

# Markov-chain-inspired search for MH370

P. Miron,<sup>1, a)</sup> F. J. Beron-Vera,<sup>1</sup> M. J. Olascoaga,<sup>2</sup> and P. Koltai<sup>3</sup>

<sup>1)</sup>*Department of Atmospheric Sciences, Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, Florida, USA*

<sup>2)</sup>*Department of Ocean Sciences, Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, Florida, USA*

<sup>3)</sup>*Institute of Mathematics, Freie Universität Berlin, Berlin, Germany*

(Dated: 18 March 2019)

Markov-chain models are constructed for the probabilistic description of the drift of marine debris from Malaysian Airlines flight MH370. En route from Kuala Lumpur to Beijing, the MH370 mysteriously disappeared in the southeastern Indian Ocean on 8 March 2014, somewhere along the arc of the 7th ping ring around the Inmarsat-3F1 satellite position when the airplane lost contact. The models are obtained by discretizing the motion of undrogued satellite-tracked surface drifting buoys from the global historical data bank. A spectral analysis, Bayesian estimation, and the computation of most probable paths between the Inmarsat arc and confirmed airplane debris beaching sites are shown to constrain the crash site, near 25°S on the Inmarsat arc.

PACS numbers: 02.50.Ga; 47.27.De; 92.10.Fj

**Application of tools from ergodic theory on historical Lagrangian ocean data is shown to constrain the crash site of Malaysian Airlines flight MH370 given airplane debris beaching site information. The disappearance of flight MH370 constitutes one of the most enigmatic episodes in the history of commercial aviation. The tools employed have far-reaching applicability as they are particularly well suited in inverse modeling, critical for instance in revealing contamination sources in the ocean and the atmosphere. The only requirement for their success is sufficient spatiotemporal Lagrangian sampling.**

## I. INTRODUCTION

The disappearance in the southeastern Indian ocean on 8 March 2014 of Malaysian Airlines flight MH370 en route from Kuala Lumpur to Beijing is one of the biggest aviation mysteries. With the loss of all 227 passengers and 12 crew members on board, flight MH370 is the second deadliest incident involving a Boeing 777 aircraft. At a cost nearing \$155 million its search is already the most expensive in aviation history.

In January 2017, almost three years after the airplane disappearance, the Australian Government's Joint Agency Coordination Centre halted the search after failing to locate the airplane across more than 120,000 km<sup>2</sup> in the eastern Indian Ocean. On May 29th 2018, the latest attempt to locate the aircraft completed after an unsuccessful several-month cruise by ocean exploration company Ocean Infinity through an agreement with the Malaysian Government.

Analysis<sup>1,2</sup> of the Inmarsat-3F1 satellite communication, provided in the form of handshakes between engines and satellite, indicated that the aircraft had lost contact along the 7th ping ring around the position of the satellite on 8 March 2014, ranging from Java, Indonesia, to the southern Indian Ocean, southwest of Australia (Fig. 1). Since then, several pieces of marine debris belonging to the airplane have been found washed up on the shores of various coastlines in the southwestern Indian Ocean (Fig. 1). The first debris piece was discovered on 29 July 2015 on a beach of Reunion Island. Locations and dates from eight confirmed beachings<sup>3</sup> are indicated in Table I.

MH370 search approaches to date have included: the application of geometric nonlinear dynamics tools on simulated surface flows or as inferred from satellite altimetry observations with a focus on the analysis of the efficacy of the aerial search<sup>4</sup>; attempts to backward trajectory reconstruction from drifter relative dispersion properties<sup>5</sup>; inspection of trajectories of satellite-tracked surface drifting buoys (drifters) and direct trajectory forward integrations of altimetry-derived currents corrected using drifter velocities<sup>6</sup>; forward trajectory integrations of various flow representations corrected to account for leeway drift<sup>7-10</sup>; Bayesian inference of debris beaching sites using trajectories from an ensemble of model velocity realizations<sup>11</sup>; backward trajectory integrations of simulated velocities<sup>12</sup>; consideration of Bayesian methods for estimating commercial aircraft trajectories using models of the information contained in satellite communications messages and of the aircraft dynamics<sup>13</sup>; and biochemical analysis of barnacles attached to debris washed ashore to infer the temperature of the water they were exposed to<sup>14</sup>.

Here we introduce a novel framework for locating the MH370 crash site. Rooted in probabilistic nonlinear dynamical systems theory, the framework uses the locations and times of confirmed airplane debris beachings and historical trajectories produced by drifters to restrict the

<sup>a)</sup>Electronic mail: [pmiron@miami.edu](mailto:pmiron@miami.edu)

Beaching site	Days since crash	Color in plots
Reunion Island (RE)	508	■
South Africa (ZA)	655	■
Mozambique (MZ)	662	■
Mozambique (MZ)	721	■
Mauritius (MU)	753	■
Madagascar (MG)	808	■
Mauritius/Rodrigues (MU/RRG)	827	■
Tanzania (TZ)	838	■

TABLE I. Beaching information for confirmed debris from Malaysian Airlines flight MH370, which crashed in the Indian Ocean on 8 March 2014.

crash site along the Inmarsat arc. An additional important aspect of our approach to MH370 search, which makes it quite different than the previous ones, is that it directly targets crash site localization while exclusively performing forward evolutions.

Even though we have chosen to carry out a fully data-based analysis, the framework may be applied on the output from a data-assimilative system, enabling operational use of it in guiding search efforts, currently suspended, if they ever are to be resumed

## II. SETUP

The probabilistic framework to be developed in this paper builds on well-established results from ergodic theory<sup>15</sup> and Markov chains<sup>16,17</sup>, which place the focus on the evolution of probability densities rather than individual trajectories in the phase space of a nonlinear dynamical system (mathematical details are deferred to Appendix A in the Supplementary Material). At the core of the measure-theoretic characterization of nonlinear dynamics is the transfer operator and its discrete version, the transition matrix. The relevant dynamical system here is that obeyed by trajectories of airplane debris pieces that are transported under the combined action of turbulent ocean currents and winds mediated by inertia<sup>18,19</sup>.

Let  $\{\xi_{t+kT}\}_{k \geq 0}$  denote the time-discrete stochastic process describing such random trajectories. Assuming that this process is time-homogeneous over a sufficient long time interval  $\mathfrak{T}$ , its transition probabilities are described by a stochastic kernel, say  $K(x, y) \geq 0$  such that  $\int_X K(x, y) \mathbb{y} = 1$  for all  $x$  in phase space  $X$ , represented by the surface-ocean domain of interest. A probability density  $f(x) \geq 0$ ,  $\int_X f(x) \mathbb{x} = 1$ , describing the distribution of  $\xi_t$  at any time  $t \in \mathfrak{T}$  evolves to

$$\mathcal{P}f(y) = \int_X K(x, y) f(x) \mathbb{x} \quad (1)$$

at time  $t + T \in \mathfrak{T}$ , which defines a Markov operator  $\mathcal{P} : L^1(X) \circlearrowleft$  generally known as a *transfer operator*<sup>15</sup>.

To infer the action of  $\mathcal{P}$  from a discrete set of trajectories one can use a Galerkin approximation referred to as Ulam's method<sup>20–22</sup>. This approach consists in covering

$X$  with  $N$  connected boxes  $\{B_1, \dots, B_N\}$ , disjoint up to zero-measure intersections, and projecting functions in  $L^1(X)$  onto a finite-dimensional space  $V_N$  spanned by indicator functions of boxes normalized by box area. The discrete action of  $\mathcal{P}$  on  $V_N$  is described by a matrix  $P \in \mathbb{R}^{N \times N}$  called a *transition matrix*. Let  $\xi_t$  be a position chosen at random from a uniform distribution on  $B_i$  at time  $t$ . Then

$$P_{ij} = \text{prob}[\xi_{t+T} \in B_j \mid \xi_t \in B_i] = \frac{\int_{B_i} \int_{B_j} K(x, y) \mathbb{y} \mathbb{x}}{\text{area}(B_i)}, \quad (2)$$

which can be estimated as (cf. Appendix A of Miron et al.<sup>23</sup>)

$$P_{ij} \approx \frac{\# \text{ points in } B_j \text{ at } t \text{ that evolve to } B_j \text{ at } t+T}{\# \text{ points in } B_i \text{ at } t}. \quad (3)$$

Note that  $\sum_j P_{ij} = 1$  for all  $i$ , so  $P$  is a (row) stochastic matrix that defines a *Markov chain* on boxes, which represent the states of the chain<sup>16,17</sup>. The evolution of the discrete representation of  $f(x)$ , i.e., a probability vector  $\mathbf{f} = (f_1, \dots, f_N)$ ,  $\sum f_i = 1$  where  $f_i = \int_{B_i} f(x) \mathbb{x}$ , is calculated under left multiplication, i.e.,

$$\mathbf{f}^{(k)} = \mathbf{f} P^k, \quad k = 1, 2, \dots \quad (4)$$

## III. CONSTRUCTION OF A SUITABLE TRANSITION MATRIX

The surface circulation of the Indian Ocean is influenced by monsoon intraseasonal variability<sup>24</sup>. An appropriate Markov-chain model for marine debris motion in the Indian Ocean must account for this variability, which we do in constructing the chain's  $P$  in a fully data-based fashion using trajectories produced by satellite-tracked drifters.

The drifter data are collected by the National Oceanic Atmospheric Administration/Global Drifter Program (NOAA/GDP)<sup>25</sup>. Trajectories sampling the world oceans including the Indian Ocean exist since 1979. For the purpose of the present analysis we restrict attention to trajectory portions during which the drogue (a 15-m-long sea anchor designed<sup>26</sup> to minimize wind slippage and wave-induced drift) attached to the spherical float carrying the satellite tracker is absent<sup>27</sup>. Undrogued drifter

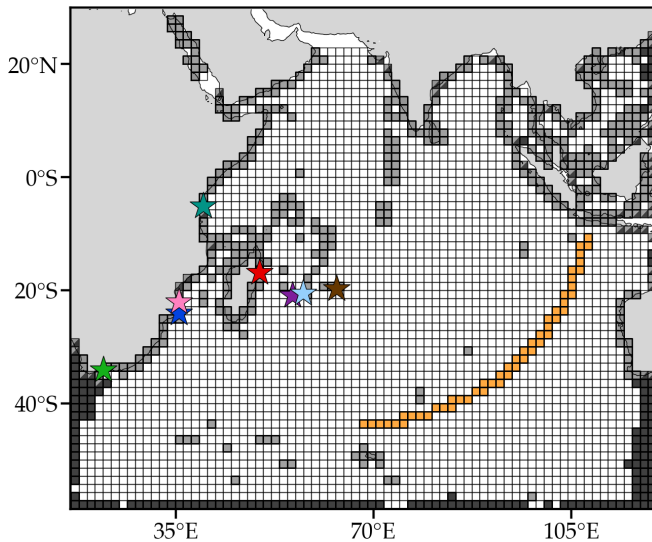


FIG. 1. Covering of the Indian Ocean domain into boxes forming the various Markov-chain models constructed using satellite-tracked undrogued drifters to describe the motion of marine debris produced by the crash of Malaysian Airlines flight MH370. Boxes with positive probability of the chain(s) to terminate outside the domain (leaky states) are indicated in dark gray, boxes including land–water interfaces (sticky states) are shown in light gray, and boxes along an arc of 7th ping ring around the Inmarsat-3F1 satellite position when communication with the MH370 flight was lost (crash states) are highlighted in yellow. Stars correspond to the airplane debris beaching sites in Table I.

motion is affected by inertial effects (i.e., those produced by buoyancy, finite-size, and shape) and thus is more representative of marine debris motion than that of drogued drifters, which more closely follow water motion<sup>19</sup>.

To construct  $P$ , we cover the Indian Ocean domain with a grid of  $0.25^\circ \times 0.25^\circ$  longitude–latitude boxes (Fig. 1). The size of the cells was selected to maximize the grid’s resolution while each individual box is sampled by enough trajectories. Similar grid resolutions in analysis involving buoy trajectory data were employed in recent work<sup>23,28,29</sup>, where sensitivity analyses to cell size variations and data amount truncations are presented. The area of the boxes varies from about 400 to 750 km<sup>2</sup>, yet the normalization by box area in the definition of the vector space  $V_N$  makes this variation inconsequential, i.e., a stochastic transition matrix is obtained without the need of a similarity transformation (e.g., Froyland and Padberg<sup>30</sup>). Ignoring time, there are 226 drifters on average per box (cf. Fig. S1 in the Supplementary Material); the number of drifters vary between 36 and 58 if the data are grouped according to season of the year. Equation (3) is then evaluated for appropriate transition time ( $T$ ) and time-homogeneity interval ( $\mathfrak{T}$ ) choices.

Using  $T = 1$  d, approximately the surface ocean Lagrangian decorrelation time<sup>31</sup>, the simple Markovian dynamics test  $\lambda(P(nT)) = \lambda(P(T))^n$ , where  $\lambda$  denotes eigenvalue, holds very well up to  $n = 10$  (cf. Fig. S2

in the Supplementary Material). Here we have chosen to use  $n = 5$  (equivalently  $T = 5$  d) as this guarantees both good interbox communication and negligible memory into the past. Similar choices have been made in recent applications involving drifter data<sup>23,28,29,32,33</sup>.

The simplest choice for the time-homogeneity interval  $\mathfrak{T}$  is one that coincides with the entire record of trajectory data<sup>23,28,29,32,33</sup>. The resulting *autonomous* Markov chain, which will be only considered for comparison purposes, ignores any mode of variability of the ocean circulation and thus is not optimal for describing debris motion in seasonally dependent environments like the Indian Ocean.

Different  $\mathfrak{T}$  intervals can be considered (e.g., van Sebille et al.<sup>34</sup>) to represent the dominant variability mode of the Indian Ocean circulation, produced by seasonal changes in the wind stress associated with the Indian monsoon<sup>24</sup>. During the northern winter, when the monsoon blows southwestward, the flow of the upper ocean is directed westward from near the Indonesian Archipelago to the Arabian Sea. During the northern summer, with the change of the monsoon direction toward the northeast, the ocean circulation reverses, with eastward flow extending from Somalia into the Bay of Bengal. Thus we consider three  $\mathfrak{T}$  intervals: January–March ( $\mathfrak{T}_W$ ), which typically corresponds to the winter monsoon season, July–September ( $\mathfrak{T}_S$ ), corresponding to the summer monsoon season, and April–June and October–December together ( $\mathfrak{T}_{SF}$ ), seasons which do not need to be distinguished from one another to represent the monsoon-induced circulation of the Indian Ocean. This results in three transition matrices,  $P_W$ ,  $P_S$  and  $P_{SF}$ , respectively, which are appropriately considered for  $t \in \mathfrak{T}_W$ ,  $\mathfrak{T}_S$  or  $\mathfrak{T}_{SF}$ , when a probability vector is evolved (pushed forward) under left multiplication. We will refer to the resulting Markov-chain model as *nonautonomous*.

Finally, if the interest is in the fate of the debris in the seasonally changing Indian Ocean environment after several years, one can more conveniently push forward probabilities using a  $P$  constructed by combining the above seasonal  $P$ s in such a way that the resulting Markov chain has a transition time  $T$  of 1 yr. Recalling that  $T = 5$  d for the seasonal  $P$ s, this is (approximately) achieved by  $P = P_W^{18} \cdot P_{SF}^{18} \cdot P_S^{18} \cdot P_{SF}^{18}$ . The resulting Markov-chain model will be referred to as *autonomous season-aware*. Similar constructions have been considered earlier (e.g., Froyland et al.<sup>35</sup>).

#### IV. CRASH SITE ASSESSMENT FROM SPECTRAL ANALYSIS

Information about the long-time asymptotic behavior of a dynamical system described by an autonomous transition matrix  $P$  can be obtained from its spectral properties<sup>36–38</sup>. Indeed, the leading eigenvector structure of  $P$  suggests a dynamical partition or *geography*<sup>23,28,35</sup> of weakly interacting sets that constrains connectivity

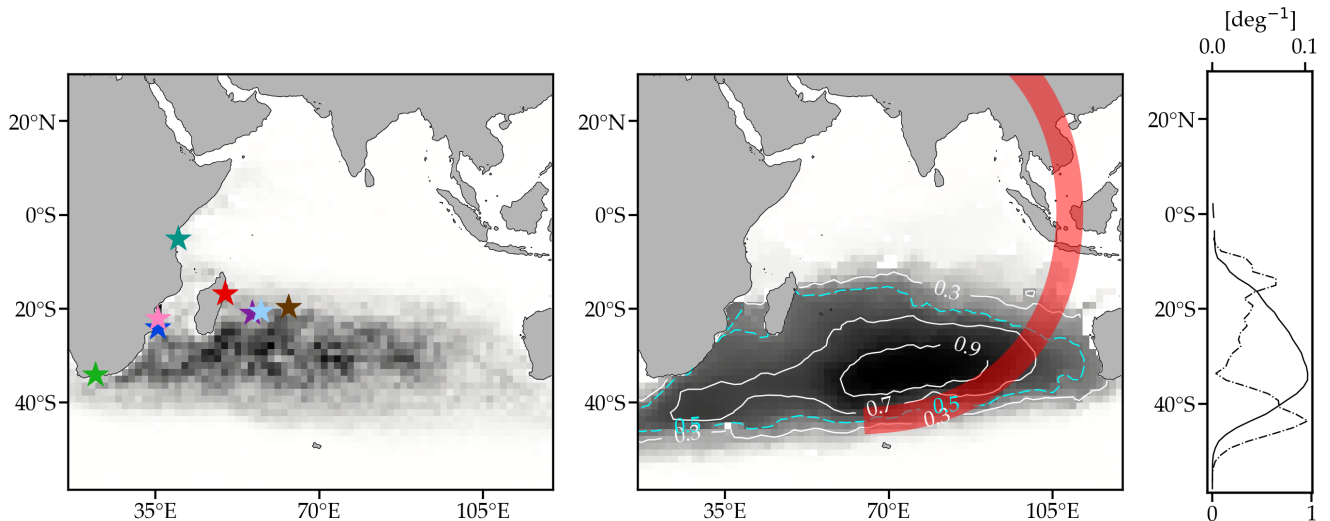


FIG. 2. (left) Dominant left eigenvector of the autonomous season-aware transition matrix with beaching sites indicated (cf. Table I). (middle) Dominant right eigenvector with the Inmarsat arc overlaid. (right) Zonally averaged right eigenvector (solid, bottom axis) and its derivative (dashed, top axis). Local maxima in the left eigenvector are regions that attract trajectories which tend to run long there before exiting the Indian Ocean. The basin of attraction roughly corresponds to the region enclosed by the 0.5 right eigenvector level set, above which the eigenvector looks closer to  $\mathbf{1}$ .

between distant points in phase space. Here we unveil such a dynamical geography from the autonomous season-aware Markov chain to gain insight into debris motion and thus the crash site.

For any irreducible and aperiodic stochastic  $P$ , its dominant *left* eigenvector,  $\mathbf{p}$ , satisfies  $\mathbf{p}P = \mathbf{p}$  and (scaled appropriately) represents a limiting invariant or *stationary* distribution, namely,  $\mathbf{p} = \lim_{k \rightarrow \infty} \mathbf{f}P^k$  for any probability vector  $\mathbf{f}$  (cf., e.g., Horn and Johnson<sup>39</sup>). Also,  $\mathbf{1} = P\mathbf{1}$ , where  $\mathbf{1}$  denotes the vectors of ones, is the *right* eigenvector corresponding to the eigenvalue 1 and  $\mathbf{p}$ .

If  $P$  is substochastic, i.e.,  $\sum_j P_{ij} < 1$  for some  $i$ , the dominant left eigenvector  $\mathbf{p} \geq 0$  decays at a rate set by the dominant eigenvalue  $\lambda_1 < 1$ , and has the interpretation of a limiting almost-invariant or *quasistationary* distribution, namely, the limiting distribution of trajectories that run long before terminating (cf. Chapter 6.1.2 of Bremaud<sup>16</sup>). Restricted to the set  $B$  where those trajectories start, i.e., the *basin of attraction*, the dominant right eigenvector is close to  $\mathbf{1}$ <sup>23,28,35,40</sup>.

The left and middle panels of Fig. 2 show respectively the dominant left ( $\mathbf{p}$ ) and right ( $\mathbf{r}$ ) eigenvectors of the autonomous  $P$  introduced in the preceding section. The right panel in turn shows again the right eigenvector, but this time zonally averaged (solid, bottom axis), along with its (meridional) derivative (dashed, top axis), which maximizes near the 0.5 level set. The dominant eigenvalue  $\lambda_1 = 0.8181$  sets an annual decay rate for  $\mathbf{p}$  of about 20%, which is nearly four times slower than that experienced by the first subdominant left eigenvector ( $\lambda_2 = 0.4012$ ). Note the structure of  $\mathbf{p}$ , taking many local maxima toward the western side of Indian Ocean,

inside the 20–40°S band. The right eigenvector is much less structured, looking closer to  $\mathbf{1}$  within the region enclosed by the 0.5 level set. This region approximately forms a basin of attraction  $B = \{\mathbf{r} > 0.5\}$  for trajectories that asymptotically distribute as  $\mathbf{p}$  conditional to staying in the domain for a long time.

The expected retention time in  $B$  is given by  $T_B = T/(1 - \lambda_B)$ , where  $\lambda_B$  is the dominant eigenvalue of  $P$  restricted to  $B$  (cf. Appendix B of the Supplementary Material for a derivation and Miron et al.<sup>23</sup> for a recent application). We compute  $\lambda_B = 0.5103$ , and noting that  $T = 1$  yr for the autonomous season-aware  $P$ , we estimate  $T_B = 2.0421$  yr, which is of the order of the mean time it took observed airplane debris to reach the African coasts (cf. Table I). This long residence time and the large area spanned by  $B$  impose a constraint on the connectivity between locations in- and outside of  $B$  by debris trajectories, which we use to make an initial assessment of the possible crash site as follows.

Note in the left panel of Fig. 2 that the observed beaching sites (stars) lie within or the border of the western region where  $\mathbf{p}$  tends to locally maximize. Note in the middle panel that the Inmarsat arc traverses the eastern side of  $B$ . This suggests a possible crash site somewhere along the Inmarsat arc sector between the latitudes of intersection with of  $B$ , roughly 20 and 40°S.

In the next sections we will show that the uncertainty (about 3600 km) of the spectral assessment above can be substantially reduced using a dedicated Bayesian analysis along with the computation of most probable paths and the inspection of a particular drifter trajectory.



## V. BAYESIAN ESTIMATION OF THE CRASH SITE

The Bayesian analysis uses the beaching events as observations to infer the probability distribution of the crash site (refer to Appendices C and D in the Supplementary Material for mathematical details). Because of the shorter-time nature of these observations, the analysis is most appropriately carried out using the nonautonomous Markov-chain model, albeit with some convenient adaptation.

As our computational domain is not the whole ocean, necessarily trajectories can leave the domain, and not have an assignable termination box. For example, the Agulhas Current transports approximately 70 Sv (1 Sv =  $10^6 \text{ m}^3\text{s}^{-1}$ ) out of the western boundary of the domain<sup>41</sup>. This leakage is represented by  $\sum_j P_{ij} < 1$  for the leaky states  $i \in \mathcal{L} \subset S := \{1, \dots, N\}$  (black boxes in Fig. 1). (Herein  $P$  is any  $P_W$ ,  $P_S$  or  $P_{SF}$ , which are applied for  $t \in \mathcal{T}_W$ ,  $\mathcal{T}_S$  or  $\mathcal{T}_{SF}$  as appropriate.) To account for the leakage and retain the open dynamics nature of the Indian Ocean, we introduce an absorbing state  $N + 1$ , commonly referred to as *cemetery* state, and consider  $1 - \sum_j P_{i \in \mathcal{L}, j}$  to be the probability of the chain to move and terminate in it, if currently being in state  $i \in \mathcal{L}$ . This amounts to augmenting  $P \in \mathbb{R}^{N \times N}$  to  $P \in \mathbb{R}^{(N+1) \times (N+1)}$  by

$$\begin{cases} P_{i \in \mathcal{L}, j=N+1} \leftarrow 1 - \sum_j P_{i \in \mathcal{L}, j}, \\ P_{i=N+1, i} = 1, \end{cases} \quad (5)$$

satisfying  $\sum_j P_{ij} = 1$ .

In addition to the dynamics being open, as we are considering floating debris which might come ashore somewhere, the transition matrix has to account for the Markov chain possibly terminating on water–land interfaces. Such events are hard to identify from the dataset because drifters can terminate for multiple reasons (e.g., malfunction, recovery, end of life). To account for beaching, we identify the set of sticky states  $\mathcal{S} \subset S$  (grey boxes in Fig. 1) with those coastal boxes that contain land–water interfaces and introduce the *land fraction* function  $\ell : \mathcal{S} \rightarrow (0, 1)$ . Note that a sticky state may also be leaky, namely,  $\mathcal{S} \cap \mathcal{L}$  need not be empty. We then denote by  $\mathcal{D} \subset \mathcal{S}$  the subset of sticky boxes where airplane debris were found to beach, to distinguish them from the others. Beaching is then modeled by augmenting  $P \in \mathbb{R}^{(N+1) \times (N+1)}$  to  $P \in \mathbb{R}^{(N+1+M) \times (N+1+M)}$ , satisfying  $\sum_j P_{ij} = 1$  and where  $M = |\mathcal{D}| (= 8)$ , according to:

$$\begin{cases} P_{i \in \mathcal{S}, j \in S \cup \{N+1\}} \leftarrow (1 - \ell(i \in \mathcal{S})) P_{i \in \mathcal{S}, j \in S \cup \{N+1\}}, \\ P_{i \in \mathcal{S}, j=N+1} \leftarrow \ell(i \in \mathcal{S}) + (1 - \ell(i \in \mathcal{S})) P_{i \in \mathcal{S}, j=N+1}, \\ P_{i \in \mathcal{D}, j=N+1+m(i \in \mathcal{D})} = \ell(i \in \mathcal{D}), \\ P_{i \in N+1+m(\mathcal{D}), i} = 1, \end{cases} \quad (6)$$

where  $\tilde{\mathcal{S}} := \mathcal{S} \setminus \mathcal{D}$ ,  $m : \mathcal{D} \rightarrow \{1, \dots, M\}$ , and  $N + 1 + m(i \in \mathcal{D})$  represents a *target cemetery* state where the chain terminates whenever beaching from  $i \in \mathcal{D}$ .

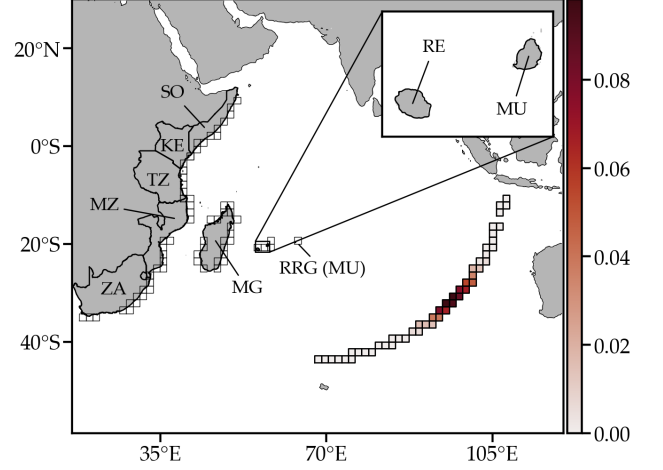


FIG. 3. Posterior probability of the crash site, for given assumed mutually independent observations of debris beaching time and based on the nonautonomous Markov-chain model. Indicated are countries and territories in the western side of the Indian Ocean domain and corresponding boxes of the domain covering. Acronyms SO and KE stand for Somalia and Kenya, respectively; cf. Table I for the rest of the acronyms.

The first line in (6) deals with sticky boxes if the chain flows on. The second line deals with sticky but nondebris beaching boxes, each of which also is terminated in the cemetery state ( $N + 1$ ). The third line deals with debris-beaching boxes if the chain terminates from them by beaching. Finally, the fourth line makes each target cemetery state absorbing.

Having settled on an appropriate  $P$  representation, we proceed to formulate the Bayesian estimation problem by first fixing some notation. Let  $\mathcal{C} \subset S$  denote the set of indices of the boxes (states) along the Inmarsat arc—the possible crash sites. We call  $b$  a state in the set  $\mathcal{B} := \{N + 2, \dots, N + M + 1\}$  of airplane debris beaching site boxes and further make  $\mathbf{b} := (b)_{b \in \mathcal{B}} \in \mathbb{N}^M$ . Let  $\{\xi_{t+kT}\}_{k \geq 0}$  be time-discrete random position variables that take values on the augmented Markov chain states  $S \cup \{N + 1\} \cup \mathcal{B}$ , and  $t$  denotes the crash time, when the chain starts. Consider then the random variable  $\tau^b$  denoting the time after which the chain gets absorbed into a particular beaching state  $b$ , namely,

$$\tau^b := \inf_{k \geq 0} \{t + kT : \xi_{t+kT} = b \in \mathcal{B}\}. \quad (7)$$

Define  $\text{prob}_c[\cdot] := \text{prob}[\cdot \mid \xi_t = c \in \mathcal{C}]$  and

$$p_c^b(k) := \mathbf{1}_c P^k \cdot \mathbf{1}_b, \quad c \in \mathcal{C}, b \in \mathcal{B}, \quad (8)$$

where  $\mathbf{1}_j = (\delta_{ij})_{i \in S \cup \{N+1\} \cup \mathcal{B}}$ , and then note that

$$\text{prob}_c[\tau^b = t + kT] = \begin{cases} p_c^b(k) = 0 & \text{if } k = 0, \\ p_c^b(k) - p_c^b(k-1) & \text{if } k > 0. \end{cases} \quad (9)$$

Let us now assume that after the crash every single piece of debris was transported mutually independently.

For each debris, we have access to two observations of random quantities: the *beaching site* and the *beaching time*. Let  $\xi$  and  $\tau$  denote these random variables, respectively, and note that if the target is absorbing, then  $\tau^b < \infty$  from (7) implies  $\xi = b$ , and thus the events  $\{\xi = b \text{ and } \tau = t^b\}$  and  $\{\tau^b = t^b\}$  are equivalent. Thus the joint probability of these observations, i.e.,

$$p(t^b|c) := \text{prob}_c[\tau = t^b \text{ and } \xi = b] \equiv \text{prob}_c[\tau^b = t^b], \quad (10)$$

is computable as in (9). The probabilities  $p(t^b|c)$  depend on the crash site  $c$ . Now, making use of the independence assumption above, the joint probability of the observations, given the crash site  $c$ , is

$$p(t^{\mathbf{b}}|c) := \prod_{b \in \mathfrak{B}} p(t^b|c). \quad (11)$$

The idea of Bayesian inversion<sup>42</sup> is to find a probabilistic characterization of the unknown parameter  $c$  (crash site) given the debris beaching times and locations. By viewing  $p(t^{\mathbf{b}}|c)$  as a function of  $c$ , one obtains a function  $L(c; t^{\mathbf{b}})$  which represents the *likelihood* of  $c$ . The *posterior distribution* of  $c$ , i.e., the probability distribution of  $c$  once  $t^{\mathbf{b}}$  have been observed, follows from Bayes' theorem:

$$p(c|t^{\mathbf{b}}) \propto p(t^{\mathbf{b}}|c) \cdot p(c), \quad (12)$$

where  $p(c)$  is the *prior distribution* of  $c$ , representing the state of knowledge about  $c$  before to data have been observed.

Assuming that  $p(c)$  is uniform (i.e., maximally uninformative) over  $\mathfrak{C}$ , the Inmarsat arc boxes, Fig. 3 shows  $p(c|t^{\mathbf{b}})$  for debris beaching times  $t^{\mathbf{b}}$  as given in Table I. The *maximum likelihood estimator*,  $c_{\max} = \arg \max_c p(t^{\mathbf{b}}|c)$ , corresponds to the index of the Inmarsat arc box centered at about 31°S. To check how reasonable this crash site estimate is, the top panel of Fig. S3 in Appendix F of the Supplementary Material shows that the joint probability  $\text{prob}_{c_{\max}}[\tau = t + kT \text{ and } \xi_{t+kT} = s]$  for each sticky state  $s \in \mathfrak{S}$  tends to maximize near the observations. The misfit is attributed to the historical drifter data not capturing all of the details of the Indian Ocean dynamics and in particular those when the crash took place and the years after it. For instance, the bottom panel of Fig. S3 shows that the misfit augments when the monsoon variability is ignored.

The Bayesian crash site estimate, 31°S on the Inmarsat arc, lies within the arc portion identified as a likely crash region using the spectral analysis of the previous section. An important difference is that the uncertainty of the assessment is constrained by the Bayesian analysis. For instance, the 95% central posterior interval length, obtained by computing the 2.5 and 97.5%-ile of the posterior distribution, is of about 12° along the arc, which is nearly twice as narrow as the spectral inference. However, the Bayesian inference is affected by a bimodality in the single posterior distributions of the crash site. This is demonstrated in Fig. 4, which shows

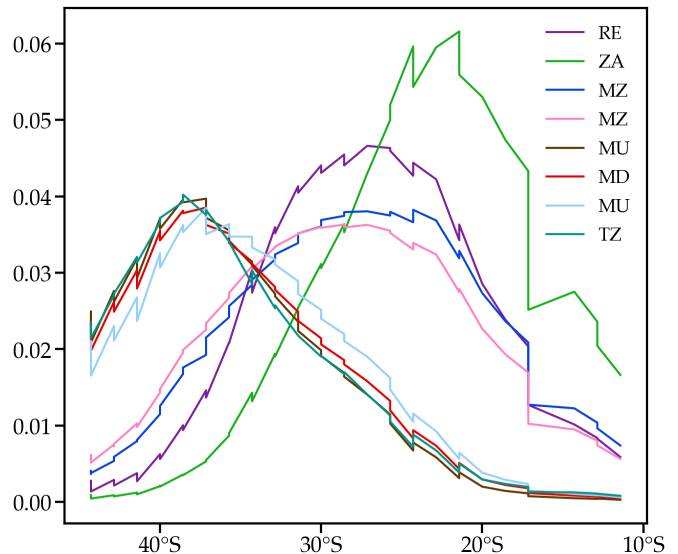


FIG. 4. Single posterior distributions of the latitude of the crash site along the Inmarsat arc, computed using the nonautonomous Markov-chain model. Refer to Table I for acronym meanings.

$p(c|t^{\mathbf{b}})$  plotted for each  $b \in \mathfrak{B}$  individually as a function of the latitude of  $c \in \mathfrak{C}$ . These can be collected into two quite consistent groups. Specifically, a southern group, which favors a most likely crash site near 36°S on the arc, and a northern group, favoring a most likely position near 25°S.

## VI. NARROWING THE UNCERTAINTY OF THE CRASH SITE DETERMINATION

We close the analysis by discussing one additional calculation and a particular observation that altogether provide means for favoring the northern of the above two possible crash sites, improving the confidence of its determination.

Drifter data does not merely give the chance to construct a Markov chain model of the surface currents, but also to study single trajectories. Thus we consider the most probable paths of our model that end up at the particular beaching sites. More precisely, as we know how long (in time) the single trajectories were, we compute most probable paths ending at a given beaching site  $b \in \mathfrak{B}$  after having ran for a fixed time  $t^b$ . The basic idea of the process is to set up and iteratively solve a dynamical programming equation relating the maximal probability of a path reaching some state  $i \in S$  after exactly  $t + T$  units of time with the maximal probabilities of paths reaching other states  $j \in S$  after  $t$  (cf. Appendix E in the Supplementary Material for details as well as for a review of standard unconstrained extremal path notions<sup>43,44</sup>). It is important to realize that the probability of *single* paths is extremely low (as there are so many possible ones), and the most probable one can sensitively

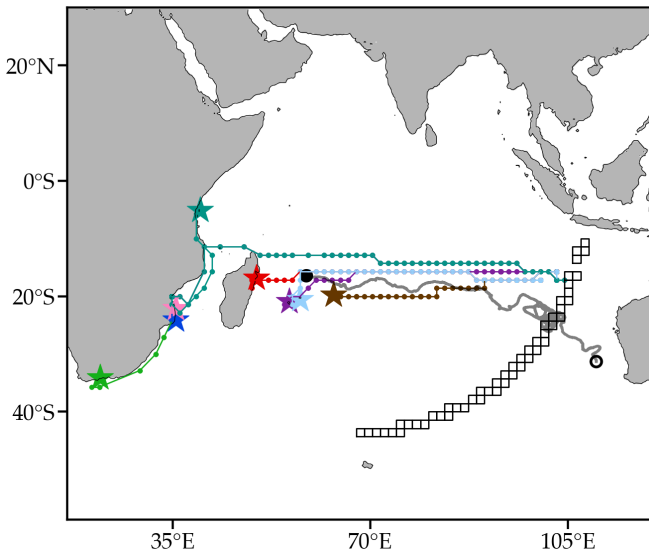


FIG. 5. Most probable paths (dotted) along the nonautonomous Markov chain between Inmarsat arc boxes and each of the debris beaching sites (cf. Table I), and trajectory (gray) of a NOAA/GDP drifter (ID 56568) launched west of the Australian coast.

depend on the slightest variations of the dynamics. Thus they are only suitable for qualitative comparisons with observations. We note that constrained paths were considered to analyze currents in the Mediterranean Sea by Ser-Giacomi et al.<sup>45</sup>, yet without excluding the possibility to hit the target before the end time.

The result is shown in Fig. 5, revealing, on the one hand, quite remarkably that the most probable paths of different fixed lengths (dotted curves) start from a common box that is the Inmarsat arc box centered at roughly (105°E, 16°S). On the other hand, Triñanes et al.<sup>6</sup> report on a drifter (ID 56568, from the NOAA/GDP database) that crossed the Inmarsat arc near (103°E, 22°S) in March 2014 and after looping for a little while about the arc near (102°E, 25°S) reached the vicinity of Reunion Island in July 2015, about the time when the first MH370 debris piece was spotted. The trajectory of the drifter in question is shown in gray in Fig. 5.

Taken together, the most probable path computation and the specific drifter trajectory observation favor the Bayesian crash site estimate near 25°. Excluding from the Bayesian analysis the debris beaching data that leads to single posterior crash site distributions peaking near 36°S, we estimate a 95% central posterior interval length of 16° along the Inmarsat arc for the refined crash site inference.

We finally note that the qualitative assessment in this section, along with the spectral analysis assessment, provide a guideline to construct a prior distribution that can be incorporated to the Bayesian analysis. More specifically, these assessments suggest an informative prior distribution narrowing near 25°N on the Inmarsat arc.

## VII. SUMMARY AND CONCLUDING REMARKS

Using historical satellite-tracked surface drifter data in the Indian Ocean, we have proposed a Markov-chain model representation of the drift of observed marine debris from the missing Malaysian Airlines flight MH370.

The results from a spectral analysis of an autonomous discrete transfer operator (transition matrix) that controls the long-term evolution of the debris in a seasonally changing environment showed that the crash region is likely restricted to the 20–40°S portion of the arc of the 7th ping ring around the Inmarsat-3F1 satellite position when the airplane lost contact on 8 March 2014. The solution of a dedicated Bayesian estimation problem that uses the locations and times of confirmed airplane debris beachings in a Markov-chain model defined by a nonautonomous transition matrix capable of resolving shorter-term details of the debris evolution, further identified two probable crash sites within the aforementioned arc portion, one near 36°S and another one near 25°S. Consideration of most probable paths between the Inmarsat arc and the debris beaching sites constrained by the observed (elapsed) beaching times, and the observation of a drifter that took a trajectory similar to the path connecting the arc and the Reunion Island beaching site, was shown when taken together to favor the 25°S crash site estimate.

Our Bayesian crash site estimate, with a 95% central posterior interval ranging from 33 to 17°S on the Inmarsat arc, is consistent with the most recently published<sup>7</sup> crash area estimate, 28–30° along the arc, but it lies north of the latest recommended search area by the Commonwealth Scientific and Industrial Research Organisation<sup>8</sup>, at around 35°S. Notwithstanding, it is consistent with the original northern definition of high-priority search zone by the Australian Transport Safety Bureau<sup>46</sup>.

The uncertainty of our crash site estimate may be attributed to unresolved nonautonomous dynamical effects by the Markov-chain model constructed from the historical drifter data. Indeed, the historical drifter data may fall short of capturing all of the details of the circulation (e.g., at the submesoscale), particularly during the years after the crash took place. Moreover, undrogued drifter motion, while different than water parcel motion, may not accurately represent the motion of airplane debris pieces with varied shapes and thus drag properties.

While our results may provide grounds for guiding search efforts, currently halted, the operational use of the probabilistic framework proposed here in such a task will require one to consider an appropriate data-assimilative system. Indeed, the probabilistic framework may well be applied on observed trajectory data, as we have chosen to do in this paper, or on numerically generated trajectory data. Clearly, the success of such an operative use of the framework will depend on the operator’s ability to appropriately modeling the effects of the inertia of the debris pieces (i.e., of their buoyancy, size and shape), which is

a subject of active research<sup>18,19,47</sup>.

We finally note that the framework here presented is well-suited for inverse modeling in a general setting and thus its utility is far reaching. Such modeling is critical, for instance, in contamination source backtracking<sup>48</sup>. Relevant oceanic problems include that of red tide early development tracing<sup>49</sup> and oil spill source detection<sup>50</sup>. In the atmosphere, examples for instance are the identification of sources of greenhouse gases emission<sup>51</sup> and toxic agents release<sup>52</sup>.

## ACKNOWLEDGMENTS

The comments by an anonymous reviewer help improved the paper. We thank Ilja Klebanov and Tim Sullivan for the benefit of discussions on Bayesian inference. The NOAA/GDP dataset is available at <http://www.aoml.noaa.gov/phod/dac/>. Some techniques employed in this paper were discussed by FJBV, MJO and PK during the Erwin Schrödinger Institute (ESI) Programme on Mathematical Aspects of Physical Oceanography. Support from ESI is greatly appreciated. The work of PK was partially supported by the Deutsche Forschungsgemeinschaft (DFG) through the Priority Programme SPP 1881 Turbulent Superstructures and through the CRC 1114 Scaling Cascades in Complex Systems project A01.

- <sup>1</sup>C. Ashton, A. S. Bruce, G. Colledge, and M. Dickinson, “The search for MH370,” *The Journal of Navigation* **68**, 1–22 (2015).
- <sup>2</sup>I. D. Holland, “MH370 burst frequency offset analysis and implications on descent rate at end of flight,” *IEEE Aerospace and Electronic Systems Magazine* **33**, 24–33 (2018).
- <sup>3</sup>Australian Transport Safety Bureau, “Assistance to Malaysian Ministry of Transport in support of missing Malaysia Airlines flight MH370 on 7 March 2014 UTC,” Investigation number: AE-2014-054 (2018).
- <sup>4</sup>V. J. García-Garrido, A. M. Mancho, S. Wiggins, and C. Mendoza, “A dynamical systems approach to the surface search for debris associated with the disappearance of flight MH370,” *Non-linear Processes in Geophysics* **22**, 701–712 (2015).
- <sup>5</sup>R. Corrado, G. Lacorata, L. Palatella, R. Santoleri, and E. Zambianchi, “General characteristics of relative dispersion in the ocean,” *Scientific Reports* **7**, 46291 (2017).
- <sup>6</sup>J. A. Trinanes, M. J. Olascoaga, G. J. Goni, N. A. Maximenko, D. A. Griffin, and J. Hafner, “Analysis of flight MH370 potential debris trajectories using ocean observations and numerical model results,” *Journal of Operational Oceanography* **9**, 126–138 (2016).
- <sup>7</sup>O. Nesterov, “Consideration of various aspects in a drift study of MH370 debris,” *Ocean Sci.* **14**, 387–402 (2018).
- <sup>8</sup>D. Griffin, P. Oke, and E. Jones, “The search for MH370 and ocean surface drift - Part II, 13 april 2017,” CSIRO Oceans and Atmosphere, Australia, Report no. EP172633, prepared for the Australian Transport Safety Bureau, available at: <https://www.atsb.gov.au> (2017).
- <sup>9</sup>N. Maximenko, J. Hafner, J. Speidel, and K. L. Wang, “IPRC Ocean Drift Model Simulates MH370 Crash Site and Flow Paths,” International Pacific Research Center, School of Ocean and Earth Science and Technology at the University of Hawaii, 4 August 2015, available at: [http://iprc.soest.hawaii.edu/news/MH370\\_debris/IPRC\\_MH370\\_News.php](http://iprc.soest.hawaii.edu/news/MH370_debris/IPRC_MH370_News.php) (2015).
- <sup>10</sup>M. van Ormondt and F. Baart, “Aircraft debris MH370 makes Northern part of the search area more likely,” *Deltares News*, 31 July 2015, available at: <https://www.deltares.nl/en/news/aircraft-debris-mh370-makes-northern-part-of-the-search> (2015).
- <sup>11</sup>E. Jansen, G. Coppin, and N. Pinardi, “Drift simulation of MH370 debris using superensemble techniques,” *Nat. Hazards Earth Syst. Sci.* **16**, 1623–1628 (2016).
- <sup>12</sup>J. Durgadoo and A. Biastoch, “Where is MH370?” GEOMAR Helmholtz Centre for Ocean Research Kiel, 28 August 2015, available at: <http://www.geomar.de> (2015).
- <sup>13</sup>S. Davey, N. Gordon, I. Holland, M. Rutten, and J. Williams, *Bayesian methods in the search for MH370*, SpringerBriefs in Electrical and Computer Engineering (Springer Open, 2016) p. 114.
- <sup>14</sup>P. de Deckker, “Chemical investigations on barnacles found attached to debris from the MH370 aircraft found in the Indian Ocean,” Appendix F in ATSB report: “The Operational Search for MH370,” AE-2014-054, 3 October 2017, available at: <http://www.atsb.gov.au>, last access: 3 October 2017. (2017).
- <sup>15</sup>A. Lasota and M. C. Mackey, *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*, 2nd ed., Applied Mathematical Sciences, Vol. 97 (Springer, New York, 1994).
- <sup>16</sup>P. Brémaud, *Markov chains*, Gibbs Fields Monte Carlo Simulation Queues, Texts in Applied Mathematics, Vol. 31 (Springer, New York, 1999).
- <sup>17</sup>J. Norris, *Markov Chains* (Cambridge University Press, 1998).
- <sup>18</sup>F. J. Beron-Vera, M. J. Olascoaga, G. Haller, M. Farazmand, J. Triñanes, and Y. Wang, “Dissipative inertial transport patterns near coherent Lagrangian eddies in the ocean,” *Chaos* **25**, 087412 (2015).
- <sup>19</sup>F. J. Beron-Vera, M. J. Olascoaga, and R. Lumpkin, “Inertia-induced accumulation of flotsam in the subtropical gyres,” *Geophys. Res. Lett.* **43**, 12228–12233 (2016).
- <sup>20</sup>S. M. Ulam, *A Collection of Mathematical Problems*, Interscience tracts in pure and applied mathematics (Interscience, 1960).
- <sup>21</sup>Z. Kovács and T. Tél, “Scaling in multifractals: Discretization of an eigenvalue problem,” *Phys. Rev. A* **40**, 4641–4646 (1989).
- <sup>22</sup>P. Koltai, *Efficient approximation methods for the global long-term behavior of dynamical systems – Theory, algorithms and examples*, Ph.D. thesis, Technical University of Munich (2010).
- <sup>23</sup>P. Miron, F. J. Beron-Vera, M. J. Olascoaga, G. Froyland, P. Pérez-Brunius, and J. Sheinbaum, “Lagrangian geography of the deep Gulf of Mexico,” *J. Phys. Oceanogr.* doi:**0.1175/JPO-D-18-0073.1** (2018).
- <sup>24</sup>F. A. Schott and J. P. McCreary, Jr., “The monsoon circulation of the Indian Ocean,” *Progress In Oceanography* **51**, 1–123 (2001).
- <sup>25</sup>R. Lumpkin and M. Pazos, “Measuring surface currents with Surface Velocity Program drifters: the instrument, its data and some recent results,” in *Lagrangian Analysis and Prediction of Coastal and Ocean Dynamics*, edited by A. Griffa, A. D. Kirwan, A. Mariano, T. Özgökmen, and T. Rossby (Cambridge University Press, 2007) Chap. 2, pp. 39–67.
- <sup>26</sup>P. P. Niiler and J. D. Paduan, “Wind-driven Motions in the north-eastern Pacific as measured by Lagrangian drifters,” *J. Phys. Oceanogr.* **25**, 2819–2830 (1995).
- <sup>27</sup>R. Lumpkin, S. A. Grodsky, L. Centurioni, M.-H. Rio, J. A. Carton, and D. Lee, “Removing spurious low-frequency variability in drifter velocities,” *J. Atm. Oce. Tech.* **30**, 353–360 (2012).
- <sup>28</sup>P. Miron, F. J. Beron-Vera, M. J. Olascoaga, J. Sheinbaum, P. Pérez-Brunius, and G. Froyland, “Lagrangian dynamical geography of the Gulf of Mexico,” *Scientific Reports* **7**, 7021 (2017).
- <sup>29</sup>M. J. Olascoaga, P. Miron, C. Paris, P. Pérez-Brunius, R. Pérez-Portela, R. H. Smith, and A. Vaz, “Connectivity of Pulley Ridge with remote locations as inferred from satellite-tracked drifter trajectories,” *Journal of Geophysical Research* **123**, 5742–5750 (2018).
- <sup>30</sup>G. Froyland and K. Padberg, “Almost-invariant sets and invariant manifolds — connecting probabilistic and geometric descriptions of coherent structures in flows,” *Physica D* **238**, 1507–1523 (2009).
- <sup>31</sup>J. H. LaCasce, “Statistics from Lagrangian observations,” *Progr.*



- Oceanogr. **77**, 1–29 (2008).
- <sup>32</sup>A. N. Maximenko, J. Hafner, and P. Niler, “Pathways of marine debris derived from trajectories of Lagrangian drifters,” *Mar. Pollut. Bull.* **65**, 51–62 (2012).
- <sup>33</sup>R. McAdam and E. van Sebille, “Surface connectivity and inter-ocean exchanges from drifter-based transition matrices,” *Journal of Geophysical Research: Oceans* **123**, 514–532 (2018).
- <sup>34</sup>E. van Sebille, E. H. England, and G. Froyland, “Origin, dynamics and evolution of ocean garbage patches from observed surface drifters,” *Environ. Res. Lett.* **7**, 044040 (2012).
- <sup>35</sup>G. Froyland, R. M. Stuart, and E. van Sebille, “How well-connected is the surface of the global ocean?” *Chaos* **24**, 033126 (2014).
- <sup>36</sup>C. S. Hsu, *Cell-to-cell mapping. A Method of Global Analysis for Nonlinear Systems*, Applied Mathematical Sciences, Vol. 64 (Springer-Verlag, New York, 1987) p. 354.
- <sup>37</sup>M. Dellnitz and O. Junge, “On the approximation of complicated dynamical behavior,” *SIAM J. Numer. Anal.* **36**, 491–515 (1999).
- <sup>38</sup>G. Froyland, “Statistically optimal almost-invariant sets,” *Physica D* **200**, 205–219 (2005).
- <sup>39</sup>R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, 1990).
- <sup>40</sup>P. Koltai, “A stochastic approach for computing the domain of attraction without trajectory simulation,” in *Dynamical Systems, Differential Equations and Applications, 8th AIMS Conference. Suppl.*, Vol. 2 (2011) pp. 854–863.
- <sup>41</sup>L. M. Beal, W. P. M. de Ruijter, A. Biastoch, R. Zahn, and SCOR/WCRP/IAPSO Working Group 136, “On the role of the Agulhas system in ocean circulation and climate,” *Nature* **472**, 429–436 (2011).
- <sup>42</sup>W. M. Bolstad and J. M. Curran, *Introduction to Bayesian statistics* (John Wiley & Sons, 2016).
- <sup>43</sup>E. W. Dijkstra, “A note on two problems in connexion with graphs,” *Numerische Mathematik* **1**, 269–271 (1959).
- <sup>44</sup>R. W. Floyd, “Algorithm 97: shortest path,” *Communications of the ACM* **5**, 345 (1962).
- <sup>45</sup>E. Ser-Giacomi, R. Vasile, E. Hernández-García, and C. López, “Most probable paths in temporal weighted networks: An application to ocean transport,” *Physical review E* **92**, 012818 (2015).
- <sup>46</sup>Australian Transport Safety Bureau, “Mh370 - definition of underwater search areas, 26 June 2014,” available at: <http://www.atsb.gov.au>.
- <sup>47</sup>J. H. E. Cartwright, U. Feudel, G. Károlyi, A. de Moura, O. Piro, and T. Tél, “Dynamics of finite-size particles in chaotic fluid flows,” in *Nonlinear Dynamics and Chaos: Advances and Perspectives*, edited by M. Thiel et al. (Springer-Verlag Berlin Heidelberg, 2010) pp. 51–87.
- <sup>48</sup>A. C. Bagtzoglou and J. Atmadja, “Mathematical methods for hydrologic inversion: The case of pollution source identification,” in *Water Pollution: Environmental Impact Assessment of Recycled Wastes on Surface and Ground Waters; Engineering Modeling and Sustainability*, edited by T. A. Kassim (Springer Berlin Heidelberg, Berlin, Heidelberg, 2005) pp. 65–96.
- <sup>49</sup>M. J. Olascoaga, F. J. Beron-Vera, L. E. Brand, and H. Koçak, “Tracing the early development of harmful algal blooms on the West Florida Shelf with the aid of Lagrangian coherent structures,” *J. Geophys. Res.* **113**, C12014 (2008).
- <sup>50</sup>B. G. Gautama, G. Mercier, R. Fablet, and N. Longepe, “Lagrangian-based Backtracking of Oil Spill Dynamics from SAR Images: Application to Montara Case,” in *Living Planet Symposium*, ESA Special Publication, Vol. 740 (2016) p. 214.
- <sup>51</sup>F. Hourdin and O. Talagrand, “Eulerian backtracking of atmospheric tracers. I: Adjoint derivation and parametrization of subgrid-scale transport,” *Q. J. R. Meteorol. Soc.* **132**, 567–583 (2006).
- <sup>52</sup>K. S. Rao, “Source estimation methods for atmospheric dispersion,” *Atmospheric Environment* **41**, 6964 – 6973 (2007).