Using Drifter Data to Estimate Ocean Diffusivity

Matthew Davison

Overview

Mathematical modelling is a vital tool to understand the world around us. Many real life natural phenomena we experience on a daily basis are complex and not completely understood. The movement of freely floating objects (such as an inflatable raft, marine debris or plankton) around the ocean is an example of a natural phenomena where research is still being done to improve understanding of it.

To predict the motion of a freely floating object we require a mathematical model that approximates the real life motion of the ocean. The model that is typically used is a so-called "advection-diffusion" model. Advection is how mass in a fluid is transported to a new region. For example, a person floating in the water at the beach will be pushed around by the tidal flow of the water. Diffusion is how mass in a fluid spreads. For example, a drop of food colouring in a glass of water will eventually spread and turn the whole glass of water that colour.

The motivation for being able to predict the movement of objects in the ocean stems from several areas. One application is in "search and rescue" missions. In 2014, the Mayalsia Airlines Flight MH370 went missing in the Indian Ocean and ocean modelling was used in an attempt to locate the wreckage. Another application is in understanding water pollution. Some cosmetic products contain small balls of plastic called "microbeads" which travel through sewer systems and eventually into rivers and oceans. By modelling the movement of these microbeads we can learn more about what impact this has on the environment.

In this report we focus only on the diffusion part of these ocean models and ignore the advection part. This is similar to assuming that water in the ocean behaves the same way water behaves in a bath tub. We present a method of estimating the diffusivity of the ocean, that is the rate at which freely floating objects spread out in the ocean.

We use data from the Global Drifter Program (GDP) ran by the National Oceanic and Atmospheric Administration (NOAA). The GDP consists of a collection of buoys called "drifters" equipped with sensors and measurement devices floating in the ocean. This is discussed further in Section 2.

Section 3 presents a method of estimating the average distance a floating object will move over time. Section 4 uses this estimate and linear regression (finding a line of best fit) to find an estimate for diffusivity. The estimate of diffusivity we arrive at is $\kappa = 4.3099 \times 10^4 m^2/s$. This value is compared with results from other research and in Section 5 we discuss how good our estimate is.

In Section 6 we suggest improvements that could be made to the method used in Section 3 and Section 4 of this report. In summary these suggestions are to:

- Modify the method to derive different diffusivity values for different regions of the ocean.
- Incorporate advection in the diffusivity calculation as it is known advection can influence the diffusivity values.

1 Introduction

The need to model the motion of particles in the ocean appears in many real life applications. One example of the need for such a model is given by Chapline (1960) where the author models the drift of a small craft. Breivik et al. (2013) gives an overview of other historical applications in search and rescue methods where the focus is predicting how freely floating objects like an inflatable emergency raft could move away from the location of a wreckage. Recently, ocean models were used to plan a search for Flight MH370, which went missing in the Indian Ocean in 2014 (Griffin et al., 2016).

Another application relates to water pollution, and pollutants such as microplastics and surface marine debris (Sebille et al., 2012). Microplastics are non-degradable and remain in the ocean for many years. This has become a large concern and sparked a discussion about the lack of regulations on the use of nanoparticles in cosmetics and foods (Hernandez et al., 2017).

Diffusivity is how particles spread in fluids and so is necessary information for a model of the ocean. In Section 2 we introduce the Global Drifter Program (GDP) and the data we will use. Use of this kind of data is common in the field of oceanography (Griffin et al., 2016; Sebille et al., 2012). In Section 3 we define an estimate of the average squared distance, which is used in Section 4 to calculate an estimate of diffusivity in the ocean.

This estimate can be used in simple models of the ocean. For example, Karapanagioti and Kalavrouziotis (2019) present a one-dimensional advection-diffusion model of the quantity of microplastics in the ocean (see Equation 1). D in Equation 1 represents the diffusivity and could be estimated using the methods in this report.

$$\frac{\partial n}{\partial t} + v\frac{\partial n}{\partial x} = D\frac{\partial^2 n}{\partial x^2} - Kn \tag{1}$$

2 Global Drifter Program

The Global Drifter Program consists of a collection of buoys equipped with sensors and measurement devices floating round the ocean collecting data. These buoys are called "drifters".

For every drifter in our data set we have the latitudinal and longitudinal position for every 6 hours from its deployment up to either the date of that drifters retirement or 5th July 2020. It also includes the surface sea temperature, eastward-westward velocity, northward-southward velocity, speed and variances for location and temperature. The data set used in this report contains data from 9988 unique drifters (Lumpkin and Centurioni, 2019). We note that many of these were released on different dates and times.

In Figure 1 we show a trace plot containing the trajectories for 100 randomly selected drifters. As can be seen from Figure 1, we have data for nearly every part of the ocean although some areas are more dense with drifters.

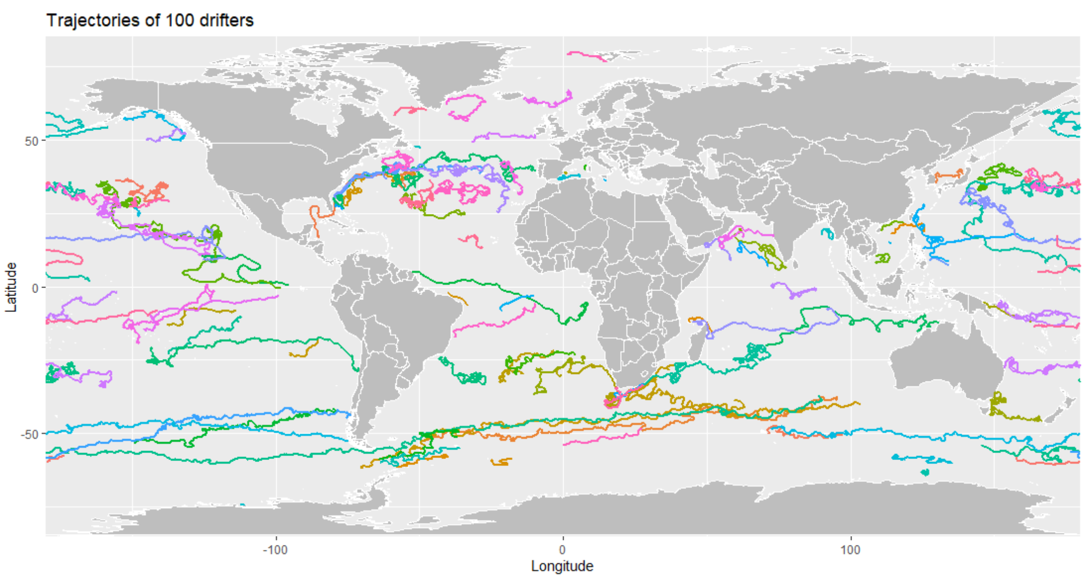

Figure 1: Trajectories for 100 random drifters in the six-hourly interpolated data

3 Average Squared Distance

An estimate of the average squared distance a particle travels over time can be measured in several valid ways. In theory, the trajectories of a particle in the ocean can be considered as positions on the complex plane.

Lilly et al. (2017) let $z(t) = u(t) + iv(t)$ be a potentially non-stationary, complex-valued, zero-mean random process, where $i = \sqrt{-1}$. In this context, $u(t)$ is the eastward-westward velocity and $v(t)$ is the northward-southward velocity. Lilly et al. (2017) define the trajectory on the complex plane as the time integral of $z(t)$, denoted as

$$r(t) \equiv \int_0^t z(\tau)d\tau. \tag{2}$$

In this report we only consider $t \geq 0$, we let $t = 0$ correspond to the date of drifter deployment and $t > 0$ be the number of days after this date. Notice also that the definition given in Equation 2 implies $r(0) = 0$. For the next part of this section we present a method to estimate $|r(t)|$ for each drifter using the discrete time GDP data.

We assume that the Earth is perfectly spherical with a radius of 6378137 metres. We also measure distance using the haversine great-circle distance between two consecutive observations. This formula is more accurate than the standard great-circle distance when considering small distances (Sinnott, 1984). We denote $h(x, y)$ as the haversine great-circle distance between coordinates x and y.

Let $D = \{d_1, \dots, d_{9988}\}$ be the set of drifters and let $p_{d,t} \in \mathbb{R}^2$ denote the coordinates of drifter $d \in D$ at time t. Because drifters were deployed at different times define T_d as the number of days where data is available for drifter d. Define $T_{max} \equiv \max_{d \in D}(T_d)$, and defining $T = \{0, 0.25, 0.5, \dots, T_{max}\}$.

$\hat{r}_d(t)$ will be our approximation to $|r(t)|$ for drifter $d \in D$ and we define it for each $t \in T$ as

$$\hat{r}_d(t) \equiv h(p_{d,0}, p_{d,t}). \tag{3}$$

This satisfies the property $\hat{r}_d(0) = 0$ for all $d \in D$. We use this to estimate average squared distance, which is denoted as $\mathbb{E}\{|r(t)|^2\}$ in Lilly et al. (2017). Let $A_t = \{d \in D : t \leq T_d\}$ be the set of all active drifters up to time t. Define average squared distance at time t as

$$ASD(t) = \frac{1}{|A_t|}\left(\sum_{d \in A_t} \hat{r}_d(t)^2\right), \tag{4}$$

where $|A_t|$ is the number of active drifters in time t. Figure 2 shows the average squared distance over time and Figure 3 shows how drifter count varies over the same time scale.

4 Estimating Ocean Diffusivity

To allow us to estimate diffusivity on a global scale, we first make two assumptions about the Earth's oceans.

1. Diffusivity is constant.

2. There is no mean flow.

Assumption 1 is assuming diffusivity doesn't depend on where in the ocean a particular drifter is whereas in reality diffusivity can depend on location for various reasons (LaCasce, 2010, 2008). This assumption also means diffusivity does not depend on time. In the context of Lilly et al. (2017) this equates to $\beta = 1$, and the motion of the drifters is referred to as diffusive ($\beta > 1$ is super-diffusive, $\beta < 1$ is sub-diffusive).

Assumption 2 uses the term "mean flow". Mean flows are caused by forces such as the spin of the Earth and gravitational forces and can be seen as tides and drifts such as the North Atlantic Drift. These are problematic as they directly affect the diffusivity values (LaCasce, 2008). De Dominicis et al. (2012) describe the drifter velocities as having a "mean flow" component and a "turbulent" component and it is the turbulent component we would isolate and use to calculate diffusivity.

To give a simple analogy for the above assumptions, imagine that the Earth's oceans behave similarly to water in a bathtub.

Diffusivity using all of the data

Continuing our adoption of the notation in Lilly et al. (2017) we introduce the isotropic diffusivity (which we call "diffusivity") as

$$\kappa \equiv \frac{1}{4}\frac{d}{dt}\mathbb{E}\left\{|r(t)|^2\right\}. \tag{5}$$

This is consistent with what is seen in other literature with varied notation (LaCasce, 2010; De Dominicis et al., 2012; LaCasce, 2008; Beron-Vera and LaCasce, 2016). Since we are assuming that κ doesn't depend on time we can multiply by four on both sides and integrate over time giving

$$4\kappa t + c = \mathbb{E}\left\{|r(t)|^2\right\}, \tag{6}$$

where $c \in \mathbb{R}$. By letting $t = 0$ and using $r(0) = 0$ it can be deduced that $c = 0$, therefore

$$4\kappa t = \mathbb{E}\left\{|r(t)|^2\right\}. \tag{7}$$

The next step is a linear regression with time t as our explanatory variable and $ASD(t)$ as our dependent variable. Precisely, we wish to find the best constant a and b such that

$$ASD(t) = bt + a. \tag{8}$$

Since $ASD(0) = 0$ we will force $a = 0$. Using all of the data we find $b = 2.2791 \times 10^4 \text{km}^2/\text{day}$ and because $4\kappa = b$ this corresponds to $\kappa = 6.5948 \times 10^4 \text{m}^2/\text{s}$. Figure 2 shows that this is a poor fit. In the next section we discuss why estimating diffusivity for such a long time horizon may not be sensible.

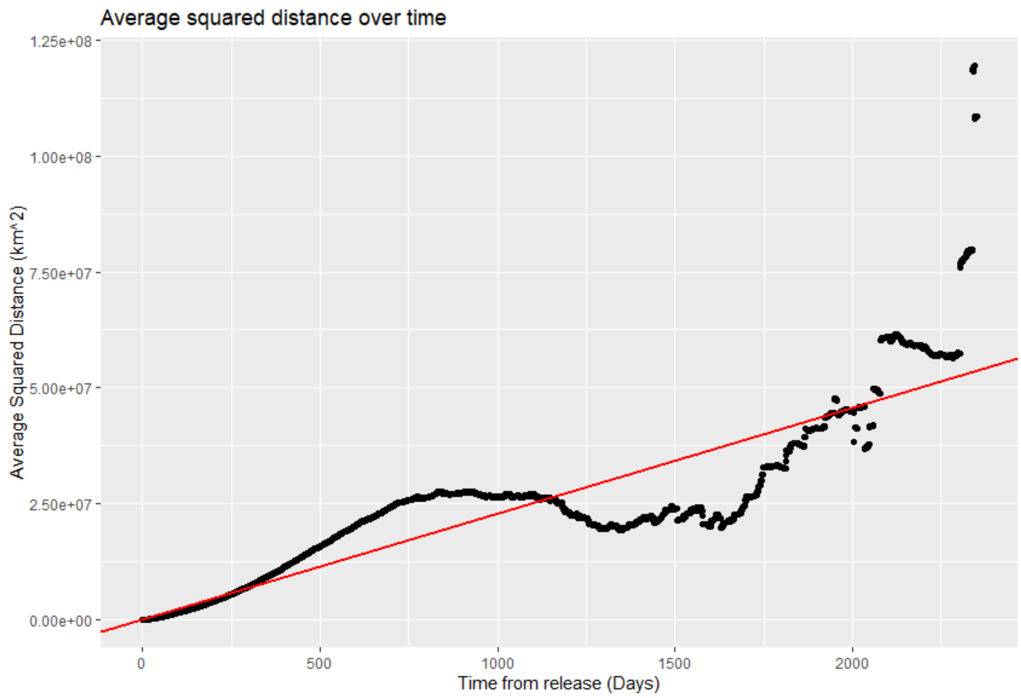

Figure 2: Average squared distance against time after release. Red linear regression line is over all of the data.

Choosing a subset of the data

We claim that it is sensible for us to ignore the first 5 days of travel in our linear regression. LaCasce (2010) found that separations smaller than 100km were better modelled as either an exponential or cubic relationship and this is consistent with the behaviour we see close to the origin in Figure 2. LaCasce (2010) suggests that separations larger than 100km increase linearly in time. For our data, average distance is greater than 100km for $t > 5$.

We also claim that it is sensible for us to ignore times greater than 75 days. As time grows the number of drifters we have data for decays exponentially as seen in Figure 3. The consequence of this for average squared distance is that the data gets noisier, individual drifters have a greater affect on the averaging. It is therefore sensible to ignore data calculated using insufficiently many drifters. Times less than 75 days use observations from at least 80% of the drifters.

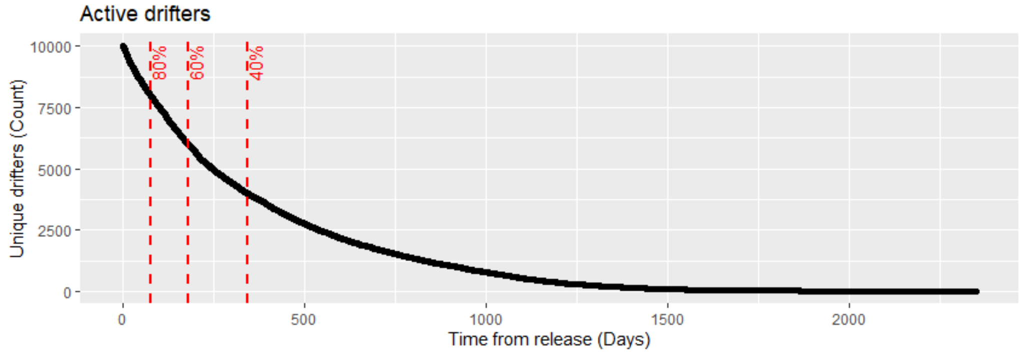

Figure 3: Drifter count over time. Red lines mark percentage of active drifters.

There is a lack of literature which study the drifter data for more than a hundred days after release. Some consider up to around 100 days (LaCasce, 2010), but some consider up to 25 days (LaCasce, 2008) or even up to only a few hours after launch (De Dominicis et al., 2012).

Finally, our assumptions about diffusivity only apply locally. For example, we know the assumption of constant diffusivity is invalid and that in reality diffusivity does vary spatially. Assumption 1 allows us to ignore this over short time windows. However, when we take longer time windows such as $t > 75$ this assumption becomes even more invalid as the drifter has likely travelled over a wide range of diffusivity values.

$ASD(t)$ for $t \in [5,75]$ looks roughly linear so fitting a linear model to only this section of the data should give us a diffusivity we can be more confident in.

Diffusivity using a subset of the data

Now we have identified a subset of the data to consider we perform a linear regression once again. This time round we don't force $a = 0$ because it is the rate of change (the gradient b) of this data that is proportional to diffusivity. Allowing a to change gives the linear regression flexibility to fit a gradient closer to the subset of data.

Using the data for $t \in [5,75]$, we find that $b = 1.4895 \times 10^4 \text{km}^2/\text{day}$ which corresponds to

$$\kappa = 4.3099 \times 10^4 \text{m}^2/\text{s}. \tag{9}$$

This κ value is smaller by $2 \times 10^4 \text{m}^2/\text{s}$ compared to what we found earlier by considering all of the data. Figure 4 confirms that the linear fit to our subset of data is a lot better than the linear fit to all of the data as the data is much closer to the line. Therefore, we are more confident in this κ value.

Comparison with other literature

Zhurbas and Oh (2004) derived lateral diffusivities in the North Atlantic ranging between $2 \times 10^3 \text{m}^2/\text{s}$ up to $2.6 \times 10^4 \text{m}^2/\text{s}$. Neither of our estimates fit within this range although they are of a similar order in magnitude to the higher diffusitivities.

De Dominicis et al. (2012) looked at specific parts of the ocean also. In particular they looked at both the Liguarian Sea and Adriatic Sea, averaging over 7 days to 4 months. They split their diffusivity calculation into two components (along mean flow and across mean flow) and achieved values between $3.1 \times 10^6 \text{m}^2/\text{s}$ and $2.7 \times 10^7 \text{m}^2/\text{s}$. These are much higher than our estimates.

We work backwards from the linear relationship given by LaCasce (2010) who used data from the POLEWARD experiment in the Nordic Seas (Koszalka et al., 2009). In the same way we calculated κ from the gradient, we believe the diffusivity proposed by LaCasce (2010) is around $2.141 \times 10^3 \text{m}^2/\text{s}$. This is similar to Zhurbas and Oh (2004) who looked at similar areas but didn't have access to POLEWARD data at the time.

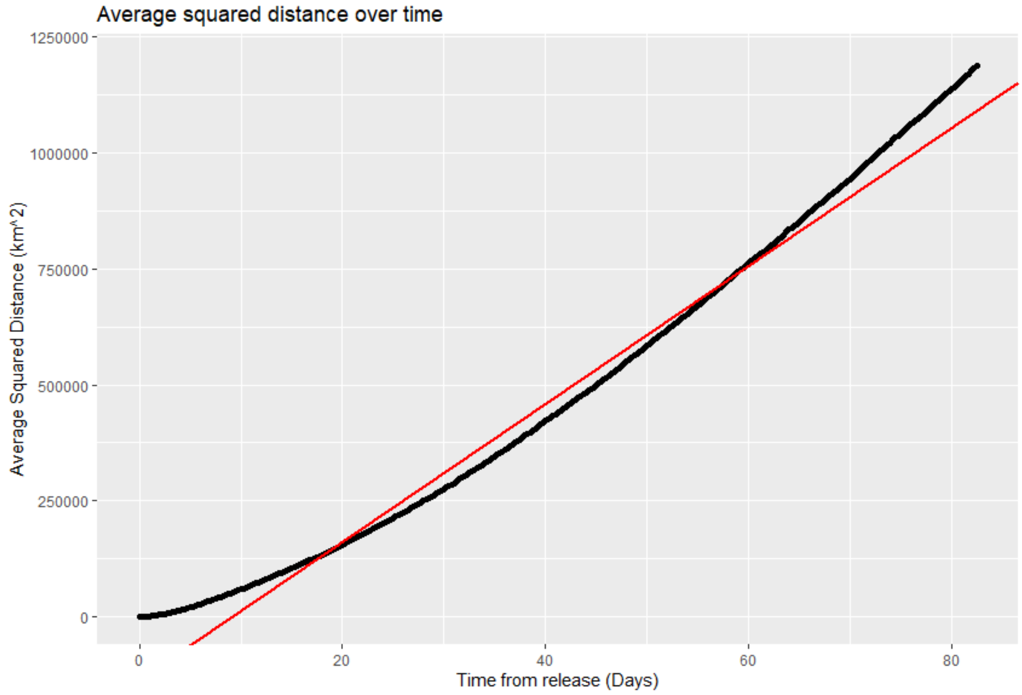

Figure 4: Average squared distance against time after release. Red linear regression line considers only data after day 5 up to day 75.

5 Conclusion

Due to the reality of large variation in diffusivity over the entire globe a single point estimate is a crude way to sum diffusivity up. Despite this, based on our review of the literature, our global estimate of diffusivity ($\kappa = 4.3099 \times 10^4 \mathrm{m}^2/\mathrm{s}$) seems to be of the correct order in magnitude. The logic behind this is the Liguarian Sea and Adriatic sea have a larger diffusivity than the North Atlantic, but the North Atlantic is much bigger. Therefore an average by area could be in the order of $1 \times 10^4 \mathrm{m}^2/\mathrm{s}$.

For a simplified model of the Earths ocean, our constant diffusivity presented in this report would be sufficient when diffusivity isn't the main focus. It would likely not be sufficient for a detailed regional study where correct estimates are critical (for example, modelling for search and rescue missions).

6 Further Research

If all of this were to be repeated, the first thing that would be improved upon is taking into account mean flow. Various literature such as LaCasce (2008); De Dominicis et al. (2012); Koszalka and LaCasce (2010) consider removing mean flow and then calculating diffusivity. The difficulty with this is that mean flow varies across the ocean. A potential approach is to bin the data and work out the mean flow for each bin but these means depend on bin size and shape, and the variances depend on number of observations within each bin. Koszalka and LaCasce (2010) propose binning using k-means clustering rather than the typical square grids. Koszalka and LaCasce (2010) claim this method is better at capturing the mean flow. These approaches use velocity data to estimate diffusivity whereas in this report distances were used.

Another aspect that could be improved is the model fitting. In this report we only considered linear regression, and only qualitatively assessed the fit. There are several measures to mathematically verify if a linear fit is good beyond looking at the figures. We could have assessed the fit using a plot of the residual values (a residual value being how far away a point is from the line of best fit) as these can expose non-linearity and outliers in the data. Similarly, we could have calculated an R-squared value for each fit. An R-squared value is between 0% and 100% and usually the closer to 100% the better the fit. Figure 4 still exhibits some non-linearity at the lower times after release.

For different lengths of time after release the average squared distance behaves in a non-linear fashion. This is recognised by LaCasce (2010) and Beron-Vera and LaCasce (2016) name three different models of diffusion for different "regimes" (see Table 1).

	Lundgren	Richardson	Rayleigh
$\mathbb{E}\{\lvert r(t)\rvert^2\}$	$r_0 \exp(8t/T)$	$\sim 5.2675\beta^3 t^3$	$\sim 4\kappa t$

Table 1: Regimes named by Beron-Vera and LaCasce (2016). This report used the "Rayleigh regime".

If we wished to relax the assumptions made in Section 4, the constant diffusion coefficient, κ, could be used to perform significance tests on how correct these assumptions are to make and how much difference there is if we don't make them.

In summary, further research on this topic should do two things:

1. Model the diffusivity both spatially and temporally, rather than assume diffusivity is constant as we did in this report.

2. Account for mean flow in an appropriate way.

References

Beron-Vera, F. and LaCasce, J. (01 Jul. 2016). Statistics of simulated and observed pair separations in the Gulf of Mexico. *Journal of Physical Oceanography*, 46(7):2183 – 2199.

Breivik, Ø., Allen, A., Maisondieu, C., and Olagnon, M. (2013). Advances in search and rescue at sea. *Ocean Dynamics*, 63(1):83–88.

Chapline, W. E. (1960). Estimating the drift of distressed small craft. *Coast Guard Alumni Association Bulletin*, 22(2):39–42.

De Dominicis, M., Leuzzi, G., Monti, P., Pinardi, N., and Poulain, P.-M. (2012). Eddy diffusivity derived from drifter data for dispersion model applications. *Ocean Dynamics*, 62(9):1381–1398.

Griffin, D., Oke, P., and Jones, E. (2016). The search for MH370 and ocean surface drift.

Hernandez, L. M., Yousefi, N., and Tufenkji, N. (2017). Are there nanoplastics in your personal care products? *Environmental Science & Technology Letters*, 4(7):280–285.

Karapanagioti, H. and Kalavrouziotis, I. K. (2019). *Microplastics in Water and Wastewater*. IWA Publishing, Ashland.

Koszalka, I. and LaCasce, J. (2010). Lagrangian analysis by clustering. *Ocean Dynamics*, 60:957–972.

Koszalka, I., LaCasce, J., and Orvik, K. (2009). Relative dispersion in the Nordic Seas. *Journal of Marine Research*, 67:411–433.

LaCasce, J. (2008). Statistics from Lagrangian observations. *Progress in Oceanography*, 77(1):1–29.

LaCasce, J. (2010). Relative displacement probability distribution functions from balloons and drifters. *Journal of Marine Research*, 68:433–457.

Lilly, J. M., Sykulski, A. M., Early, J. J., and Olhede, S. C. (2017). Fractional Brownian motion, the Matérn process, and stochastic modeling of turbulent dispersion. *Nonlinear Processes in Geophysics*, 24(3):481–514.

Lumpkin, R. and Centurioni, L. (2019). NOAA Global Drifter Program quality-controlled 6-hour interpolated data from ocean surface drifting buoys. [buoydata_15001_jun20].

Sebille, E., England, M., and Froyland, G. (2012). Origin, dynamics and evolution of ocean garbage patches from observed surface drifters. *Environmental Research Letters*, 7:044040.

Sinnott, R. W. (1984). Virtues of the Haversine. *Sky and Telescope*, 68(2):158.

Zhurbas, V. and Oh, I. (2004). Drifter-derived maps of lateral diffusivity in the pacific and atlantic oceans in relation to surface circulation patterns. *Journal of Geophysical Research: Oceans*, 109(C5).