OCEAN SURFACE TRANSPORT METHODOLOGY REPORT

Authors: A. C. Savage, V. Balza Pineda, B. Ball, T.-H. Chung, F. Heng, and T. N. Beatty

ABSTRACT: This report details the methodology used for the surface transport section of the overall quantification methodology.

1 Introduction

A critical component of Running Tide's carbon removal quantification process is modeling the surface transport of deployed material. This process provides the required information to determine whether or not the deployed material was durably sequestered on the ocean floor below a $1,000\,m$ depth. The surface transport methodology incorporates both laboratory and in-situ data on the float time of material, in-situ GPS data collected from a fleet of surface drifters, and third-party data on ocean surface currents, waves, and wind to estimate the terminal location of deployed material. This is necessary not only for carbon removal quantification, but is also an important part of ecological impact monitoring and evaluation. The remainder of this report is structured as follows: section two describes the computational implementation of surface transport modeling in detail while section three presents results of modeled surface transport trajectories.

2 Methodology

2.1 Material float times

The float time of Running Tide carbon removal material is primarily dependent on four parameters: form factor, coating recipe, moisture content, and sea surface temperature (SST). Since form factor is relatively consistent across the deployed material, and since sorting deployed material by form factor is impractical, coating recipe, moisture content, and SST have more measurable effects on the floating characteristics of material. Material float times are measured in two ways:

Laboratory float time testing Float time tests are conducted using a sample of the material deployed for each deployment and are conducted in a laboratory setting in the Ocean Hub in Portland, ME. These tests are conducted in tanks, which can either be static or have an induced standing wave with a magnitude of several centimeters. Typically, these tests have been conducted at room temperature using the standardized salinity profile of Instant Ocean, but can be conducted in a refrigerated setting to reduce the water temperature. Once per day, the material that remains floating in the tank is manually retrieved and weighed, and the fractional mass is reported.

In-situ machine vision float time observations Float time tests are also conducted in-situ using a drifting camera system that records and transmits photos three times daily. The images are processed to measure the fraction of biomass in the sample that remains floating on the surface. In order to compute floating fraction, we automatically segment each biomass within the image using a convolutional neural network model and compute the 2-dimensional pixel area for each detected biomass. As these tests are conducted in-situ, the material is subject to sea surface temperature, salinity, pressure, and waves. Therefore, in order to minimize these effects, we discard noisy images using a binary classification model.

In both cases, the data demonstrates that a decreasing fraction of material remains on the surface over time, confirming that, eventually, all material becomes saturated and inevitably sinks to the

seafloor. Both datasets are fit to an exponential decay curve

$$f(t) = ab^t + c \tag{1}$$

where t is time since deployment, $f(t)$ is the fraction of material that remains floating on the surface at time t, a is the fraction of material that does not sink immediately upon deployment, b is the decay rate, and c is the fraction of material that remains floating at the end of the test. The coefficients are determined with non-linear least squares fit for each dataset separately, generating two distinct float time curves. There is considerable variability between the two curves, most likely due to differences in measurement technique as well as moisture content and SST, but more testing is needed to fully map the sensitivity of float time curves to such parameters.

2.2 Data integration for surface transport

With every Running Tide carbon removal deployment, a fleet of 27 drifters is deployed to track the movement of the deployed material. Of the deployed drifters, 25 send back GPS data at 4 hourly intervals and two drifters are the camera systems used to record material float times as described in Sec. 2.1, which also record and report GPS location. The GPS data is automatically QA/QC'd, and additionally checked for accuracy after the QA/QC process.

To model the surface transport of the material, both third-party datasets as well as in-situ data are integrated into a Lagrangian simulator that models the trajectories of simulated deployed material. We use an open source library called OceanPARCELS (Kehl et al., 2023), which uses an efficient implementation of a Runge-Kutte integration scheme to simulate multidimensional Lagrangian transport with a provided velocity vector field. The total horizontal velocity vector field used to force the Lagrangian simulations can be written as a superposition of multiple velocity components

$$\mathbf{u}_t = \mathbf{u}_h + \mathbf{u}_s + \mathbf{u}_w \tag{2}$$

where $\mathbf{u}_t$ is the total velocity, $\mathbf{u}_h$ is the ocean dynamics velocity, $\mathbf{u}_s$ is the Stokes drift velocity, and $\mathbf{u}_w$ is the windage velocity. The ocean dynamics velocity fields are sourced from HYbrid Coordinate Ocean Model [HYCOM; Cummings and Smedstad (2013)] and the Stokes drift and windage velocities are sourced from the European Centre for Medium-Range Weather Forecasts Reanalysis dataset (Hersbach et al., 2023). Comparison of modeled trajectories to those reported from the GPS drifters shows the necessity for weighting the components of the total velocity to account for their varying influence on the horizontal transport of material as well as for subtle errors in the HYCOM and ERA5 data used. Therefore, the total velocity can more accurately be written as

$$\mathbf{u}_t = \alpha_h(t)\mathbf{u}_h + \alpha_s(t)\mathbf{u}_s + \alpha_w(t)\mathbf{u}_w \tag{3}$$

where α_h, α_s, and α_w are positive scalars that depend on time and can be adjusted for the most accurate fit of simulated trajectories to observed trajectories. We then solve for the values of $\alpha(t)$ that provide the most accurate fit against the in-situ GPS data.

To solve for the optimal α values, we've implemented a gradient descent algorithm that iteratively decreases the error between modeled trajectories and in-situ trajectories until the error converges to a minimum. This first requires building a 3T dimensional parameter space, where T is the number of discrete time blocks that define α between the start and end of the deployment simulation. This can be represented as

$$\begin{bmatrix} \alpha_h(t) \\ \alpha_s(t) \\ \alpha_w(t) \end{bmatrix} = \begin{bmatrix} \alpha_{h,1} & \alpha_{h,2} & \cdots & \alpha_{h,T} \\ \alpha_{s,1} & \alpha_{s,2} & \cdots & \alpha_{s,T} \\ \alpha_{w,1} & \alpha_{w,2} & \cdots & \alpha_{w,T} \end{bmatrix}. \tag{4}$$

We then compute the error associated with this parameter space by running our trajectory simulation and computing an RMSE using the weighted centers of the simulated drifters and the weighted centers of the in-situ drifters at each time step, defined as

$$RMSE = \sqrt{\frac{\Sigma_0^N (x_{simulated} - x_{oberved})^2}{N} - \nu (D_{simulated} - D_{observed})^2} \tag{5}$$

where $x_{simulated}$ and $x_{observed}$ are the simulated and observed centroid locations (mean trajectories) of the trajectory clusters, respectively, $D_{simulated}$ and $D_{observed}$ are the total distances traveled by the simulated and observed mean trajectories, respectively, and ν is a weighting factor. Thus far it has been necessary to test multiple values of ν for a singular deployment in order to find the definition of RMSE that produces the best results. More work is needed to understand how an appropriate value of ν may be automatically generated.

Gradient descent, or steepest descent, (Polyak, 1987) is a widely used optimization algorithm that finds the local minima of a differentiable loss or error function. It achieves this by computing all partial derivatives of the loss function and moving the parameter space in the direction of steepest descent until the minimum is found, i.e. where the gradient of the loss function approaches zero. Since the loss function, Eq. 5, is computed by comparing the output from a simulation against in-situ GPS data, the gradient of this loss function is computed by calculating the value of the loss function at a nearby point and evaluating the derivative traditionally as

$$\frac{\partial RMSE}{\partial \alpha_{h,1}} = \frac{RMSE(\alpha_{h,1}) - RMSE(\alpha_{h,1} + \epsilon)}{\epsilon}. \tag{6}$$

The partial derivatives are computed for each of the 3T parameters in the parameter space defined in Eq. 4. The perturbation parameter, ϵ, is set to a small constant number.

After the alpha values have been determined for each velocity component, a stochastic dispersion parameter, k_{diff}, is also tuned. This is tuned simply by testing a series of values of k_{diff}, typically $10 \leq k_{diff} \leq 200$, and using the value that best estimates the spread of the observed drifters over time. In this way, k_{diff} can be used to estimate the effect of subgridscale dynamics that are not resolved in the modeled velocity fields. The spread of modeled drifters is relatively insensitive to small changes in k_{diff}, and thus more sophisticated methods for optimizing k_{diff} are currently not required.

2.2.1 Parameter Uncertainties

The parameter uncertainties are determined following the methodology in (Turley, 2018). Additionally, we make the approximation that $\mathbf{J}^T\mathbf{J} \approx \mathbf{H}$ (Jorge Nocedal, 2006, p. 246). $\mathbf{H}$ here is the Hessian matrix defined as

$$H(f)_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j} \tag{7}$$

Thus the covariance matrix $\mathbf{C}$ is

$$\mathbf{C} = \sqrt{\frac{\Sigma_0^N (x_{simulated} - x_{oberved})^2}{N - 3T}} \mathbf{H}^{-1} \tag{8}$$

The technical calculation of the Hessian matrix follows the same methodology as Eq. 6.

2.3 Monte Carlo simulations

Once the best fit parameter space has been determined through the optimization described in Sec. 2.2, a Monte Carlo sequence is run with a set of 1000 simulations. In each simulation, two parameters are perturbed:

Float time coefficients The float time coefficients are perturbed by taking random curves between the laboratory and in-situ machine vision curves.

Alpha values The alpha values are perturbed by generating Gaussian noise using the covariance matrix (see Eq. 8), centered at the optimal alpha values.

When summarizing, we present the median value as a "representative case" and the fifth percentile as the confidence interval from which we derive uncertainty.

3 Results

The surface-transport optimization methodology described in Sec. 2.2 demonstrates a consistent ability to represent the position and spread of the in-situ GPS buoys. The distance between the average location of the modeled trajectories and the average location of the GPS buoys is typically less than $50\,\mathrm{km}$. The figure below shows the modeled and in-situ trajectories from an example deployment, conducted in summer of 2023.

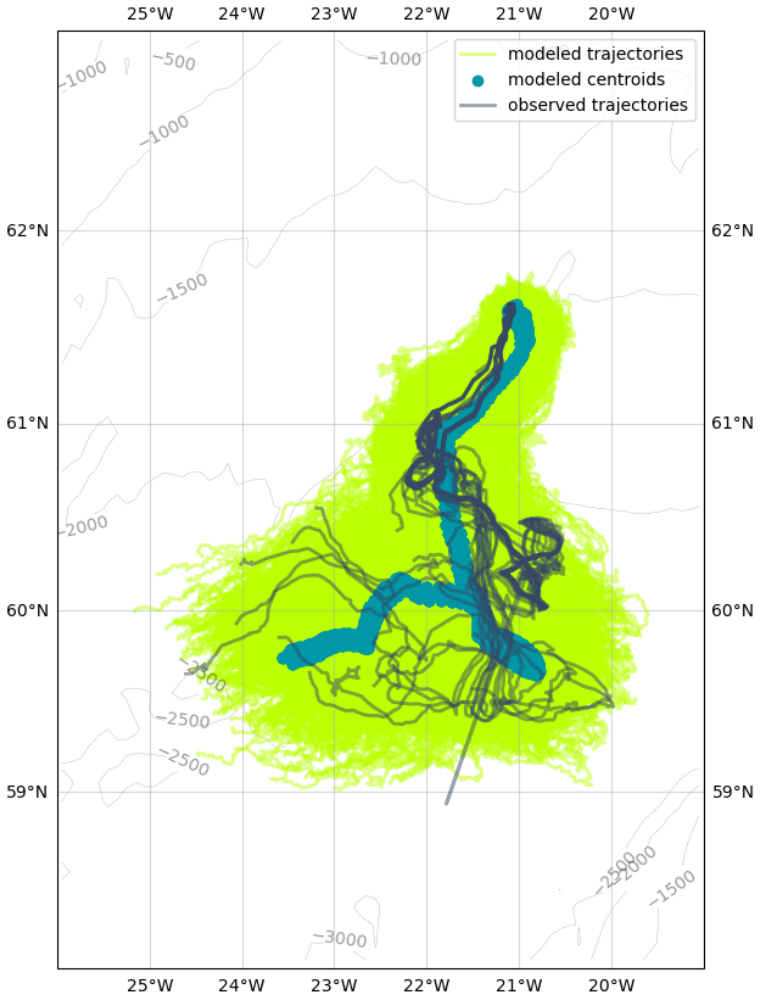

Figure 1: The optimized trajectories for IS-CD-6. Dark blue lines are the trajectories from the in-situ GPS buoys, green lines are the modeled trajectories, and blue dots are the weighted averages of the modeled trajectories.

Figure 1 demonstrates that the modeled trajectories accurately simulate the trajectories of the in-situ GPS buoys, and can be used to fully describe the terminal location of deployed material.

References

Cummings, J. A. and Smedstad, O. M. (2013). Variational data assimilation for the global ocean. In Park, S. and Xu, L., editors, *Data Assimilation for Atmospheric, Oceanic, and Hydrological Applications*, volume II. Springer, Berlin, Heidelberg.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Savater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., and Thépaut, J.-N. (2023). ERA5 hourly data on single levels from 1940 to present. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*.

Jorge Nocedal, S. J. W. (2006). *Numerical Optimization*. Springer New York, NY, 2 edition.

Kehl, C., Nooteboom, P. D., Kaandorp, M. L. A., and van Sebille, E. (2023). Efficiently simulating lagrangian particles in large-scale ocean flows – data structures and their impact on geophysical applications. *Computers and Geosciences*, 175.

Polyak, B. T. (1987). *Introduction to Optimization*. Optimization Software, Inc., New York.

Turley, R. S. (2018). Fitting parameter uncertainties in least squares fitting. *Faculty Publications*, 2236.