



RESEARCH ARTICLE

# Towards data assimilation of ship induced aerosol-cloud interactions

Lekha Patel<sup>1</sup> \* and Lyndsay Shand<sup>1</sup>

<sup>1</sup>Sandia National Laboratories, PO Box 5800 MS 0829, Albuquerque, New Mexico, 87185, United States

\*Corresponding author. Email: [lpatel@sandia.gov](mailto:lpatel@sandia.gov)

**Received:** 18 January 2022

**Keywords:** Atmospheric modeling, aerosol-cloud interactions, data assimilation, stochastic simulation-based learning.

## Abstract

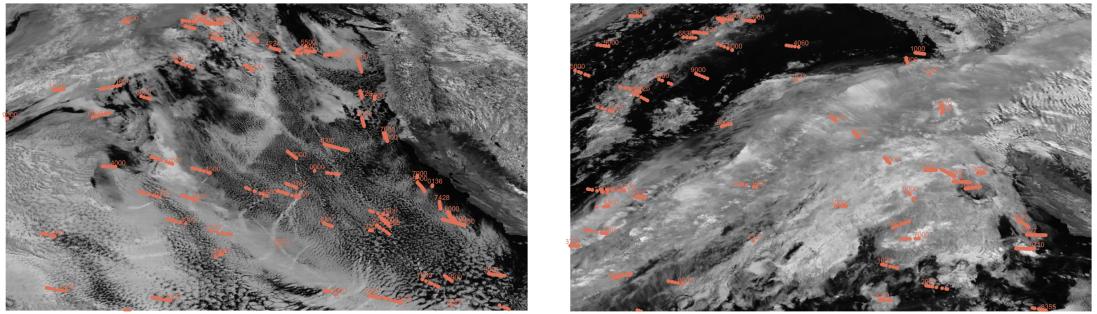
Satellite imagery can detect temporary cloud trails or *ship tracks* formed from aerosols emitted from large ships traversing our oceans, a phenomenon that global climate models cannot directly reproduce. Ship tracks are observable examples of marine cloud brightening, a potential solar climate intervention that shows promise in helping combat climate change. In this paper, we demonstrate a simulation-based approach in learning the behavior of ship tracks based upon a novel stochastic emulation mechanism. Our method uses wind fields to determine the movement of aerosol-cloud tracks, and uses a stochastic partial differential equation (SPDE) to model their persistence behavior. This SPDE incorporates both a drift and diffusion term which describes the movement of aerosol particles via wind and their diffusivity through the atmosphere, respectively. We first present our proposed approach with examples using simulated wind fields and ship paths. We then successfully demonstrate our tool by applying Approximate Bayesian Computation method - Sequential Monte Carlo (ABC-SMC) for data assimilation.

## Impact Statement

The impact of aerosol injections as potential solar climate intervention strategies is poorly understood due to largely unobserved aerosol-cloud interactions. One of few observable examples are temporary cloud trails produced by ship emitted aerosols. This work focuses on mathematically modeling satellite observations of ship induced aerosol injections and leveraging it to learn underlying parameters characterizing their behavior.

## 1. Introduction

For decades, satellite imagery has been able to detect *ship tracks*, temporary cloud trails created via cloud seeding by emitted aerosols of large ships traversing our oceans. Ship tracks are of interest because they are unintentional and observable examples of marine cloud brightening, a potential solar climate intervention (e.g. [Latham, 1990](#); [Council, 2015](#); [Gunnar et al., 2013](#)). Ship tracks portray the ability of anthropogenic aerosols to perturb boundary layer clouds enough to alter the albedo of the atmosphere, usually brightening the surrounding clouds ([Twomey Effect, Twomey et al., 1966](#); [Diamond et al., 2020](#)), and thus contribute to indirect radiative forcing ([Capaldo et al., 1999](#); [Eyring et al., 2010](#)). This phenomenon has become more frequently observed as satellite technology has significantly improved since ship tracks were first observed by [Conover \(1966\)](#) and [Twomey et al. \(1966\)](#). Using the



**Figure 1.** Visible ship tracks (left) on April 12, 2019 compared with no visible tracks (right) on April 7, 2019 with 3 hours of known ship locations (shown in red). Images ( $5000\text{km} \times 3000\text{km}$ ) taken at 12:00 GMT with ABI spectral band C06 off the coast of California.

recently deployed GOES-R geostationary satellite series<sup>1</sup>, tracks have been observed in the atmosphere throughout the year, sometimes persisting more than 24 hours before mixing back into the atmosphere.

Although ship tracks have been actively studied since the 1960s, indirect radiative forcing, amongst other differences to cloud properties, contribute to the largest sources of uncertainty regarding overall radiative forcing in climate models (Carslaw et al., 2013). Current understanding of the specific cloud effects from aerosol injections come from physical simulations under pristine conditions, not representing reality. In climate simulations of this phenomenon, aerosol injections are initiated by the user at a known location in fully defined environments (Wang et al., 2011; Berner et al., 2015; Possner et al., 2018; Blossey et al., 2018). Satellite-observed tracks, however, form in complex dynamic environments that are challenging and expensive to replicate in physical simulations.

Inferring track behavior from observations comes with many challenges. Track visibility and persistence are highly dependent on atmospheric conditions (Possner et al., 2018) and inconsistent data availability can also cause tracks to be poorly observed. Due to complex atmosphere dynamics along with variations in fuel types and concentrations, not all emission bursts will produce visible tracks (see Figure 1). Interruptions in the visibility of an existing ship track can occur when ships pass under different atmospheric conditions. Further, vertical transport of the aerosols between the ship's smoke stack and the boundary clouds is largely unknown and unobserved. The exact altitude of the boundary clouds in which ship tracks are visible and the time lag between an aerosol burst released from a ship, largely depends on complex weather and cloud dynamics. While the height can be approximated using satellite retrieval or reanalysis products, the time lag is difficult to infer from satellite images whose spatial resolution is greater than a kilometer. To the naked eye, new ship track observations appear in images directly above known ship locations due to the imaging resolution. Hence, it is reasonable to assume that the vertical transport path from ship to boundary layer occurs at the same latitude/longitude. We thus implicitly impose a known time lag between ship emissions and their first detection.

In this work, we present a computationally efficient, statistical approach to emulating and learning the observed formation and behavior of ship tracks, based upon an advection-diffusion model. Aerosol diffusion models are either driven by the chemical evolution of aerosol composition (e.g., Riemer et al., 2008; Sofiev et al., 2009) or rely upon physical intuition and/or numerical discretization for evaluating a diffusion coefficient (e.g., Stein et al., 2015; Wang et al., 2020). These methods, however, are validated using limited data collected by targeted campaigns leading to high model sensitivity to cloud and aerosol properties. To the best of our knowledge, few attempts have been made to simulate direct aerosol injections in non-pristine modeling environments, limiting our ability to drive down the known

<sup>1</sup>GOES-R is operated by the National Oceanic and Atmospheric Administration (NOAA) and the National Aeronautics and Space Administration (NASA).

model uncertainties. Importantly, our approach's ability to learn underlying parameters from observations highlights the power of passively observed data. Specifically, applying our approach to simulated data under controlled settings will enable studying cloud changes from track formation in imagery.

For a given ship, we consider modeling each aerosol *emission burst* as a single target and assume the ship is continuously emitting bursts. Each target is transported vertically from the ship through the atmosphere until it reaches a specific altitude near the cloud top height at which the target can become visible to orbital satellites and form linear tracks in a cloud. Ship track formations will then move with wind dynamics, a variable that is straight-forward to simulate and is independent of ship movement. The visible tracks then persist in the clouds for an unknown time as ship tracks, until the aerosols are fully diffused into the atmosphere and are no longer distinguishable from the surrounding clouds. Figure 2 outlines the general behaviors of the aerosols that are (un)observed via satellite. In this figure, the green box represents the portion of the track formation process that is visible via satellite.

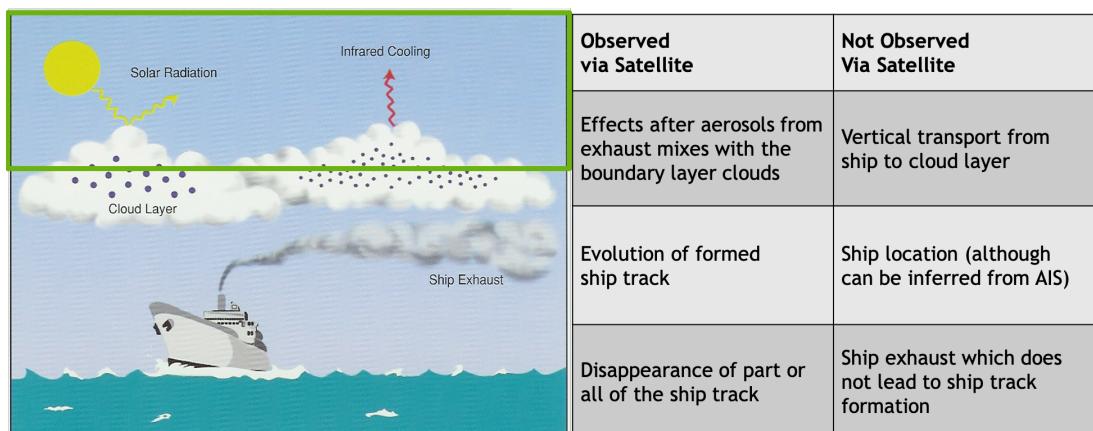
The remainder of this paper is organized as follows. Section 2 outlines our emulation approach. Section 3 discusses parameter learning. Section 4 provides simulated examples. Lastly, Section 5 considers potential impacts of this research and directions for future work.

## 2. Modeling aerosols using a Hidden Markov Model (HMM)

To model the formation and behavior of the aerosol tracks, we construct a state-space point process model relating imaging observations of ship tracks to true aerosol emission bursts.

### 2.1. State-space representation

The true emission path is generated by continuously emitted aerosol emission bursts or *particles* by a single ship over the spatial window  $X \subset \mathbb{R}^2$  up to time  $t \in [0, N\Delta]$  where  $N$  is the number of frames and  $\Delta > 0$  is the time between frames  $n$  and  $n + 1$  (typically between 5 and 15 minutes). A (Lagrangian) particle here is a mathematical object that carries mass in space at a specific time; it models a group of aerosol molecules with distinct mass. The unobserved spatio-temporal point process denoted  $\{X_{t_n} : (x, y, t_n) \in \mathbb{R}^2 \times \mathbb{R}\}$  characterizes the set of unknown cardinality and true positions of aerosol emission particles, continuously released prior to (and still visible at) time  $t_n$ . The observed point process  $\{Y_{t_n} : (x, y, t_n) \in \mathbb{R}^2 \times \mathbb{R}\}$  characterizes positions in satellite imagery containing observed ship tracks in image frame  $n$  generated by  $X_{t_n}$ . For ship  $k = 1 \dots K$  which produces a track, we assume that its entire emission path is comprised of  $P_k > 0$  aerosol particles, all of which reach the boundary



**Figure 2.** Observable and unobservable behaviors of aerosol emissions from satellite sensors<sup>2</sup>.

<sup>2</sup>Image available at <https://ral.ucar.edu/staff/jwolff/aerosols.html/intro.html>.

layer clouds. Assuming that only  $k_{t_n}$  of  $K$  ships have entered the window  $X$  by time  $t_n < T$ , for an arbitrary single track  $k \leq k_{t_n}$ , only  $p_k \leq P_k$  particles are expected to become visible. We denote the unobserved point process of true particle positions at time  $t_n$  as  $X_{t_n} = \{\{\mathbf{x}_{i,t_n}\}_{i=1}^{p_k}\}_{k=1}^{k_{t_n}}$ .

Existing ship tracks are modified at the next time step  $t_{n+1}$  in three possible ways: The oldest aerosol emissions at the end of the track diffuse completely and mix back into the atmosphere (leaving no detectable trace), surviving particles diffuse and become less distinguishable as part of the track, and new particles appear at the front of the track in the direction of ship movement. These situations result in  $p_{k_{t_{n+1}}}$  new locations  $\{\mathbf{x}_{i,t_{n+1}}\}_{i=1}^{p_{k_{t_{n+1}}}}$  in each of the new and existing emission tracks at  $t_{n+1}$ . In practice, the full lifespan (from first appearance to permanent disappearance) of each emission particle is *unknown*. Instead, at each observed image frame  $n$ , a partial observation of the track within surrounding clouds is captured by the satellite sensor without information on age of its particles. For track  $k_{t_n}$  therefore, a set of  $o_{k_{t_n}} \leq p_{k_{t_n}}$  observations  $Y_{t_n} = \{\{\mathbf{y}_{i,t_n}\}_{i=1}^{o_k}\}_{k=1}^{k_{t_n}}$  is recorded, where  $\mathbf{y}_{i,t_n} \in X$ .

At time  $t_n$ , a newly observed track can be generated from newly released emission particles from a new ship. Due to complex atmospheric dynamics, it is not often possible to link new observations to a source; an observed track is not always visible directly above a ship. Thus, we assume that there is no information about which emission particle generates which observation. We now specify a simulation model  $\mathcal{M}_\theta$  relating states  $\{X_{t_i}\}_{i=1}^N$  from their observation sets  $\{Y_{t_i}\}_{i=1}^N$  generated by parameters  $\theta$ .

### 2.1.1. Multi-target state model

First, we define the three cycles of an aerosol particle: survival ensuing motion and diffusion through the atmosphere, birth of new particles and death, a particle's permanent disappearance.

After an aerosol track has already formed at time  $t_{n-1}$ , if an arbitrary associated emission particle  $\mathbf{x}_{t_{n-1}} \in X_{t_{n-1}}$  survives to time  $t_n > t_{n-1}$ , its subsequent state is determined by a *drift* function which is described by the wind velocity  $\mu(\mathbf{x}_{t_{n-1}})$  at  $\mathbf{x}_{t_{n-1}}$ , and a *diffusion* term  $\sigma(\mathbf{x}_t)$  which describes the diffusion of the emission particle within the clouds it is situated in. This evolution describes a **Markov diffusion process** and is described by the (continuous time) stochastic partial differential equation (SPDE):

$$d\mathbf{x}_t = \mu(\mathbf{x}_t, t)dt + \sigma(\mathbf{x}_t)dB_t, \quad (1)$$

where  $B_t \sim \mathcal{N}_2(\mathbf{0}, t\mathbb{I}_2)$  denotes a standard Brownian motion in two dimensions, with  $\mathbb{I}_2$  denoting the 2-dimensional identity matrix. While the drift (wind velocity) is generally known, the diffusion function  $\sigma(\mathbf{x}_t) \equiv \sigma_x$  is set to be an unknown constant that describes the diffusivity of an aerosol parcel within the atmospheric boundary layer, and is to be learned from data. To avoid solving (1) with computationally expensive numerical methods, we assume that  $\Delta$  is taken small enough so that the wind velocity within each interval  $I_n := (n\Delta, (n+1)\Delta]$   $n \in \mathbb{Z}^+$  is approximately constant. With this and given state  $\mathbf{x}_s$  at  $s \in I_n$ , the SPDE  $d\mathbf{x}_t = \mu(\mathbf{x}_t)dt + \sigma_x dB_t$  for  $s < t \in I_n$ , can be solved via  $\mathbf{x}_t - \mathbf{x}_s = \int_s^t \mu(\mathbf{x}_w) dw + \sigma_x(B_t - B_s)$

$$\mathbf{x}_t - \mathbf{x}_s = \mu(\mathbf{x}_s)(t-s) + \sigma_x B_{t-s} \quad B_{t-s} \sim \mathcal{N}_2(\mathbf{0}, (t-s)\mathbb{I}_2). \quad (2)$$

In particular, the particle's transition density from  $\mathbf{x}_{t_{n-1}}$  to  $\mathbf{x}_{t_n}$  is given by  $f_{t_n|t_{n-1}}^M(\mathbf{x}_{t_n} | \mathbf{x}_{t_{n-1}}) = \mathcal{N}_2(\mathbf{x}_{t_{n-1}} + \mu(\mathbf{x}_{t_{n-1}})\Delta, \sigma_x^2 \Delta \mathbb{I}_2)$ , modeled by the point process  $S_{t_n|t_{n-1}}(\mathbf{x}_{t_{n-1}})$

$$S_{t_n|t_{n-1}}(\mathbf{x}_{t_{n-1}}) = \begin{cases} \mathbf{x}_{t_n} & \text{where } \mathbf{x}_{t_n} \sim f_{t_n|t_{n-1}}^M(\cdot | \mathbf{x}_{t_{n-1}}) \text{ with probability } p_{S,t_n}(b_{\mathbf{x}_{t_{n-1}}}) \\ \emptyset & \text{otherwise,} \end{cases} \quad (3)$$

where  $p_{S,t_n}(b_{\mathbf{x}_{t_{n-1}}})$  denotes the probability that particle  $\mathbf{x}_{t_{n-1}}$  will survive (be visually trackable) at  $t_n$ .

A new emission particle at time  $t_n \in \mathbb{R}$  can arise either as a spontaneous birth (of a newly risen emission) independent of any existing tracks, or via *spawning* from an existing emission source (e.g. at the head of an existing track), resulting in a newly visible emission particle. The *birth time* of particle  $\mathbf{x}_{t_n}$  observed at time  $t_n$  is denoted  $b_x$ . Spontaneous births at time  $t$  are denoted by the finite point process  $\Gamma_t$ , modeled as a **Poisson point process** (Mahler, 2007) with *intensity function*  $\gamma_t(\mathbf{x}) = \lambda_{\gamma_t} f_{b,t}(\mathbf{x})$ ,

where for  $\mathbf{x} \in \mathcal{X}$ ,  $\Gamma_t \sim \text{Poisson}(\lambda_{\gamma_t} f_{b,t}(\mathbf{x}))$ . Here,  $N_{b,t} \sim \text{Poisson}(\int_{\mathcal{X}} \lambda_{\gamma_t} f_{b,t}(\mathbf{x}) d\mathbf{x})$  denotes the number of births occurring in  $\mathcal{X}$  at time  $t$  and  $f_{b,t}(\mathbf{x})$  denotes their spatial distribution. We may assume that if  $\mathbf{x}_{b,t_n}$  is the position of a new ship at time  $t_n$ , then  $f_{b,t_n+\epsilon_b}(\mathbf{x}) = \mathcal{N}_2(\mathbf{x}_{b,t_n}, \sigma_b^2 \mathbb{I}_2)$  where  $\epsilon_b$  denotes the time lag between ship emission and aerosol observation at the boundary layer. On the other hand, spawned births occurring within  $I_{n-1}$  denote newly visible particles from existing tracks that reach the cloud top layer at time  $t_n$ . These can only be spawned by particles birthed in  $[t_n - \epsilon_b, t_n]$ , as this models the continuous emission of aerosol particles in a single stream. In this paper, we model the set of spawned births  $B_{t_n|t_{n-1}}(\mathbf{x}_{t_{n-1}})$  at time  $t_n$  from a particle  $\mathbf{x}_{t_{n-1}}$  as a **Bernoulli point process** (Mahler, 2007)

$$B_{t_n|t_{n-1}}(\mathbf{x}_{t_{n-1}}) = \begin{cases} \{\mathbf{x}\}; \mathbf{x} \sim f_{t_n|t_{n-1}}^{\beta}(\mathbf{x}|\mathbf{x}_{t_{n-1}}) \text{ with probability } p_{\beta,t_n} & t_{n-2} < b_{\mathbf{x}_{t_{n-1}}} \leq t_{n-1} \\ \emptyset & \text{otherwise.} \end{cases}$$

Knowledge of ship positions (e.g. via SeaVision<sup>3</sup>) when observing tracks being formed, motivates using the spawning density  $f_{t_n|t_{n-1}}^{\beta}(\mathbf{x}|\mathbf{x}_{t_{n-1}}) = \mathcal{N}_2(\mathbf{x}_{b,t_n-\epsilon_b}, \sigma_{\beta}^2 \mathbb{I}_2)$ . Since the spawning probability  $p_{\beta,t_n}$  is directly related to the number of aerosol particles emitted, for simulation purposes, we assume that each ship continuously emits aerosols in  $\mathcal{X}$  up to time  $T = N\Delta$ , enabling  $p_{\beta,t_n} = \mathbb{1}(t_n \leq T)$ .

Given  $X_{t_{n-1}}$  at time  $t_{n-1}$ , each particle  $\mathbf{x} \in X_{t_{n-1}}$  with birth time  $b_{\mathbf{x}}$ , either continues to be visually trackable to  $t_n$  with probability  $p_{S,t_n}(b_{\mathbf{x}})$ , or “dies” with probability  $1 - p_{S,t_n}(b_{\mathbf{x}})$ . Here, a “death” of an emission particle occurs when it sinks back through the atmosphere and ceases to be visible. Further, its survival probability is solely a function of its persistence time, since the effects of up and downward drafts in the atmosphere on each particle render spatial effects negligible. We assume that each ship produces a cloud-aerosol track that has an *average* lifetime  $T_d \sim \text{Exp}(\lambda_T)$  from birth. Given  $T_d$ , individual aerosol particles contained in its emission then each have an independently and identically distributed (i.i.d.) death time  $d \sim \text{Log-normal}(\mu_d = \log T_d (\sigma_{p_d}^2 + T_d^2)^{-1/2}, \sigma_d^2 = \log(\sigma_{p_d}^2 + T_d^2)/T_d^2)$ , where  $\sigma_{p_d}^2$  is the variance of particle death time, a fixed simulation input requiring estimation from data.

Altogether, we have  $X_{t_n} = [\cup_{\mathbf{x} \in X_{t_{n-1}}} S_{t_n|t_{n-1}}(\mathbf{x})] \cup [\cup_{\mathbf{x} \in X_{t_{n-1}}} B_{t_n|t_{n-1}}(\mathbf{x})] \cup \Gamma_{t_n}$  over independent unions.

### 2.1.2. Multi-target observational model

Here, we describe a finite point process model for the time evolution of the set  $\{Y_{t_n}\}_{n=1}^N$  generated from emission tracks  $\{X_{t_n}\}_{n=1}^N$  and observed over images  $\{\mathcal{Y}_{t_n}\}_{n=1}^N$ .

An arbitrary observation  $\mathbf{y}_{t_n} \in Y_{t_n}$  of unknown particle  $\mathbf{x}_{t_n} \in X_{t_n}$  is generated from a Gaussian density centered at  $\mathbf{x}_{t_n}$ , with covariance taken to be the *marginal* covariance of  $\mathbf{x}_{t_n}$ . Its marginal density can be calculated via  $f(\mathbf{x}_{t_n}) = \int_{\mathcal{X}} f_{t_n|b_{\mathbf{x}}}^M(\mathbf{x}_{t_n}|\mathbf{x}_{b_{\mathbf{x}}}) \pi(\mathbf{x}_{b_{\mathbf{x}}}) d\mathbf{x}_{b_{\mathbf{x}}}$  with  $\pi(\mathbf{x}_{b_{\mathbf{x}}})$  being the initial probability density of particle  $\mathbf{x}_{t_n}$  at the time of its birth. In this paper, we take  $\pi(\mathbf{x}_{b_{\mathbf{x}}}) = \delta(\mathbf{x}_{b_{\mathbf{x}}})$ , the dirac delta function centered at  $\mathbf{x}_{b_{\mathbf{x}}}$ , yielding  $\mathbf{y}_{t_n}|\mathbf{x}_{t_n} \sim \mathcal{N}_2(\mathbf{x}_{t_n}, \sigma_x^2(t_n - b_{\mathbf{x}})\mathbb{I}_2)$ . For image observations, we discretize this such that the pixel intensity of a pixel  $\mathcal{Y}_{t_n}(P)$ , denoted  $\mathcal{I}_{t_n}(P)$ , follows  $\mathcal{I}_{t_n}(P) \propto \sum_{\mathbf{x} \in X_{t_n}} \int_P f(\mathbf{y}|\mathbf{x}) d\mathbf{y}$  with normalization constant given by the highest pixel intensity across the video.

A particle  $\mathbf{x} \in X_t$ , at time  $t$  is only detected by a satellite sensor with probability  $p_{D,t}(\mathbf{x})$ . This *detection probability* has a spatio-temporal dependence structure which is needed to first, model the spatial randomness of cloud humidity and second, to account for cloud movement across  $\mathcal{X}$ . If the cloud humidity is too low or too high, emission particles cannot be detected. In the former case, particles cannot be observed since clouds cannot form to produce the necessary observations. In the latter, the cloud density may be too high, or may already be contaminated with existing aerosols which would subsequently not produce observations of new particles. To deal with this, we choose to model  $p_{D,t}(\mathbf{x})$  as a function of the existing cloud humidity, formulated by modeling *baseline* pixel intensities measured by the satellite sensor and utilizing a lower and upper threshold  $\iota_L, \iota_U$ . In particular, setting  $p_{D,t_n}(\mathbf{x}) = \mathbb{1}(\iota_L < \mathcal{I}_{t_n}(\mathbf{x}) < \iota_U)$ , enables a particle to be observed with probability one if its true location  $\mathbf{x}$  lies within a pixel of the  $n$ th frame with an intensity  $\mathcal{I}_{t_n}(\mathbf{x}) \in (\iota_L, \iota_U)$ , sufficient for its observation.

---

<sup>3</sup>SeaVision data based upon Automatic identification systems (AIS) available at <https://info.seavision.volpe.dot.gov/>.

Subsequently, the **observational point process**  $\Theta_{t_n}(\mathbf{x}_{t_n})$  from an emission particle  $\mathbf{x}_{t_n} \in \mathcal{X}$  follows

$$\Theta_{t_n}(\mathbf{x}_{t_n}) = \begin{cases} \{\mathbf{y}\} \text{ where } \mathbf{y} \sim \mathbf{y}_{t_n} | \mathbf{x}_{t_n} & \text{with probability } p_{D,t_n}(\mathbf{x}_{t_n}) \\ \emptyset & \text{with probability } 1 - p_{D,t_n}(\mathbf{x}_{t_n}), \end{cases} \quad (4)$$

specifying multi-target observations  $Y_{t_n} = \bigcup_{\mathbf{x} \in X_{t_n}} \Theta_{t_n}(\mathbf{x})$  observed within pixelated images  $\{\mathcal{Y}_{t_n}\}_{n=1}^N$ .

### 3. Data assimilation and parameter learning

In this section, we discuss a method for data assimilation and learning underlying parameters of the model from satellite observations, based upon the ship track emulation tool proposed here.

The emulation model  $\mathcal{M}_\theta$  comprises of parameters  $\theta$  that require estimation from data. Typical methods of data-driven estimation extend standard filtering approaches (Kalman, 1960) to the point process domain by utilizing Cardinalized Probability Hypothesis Density (CPHD) filters (e.g., Vo et al., 2006). These not only allow for efficient estimation of the cardinality and states of the underlying point process, but can be leveraged for statistical parameter learning (e.g., Jiang et al., 2015). Unfortunately, utilizing filtering approaches for this problem requires extracting the observations  $\{Y_{t_n}\}_{n=1}^N$  from relatively coarse satellite images with pixels containing overlapping tracks, a highly challenging problem. To alleviate this, we propose accurate simulation-based learning via Sequential Monte Carlo within an Approximate Bayesian Computation (ABC-SMC) approach, described in detail in Toni et al. (2009).

ABC-SMC works as follows. At iteration  $\tau = 0$ , parameters are sampled from a prior density  $\theta_j^{(0)} \sim \pi(\theta)$  until  $M$  datasets  $\mathcal{Y}^j \sim \mathcal{M}_{\theta_j^{(0)}}$  are within a tolerance  $\epsilon_\tau$  of the observed dataset  $\mathcal{Y}^d$ . This is computed over the sequence of images via the chosen distance function

$$\Delta(\mathcal{Y}^d, \mathcal{Y}^j) = \sum_{n=1}^N \sum_P (\mathcal{Y}_n^d(P) - \mathcal{Y}_n^j(P))^2.$$

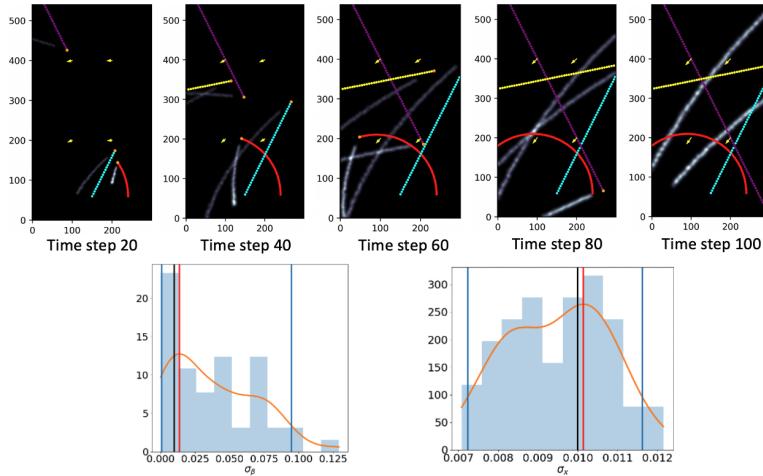
For  $\tau = 1, \dots, N_{MC}$ , tolerances  $\epsilon_\tau < \epsilon_{\tau-1}$  are chosen and  $\theta_m^* \sim \{\theta_m^{\tau-1}\}_{m=1}^M$  sampled with replacement using importance weights  $w_j^{(\tau)} = \pi(\theta_j^{(\tau)}) / \sum_{n=1}^N w_n^{(\tau-1)} K_\tau(\theta_j^* | \theta_n^{(\tau-1)})$  and perturbed via  $K_\tau$  to generate  $\theta_j^{(\tau)}$ , until  $M$  datasets  $\mathcal{Y}^j \sim \mathcal{M}_{\theta_j^{(\tau)}}$  satisfy  $\Delta(\mathcal{Y}^d, \mathcal{Y}^j) < \epsilon_\tau$ . Parameter values are therefore sequentially updated, enabling the posterior  $p_{\epsilon_\tau}(\theta | \mathcal{Y}^d) \approx \sum_{j=1}^M w_j^{(\tau)} \delta(\theta - \theta_j^{(\tau)}) / M$  at  $\tau = N_{MC}$  to be used to infer  $\theta$ .

### 4. Imaging simulation

Here, we present a ship track emulation example and parameter learning with ABC-SMC.

A snapshot of five simulated images taken 20 time steps apart with time step  $\Delta = 0.2$  hours are shown in Figure 3. In this study, four ships were simulated in  $300 \times 500$  longitude/latitude units appearing at staggered times. The tracks were generated using ship positions, spawning, persistence and death processes within a simulated circular wind motion where  $\mu(\mathbf{x}_t) = 10\pi\sqrt{(x - 0.2t)^2 + (y + 0.1t)^2} / 4N\Delta$ . An initial realistic cloud background image, where cloud pixels also moved with the simulated wind field, was omitted for illustration purposes. In particular, tracks are observed to follow both the ships' directions and wind field, with diffusivity emphasized by the broadening of each track through time. Further, pixel intensities are observed to be higher when cloud tracks overlap, highlighting the expected increase in pixel intensity when multiple sources are present.

Figure 3 also shows posterior distribution samples drawn using the ABC-SMC algorithm (described in Section 3) targeting parameters  $\theta = (\sigma_\beta \ \sigma_x)$  from the track produced by the red ship, with other input parameters fixed at their simulated values. Here, it is seen that the algorithm is accurate in learning the underlying simulation parameters, with true values contained within 95% posterior credible intervals.



**Figure 3.** **Top:** Simulation snapshots taken 4 hours apart, with  $N = 100$ ,  $\Delta = 0.2$  hours,  $\epsilon_b = 5$  hours,  $\sigma_\beta = \sigma_x = 0.01$ ,  $\lambda_T = 80$  hours,  $\sigma_{p_d} = 0.2$  hours,  $\iota_L = 0.25$  and  $\iota_U = 0.75$ . Ships (red, blue, purple, yellow) indicated by colored dotted trajectories have initial conditions  $\mathbf{b}_t = [5 \cos(\pi t / 10N\Delta) + 3, 5 \sin(\pi t / 2N\Delta) + 2], [1 + 5t / N\Delta, 18 - 2t / N\Delta], [1 + 5t / N\Delta, 18 - 10t / N\Delta], [-4 + 10t / N\Delta, 10 + 2t / N\Delta]$  respectively, with heads (orange). Wind direction shown via yellow arrows and tracks indicated by white trajectories. **Bottom:** Approximate posterior densities for  $\theta_1 = \sigma_\beta$  (left) and  $\theta_2 = \sigma_x$  (right) shown with estimated values (red), true values (black) and 95% credible intervals (blue). Here,  $N_{MC} = 4$ ,  $M = 50$ ,  $\theta_i \sim \text{Lognormal}(-5, 1)$ ,  $K_\tau(\theta_i | \theta_i^*) = \text{Uniform}(\max(0, \theta_i^* - \sigma_i^{(\tau)}), \theta_i^* + \sigma_i^{(\tau)})$  component wise, with  $\sigma_i^{(\tau)} = 0.5(\max_{1 \leq k \leq M} \{\theta_i^{(k, \tau-1)}\} - \min_{1 \leq k \leq M} \{\theta_i^{(k, \tau-1)}\})$ ;  $\epsilon_0 = 1$  and  $\epsilon_{\tau > 0}$  computed at the 80% quantiles of accepted parameter distances at the previous iteration (Filippi et al., 2013).

## 5. Discussion and follow-up work

In this paper, we have described and demonstrated a computational method to emulate aerosol-cloud tracks given wind and ship fields, using an SPDE that incorporates aerosol packet birth, motion, diffusion and death. A demonstration of data assimilation and parameter learning using simulation-based ABC-SMC, highlights the power of our algorithm in learning aerosol-cloud behavior from simulated inputs. Incorporating parameters that account for observed wind data from ERA-5 reanalysis and available atmospheric information that are well-documented to contribute to cloud track formation, such as cloud condensation nuclei (CCN), liquid water path (LWP), and low-lying cloud cover, would also improve emulation of more realistic aerosol-cloud behaviors, for comparison with satellite imagery.

Using this methodology, future work is to verify that this surrogate model is accurate in representing aerosol-cloud paths using satellite imagery and Large Eddy Simulations (LES). This is challenging as real tracks have poorly identifiable sources and the relationship between observed atmospheric properties and track behavior is not trivial to infer from imagery alone. Currently, methods are being pursued to validate against real and LES data through model calibration, matching satellite-observed lateral behavior and the vertical dispersion of tracks in LES. LES runs will also enable incorporation of dependencies on physical dependencies such as the aforementioned CNN, LWP and low-lying cloud cover.

While this would allow application of the presented methodology, the high computational burden that is expected when using simulation-based learning for high-resolution feature-rich satellite imagery, must be taken into account. For example, leveraging (pre-processing) image compression and feature extraction algorithms such as Convolution Neural Networks (CNNs) already developed for this problem (Yuan et al., 2019), could result in improved structured sparsity and reduce computational complexity.

**Acknowledgments.** We would like to thank Dr. Erika Roesler (Sandia National Laboratories), Dr. Michael Schmidt (Sandia National Laboratories) and Dr. Golo Wimmer (Los Alamos National Laboratory) for their valued time and input in this work.

**Funding Statement.** This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government. This work was supported by the Laboratory Directed Research and Development program at Sandia National Laboratories, a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

**Competing Interests.** None declared.

**Data Availability Statement.** GOES-R imagery data available at <https://www.bou.class.noaa.gov/saa/products/welcome> and AIS data available at <https://info.seavision.volpe.dot.gov>.

**Ethical Standards.** The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

**Author Contributions.** Conceptualization: L.P; L.S. Methodology: L.P. Data curation: L.P; L.S. Data visualisation: L.P; L.S. Writing original draft: L.P; L.S. All authors approved the final submitted draft.

## References

- Berner, A. H. et al. (2015). Large eddy simulation of ship tracks in the collapsed marine boundary layer: a case study from the monterey area ship track experiment. *Atmospheric Chemistry and Physics* 15, 5851–5871.
- Blossey, P. N. et al. (2018). Locally enhanced aerosols over a shipping lane produce convective invigoration but weak overall indirect effects in cloud-resolving simulations. *Geophysical Research Letters*, 9305–9313.
- Capaldo, K. et al. (1999). Effects of ship emissions on sulphur cycling and radiative climate forcing over the ocean. *Nature* 400(6746), 743–746.
- Carslaw, K. S. et al. (2013). Large contribution of natural aerosols to uncertainty in indirect forcing. *Nature* 503, 67–80.
- Conover, J. H. (1966). Anomalous cloud lines. *Journal of Atmospheric Science* 23, 778–785.
- Council, N. R. (2015). *Climate Intervention: Reflecting Sunlight to Cool Earth*. Washington, DC: The National Academies Press.
- Diamond, M. S. et al. (2020). Substantial Cloud Brightening from Shipping in Subtropical Low Clouds. *AGU Advances* 1(1).
- Eyring, V. et al. (2010). Transport impacts on atmosphere and climate: Shipping. *Atmospheric Environment* 44(37), 4735–4771.
- Filippi, S. et al. (2013). On optimality of kernels for approximate bayesian computation using sequential monte carlo. *Statistical Applications in Genetics and Molecular Biology* 12(1), 87–107.
- Gunnar, M. et al. (2013). Anthropogenic and natural radiative forcing. In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Chapter 8. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.
- Jiang, L. et al. (2015). Bayesian Tracking and Parameter Learning for Non-Linear Multiple Target Tracking Models. *IEEE Transactions on Signal Processing* 63(21), 5733–5745.
- Kalman, R. (1960). A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME-Journal of Basic Engineering* 82(Series D), 35–45.
- Latham, J. (1990). Control of global warming? *Nature* 347(6291), 339–340.
- Mahler, R. (2007). *Statistical Multisource-Multitarget information fusion*. Artech house.
- Possner, A. et al. (2018). The efficacy of aerosol–cloud radiative perturbations from near-surface emissions in deep open-cell stratocumuli. *Atmospheric Chemistry and Physics* 18(23), 17475–17488.
- Riemer, N. et al. (2008). Simulating the evolution of soot mixing state with a particle-resolved aerosol model. *Journal of Geophysical Research: Atmospheres* 114, D09202.
- Sofiev, M. et al. (2009). Turbulent diffusion and turbulent thermal diffusion of aerosols in stratified atmospheric flows. *Journal of Geophysical Research: Atmospheres* 114, D18209.
- Stein, A. et al. (2015). NOAA's HYSPLIT Atmospheric Transport and Dispersion Modeling System. *Bulletin of the American Meteorological Society* 96(12), 2059–2077.
- Toni, T. et al. (2009). Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. R. Soc. Interface* 6, 187–202.
- Twomey, S. et al. (1966). Comments on "Anomalous Cloud Lines". *Journal of Atmospheric Science* 25, 333–334.
- Vo, B. et al. (2006). The Cardinalized Probability Hypothesis Density Filter for Linear Gaussian Multi-Target Models. In *2006 40th Annual Conference on Information Sciences and Systems*.
- Wang, H. et al. (2011). Manipulating marine stratocumulus cloud amount and albedo: a process-modelling study of aerosol–cloud-precipitation interactions in response to injection of cloud condensation nuclei. *Atmospheric Chemistry and Physics* 11, 4237–4249.
- Wang, H. et al. (2020). Aerosols in the E3SM Version 1: New Developments and Their Impacts on Radiative Forcing. *Journal of Advances in Modeling Earth Systems* 12(1), e2019MS001851.
- Yuan, T. et al. (2019). Automatically finding ship tracks to enable large-scale analysis of aerosol–cloud interactions. *Geophysical Research Letters* 46(13), 7726–7733.