions are required. These include variations of dust composition due to altitude and other factors such as chemical reactions of the mineral dust within a jet engine. This is proving extremely difficult to model and include in their current work. Instead, we can consider a simpler setup of the challenge.

Rolls-Royce provided us with a dataset consisting of 20 real decommissioned jet engines with all their respective flight paths and details, including the individual flights with their departure and arrival airports. Additionally, we had synthetic data on the different dust-type concentrations present for each of these engines when decommissioned. Note that Rolls-Royce is currently obtaining true values for these engines, so synthetic values for the different dust-type concentrations in the engines will be used. To simplify the task, the dust types (composites) were split into **five** categories to make up the full dust concentration, consisting of four main mineral oxides [Clarkson, 2021]: Calcium oxide (**CaO, C**), magnesium oxide (**MgO, M**), aluminium oxide ($\mathbf{Al_2O_3}$**, A**), silicon oxide ($\mathbf{SiO_2}$**, S**) and other dust minerals (**O**). Note that, we do not consider any chemical reactions or mixtures between these chemical compounds that may occur within an engine. There were synthetic ground truth values for the dust-type concentrations present at 298 airports. The aim was to construct a reliable method to estimate the five different dust-type concentrations for 298 airports, which the engines have flown to and from.

We propose the use of a linear system of equations for this dataset and problem, each describing the known dust compositions sample in an engine as a function of dust compositions taken from the airports the engine has been to. We draw from the field of inverse problems and intend to use a generalised inverse via least-squares to solve the system and try to extend it with Bayesian statistics. Primarily, we will work on this 20-engine dataset. We hope these methods can then be extended to a more realistic scenario to find more complex dust composite concentrations at airports.

3 Methodology and approaches

In this section, first we introduce the main mathematical setup, with the dataset mentioned in Section 2, alongside a forward model to form a linear system of equations. Then we discuss two approaches in trying to solve the system to obtain the different dust-type concentrations at the airports.

As mentioned in Section 2, we have this 20-engine dataset, with over 5800 flights taken across the engines, and would like to investigate the five different dust composites and their respective concentrations for 298 airports. Analysing the data, we found on average each engine takes 2000 - 3000 flights during its operating life. Also, a high portion of these flights are between certain airports. For example, engine 1 travelled from London Heathrow 1506 times out of its 3057 flights (49% of total flights, see Figure 2). This tended to be the case with other engines with many having a main airport flying from, indicating an airport base for that engine (e.g. airport-specific base used by an airline, see Figure 3). Also, the number of airports visited by engines frequently was much lower than the total number of airports in the dataset.

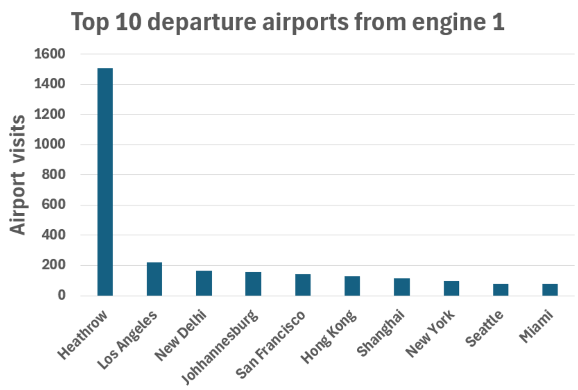

Figure 2: Top 10 departure airports for engine 2 (3057 total flights)

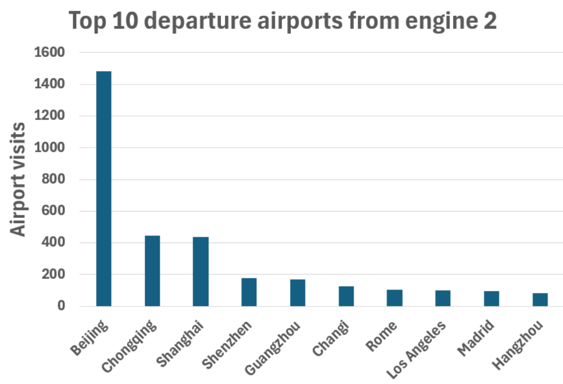

Figure 3: Top 10 departure airports for engine 2 (3696 total flights)

We can set up the problem as a linear system of equations involving vectors and matrices. The formulation of linear systems using linear algebra has been beneficial in a variety of applications of mathematics [Shafarevich and Remizov, 2012]. Often it is hard to solve systems analytically, so numerical and iterative methods are used to estimate solutions [Greenbaum, 1997]. In our case, we consider the matrix system of equations as

$$AX = Y \tag{3.1}$$

where X has dimensions $(N_\mathrm{a} \times 5)$ representing the five dust-type concentrations present at N_a airports and Y has dimensions $(N_\mathrm{e} \times 5)$ representing the five dust-type concentrations in N_e engines. We denote each row vector of X and Y by $\boldsymbol{x}_i$ and $\boldsymbol{y}_j$, where both are vectors with dimensions (1×5) for the concentration of dust types for airport $i = 1, 2, \ldots, N_\mathrm{a}$ and engine $j = 1, 2, \ldots, N_\mathrm{e}$ respectively. Note that the sum of the elements of $\boldsymbol{x}_i$ and $\boldsymbol{y}_j$ are 1 representing the full atmospheric dust content (overall dust concentration) at the airports and engines respectively. The matrix A with dimensions $(N_\mathrm{e} \times N_\mathrm{a})$ takes $X \to Y$, transforming the airport to the engine dust-type concentrations. We wish to specify a suitable A as our forward model for (3.1). Then we can solve the inverse problem of going from $Y \to X$ (Section 3.2).

From the dataset $N_\mathrm{a} = 298$ and $N_\mathrm{e} = 20$, so we currently have a severely under-determined system where we have 298 unknowns and only 20 equations. Note that, if a solution to an under-determined system exists then there exists an infinite number of solutions [Donoho et al., 2006], which is a major problem in our current setup. This brings the motivation to tackle this using inverse problems techniques. as we will see in Section 3.2.

3.1 Forward model

Before investigating ways to solve (3.1), we construct a relevant forward model to represent A. First, we assume that at each airport the overall dust composite concentration stays the same over time. In reality, this is not the case with changing levels of dust composites found across the world due to weather seasons and events such as sandstorms throughout the year [Bojdo et al., 2020]. For now, we consider X as time independent and further extensions to implement seasonality with time dependence are discussed in Section 5.

We now break down how each flight from a given engine will contribute to the overall dust concentration in that engine. Consider a given flight i taken by an engine, the mass accumulated in that engine is

$$\underbrace{m_{i,out} \cdot \boldsymbol{x}_{i,out}}_{\text{take-off dust}} + \underbrace{m_{i,in} \cdot \boldsymbol{x}_{i,in}}_{\text{arrival dust}} \tag{3.2}$$

where $m_{i,out}$ and $\boldsymbol{x}_{i,out}$ are the mass of dust and different dust-type concentrations taken from the departing airport respectively. Similarly, $m_{i,in}$ and $\boldsymbol{x}_{i,in}$ represent the same quantities but for the arrival airport. As noted in Section 2, we are only concerned with dust ingestion at airports and not at high altitudes.

We assume that the amount of dust accumulated from the first flight does not affect the dust accumulation in the later flights. Therefore, we extend (3.2) to sum over all the flights taken by engine j over its operating life and consider the different dust-type concentrations

$$\boldsymbol{y}_j = \frac{\sum_{i=1}^{N_\mathrm{j}} (m_{i,out} \cdot \boldsymbol{x}_{i,out} + m_{i,in} \cdot \boldsymbol{x}_{i,in})}{\sum_{i=1}^{N_\mathrm{j}} (m_{i,out} + m_{i,in})} \tag{3.3}$$

where N_j is the total flights taken by engine j. Here (3.3) is similar to a weighted average and is constructed in this way because Rolls-Royce does not measure the dust type masses accumulated in their engines, instead they measure the relative concentrations of the dust types. The equation (3.3) essentially sums all the different types of dust mass intake for all of the flights taken by engine j divided by the total dust mass intake by engine j.

We simplified (3.3) by assuming the mass contribution of the arrival airport dust $m_{i,in}$ was negligible compared to the take off airport dust mass $m_{i,out}$, as mentioned in Section 2. Also, we assumed that the mass intake $m_{i,out}$ was known and $m_{i,out} = 1$ across all airports. Mass consideration links to challenge 1 with the CAMS dataset and we will discuss how this could be implemented within our model in Section 5.

Using these additional assumptions, (3.3) is simplified to

$$\boldsymbol{y}_j = \frac{\sum_{a=1}^{N_\mathrm{a}} n_{a,j} \boldsymbol{x}_a}{\sum_{a=1}^{N_\mathrm{a}} n_{a,j}} \tag{3.4}$$

where $n_{a,j}$ is the number visits from engine j to airport a and $\boldsymbol{x}_a$ is the dust-type concentrations at airport a. This formulation (3.4) only considers the departure airports of flights taken by engine j, $j = 1, \ldots, N_\mathrm{e}$ across a airports, $a = 1, 2, \ldots, N_\mathrm{a}$ for the contribution of $\boldsymbol{y}_j$. Unlike (3.3), there is no indexing over the total number of flights N_j of engine j for (3.4).

To get into the equivalent form of (3.1), we have $X = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{N_\mathrm{a}}], Y = [\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{N_\mathrm{e}}]$ and

$$A = \underbrace{\begin{bmatrix} n_{1,1} & \cdots & n_{1,N_\mathrm{e}} \\ \vdots & \ddots & \vdots \\ n_{N_\mathrm{a},1} & \cdots & n_{N_\mathrm{a},N_\mathrm{e}} \end{bmatrix}}_{\text{airport frequency matrix}} \otimes \underbrace{\boldsymbol{I}_5}_{\text{dust interaction}} \tag{3.5}$$

where $n_{i,j}$ is the number of visits to airport i for engine j as a fraction of the total number of airports visited for engine j. Indexing i and j to N_a and N_e respectively forms the airport frequency block matrix, where each $n_{i,j}$ has dimensions (5×5) to set up for the five dust types. Also, $\boldsymbol{I}_5$ is the (5×5) identity matrix which models the interaction between these five dust types ($\mathbf{C}, \mathbf{M}, \mathbf{A}, \mathbf{S}, \mathbf{O}$ from Section 2). Note that $\otimes$ is the *Kronecker product* between the matrices. The identity matrix could be adapted to represent chemical reactions between these five dust types and is discussed in Sections 5 and 6.

We see in Figure 4 an example of A for $N_\mathrm{e} = N_\mathrm{a} = 19$ with the stacked elements. Each 5×5 small block represents the fraction contribution of an airport to a particular engine and the colour bar on the right of Figure 4 measures the contribution. This particular A is further discussed in Section 3.2.

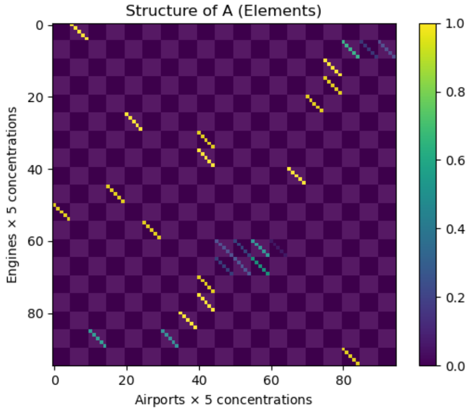

Figure 4: Elements of A stacked with the 5 dust-type concentrations for 19 airports and 19 engines

3.2 Standard inverse problems approach: least-squares method with a generalised inverse

We now try to solve (3.1) using A from (3.5). Note, as before $N_\mathrm{a} = 298$ and $N_\mathrm{e} = 20$, we have an under-determined linear system. The method of least-squares using a generalised inverse (Moore-Penrose inverse, [Prasad and Bapat, 1992]) of A as $A^\dagger$, can be used to give an approximate solution for the dust-type concentration X at all of the airports. This least-squares method is usually a convenient starting point when trying to solve such a system [Donoho et al., 2006].

Looking at (3.1), the least-squares method minimises

$$\min_X \left\{ \|AX - Y\|^2 \right\} \tag{3.6}$$

under any valid matrix norm $\|\cdot\|$, in our case we use the Frobenius norm. The generalised inverse $A^\dagger$ gives a solution to (3.6) as $X^\dagger$

$$X^\dagger = A^\dagger Y. \tag{3.7}$$

Depending on the structure of the problem setup, this solution $X^\dagger$ can be a reasonable approximation of the ground truth X.

However in our case, due to the large disparity between N_e and N_a with an under-determined system, the least-squares method would yield inaccurate estimates of the different dust composite concentrations at all the airports. Here A would be severely ill-conditioned. So, we have adjusted the problem to make it easier to accurately predict airports that are visited very frequently by engines over trying to predict all 298 airports. To do this, we only considered airports in the dataset that are visited at least 350 times across all the engines. Also, we found that one engine in the dataset did significantly fewer flights compared to the rest of the engines, so we discarded this engine in our mathematical setup. This results in using $N_\mathrm{e} = 19$ engines to predict the dust-type concentrations at $N_\mathrm{a} = 19$ airports. Therefore we now have a square matrix A with dimensions (19×19) and a fully determined system to solve. We see our A in Figure 4, which is very sparse.

In Section 4.1, the numerical simulations and results of applying $A^\dagger$ are outlined.

3.3 Bayesian approach: sampling distributions (MCMC)

Another approach to consider is to use a Bayesian setup for (3.1). This allows us to predict a range of common values (i.e. a probability distribution) for the different dust-type concentrations of $\boldsymbol{x}_i$ rather than just a singular point estimate for $\boldsymbol{x}_i$ when using the least-squares method (Section 3.2). It could be seen as an improvement of the generalised inverse solution, where if one could not solve the linear system (3.1) analytically, the Bayesian setup can give a reasonable approximation for the ground truth X. Also, the Bayesian setup can be flexible to incorporate spatial dependence between nearby airports and help infer different dust composite concentrations that link between these airports. This is discussed further in Section 5.

Using the linear system formulation (3.1) and forward model (3.5), we use Bayes' rule for $\boldsymbol{x}_i$ and $\boldsymbol{y}_j$

$$\underbrace{\mathbb{P}(\boldsymbol{x}_i|\boldsymbol{y}_j)}_{\text{posterior}} \propto \underbrace{\mathbb{P}(\boldsymbol{y}_j|\boldsymbol{x}_i)}_{\text{likelihood}} \times \underbrace{\mathbb{P}(\boldsymbol{x}_i)}_{\text{prior}} \tag{3.8}$$

where we seek the posterior distribution $\mathbb{P}(\boldsymbol{x}_i|\boldsymbol{y}_j)$ to give us the overall dust composite concentrations at airport i given the overall dust composite concentrations in engine j. This posterior is complex and difficult to find exactly, due to the nature of our problem setup, so we use sampling techniques. We opted to use Markov chain Monte Carlo (MCMC) which is a widely used sampling method in Bayesian statistics [Mira, 2005]. MCMC samples from the posterior distribution and incorporates the likelihood and prior distributions from (3.8) to converge to the true posterior distribution that we are seeking.

To model the likelihood distribution for $\boldsymbol{y}_j|\boldsymbol{x}_i$, we use a multivariate normal distribution

$$\boldsymbol{y}_j|\boldsymbol{x}_i \sim \mathcal{N}(A\boldsymbol{x}_i, \sigma^2\boldsymbol{I}_5) \tag{3.9}$$

where σ would model the noise in the measurements of the relative dust concentrations among dust types across the engines and $\boldsymbol{I}_5$ is the (5×5) identity matrix. For the prior distribution, as each dust concentration is between 0 and 1 and the whole dust composite concentrations sum to 1, we use a Dirichlet distribution (a multivariate version of the beta distribution)

$$\boldsymbol{x}_i \sim \textbf{Dirichlet}(\boldsymbol{\alpha}_i). \tag{3.10}$$

The parameter $\boldsymbol{\alpha}_i$ for an airport i is a (1×5) dimensional vector used to represent the five different dust-type characteristics unique to that airport i. We consider $\boldsymbol{\alpha}_i$ to be independent across all airports, but to incorporate spatial dependence between nearby airports, certain $\boldsymbol{\alpha}_i$ can be linked together. Note $\boldsymbol{\alpha}_i$ is considered as a hyperprior on $\boldsymbol{x}_i$, since we model $\boldsymbol{\alpha}_i$ as part of a continuous uniform distribution

$$\boldsymbol{\alpha}_i \sim \textbf{Uniform}([0,1]^5). \tag{3.11}$$

An interesting point to note is how best to determine what the realistic bounds and estimates are for the hyperpriors of $\boldsymbol{\alpha}_i$, to represent a realistic model for the different dust composite concentrations at airports. We touch on this in the future work and discussion sections (Sections 5 and 6).

For the simulations in Section 4.2, we focus on a toy problem separate from the 20-engine dataset that we have discussed mainly in this report. This is to trial the Bayesian approach and the hope is that we can upscale it to the main dataset with more realistic and complex setups. The toy setup (see Figure 5) consists of three airports, each with two dust types where their dust concentrations ($X = [\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3]$ with dimensions (3×2)) would be estimated by the Bayesian and MCMC method setup. We also consider two jet engines with their concentrations of the two dust types ($Y = [\boldsymbol{y}_1, \boldsymbol{y}_2]$ with dimensions (2×2)). These engines travel between different pairs of airports but both travel to and from airport 2.

Figure 5: Toy setup with the true concentrations of dust types $(1, 2)$ for the airports (in grey) and the two jet engines (in **blue** and **red**)

4 Preliminary results

In this section, we discuss the preliminary results of the two approaches and see numerical simulations for estimating the concentration of dust types at different airports. The main objective is to show how these approaches are valid and promising when predicting the different concentrations of dust types that help give valuable information to RR. All simulations were done in Python and the code is available via Appendix A.

4.1 Least-squares method

The setup with our A (Figure 4) and considering 19 engines across 19 airports yielded some interesting preliminary results (Figure 6). Note, in addition to using the least-squares method with the generalised inverse (Section 3.2), we constrained the predictions for $\boldsymbol{x}_i$ to be between 0 and 1 and the sum of each row elements for $\boldsymbol{x}_i$ to equal 1, to ensure the full dust concentration is accounted for. Extensions for solving the least-squares minimisation problem (3.6) with Tikhonov (L2) regularisation were explored but only showed marginal improvements. Other setups could be considered as discussed in future work (Section 5).

In Figure 6 we see among the different airports, with their respective airport codes, that the absolute error $\|X^\dagger - X\|$ was fairly low across most of the five dust categories. In particular, for **M** (MgO) the absolute error was generally the lowest of the five dust type categories, with errors in its dust concentration around 1 - 3% from its true value (see MgO's error EPDF in Figure 6). The empirical probability density functions (EPDFs in Figure 6) of the errors in the four main dust types (excluding other minerals, **O**) generally had a normal distribution-like structure apart from a few exceptions. This normal distribution structure would allow us to model the errors, across the dust types, for our predictions $X^\dagger$ accurately to account for any small discrepancies we may have when solving the linear system (3.1).

However, there were some big discrepancies for **C** (CaO) and **S** (SiO_2) with some of their absolute errors out by around 20%. Upon closer inspection, we found that for all the dust types, especially **S**, at airport RJCC (New Chitose, Japan) only one of the engines visited this airport more than 350 times. This engine visited many other airports as well but there was no overlapping data of airport RJCC among the other engines. This meant due to our method setup (Section 3.2) with A, the knowledge of the dust types at RJCC could only be taken from that one engine and was therefore skewed massively giving inaccurate predictions. This issue and other factors are addressed in future work (Section 5).

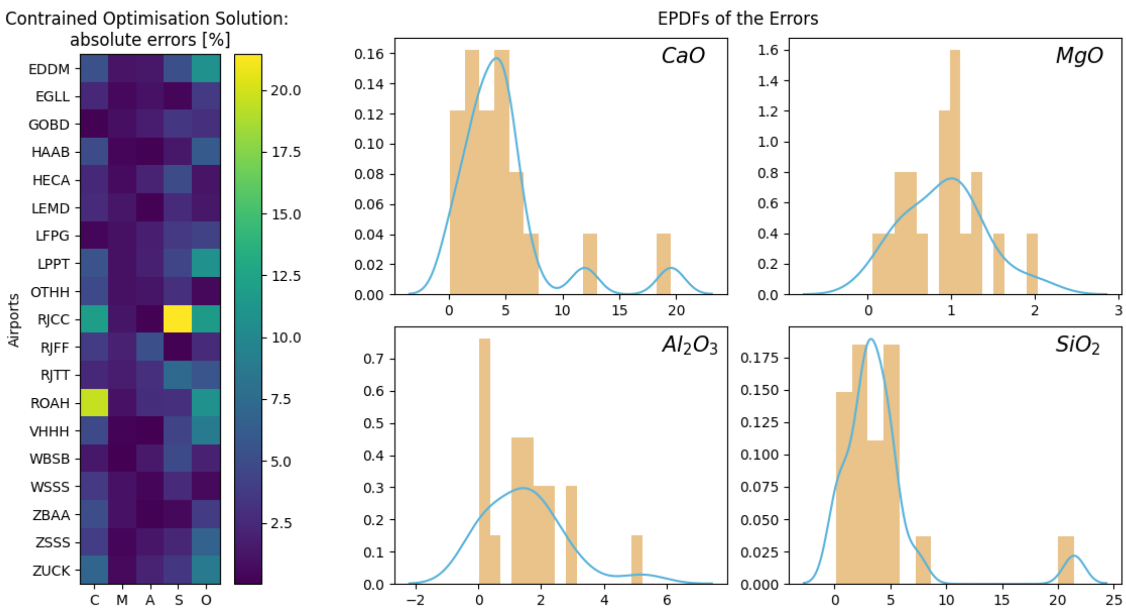

Figure 6: **Left** - absolute error $\|X^\dagger - X\|$ for the five dust types across 19 airports for 19 engines
Right - empirical probability density functions (EPDFs) for the absolute error of the 4 main dust types

4.2 Bayesian approach on the three airports toy problem

We considered the toy setup (Figure 5) and applied the Bayesian with MCMC (Section 3.3) to the problem. We looked at two different scenarios. One in predicting the concentration of the two dust types across all three airports and the other, given the first airport's concentrations of dust types are known, predicting the concentrations of dust types for the other two airports. Both scenarios sampled the posterior distributions, via MCMC, 1000 times (for more details see Appendix A).

For the first scenario, we see the results in Figure 7. The estimated posterior PDFs among the three airports varied compared to the true dust-type concentration values. For airport 1, the peaks of the posterior were centred mostly at the true dust-type concentration values $[0.3, 0.7]$ but the range of potential values was large. For airports 2 and 3, the posterior PDFs also contain the true dust-type concentrations ($[0.4, 0.6]$, $[0.1, 0.9]$ for airports 2 and 3 respectively), but they were both heavily skewed to underestimate or overestimate the true concentrations of the two dust types. We could not accurately predict the true dust-type concentration values for all three airports. This shows us that more information is needed to make more informed and accurate estimates.

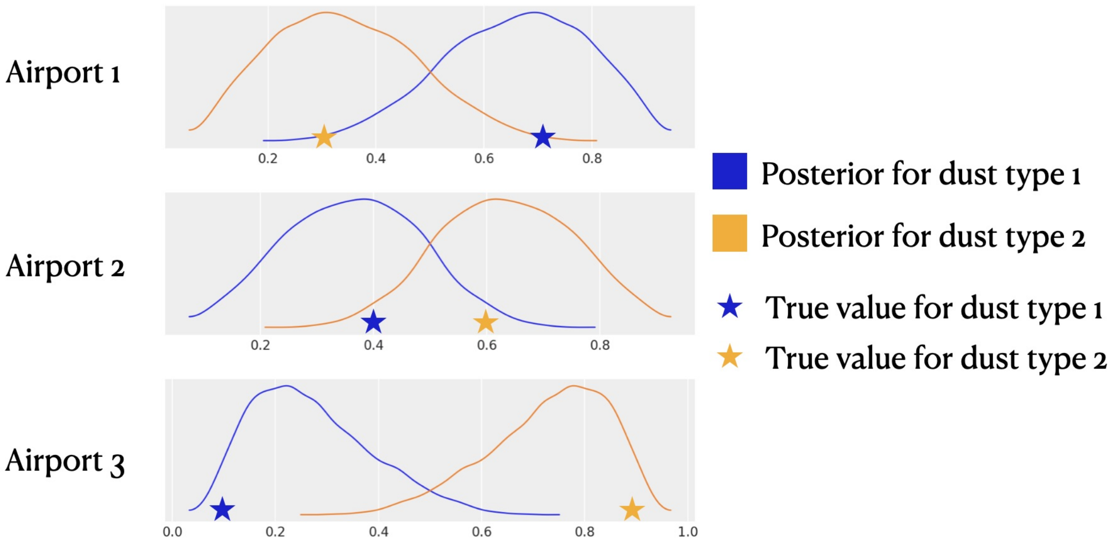

Figure 7: MCMC prediction for the posterior distributions for the concentrations of two dust types at the three airports compared to their true values

This motivated the second scenario with airport 1's true information now known about its dust-type concentrations. We see the results in Figure 8, with much more informative and accurate estimates for the dust type concentrations at airports 2 and 3. We see that like in Figure 7, the true concentrations for these airports are included in the posterior, but this time the distributions in Figure 8 are centred around these true values and give a much more certain estimate.

From this, we can infer the dust-type concentrations at the airports, given some known information at another airport, with a high degree of accuracy in this Bayesian with MCMC framework. How this would work in a more realistic scenario with many more engines and airports is an interesting question and is discussed in Sections 5 and 6.

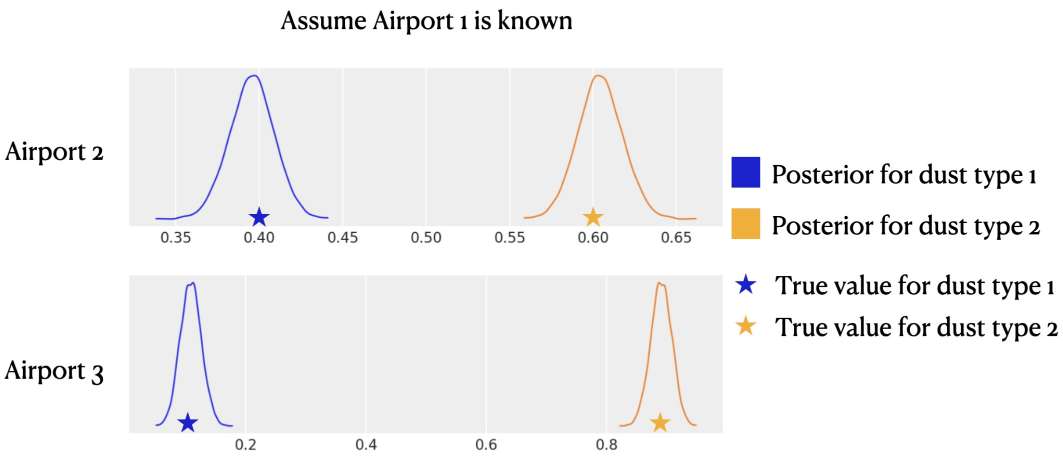

Figure 8: MCMC prediction for the posterior distributions for the concentrations of two dust types at airports 2 and 3 given airport 1 is known, compared to their true values

5 Future work

In this report, we have considered the use of linear systems of equations and Bayesian statistics in tackling challenge 2. These have largely depended on constructing a valid forward model for the problem (3.1) - (3.5), solving by least-squares via the generalised inverse of A (3.6) and using Bayesian statistics (3.8) - (3.11). We mainly worked on the 20-engine dataset that Rolls-Royce provided us with but also used a toy setup and dataset for the Bayesian method. There have been some promising preliminary results and opportunities to move forward.

5.1 Objectives and plans

1. Test the methods on real dust-type concentration data for the engines and airports. Refine and link the methods.

Rolls-Royce, in collaboration with the University of Manchester, is in the process of obtaining true values for dust-type concentrations for over 70 decommissioned engines with some knowledge of the dust-type concentrations at a few airports. Once this data is available, we can test out the least-squares method on this and see how well the method will predict the dust-type concentrations across these airports.

By having more data across more engines and airports, the least-squares method could be extended to predict for more airports (compared to the 19 airports in Section 4.1), as long as there is sufficient overlap between the engines and the airports they have visited. However, the effective performance of the method is not a guarantee, since A could be so severely ill-conditioned for (3.1), that using $A^{\dagger}$ from (3.7), produces inadequate predictions. So other more sophisticated techniques for inverse problems should be investigated. These include using iterative methods [Greenbaum, 1997] and different regularisation techniques such as variational methods, which have proven to be very useful for many inverse problems [Benning and Burger, 2018].

When formulating the Bayesian method in Section 3.3, we mentioned how the least-squares method via the generalised inverse may not be possible in certain situations and that the Bayesian method can help give a more reasonable approximation for X. To help link the two methods, we could therefore extend the linear system (3.1) as a complete Bayesian inverse problem [Matthies et al., 2016], which could understand the uncertainty in the predictions for the dust-type concentrations across the airports. For instance, the least-squares method could give suitable initial guesses for prior parameters such as the Dirichlet parameters $\boldsymbol{\alpha}_i$ (3.10) - (3.11) that could then speed up the sampling for the Bayesian MCMC and its predictions.

2: Extend forward model with mass considerations, linking to challenge 1, and temporal dependence (seasonality).

Many assumptions were taken for the forward model (3.4) - (3.5), including constant mass intake across the airports and time-independence over the dust-type concentrations at airports and in engines. We can extend our forward model to relax these assumptions.

For the mass consideration, this links closely to challenge 1 with the CAMS dataset [Inness et al., 2019], although that challenge was more concerned with the mass dust intake across the whole flight path and not just at the airports. However, challenge 1 was not worked on during the ITT. We could study the CAMS dataset and see how the mass intake varies for typical flights at different airports and incorporate it in our forward model (3.5). We would model the mass intake as a statistical random variable in (3.3), with the random variable calibrated to replicate the conditions based at each airport. Rolls-Royce mentioned that the mass dust intake variation from different airports was small [Ryder et al., 2023], but modelling the mass dust intake as a random variable can account for these variations. This can make the forward model more reliable over time and add elements of statistical modelling. Whereas a completely deterministic forward model is more complicated and harder to account for these mass variations.

To include temporal dependence (seasonality) in the forward model for the dust-type concentrations of airports, we could weigh the flights for an airport, that an engine takes over its operating life. One option is to use exponential decay in time series [De Silva et al., 2010]. We could consider flights taken from an airport early in an engine's life to be less important than flights taken from that airport later, for that engine. The overall dust-type concentration for that airport would change over time. We would have a modified $\boldsymbol{x}_a$ in (3.4) as $\tilde{\boldsymbol{x}}_{\boldsymbol{a}}$

$$\tilde{\boldsymbol{x}}_{\boldsymbol{a}} = \sum_{t=1}^{T} \beta^{T-t} \boldsymbol{x}_{a,t} \tag{5.1}$$

where T is the total number of flights taken to airport a over time by an engine j, with $t = 1$ being the first flight taken and $t = T$ being the most recent flight taken at airport a. The time series of flights for an airport and engine are encoded in T. The dust-type concentrations at t for airport a are denoted by $\boldsymbol{x}_{a,t}$. The β parameter is bounded, $0 < \beta < 1$ and could be estimated in various ways, including a similar Bayesian approach we used to estimate the dust-type concentrations at the airports. This follows the time series framework [De Silva et al., 2010] and could extend to incorporate seasonality in weather patterns across airports. For example, severe sandstorms or higher levels of dust-type concentrations at different times of year [de Villiers and Heerden, 2007, Bojdo et al., 2020].

Note that, similar temporal considerations can be taken to the dust accumulated in an engine over time and we could simulate how some dust may leave the engine from an airport, long after its flight there.

3. Upscale Bayesian model to the main dataset and decide on which dust-type concentrations at airports to empirically measure. Include spatial dependence between nearby airports.

An immediate step is to upscale the Bayesian model from the toy setup (Figure 5) to the 20-engine dataset that the least-squares method used and then hopefully to the real dataset as outlined in objective 1. The Bayesian setup (Section 3.3) is flexible to extend to these setups but requires careful consideration when computing the dust-type concentrations across the airports. As we saw in the two results (Figures 7 and 8), having prior dust-type concentration information at some airports helps in the overall predictions. So, having empirical measurements for dust-type concentrations at airports that best reduce the uncertainty in the Bayesian posterior distribution estimates is crucial.

Spatial dependence can be included in the Bayesian model with the pooling of Dirichlet parameters (3.10) based on the geographical location of nearby airports [Jo et al., 2021]. For example, among airports across London their corresponding Dirichlet parameters can be coupled via different length scales [Gelman and Pardoe, 2006]. This would reduce the uncertainty in dust-type concentration estimates and help give more accurate information for Rolls-Royce to use. Other knowledge-based constraints can be applied, such as when there is minimal concentration for a certain dust type at an airport, giving useful information that can then be used in the Dirichlet parameters for the hyperpriors (3.11).

6 Discussion

Overall, this report has shed some light on the use of linear systems of equations along with Bayesian statistics to help accurately predict the dust-type contents at airports from decommissioned jet engine dust samples. First, through the investigation of the problem background (Section 2). Then, we proposed a mathematical setup of the problem alongside two methods (Section 3) with some promising preliminary results (Section 4).

There are a variety of interesting next steps outlined in Section 5. These are not without difficulties. For instance, using real data for the proposed methods could be more complex. There would be more dust types and chemical reactions between them to consider, compared to the five mineral oxide dust types that our methods used. Careful consideration will be required to see how the new dust mineral dust types with reactions can be fully captured by the proposed methods (e.g. updating the dust interaction matrix in (3.5)). Additionally, it will be necessary to determine how easy it is to map between the more complex dust types to the simpler ones the methods currently use. There is potential for interdisciplinary projects or workshops between academics and engineers at Rolls-Royce. Academics could be on-site and test and implement extensions of the proposed methods. The validity and limitations of the methods can be found and discussions between the groups can help accelerate the research.

For extending the forward model, challenge 1 with the CAMS dataset may be very difficult to implement well without having a more complex forward model than the one we currently have. This could be very time-consuming and may only achieve marginal gains compared to the constant mass considerations we assumed before. Also, it is tough to model the correct uncertainty in the model, but developing a base model as a benchmark can give a good indication of how Rolls-Royce can more accurately charge for maintenance and servicing contracts of their engines.

A major concern for the upscaling of the Bayesian approach is the computational challenges in implementing the MCMC sampling to model realistic scenarios. High-performance computing (HPC) systems and further refinements of the approach can be taken through discussions among academics including with mineral dust geologists and engineers at Rolls-Royce. These discussions can give expert opinions providing valuable information to help model realistic priors for the Bayesian model (Section 3.3). Similarly, getting all major airports to install real-time dust-type monitoring sensors is a huge logistical task and is likely to be very expensive. Instead, strategically placing sensors among certain airports that are high dust contributors to the engines and suffer from large changes in dust-type concentrations (e.g. New Delhi, India) can optimise the Bayesian approach while staying within budget. Ultimately, if the Bayesian approach is too costly to upscale, it can still give beneficial knowledge on dust-type concentrations for a subset of airports from a small group of engines.

Finally, we can link to challenge 3, where here we were more concerned with developing a robust mathematical model for dust accretion and shedding inside a jet engine across flights. The mathematical modelling group at the ITT considered both discrete and continuum partial differential equations (PDEs) models with viscoelastic and thin-film fluid approximations. Our proposed methods in helping to find the dust-type concentrations at airports could help the modelling group incorporate what type of dust to include in their PDEs. For instance, how different levels of certain dust types present at airports could affect the overall dust accretion and shedding in the engine over time (e.g more MgO than SiO_2 present could cause more/less shedding).

In conclusion, with our mathematical formulation and proposed methods there are ample opportunities to help tackle the reverse engineering dust present at airports from engine dust samples challenge. The hope is these methods can give a better indication and sufficient evidence to Rolls-Royce, particularly their EEP team, on how to more accurately charge for engine maintenance and service contracts to their customers.

References

[Benning and Burger, 2018] Benning, M. and Burger, M. (2018). Modern regularization methods for inverse problems. *Acta numerica*, 27:1–111.

[Bojdo et al., 2020] Bojdo, N., Filippone, A., Parkes, B., and Clarkson, R. (2020). Aircraft engine dust ingestion following sand storms. *Aerospace Science and Technology*, 106:106072.

[Clarkson, 2021] Clarkson, R. (2021). Investigation of different dust composites: typical mineral dust in the atmosphere and jet engines. *Rolls-Royce, University of Manchester*.

[De Silva et al., 2010] De Silva, A., Hyndman, R. J., and Snyder, R. (2010). The vector innovations structural time series framework: a simple approach to multivariate forecasting. *Statistical Modelling*, 10(4):353–374.

[de Villiers and Heerden, 2007] de Villiers, M. P. and Heerden, J. v. (2007). Dust storms and dust at Abu Dhabi international airport. *Weather*, 62(12):339–343.

[Donoho et al., 2006] Donoho, D., Kakavand, H., and Mammen, J. (2006). The simplest solution to an underdetermined system of linear equations. In *2006 IEEE International Symposium on Information Theory*, pages 1924–1928. IEEE.

[Gelman and Pardoe, 2006] Gelman, A. and Pardoe, I. (2006). Bayesian measures of explained variance and pooling in multilevel (hierarchical) models. *Technometrics*, 48(2):241–251.

[Greenbaum, 1997] Greenbaum, A. (1997). *Iterative methods for solving linear systems*. SIAM.

[Inness et al., 2019] Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., Dominguez, J. J., Engelen, R., Eskes, H., Flemming, J., et al. (2019). The CAMS reanalysis of atmospheric composition. *Atmospheric Chemistry and Physics*, 19(6):3515–3556.

[Jo et al., 2021] Jo, A., Nandram, B., et al. (2021). Bayesian pooling for analyzing categorical data from small areas. *Survey Methodology*, 47(1):191–214.

[Matthies et al., 2016] Matthies, H. G., Zander, E., Rosić, B. V., Litvinenko, A., and Pajonk, O. (2016). Inverse problems in a bayesian setting. *Computational Methods for Solids and Fluids: Multiscale Analysis, Probability Aspects and Model Reduction*, pages 245–286.

[Mira, 2005] Mira, A. (2005). MCMC methods to estimate Bayesian parametric models. *Handbook of statistics*, 25:415–436.

[Prasad and Bapat, 1992] Prasad, K. M. and Bapat, R. (1992). The generalized Moore-Penrose inverse. *Linear Algebra and its Applications*, 165:59–69.

[Ryder et al., 2023] Ryder, C. L., Bézier, C., Dacre, H. F., Clarkson, R., Amiridis, V., Marinou, E., Proestakis, E., Kipling, Z., Benedetti, A., Parrington, M., et al. (2023). Aircraft engine dust ingestion at global airports. *EGUsphere*, 2023:1–34.

[Shafarevich and Remizov, 2012] Shafarevich, I. R. and Remizov, A. O. (2012). *Linear algebra and geometry*. Springer Science & Business Media.

A Code availability and description

The code folder for this report is available in `https://github.com/hajdan42/ITT20_Airportdust` called `ITT20_Airportdust` that reproduces all the plots and results in this report, including the Python files with the construction of methods. In the folder are a `README.md` and a `dependencies.yml`, which outline what this code folder entails and a guide to download the required Python libraries. The main Python files used for the Section 4 were `limit_data.py` (Section 4.1), `flight_toy_two_one.ipynb` and `flight_toy_two_two.ipynb` (Section 4.2).