Author

Andrea Steinberger has always been fascinated by the ocean. Her favorite hobbies include surfing, sailing and scuba diving, so it's no surprise that her love of the water has influenced her undergraduate studies. After taking Physical Oceanography with Dr. Francois Primeau, she started working with him on the relatively new area of forecasting and now-casting the ocean. Since completing this research project, she feels prepared and enthusiastic for her graduate studies in physical/chemical oceanography, which she hopes to begin next fall. Andrea has also been conducting studies for the Southern California Coastal Water Research Project since 1999, and is currently working on acquiring a skipper's license and becoming a master diver.

Key Terms

- Advective/Diffusive Flow
- Data Assimilation
- Eulerian Circulation Models
- Global Climate Models
- Lagrangian Drifters
- Method of Maximum Likelihood
- Ocean Circulation

Estimating Eddy Diffusivity and Advective Flow from Simulated Lagrangian Drifter Data Using the Method of Maximum Likelihood

Andrea Steinberger
Chemistry and Earth & Environmental Science

Abstract

A new method for estimating advection and diffusion in an ocean basin from Lagrangian data is developed using a simple box-model approach and the method of maximum likelihood. This method estimates both advective and diffusive flow concurrently, which is not possible using existing methods. Computer simulations of one- and two-box ocean models are used to generate synthetic Lagrangian data on which the method is tested. In these models, true flow parameters are defined then recovered by maximizing the likelihood function of the one- and two-box probability models. The relative likelihood function is used to extract information regarding the consistency of the parameter estimates, while ensemble means are used to determine the method bias. For both the one- and two-box probability models, the method returned flow estimates that were consistent and unbiased in the limit that the number of Lagrangian drifters, N, became large. Limitations to concurrent estimation of the advective and diffusive flows were observed for small N, but were absent for large N. Techniques used in this study show capacity for use in planning Lagrangian drifter studies. The method shows potential for use in optimizing data assimilation in ocean circulation models, which are subsequently used in climate models to study global climate change.

Faculty Mentor

The positions of thousands of surface and subsurface drifters deployed throughout the oceans are tracked daily in order to monitor the oceans' circulation. The drifter trajectories, when combined with an ocean circulation model during data assimilation, provide insight into such significant variables as temperature and velocity. Data assimilation methods require the Lagrangian drifter position data to be converted into an Eulerian velocity field on the fixed model grid. An unfortunate result of this conversion is that some information within the data gets thrown away. In an effort to extract more of the useful information, Andrea Steinberger successfully tested a new method, using the drifter data to estimate a sub-gridscale eddy-diffusivity in addition to the large-scale flow field. Focusing on idealized models, Andrea developed an understanding of the limitations and pitfalls of the method's implementation. Encouraged by Andrea's results, our group is currently extending this method to more complex ocean models.

Francois W. Primeau
School of Physical Sciences

Introduction

The ability to describe and model the world's oceans is limited by the logistical challenge of observing the ocean on a global scale (WOCE, 1997). As a result, ocean models are much less advanced than their atmospheric counterparts (Pinardi et al., 2002). Atmospheric modeling enjoyed a veritable renaissance age in the latter half of the 20th century through the advance of observational and computational technology that allowed the development of realistic atmospheric models. However, similar advances have yet to be made in the field of oceanography, though efforts have accelerated since the 1990s, bringing ocean modeling to a similar caliber as atmospheric modeling. Ocean models have lagged behind atmospheric models for two primary reasons. First, the size and turbulance of the ocean limits the amount of spatially and temporally representative observations that can be made (NRC, 1999). Second, the geometry of the ocean is more complex than the atmosphere and a fine-grid scale is necessary to resolve the turbulent eddy motions that characterize ocean dynamics (Hartmann, 1994). Consequently, ocean models require greater computational power than atmospheric models, which have only recently become available.

Observational and computational technology has now advanced to a degree that reliable modeling of the ocean can become a reality. The impetus to develop these models, however, is not purely esoteric. Current scientific focus on climate change and the potential impact of global warming has served to bolster the field of physical oceanography. Because the ocean is a major reservoir of heat, freshwater and carbon, it plays an integral role in the climate system (Haine and Hall, 2002; Falco et al., 2000; WOCE, 1997). Developing valid ocean models could help scientists understand, for example, the ability of the ocean to act as a sink for atmospheric carbon dioxide (CO_2), one of the leading causes of global warming. This would would help scientists develop more accurate predictions of the rate and magnitude of potential global warming due to anthropogenic emissions of CO_2.

Purely atmospheric models are incapable of predicting changes in atmospheric conditions more than a week into the future (Hartmann, 1994), and are therefore not reliable for predicting the long-term effects of global warming. Long-term predictions require incorporating the ocean's influence into climate models. This influence had been neglected until recently because the ocean was not believed to have a significant influence on the atmosphere outside tropical regions (Colling, 2001), global observations of the ocean were limited spatially and temporally, and adequate models coupling the ocean and atmosphere had not been developed (Colling, 2001). A thorough understanding of ocean dynamics could serve to extend the predictive capabilities of climate models.

Observational data is essential for developing and validating realistic ocean models. Between 1990 and 1997, scientists from 30 nations collected shipboard measurements and remote satellite observations of the world ocean during an effort known as the World Ocean Circulation Experiment (WOCE). A major component of this project was the development and deployment of autonomous Lagrangian ocean drifters designed to collect information about ocean currents and various physical parameters, such as temperature and salinity. This information was gathered for assimilation into ocean models, which could then be coupled with atmospheric models to study global climate change. The success of WOCE served as a catalyst for the field of physical oceanography.

Current ocean circulation models are based exclusively on the fundamental equations of motion that govern fluid dynamics in the ocean. Because fluid motions in the ocean are turbulent and chaotic, a modeled state of the ocean will eventually diverge from the true state. Therefore, ocean models require the assimilation of observational data to constrain the model to remain consistent with the true state of the ocean. Assimilation is essentially a technique used to "remind" the model to stay on track. Two types of data can be used for assimilation into ocean circulation models, Eulerian and Lagrangian, and both have advantages and disadvantages.

Eulerian data is logistically difficult to obtain because this approach involves measuring current velocity in the ocean at a fixed point using stationary buoys or directly from research vessels. The Eulerian method is expensive in terms of equipment and time since making current measurements across the ocean breadth and depth is difficult and achieving spatial homogeneity is nearly impossible (Rhein, 2000). The advantage of using Eulerian data is that it can be directly assimilated into ocean models, which are most easily formulated in terms of an Eulerian representation of the flow. In fact, existing data assimilation procedures can only make use of this data.

Lagrangian data is logistically more feasible to obtain. The Lagrangian method involves observing a material fluid element that moves freely with the currents in the ocean. The path of a point particle is known as a Lagrangian trajectory,

thus Lagrangian ocean drifters are used to obtain these trajectories. These drifters are deployed in the ocean and allowed to move with the currents, intermittently relaying positional information via satellite to a land-based station. Lagrangian drifters allow for greater spatial coverage of the ocean with lower operational expense. Technologically advanced drifters may also have the capacity to study the ocean at varying depths since many are designed to drift at pre-determined depths. The disadvantage to using Lagrangian data is the complexity of assimilating this mathematically disparate data into Eulerian circulation models.

The relative ease of obtaining Lagrangian data and the greater spatial resolution this data offers outweigh the benefit of using Eulerian data to validate ocean circulation models, which is the simplicity of data assimilation Using Lagrangian data necessitates the development of methods for converting Lagrangian position data into an Eulerian velocity field for assimilation into circulation models. Accepted methods of data transformation include optimal linear interpolation of Lagrangian position data and projection onto an Eulerian grid (Molcard et al., 2003; Lavender et al., 2000), and/or normal mode analysis for reconstruction of the Eulerian velocity field from the Lagrangian data (Poje et al., 2002; Toner et al., 2001). These methods fail to distinguish between advective and diffusive circulation–two types of fluid motion that define circulation in ocean models. However, an indirect method for estimating diffusivity from Lagrangian data has been studied (Belluci et al., 2001; Falco et al., 2000; Buffoni et al., 1997). This method involves reconstructing the Eulerian velocity field and then using the velocity field to integrate the advection-diffusion equation. Estimation of diffusivity, in addition to the advective flow, is important for resolving the smaller-scale transport properties of the ocean.

In this study, a new method is developed for estimating advective flow and diffusivity fields from Lagrangian data as a first step toward assimilating Lagrangian data into ocean models. This method differs from the standard approach in that the advective and diffusive flows are estimated directly and concurrently without first converting the Lagrangian position data into velocity data. The advective flow and eddy diffusivity are then parameters within the model that can be estimated using the method of maximum likelihood.

The method of maximum likelihood is a statistical technique for determining the most probable parameter, in this case the eddy diffusion and advective flow, based on observed data (Wilks, 1995). Some of the benefits of this method include a sensitivity to sample size, data dependency, and simple extension of the theory to multiparameter estimations (Whittaker, 2002). Using the method of maximum likelihood, the most probable parameter is determined by maximizing the likelihood function, which is the probability that a drifter will move from point A to point B in a given time. This method has the added benefit of being universally applicable to any region of the ocean.

To develop and test this method, one- and two-box models were formulated in which a marginal sea exchanges fluid with the open ocean (Figure 1). These models were used to generate synthetic drifter data for a fixed set of flow parameters. The method of maximum likelihood to return these flow parameters from the simulated drifter data. The method's success was tested by comparing the estimated flow parameters to the true values.

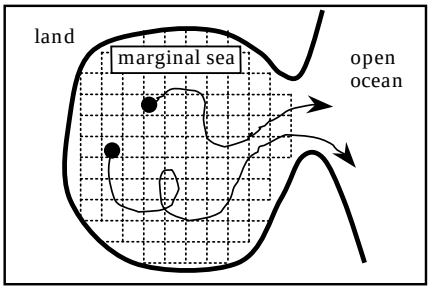

Figure 1
An *n*-box marginal sea showing examples of Lagrangian drifter trajectories as they escape to the open ocean due to circulation in the marginal sea.

Theory

One-Box Model

One-Box Probability Model

We modeled a marginal sea as a well-mixed box of uniform depth and volume V in steady state with the open ocean (Figure 2). A steady exchange of fluid between the marginal sea and the open ocean is assumed, and is designated by the flow parameter q.

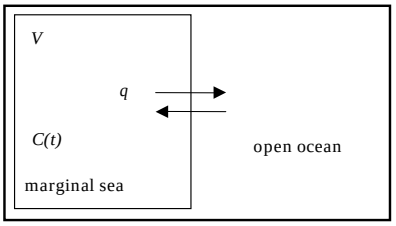

Figure 2
One-box model of a marginal sea in exchange with the open ocean via recirculation flow q, where $C(t)$ is the probability density per unit volume of drifters in the marginal sea.

Assuming a drifter is deployed in the marginal sea at time $t = 0$, the probability that it is in the marginal sea at time $t = 0$ is equal to unity. The initial condition expressed in terms

of the probability density (per unit volume) is then given by:

$$C_0 = C(t=0) = \frac{1}{V} \tag{1.1}$$

As time progresses, the float or drifter circulates out of the box under the influence of the exchange flow q. The rate of change of float density in the marginal sea can be described by the following differential equation:

$$\frac{dC(t)}{dt} = -\frac{q}{V} C(t) \tag{1.2}$$

Equation 1.2 is a first-order differential that can be solved analytically. Subject to the initial condition (Equation 1.1), its solution is:

$$\mathrm{C}(t) = \mathrm{C}_0 e^{-qt/V} = \frac{1}{V} e^{-qt/V} \tag{1.3}$$

The probability that a float is still in the box at a later time, $t > 0$, given that it was in the box at time $t = 0$ is:

$$\Pr(t \mid 0) = \mathrm{C}(t)V \tag{1.4}$$

Therefore, the probability that the float escaped from the marginal sea before time t is:

$$M(t) = 1\text{-}\Pr(t \mid 0) = 1\text{-}\mathrm{C}(t)V \tag{1.5}$$

where $M(t)$ is the cumulative distribution function for the times at which each individual float escaped from the marginal sea (escape times). Taking the derivative of $M(t)$ with respect to time gives the probability density (per unit time) that a given float will escape from the marginal sea at time t,

$$\frac{dM(t)}{dt} = -\frac{dC(t)}{dt} V = \frac{q}{V} e^{-qt/V} \tag{1.6}$$

where the last equality follows from Equation 1.3. The probability model for the float escape times is then:

$$\Pr(t \mid q) \approx \frac{dM(t)}{dt} dt \tag{1.7}$$

Parameter Estimation in the One-Box Model

The likelihood function for the one-box model is introduced to estimate the exchange flow parameter q from a set of observed float escape times $\{t_1, t_2, t_3, \ldots t_n\}$. The likelihood function, L(q), is proportional to the probability of observing the data for a given value of q:

$$L(q) \equiv K \Pr(data \mid q) = K \Pr(t_1, t_2, t_3, \ldots t_n \mid q) \tag{1.8}$$

where K is an arbitrary constant of proportionality. Since the escape times are independent, the probability relationship in Equation 1.8 can be expressed as a product of the individual probabilities:

$$L(q) = K \Pr(t_1 \mid q) \Pr(t_2 \mid q) \ldots \Pr(t_n \mid q) \tag{1.9}$$

Using Equations 1.6 and 1.7, the likelihood function for this system then becomes:

$$L(q) = K \frac{q}{V} e^{-qt_1/V} \frac{q}{V} e^{-qt_2/V} \cdots \frac{q}{V} e^{-qt_n/V} (dt)^n \tag{1.10}$$

which simplifies to:

$$L(q) = K \left(\frac{q}{V}\right)^n e^{\left(-q/V\right) \sum_{i=1}^{n} t_i} \tag{1.11}$$

Note that the factor of $(dt)^n$ does not depend on q and can be folded into the constant K. The most likely value for q is then the maximum of $L(q)$. To maximize the likelihood function in Equation 1.11, it is convenient to first take the natural logarithm. Because the log-likelihood is a monotonic function, both the likelihood and log-likelihood functions are maximized by the same parameter value. The natural logarithm of $L(q)$ is:

$$\ln L(q) \equiv l(q) = \ln(K) + n \ln(q) - n \ln(V) - \left(\frac{q}{V}\right) \sum_{i=1}^{n} t_i \tag{1.12}$$

and the derivative of Equation 1.12 with respect to q is then:

$$\frac{dl(q)}{dq} = \left(\frac{n}{q}\right) - \left(\frac{1}{V}\right) \sum_{i=1}^{n} t_i \tag{1.13}$$

Setting Equation 1.13 equal to zero and solving for q gives the maximum likelihood estimate (MLE) of the exchange flow between the marginal sea and the open ocean

$$\hat{q} = \frac{nV}{\sum_{i=1}^{n} t_i} \tag{1.14}$$

where $\hat{q}$ is the MLE of the flow parameter q expressed as a function of the volume, the number of floats deployed n, and a set of observed escape times t_i.

The confidence in the MLE can be assessed from the relative likelihood function (RLF), which is defined as:

$$R(q) = \frac{L(q)}{L(\hat{q})} \tag{1.15}$$

This function gives the likelihood of alternative flow values relative to the MLE. For example, an estimate for q such that $R(q) = 0.10$ is one-tenth as likely as the MLE, suggest-

ing that it is an implausible parameter. Estimates for q that fall in the range bounded by $0.10 \leq R(q) \leq 1$ are considered plausible estimates. The relative likelihood is best understood in terms of its inverse relationship to the confidence interval. A confidence interval of 90% describes the most probable parameter estimates whereas a 10% likelihood interval describes the most likely parameter estimates. The 10% likelihood interval is defined at $R(q) = 0.10$ and is the benchmark used for measuring the certainty of the flow parameter estimates.

Two-Box Models

Developing two-box models is the next step to achieving greater resolution of circulation in a marginal sea. The method for developing the one-box model can be applied to the two-box model, though the complexity of parameter estimation is greater.

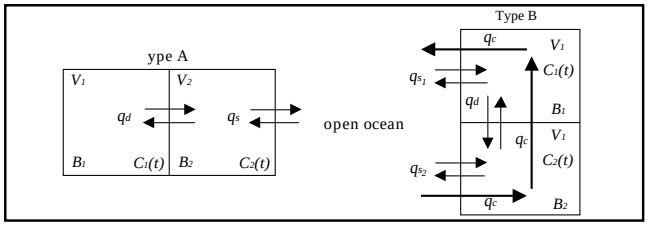

Figure 3

Two-box models of a marginal sea with exchange flows to the open ocean via recirculation flows q_d and q_s in the type A model, where only q_s exchanges with the open ocean; and via recirculation flows q_d, q_{s1} and q_{s2}, and net current q_c, in the type B model, where both boxes exchange with the open ocean. Note that the type B model reduces to the type A model for $q_c = 0$ and $q_{s1} = 0$.

Two types of two-box models were developed (Figure 3). Both use essentially the same notation as that used for the one-box model, and with subscripts to differentiate between the two boxes, the additional flow parameters, and the probability densities. Boxes one and two are denoted, respectively, as B_1 (with volume V_1) and B_2 (with volume V_2); the flow between B_1 and B_2 in the type A model is denoted as q_d, and the flow between B_2 and the open ocean is denoted as q_s. For the type B model, both B_1 and B_2 have exchange flows to the open ocean denoted as q_{s1} and q_{s2}, respectively. The type B model also contains a recirculation flow q_d between B_1 and B_2, and a net current q_c that flows from the open ocean through the marginal sea and back to the open ocean. The probability density of floats (per unit volume) in B_1 and B_2 is denoted by $C_1(t)$ and $C_2(t)$ respectivel, for both two-box models.

The type A and type B models offer differing perspectives on circulation in a marginal sea. The type A model is split into two boxes, only one of which exchanges directly with the open ocean. This simulation mimics deep ocean basins in which fluid furthest from the outlet may be hindered from flowing directly to the open ocean by turbulent eddy motions within the basin. Similar to the one-box model, the type A model depicts only recirculating flows (q_d and q_s), in which there is no net flow to the open ocean. The type B model contains recirculating flows (q_d, q_{s1}, and q_{s2}) as well as a net current (q_c). In this model, both boxes communicate with the open ocean, offering two means by which floats can leave the marginal sea. There is also a net flow through the boxes of the marginal sea representing a general directional current.

At the boundaries between the boxes there is a given flow in and a given flow out. When there is a net flow across a given boundary it is referred to as an advective flow. Balanced flows in and out of a box are considered recirculating, or diffusive, flows. Advective flows can either be positive or negative, while diffusive flows are always greater than zero. Ultimately, the parameters estimated in these models require further manipulation based on more explicit characterization of the marginal sea of interest (i.e. cross-sectional area of the fluid flux), before they can be exactly labeled as diffusive and/or advective. To avoid unnecessary complexity in developing this method, the models were not parameterized to the extent necessary to define the precise diffusive and advective flows. It suffices, however, to say that the estimated flows can be converted to these respective values later.

Type A Probability Model

It is assumed that a float is deployed in the marginal sea at time $t = 0$ to establish the probability that it will circulate to the open ocean at time t. For a two-box model, a float can be deployed in either B_1 or B_2. The initial condition, in terms of float densities, that the float was deployed in B_1 is given by:

$$\text{(a) } C(s \in B_1, t=0 \mid s \in B_1, t=0) = \frac{1}{V_1} \qquad (2.1)$$

$$\text{(b) } C(s \in B_2, t=0 \mid s \in B_1, t=0) = 0$$

where s is the position of the float. If the float is in B_1, then s is an element of B_1 ($s \in B_1$). This condition is denoted as $i = 1$ (or $j = 1$) (Table 1). Hence the initial condition, in terms of float densities, that the float was deployed in B_2 is given by:

$$\text{(a) } C(1,0 \mid 2,0) = 0$$

$$\text{(b) } C(2,0 \mid 2,0) = \frac{1}{V_2} \qquad (2.2)$$

where Equation 2.1a is the initial condition that the float is in B_1 at time $t = 0$ given that it was deployed in B_1, and Equation 2.1b is the initial condition that the float is in B_2 at time $t = 0$ given that it was deployed in B_1. Similarly, Equation 2.2a is the initial condition that the float is in B_1 at time $t = 0$ given that it was deployed in B_2, and Equation 2.2b is the initial condition that the float is in B_2 at time $t = 0$ given that it was deployed in B_2. The simplified notation in Equation 2.2 will be used for the remainder of this paper.

Table 1
Key to notation used in float density expressions $C(i,t \mid j,t)$

Position of Float (i or j)	Time at Location (t)
$(s \in B_1) \rightarrow 1$	$(time = 0) \rightarrow 0$
$(s \in B_2) \rightarrow 2$	$(time = t) \rightarrow t$
$(s \in$ open ocean$) \rightarrow 0$	--

The conservation equations for the probability of finding a float in B_1 or B_2, expressed in terms of the float densities ($C(i,t \mid j,0)$) are:

$$\frac{dC(1,t \mid j,0)}{dt} = -\frac{q_d}{V_1} C(1,t \mid j,0) + \frac{q_d}{V_1} C(2,t \mid j,0) \qquad (2.3)$$

and

$$\frac{dC(1,t \mid j,0)}{dt} = \frac{q_d}{V_2} C(1,t \mid j,0) - \frac{(q_d + q_s)}{V_2} C(2,t \mid j,0) \qquad (2.4)$$

where $j = 1$ or 2. Given either Equation 2.1 or Equation 2.2, the general solution to Equation 2.3 and Equation 2.4, respectively, is:

$$\begin{bmatrix} C(1,t \mid j,0) \\ C(2,t \mid j,0) \end{bmatrix} = e^{-At} \begin{bmatrix} C(1,0 \mid j,0) \\ C(2,0 \mid j,0) \end{bmatrix} \qquad (2.5)$$

where A is the 2 x 2 matrix:

$$A = \begin{bmatrix} -\frac{q_d}{V_1} & \frac{q_d}{V_1} \\ \frac{q_d}{V_2} & -\frac{(q_d + q_s)}{V_2} \end{bmatrix} \qquad (2.6)$$

The probability that a float is still in the marginal sea at some later time t is given by the sum of the probabilities that it is in B_1 at time t or B_2 at time t. For the case where the float started in B_1 at time $t = 0$, the probability is obtained by:

$$\Pr(\text{float in marginal sea}, t \mid 1,0) = C(1,t \mid 1,0)V_1 + C(2,t \mid 1,0)V_2 \qquad (2.7)$$

Similarly the probability the float started in B_2 at time $t = 0$ is obtained by:

$$\Pr(\text{float in marginal sea}, t \mid 2,0) = C(1,t \mid 2,0)V_1 + C(2,t \mid 2,0)V_2 \qquad (2.8)$$

Given that the probability a float is in the marginal sea at time $t = 0$ is equal to unity, the probability that the float escapes to the marginal sea at time t given Equation 2.7 and Equation 2.8 is:

$$\Pr(0,t \mid 1,0) = M_1(t) = 1 - C(1,t \mid 1,0) - C(2,t \mid 1,0) \qquad (2.9)$$

and

$$\Pr(0,t \mid 2,0) = M_2(t) = 1 - C(1,t \mid 2,0) - C(2,t \mid 2,0) \qquad (2.10)$$

where Equation 2.9 represents the probability that the float escapes from the marginal sea before time t given that it was in B_1 at time $t = 0$, and Equation 2.10 represents the probability that the float escapes from the marginal sea before time t given that it was in B_2 at time $t = 0$. It was assumed that once a float escaped from the marginal sea it would stay in the open ocean and not return to the marginal sea.

The derivatives of Equation 2.9 and Equation 2.10 with respect to time give the probability density per unit time that the float escapes from the marginal sea before time t:

$$\frac{dM_1(t)}{dt} = -\frac{dC(1,t \mid 1,0)}{dt} V_1 - \frac{dC(2,t \mid 1,0)}{dt} V_2 \qquad (2.11)$$

$$\frac{dM_1(t)}{dt} = -\frac{dC(1,t \mid 2,0)}{dt} V_1 - \frac{dC(2,t \mid 2,0)}{dt} V_2 \qquad (2.12)$$

Because of the greater complexity of a two-parameter estimation, it is not practical to solve these equations analytically, as was done with the one-box model. Instead, a numerical approach is necessary. In developing the two-box models, both were tested using the analytical and numerical approaches.

Type B Probability Model

The method for solving the type B two-box model is the same as that for the type A two-box model. The only difference is the addition of a net advective current in the type B model and the fact that floats can escape to the open ocean from both boxes. Only the recirculation flows were estimated in the one-box and type A two-box models. The type B two-box model shows the utility of this method for estimating both the diffusive and advective parameters.

The conservation equations for the probability of finding a float in B_1 or B_2 in the type B model are:

$$\frac{dC_1(t)}{dt} = \frac{1}{V_1}\left[-q_d - q_{s_1} - \frac{(q_c + |q_c|)}{2} + \frac{(q_c - |q_c|)}{2}\right] C_1 + \frac{1}{V_1}\left[q_d + \frac{(q_c + |q_c|)}{2}\right] C_2 \qquad (3.1)$$

and

$$\frac{dC_2(t)}{dt} = \frac{1}{V_2}\left[q_d - \frac{(q_c - |q_c|)}{2}\right]C_1 + \frac{1}{V_2}\left[-q_d - q_{s_2} - \frac{(q_c + |q_c|)}{2} + \frac{(q_c - |q_c|)}{2}\right]C_2 \quad (3.2)$$

The expressions in Equation 3.1 and Equation 3.2 allow for the advective flow q_c to reverse direction, as might be the case under natural conditions, such as during the monsoon season in the Indian Ocean, when a reversal of the general circulation forces moisture over the Indian subcontinent. This formulation is important when there is not *a priori* knowledge of the direction of the advective flow.

Again there were two possible initial conditions in which the floats were deployed in B_1 or B_2. The mathematical expressions for these initial conditions are the same as for the type A model. The general solution to this system is formally also the same as Equation 2.5, except that A is given by the 2 x 2 matrix:

$$A = \begin{bmatrix} \frac{1}{V_1}\left(-q_d - q_{s_1} - \frac{1}{2}\overleftarrow{q_c} + \frac{1}{2}\overrightarrow{q_c}\right) & \frac{1}{V_1}\left(q_d + \frac{1}{2}\overleftarrow{q_c}\right) \\ \frac{1}{V_2}\left(q_d + \frac{1}{2}\overrightarrow{q_c}\right) & \frac{1}{V_2}\left(-q_d - q_{s_2} - \frac{1}{2}\overleftarrow{q_c} + \frac{1}{2}\overrightarrow{q_c}\right) \end{bmatrix} \quad (3.3)$$

where $\overleftarrow{q_c} = (q_c + |q_c|)$ and $\overrightarrow{q_c} = (q_c - |q_c|)$. In the general case, the type B two-box model is a four-parameter estimation that cannot be solved analytically. For simplicity, the type B model is reduced to a two-parameter estimation by setting the recirculating flows equal to each other, $q_d = q_{s1} = q_{s2}$. We also set the advective current q_c one order of magnitude larger than the recirculating flows to more realistically represent the relative magnitude of these flows in a true ocean basin. The two unknown parameters q_d and q_c are determined.

Parameter Estimation in the Two-Box Models

The two-box model likelihood function is the probability of generating the observed escape times given the unknown parameters (q_d and q_s; q_d and q_c). Therefore, the likelihood function for the type A two-box model is:

$$L(q) \equiv K \Pr(t_1, t_2, \ldots t_n, t_{n+1}, t_{n+2}, \ldots t_{n+m} \mid q_d, q_s) \quad (4.1)$$

and for the type B two-box model is:

$$L(q) \equiv K \Pr(t_1, t_2, \ldots t_n, t_{n+1}, t_{n+2}, \ldots t_{n+m} \mid q_d, q_c) \quad (4.2)$$

The float escape times given by t_1 through t_n represent deployment of the floats in B_1, while the escape times t_{n+1} through t_{n+m} represent the float escape times given that the floats were deployed in B_2.

Because the escape times are independent, Equations 4.1 and 4.2 can be rewritten as the product of the probability densities evaluated at the observed escape times:

$$L(q) = K \frac{dM_1(t_1)}{dt}\frac{dM_1(t_2)}{dt}\ldots\frac{dM_1(t_n)}{dt}\frac{dM_2(t_{n+1})}{dt}\frac{dM_2(t_{n+2})}{dt}\ldots\frac{dM_2(t_{n+m})}{dt}(dt)^{n+m} \quad (4.3)$$

The log-likelihood functions for both two-box models are then given by:

$$l(q) = \ln\sum_{i=1}^{n}\frac{dM_1(t_i)}{dt}(dt)^n + \ln\sum_{i=n+1}^{n+m}\frac{dM_2(t_i)}{dt}(dt)^m \quad (4.4)$$

The maximum of Equation 4.4 gives the MLEs of q_d and q_s for the type A model, and the MLEs of q_d and q_c for the type B model.

Testing the Method

The method of maximum likelihood was tested on the one- and two-box models. For each box-model type, a set of escape times was generated by simulated random exchange of floats between the marginal sea and the open ocean. For the two-box models, two possible conditions of initial float deployment were simulated, and the escape times from one simulation of each were pooled for the estimation step. Ensembles of 10 simulations were conducted with N = 10, 50, 100, 1,000, and 10,000 floats per box. The MLE was calculated for each run in an ensemble, and the mean and standard deviation of each ensemble were calculated. Ensemble averages yield valuable information about the nature and magnitude of uncertainty in an estimation method (Wilks, 1995). Ensemble averages were used to evaluate the bias, if any, associated with using the method of maximum likelihood to estimate the most likely flow parameter(s) given a set of data. An unbiased method returns a mean that approaches the true value of a parameter as long as the number of ensembles used to compute the mean is large. In addition to using ensemble averages and standard deviations to assess the bias, three ensembles of 1,000 at N = 10, 100, and 1,000 were run for the one-box model only. Histograms of the MLEs computed for each run in an ensemble were used to assess the distribution of the estimates, and the ensemble average was compared to the true q.

To evaluate the accuracy of the method, the RLF was computed for each individual run within an ensemble. The log-RLF was plotted for a range of q values, which included $\hat{q}$ (the MLE) and the true q value, to graphically assess convergence on the true value, with the peak of the log-RLF curve ($R(q) = 0$) representing the MLE. This was done by superimposing the 10% likelihood interval, given by ($R(q) = 0.10$), on the plots of the log-RLF versus q. The RLF pro-

vided insight to the certainty of a single estimate calculated by the method of maximum likelihood.

Estimation of the MLE for the one-box model utilized the analytical solution to the likelihood function. The numerical and analytical solutions to the likelihood functions for the two-box models were both developed to calculate the MLE. However, when it was verified that the numerical solution produced the same results as the analytical, only the analytical solution was used in order to maximize computational efficiency. A grid-search method was used to find the maximum of the likelihood function. This involved defining a range of q values that bounded the true q value and computing the log-likelihood using the analytical (or numerical) solution at each of these discrete values. The MLE was then found to be the discrete q that best maximized the likelihood function. As a consequence of using a grid-search, however, the accuracy of this method was limited to finding the discrete value of q that fell closest to the true value, as opposed to the true value itself. This was not an inherent problem with the method since the mesh of the grid could easily be made finer to estimate parameters of a more realistic model.

Results and Discussion

One-Box Model

As the number of floats used in a simulation of the one-box model increased, the MLE for q converged to the true value (Figure 4). The relative error of the ensemble mean MLE, using the $N = 10$ simulations, was approximately 56% with a coefficient of variation (CV) of approximately 39%. By increasing the number of floats from 10 to 50, the relative error of the ensemble mean MLE decreased to 5% with a CV of 13%. There was little change in the relative error between the simulation that used 100 floats and the simulation that used 10,000 floats. However, the CV for a 100 float run was approximately 10% compared to 1% for the 10,000 float simulation. With a relative percent error of about 2% and a CV of just over 3%, the 1,000 float simulation returned an MLE sufficiently close to the true value an acceptable percentage of the time. If the one-box model was a realistic representation of the marginal sea, these results would help determine the optimal number of floats needed to achieve a desired level of accuracy and precision in the estimation.

Figure 5 shows how uncertainty in a single estimate can be reduced by using a larger number of floats. The 10% likelihood interval, given by the intersection points of the log-RLF curve and the $R(q) = 0.10$ ($\ln R(q) = -2.30$) line, gave the range of q values that were considered plausible estimates of q. At $N = 10$ this interval was broad, and a wide

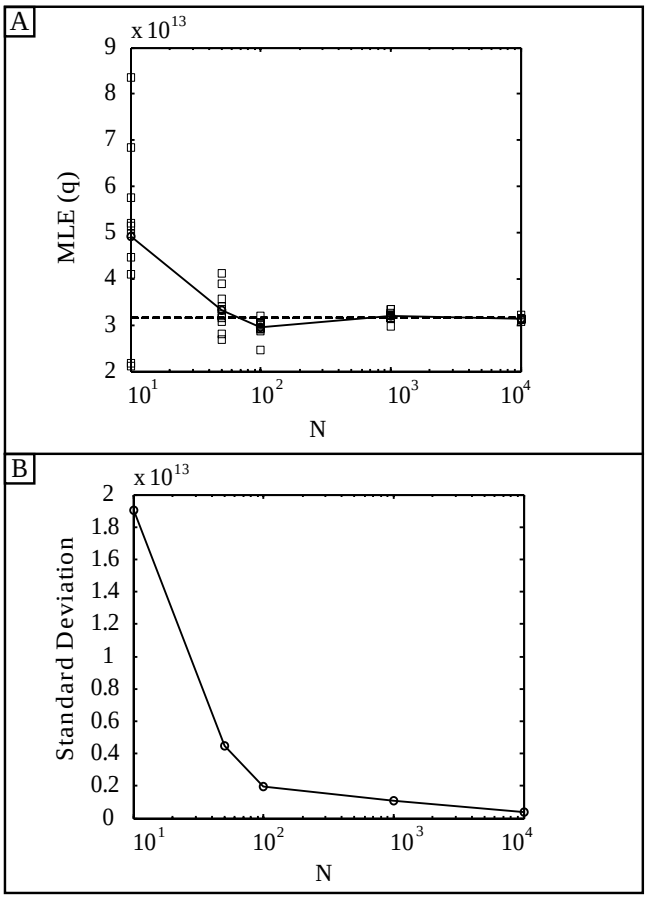

Figure 4
One-box model plots of the (A) ensemble mean MLE for q (solid line), the true q value (dashed line), and the individual MLEs from each run in the 10-member ensemble (squares), versus the number of floats used in the simulation N; and (B) the ensemble standard deviation versus N. Units of q are cubic meters per year.

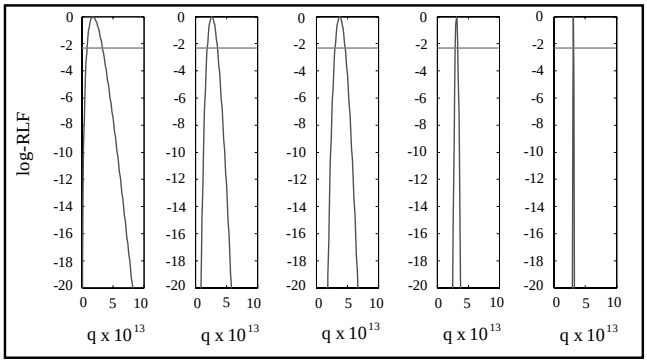

Figure 5
Log-RLF (curve) with the 10% likelihood level (line) plotted versus a range of q values, for N = 10, 50, 100, 1,000 and 10,000 floats, from left to right. Plausible estimates of q lie on the curve above the likelihood level. The units of q are cubic meters per year.

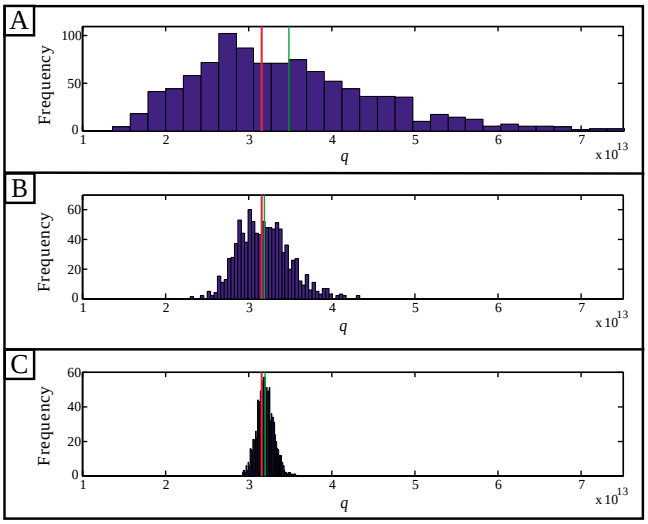

Figure 6
One-box model histograms of MLEs of q for 1,000-member ensembles at N = 10 , 100, and 1,000 floats — (A), (B), and (C), respectively. Also plotted is the true q (red line) and the mean MLE of q (green line). Units of q are cubic meters per year.

range of q values were plausible estimates of the true value, while at N = 10,000 the estimate became more accurate.

Simulations using a smaller number of floats seemed to have a slightly positive bias to the MLE for q (Figure 6). One thousand-member ensembles were run at at N = 10, 100 and 1,000 floats to see if there was a bias in the method or if the apparent bias was simply due to random variability. Histograms of these ensembles showed a positive bias for N = 10 floats, where the distribution of the estimates was right-skewed and the ensemble mean was noticeably greater than the true value. This bias decreased for the larger float simulations, where the spread of distribution narrowed and its shape became more Gaussian. Despite the slight positive bias to the histogram of the 10-float ensemble, only 54% of these runs yielded a $\hat{q}$ that was greater than the true value, which was not significant enough to indicate a true bias in the method.

Two-Box Models

Type A Two-Box Model

Two flow parameters, q_d and q_s, were estimated in the type A two-box model. Again ensemble averages were used to assess the bias of the method. Based on the ensemble means for q_d and q_s, the method did not show any strong bias, and the true value was well approximated for increasing float numbers (Figure 7). For both parameters, the ensemble standard deviation decreased exponentially with increasing numbers of floats (Figure 8), with the range of

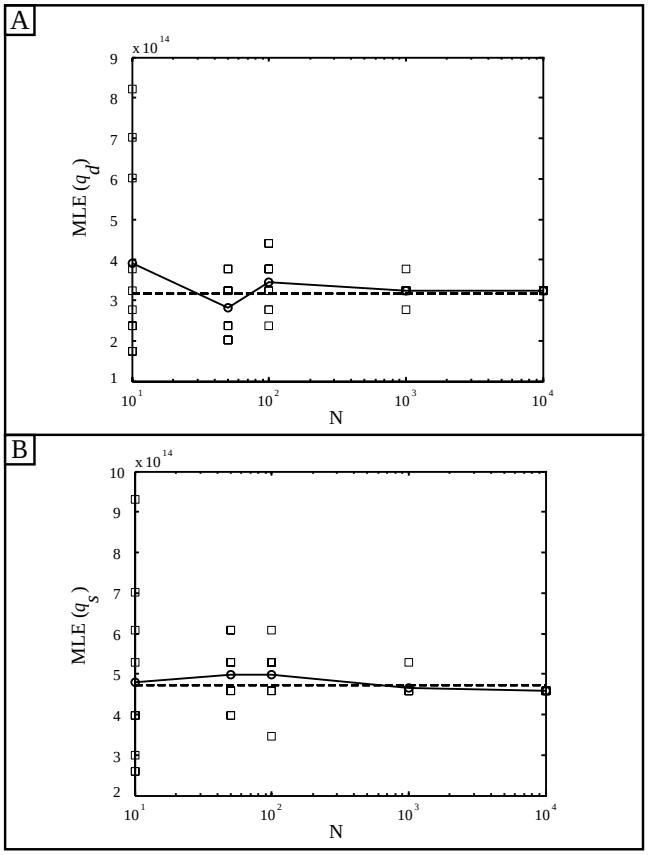

Figure 7
Type A two-box model plots of the ensemble mean MLE (solid line) for (A) q_d and (B) q_s, the true q_d and q_s values (dashed line), and the individual MLEs from each run in the 10-member ensembles (squares), versus N. Units of q are cubic meters per year.

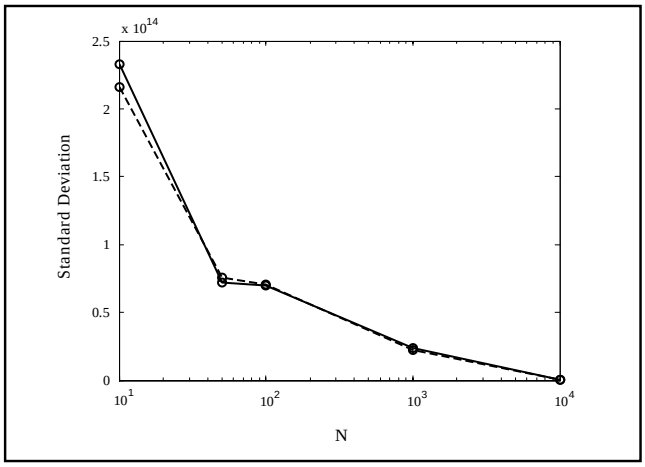

Figure 8
Type A two-box model plots of the ensemble standard deviation versus N for q_d (solid line) and q_s (dashed line). Units of standard deviation are cubic meters per year.

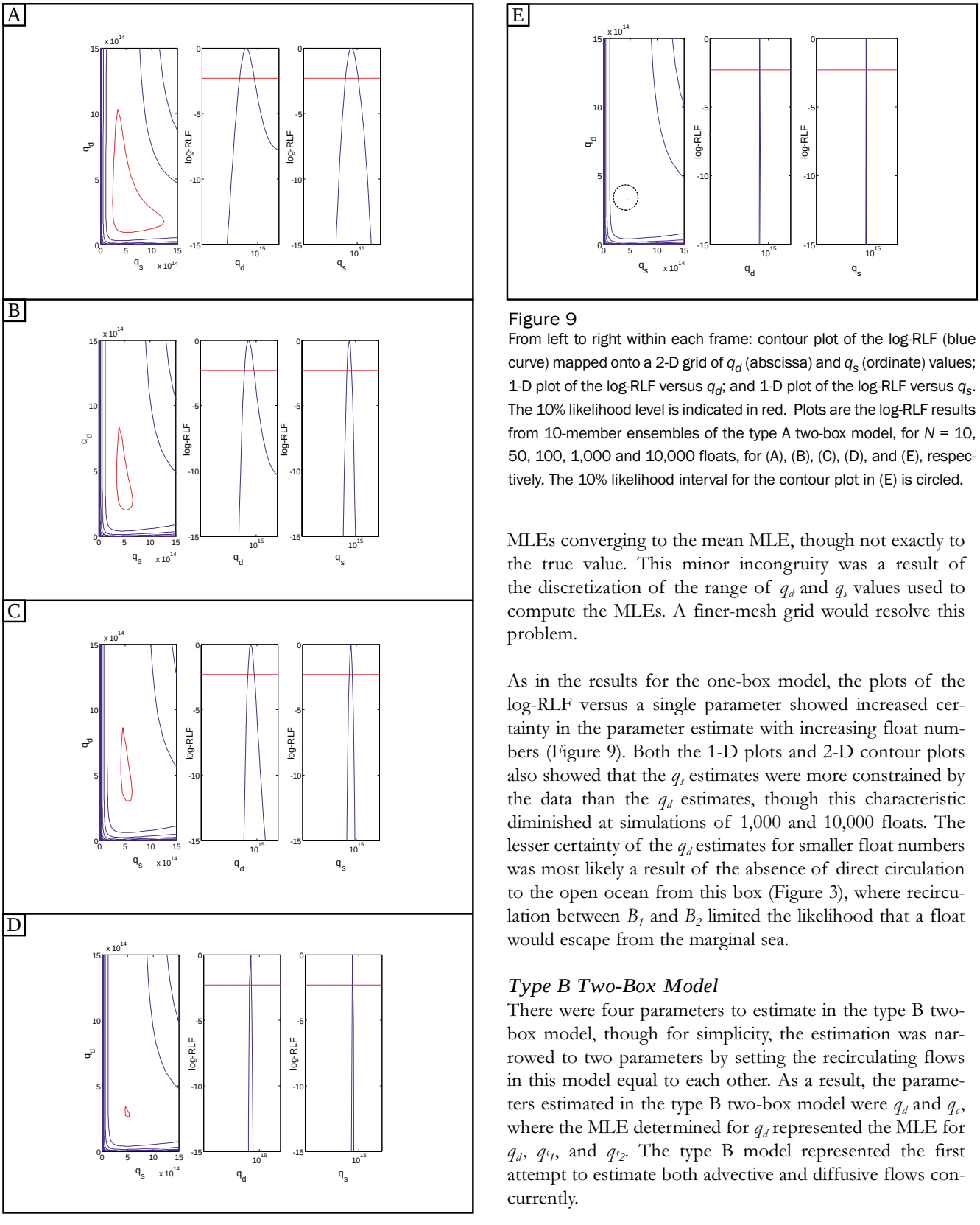

Figure 9

From left to right within each frame: contour plot of the log-RLF (blue curve) mapped onto a 2-D grid of q_d (abscissa) and q_s (ordinate) values; 1-D plot of the log-RLF versus q_d; and 1-D plot of the log-RLF versus q_s. The 10% likelihood level is indicated in red. Plots are the log-RLF results from 10-member ensembles of the type A two-box model, for N = 10, 50, 100, 1,000 and 10,000 floats, for (A), (B), (C), (D), and (E), respectively. The 10% likelihood interval for the contour plot in (E) is circled.

MLEs converging to the mean MLE, though not exactly to the true value. This minor incongruity was a result of the discretization of the range of q_d and q_s values used to compute the MLEs. A finer-mesh grid would resolve this problem.

As in the results for the one-box model, the plots of the log-RLF versus a single parameter showed increased certainty in the parameter estimate with increasing float numbers (Figure 9). Both the 1-D plots and 2-D contour plots also showed that the q_s estimates were more constrained by the data than the q_d estimates, though this characteristic diminished at simulations of 1,000 and 10,000 floats. The lesser certainty of the q_d estimates for smaller float numbers was most likely a result of the absence of direct circulation to the open ocean from this box (Figure 3), where recirculation between B_1 and B_2 limited the likelihood that a float would escape from the marginal sea.

Type B Two-Box Model

There were four parameters to estimate in the type B two-box model, though for simplicity, the estimation was narrowed to two parameters by setting the recirculating flows in this model equal to each other. As a result, the parameters estimated in the type B two-box model were q_d and q_c, where the MLE determined for q_d represented the MLE for q_d, q_{s1}, and q_{s2}. The type B model represented the first attempt to estimate both advective and diffusive flows concurrently.

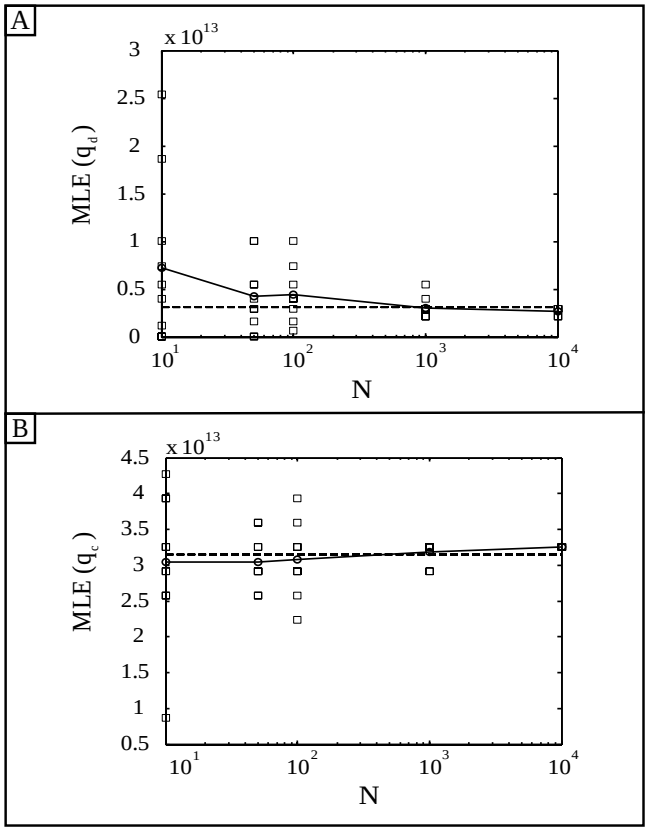

Figure 10
Type B two-box model plots of the ensemble mean MLE (solid line) for (A) q_d and (B) q_c, the true q_d and q_c values (dashed line), and the individual MLEs from each run in the 10-member ensembles (squares), versus N. Units of q are cubic meters per year.

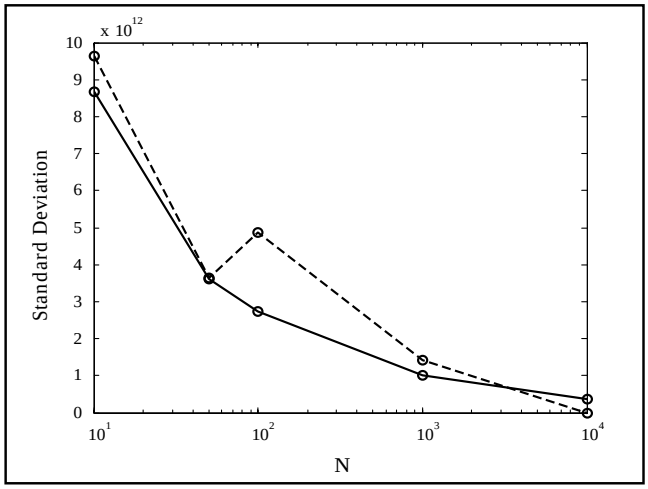

Figure 11
Type B two-box model plots of the ensemble standard deviation versus N for q_d (solid line) and q_c (dashed line). Units of standard deviation are cubic meters per year.

Plots of the ensemble mean MLEs, and the ensemble standard deviations, for both q_d and q_c showed that the method was unbiased for large float numbers (Figure 10). A 10-member ensemble returned the approximate true value with little variance for simulations using 1,000 and 10,000 floats. Again, the MLEs converged to the ensemble mean, as opposed to the true value. As with the type A model this was a result of the discrete set of q_d and q_c values used to find the maximum.

Plots of the ensemble standard deviations showed that, in general, the precision of the MLEs for the advective flow parameter q_c was greater than the precision for the diffusive flow parameter q_d (Figure 11). While these plots seemed to show similar results for the ensemble standard deviations, it is important to note that the true value of q_d was on the order of 10^{12}, in which case the standard deviation was of the same magnitude as the parameter value for float numbers less than 1,000. In fact, for $N = 10$ the ensemble CV for q_d was more than 100%; whereas the ensemble CV at $N = 10$ for q_c was approximately 32%. This discrepancy was not noted in the two-parameter estimation for the type A model. In the type A model, however, the two flows were of the same order of magnitude, since they were both diffusive flows. In the type B model, it may be difficult to estimate flows with equal precision when their magnitudes are very different. This problem vanished, however, as N became large.

One-dimensional plots of the log-RLF versus each individual parameter showed a reduction in the uncertainty of the parameter estimate with increasing N (Figure 12). The 2-D contour plots showed that the advective flow was better constrained by the data than the diffusive flow for smaller N, corroborating the results shown by the ensemble data that indicate that at small N the diffusive flow was difficult to constrain in the presence of the larger advective flow.

Two features are worth noting in the log-RLF plots for the type B model (Figure 12). First, for small N the plot of the log-RLF leveled off for small values of q_d. This indicated that a small diffusive flow was not significantly different from a zero value in the presence of the larger advective flow. In fact, for $N = 10$, a value of zero for q_d was a plausible estimate since this portion of the plot was positioned above the 10% likelihood interval. The second interesting feature can be seen in the 1-D log-RLF plots for q_c. The double maxima in the plot for $N = 10$ occurred symmetrically on either side of $q_c = 0$, where the negative value indicated a flow in the opposite direction than depicted in Figure 3. The negative local maximum, however, fell below

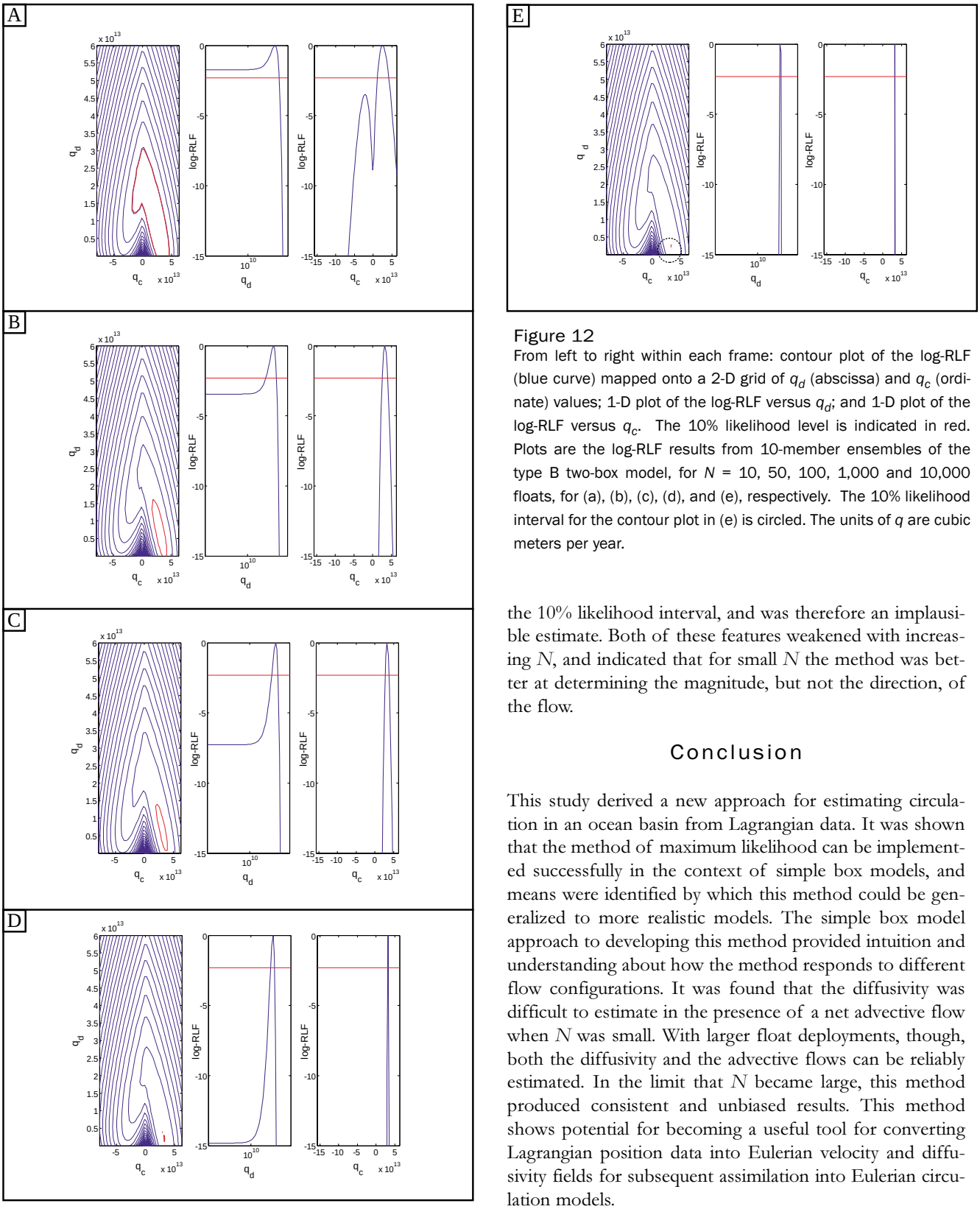

Figure 12

From left to right within each frame: contour plot of the log-RLF (blue curve) mapped onto a 2-D grid of q_d (abscissa) and q_c (ordinate) values; 1-D plot of the log-RLF versus q_d; and 1-D plot of the log-RLF versus q_c. The 10% likelihood level is indicated in red. Plots are the log-RLF results from 10-member ensembles of the type B two-box model, for N = 10, 50, 100, 1,000 and 10,000 floats, for (a), (b), (c), (d), and (e), respectively. The 10% likelihood interval for the contour plot in (e) is circled. The units of q are cubic meters per year.

the 10% likelihood interval, and was therefore an implausible estimate. Both of these features weakened with increasing N, and indicated that for small N the method was better at determining the magnitude, but not the direction, of the flow.

Conclusion

This study derived a new approach for estimating circulation in an ocean basin from Lagrangian data. It was shown that the method of maximum likelihood can be implemented successfully in the context of simple box models, and means were identified by which this method could be generalized to more realistic models. The simple box model approach to developing this method provided intuition and understanding about how the method responds to different flow configurations. It was found that the diffusivity was difficult to estimate in the presence of a net advective flow when N was small. With larger float deployments, though, both the diffusivity and the advective flows can be reliably estimated. In the limit that N became large, this method produced consistent and unbiased results. This method shows potential for becoming a useful tool for converting Lagrangian position data into Eulerian velocity and diffusivity fields for subsequent assimilation into Eulerian circulation models.

The principal advantage of this method is the ability to deduce information about parameter uncertainties using the relative likelihood function. Understanding the uncertainty of the parameter estimate allows oceanographers to determine the optimal weights needed for blending observation with model output when assimilating the observational data into an ocean model.

A second benefit of this study is the potential for using *n*-member ensembles to optimize Lagrangian drifter studies. Pre-field testing using this technique would allow oceanographers to determine the optimum number of floats needed to achieve a desired level of accuracy and precision in the flow estimates, thereby minimizing the cost of sampling and the potential for aliasing due to underestimation of the number of floats necessary to characterize the circulation.

As the method is developed for more realistic *n*-box models, several issues of refinement will need to be addressed. First, the grid search method for finding the MLE should be replaced by a more sophisticated optimization procedure. Other possible methods are already being considered and will be tested in the future. Second, generalization to an *n*-box model must be tested to be certain that multiparameter estimations are indeed numerically feasible. It was assumed that because in theory the method of maximum likelihood extends readily to multiparameter estimation, this transition would be straightforward though possibly computationally expensive. This generalization will require an exclusively numerical approach for integrating the differential equations that define the probability models. The numerical methods have already been developed, though the probability model has yet to be built for an *n*-box system. The final step to fully developing this method is to test it using real Lagrangian data.

Ultimately, this method offers an alternative procedure for converting Lagrangian position data into Eulerian flow fields; one that is reliable, accurate, and provides greater resolution of the two types of fluid motion that describe ocean circulation. Improved resolution of the observational data ensures the development of higher quality and more reliable ocean circulation models that will produce more realistic results when incorporated into climate models for studying global climate change.

Acknowledgements

I would like to thank Debbie Elmore and Les Steinberger for editing this document for clarity; Linda Steinberger for her technical suggestions; Dr. Kerry Ritter for reviewing the mathematics and statistics; and, my advisor, Dr. Francois Primeau, for his constant encouragement, extraordinary patience, and valuable mentorship. I am grateful to him for guiding my research, and for giving me the opportunity to explore the fascinating science of physical oceanography. I would also like to thank my neighbor and friend, Dianne Chaffee, for her unceasing support. I am also grateful to my mother, Diana Steinberger, for always nudging me lovingly in the right direction.

Works Cited

Belluci, A., G. Buffoni, A. Griffa, and E. Zambianchi. "Estimation of Residence Times in Semi-enclosed Basins with Steady-flows." Dynamics of Atmospheres and Oceans 33 (2001): 201-208.

Buffoni, G., P. Falco, A. Griffa, and E. Zambianchi. "Dispersion Processes and Residence Times in a Semi-enclosed Basin with Recirculating Gyres: An Application to the Tyrrhenian Sea." Journal of Geophysical Research 102 (1997): 18,699-18,713.

Colling, Angela, ed. Ocean Circulation, Second Edition. Oxford: Butterworth-Heinemann/The Open University, 2001. 243-249.

Falco, P., A. Griffa, P. Poulain, and E. Zambianchi. "Transport Properties in the Adriatic Sea as Deduced from Drifter Data." Journal of the American Meteorological Society 30 (2000): 2055-2071.

Haine, T.W.N and T.M. Hall. "A Generalized Transport Theory: Water-Mass Composition and Age." Journal of Physical Oceanography 32 (2002): 1987-1991.

Hartmann, Dennis L. Global Physical Climatology. San Diego: Academic Press, 1994. 254-285.

Lavender, K.L., R.E. Davis, and W.B. Owens. "Mid-depth Recirculation in the Interior Labrador and Irminger Seas by Direct Velocity Measurements." Nature 407 (2000): 66-69.

Molcard, A., L.I. Piterbarg, A. Griffa, T.M. Ozgokmen, and A.J. Mariano. "Assimilation of Drifter Observation for the Reconstruction of the Eulerian Circulation Field." Journal Geophysical Research 108.C3 (2003): 3056.

National Research Council (NRC). "Global Ocean Science: Toward an Integrated Approach." National Academy of Sciences. Washington, DC: National Academy Press, 1999.

Pinardi, N. and J. Woods, eds. Ocean Forecasting: Conceptual Basis and Applications. Berlin, Germany: Springer-Verlag, 2002.

Poje, A.C., M. Toner, A.D. Kirwan, and C.K.R.T. Jones. "Drifter Launch Strategies Based on Lagrangian Templates." Journal of Physical Oceanography 32 (2002): 1855-1869.

Rhein, Monika. "Drifters Reveal Deep Circulation." Nature 407 (2003): 30-31.

Toner, M., A.C. Poje, A.D. Kirwan, C.K.R.T. Jones, B.L. Lipphardt, and C.E. Grosch. "Reconstructing Basin-scale Eulerian Velocity Fields from Simulated Drifter Data." Journal of Physical Oceanography 13 (2001): 51-56.

Whittaker, Joe. Math 235 Statistics. 2002. Department of Math and Statistics, Lancaster University, England. http://www.maths.lancs.ac.uk.

Wilks, Daniel S. Statistical Methods in the Atmospheric Sciences. San Diego: Academic Press, 1995.

World Ocean Circulation Experiment (WOCE). Ocean Circulation and Climate: World Ocean Circulation Experiment. Southampton: WOCE International Project Office [154/94], 1997.