Numerical Evaluation of Pathline Predicates of the Benguela Upwelling System

Pascal Nardini*, Michael Böttinger†, Hans Pogrzeba*, Lydia Siegfried‡, Martin Schmidt‡ and Gerik Scheuermann*

Image and Signal Processing Group, Institute for Computer Science, Leipzig University, Leipzig, Germany
E-mail: nardini@informatik.uni-leipzig.de, kontakt@hanspog.de scheuermann@informatik.uni-leipzig.de

†German Climate Computing Center, Hamburg, Germany
E-mail: boettinger@dkrz.de

‡Leibniz Institute for Baltic Sea Research, Warnemünde, Germany
E-mail: lydia.siegfried@io-warnemuende.de, martin.schmidt@io-warnemuende.de

Index Terms—component, formatting, style, styling, insert

Abstract—Using simulation data of a regional ocean model, Nardini et al. [1] applied pathline predicates for a detailed post-hoc analysis of the Benguela upwelling system. In this work, we evaluate the accuracy of this technique. Using different temporal samplings, we aim at finding minimum requirements for the temporal resolution of the flow data in the context of retroactive particle pathline techniques. Besides the flow field, our simulation data contains synthetic tracer fields for different tracer source regions. Using the flow data, dense trajectories are computed to enable deriving "emulated tracer fields" based on the local ratio of pathline particles originating from tracer source regions to other ones, which can then be compared to the original tracer fields. We find that the emulated tracer concentrations are overestimated in comparison to the original ones. However, the shape of the regions with high tracer concentration can be reproduced.

I. INTRODUCTION

Driven by surface winds, coastal upwelling systems transport cool and nutrient-rich water from greater depths to shallower layers of the ocean and supply in this way nutrients to the marine ecosystem. Beside a lack of understanding of upwelling in detail, its importance to marine life, its variability and potential future changes in a changing climate further motivate research activities.

At the example of simulation data for the Benguela upwelling system, one of the four major eastern boundary upwelling systems (BUS) [2], [3], we apply Lagrangian techniques to complement analysis methodologies already applied by the domain scientists in order to gain a deeper understanding of the complex spatio-temporal processes characterizing the phenomenon. Among the methodologies used by domain scientists is the computation of passive tracer fields with geographically differing source regions to estimate contributions of different inflows to the upwelling system [4]. For this technique, additional 3D scalar fields are defined in the model code and initialized with a concentration of 0.0. From a user defined starting time on, grid cells within a source region (that has to be defined before starting the simulation) are set to a concentration of 1.0 at each simulation time step. Within the simulation, advection and diffusion processes transport and dilute the passive tracer. Given that the most important source regions contributing to the upwelling system are covered by the different synthetic tracer source regions, analysis of the spatio-temporal evolution of the concentrations of each tracer, mostly accomplished by studying horizontal and vertical sections at different time steps, allows tracking related flow and estimating their individual roles in the system.

However, with respect to more focused analyses, we aim at a 3D-methodology that allows to identify actual upwelling of individual water particles and a detailed analysis of their pathlines. Nardini et al. [1] presented the use of the pathline predicates methodology developed by Salzbrunn et al. and Born et al. [5]–[8] for a study of the BUS. Based on the time-dependent 3D flow field simulated with a regional ocean model, they computed space-filling trajectories covering the full model grid and used subsequent pathline predicates to select only trajectories of particles encountering upwelling, which enabled creating direct visualizations of the spatio-temporal structure of the BUS. Furthermore, analyses of the selected pathlines showed a distinct annual cycle in the upwelling activity, which fits well to observation-based analyses found in literature (cf. figure 4 in [9]).

The major advantage of the pathline predicates method over passive tracer advection calculated directly within the simulation is the ability "to look back in time". For any area of interest, traversing trajectories can be examined with respect to their past locations, and source regions of flows are identified relatively straightforward, while the optimal source regions for synthetic tracers need to be determined in a tedious and compute-intensive trial and error process. However, to save storage space and reduce the simulation costs, ocean model data is typically stored at time intervals much larger than the model time step. In this work, we demonstrate the impact of different temporal output samplings on the possible accuracy of post-hoc Lagrangian techniques.

LEVIA’18: Leipzig Symposium on Visualization in Applications 2018
This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).
Our analysis includes two parts. The first part compares, using mean absolute distance, the trajectory positions derived using data with coarser temporal samplings with those of a 3-hourly dataset. The second part of our analysis is using one of several passive tracer fields directly computed within the ocean simulation using the original short model time steps as gold standard. We use the calculated trajectories from one of the datasets and simulate a trajectory-based tracer field with a technique similar to the upwelling particle ratio (UPR) technique described in Nardini et al. [1]. After that, we can compare the different resulting emulated tracer fields with the passive tracers of the simulation. For our evaluation, we aim at determining minimum requirements on the temporal resolution of flow data for retroactively applying particle pathline techniques for visual and statistical analysis. By analyzing the skill in reproducing a synthetic model tracer field, we can set up rules for using particle pathline methods in general.

II. RELATED WORK

As mentioned before, we evaluate the method described by Nardini et al. [1] that builds upon the concept of pathline predicates developed by Salzbrunn et al. and Born et al. [5]–[8]. One part of the problem, the different precision of particle integration methods, has been well studied in numerical mathematics and received attention in visualization literature. An early work by Pang et al. in 1997 [10] visualized the difference between Euler and Runge-Kutta integration as an example of uncertainty visualization. Another aspect of our work is a comparison between Eulerian methods (e.g. passive tracers) and Lagrangian methods (e.g. trajectories). All of fluid mechanics is concerned with these two fundamental lines of thought and both approaches have their strength.

A comprehensive state-of-the-art survey on Lagrangian techniques and tools for oceanographic applications was recently given by Van Sebille et al. [11]. Stohl [12] gave a well-received overview of errors in trajectory calculation for the atmosphere. Physical experiments with balloons or smoke were compared with computed trajectories to estimate the precision of this method. Lagrangian particle dispersion models (LPDM) were suggested to account for small-scale turbulent processes.

The problem of reconstruction of pathlines from vector fields at a coarse temporal resolution has also been addressed by Chandler et al. [13]. They looked into the case of particle based flow simulation such as smoothed particle hydrodynamics, and by study of Qin et al. [14], who evaluated the impact of coarse temporal samplings on the accuracy of Lagrangian techniques in the field of ocean model analysis. In contrast to our study, they used, however, data of a global high resolution ocean model with a minimal temporal resolution of 3 days. For coarser temporal resolutions of 9 to 30 days they found mean changes in transports of up to 25 percent; the largest errors occurred in regions with high eddy kinetic energy (EKE), which also applies to the northern part of our simulation domain.

III. SIMULATION DATA

For this work, we use ocean current data and passive tracer fields from numerical simulations carried out with a regional hydrodynamical model based on the modular ocean model MOM version 5 [15]–[17]. The model domain covers the southeast Atlantic, as shown in Figure 1, and is represented by a stretched rectilinear grid of the size 261 x 351 x 62. The cell size varies horizontally between 8 and 18km and vertically between 1.5m at the surface and 500m at greater depths. "Dry" grid cells (land) are masked out with a special value to prevent them from being used in the analysis. The maximal velocity magnitude simulated for this domain is of the order of less than 1 m/s.

Due to storage constraints, ocean model data is often stored at relatively coarse time intervals. Nardini et al. [1] used ocean model data stored as 5-day means - a quite long time period in spite of the original simulation time step of 20 minutes. For this work, a simulation was carried out at an output frequency of 3 hours in order to produce a data set that is suitable for retrospective pathline computations; the maximal distance traversed by particles within 3 hours is of the order of the minimal cell size (~10 km). This temporal sampling was the finest possible without encountering unacceptable I/O times on the supercomputer. For a systematic evaluation using different temporal samplings we derive datasets with coarser temporal samplings by computing temporal means based on the 3-hourly data for intervals of 6-hours, 12-hours, 1-day, 2-days, 3-days and 5-days. In addition, we use a passive tracer field that was computed within the simulation at the original simulation time step of 20 minutes.

![Fig. 1. Example for filtered tracer pathlines visualized with draw line set method. The image shows trajectories that contribute to upwelling in the Benguela region between 15°S and 27°S, which are colored according to different tracer source regions. Inset: Full simulation domain south-west of Africa (black rectangle).](image-url)
IV. METHODS

We evaluate the pathline computations for the temporally differently sampled data with regard to two aspects: First, we derive and compare the positional differences between the trajectories. Secondly, we emulate passive tracer fields by selecting only particles with pathlines that started from the same grid cells as the passive tracers. We derive the ratio between the corresponding particles and other particles in each cell. The resulting density fields are then compared with the original tracer fields.

For the comparison of pathlines computed on the basis of different temporal resolutions, we emulate a 6-hourly, 12-hourly, 1-daily, 2-daily, 3-daily and 5-daily data set by temporal smoothing and downsampling of a given 3-hourly data set, or more specifically, by computation of according temporal means.

A. Trajectory Density Field

For our evaluation, we need a dense set of trajectories covering the whole simulation domain. In order to calculate such a set, we follow the method developed by Salzbrunn et al. [5], [6] and initialize a trajectory set by seeding $N$ weightless particles at time $t = 0$ in each of the models grid cells. Following a Runge-Kutta 4th-order scheme (RK), we compute their pathlines for the remaining time steps. To cover the complete simulation time with a space filling set of trajectories, we check at each time step the number of particles contained in each cell. Whenever a grid cell is found with less then $N$ particles, new particles are seeded in this cell at this time step. In case of $N = 1$, the new particle will be seeded in the center of the cell. In the other cases, the cell is split into $N$ subcells and if one of these subcells is found empty, a new particle is seeded in the center of this subcell. As these particles only propagate forward from the current time step through all remaining time steps in this first path, we also perform a backward integration for each particle added (to the very first timestep or until they leave the domain).

For the different temporal samplings (3h, 6h, 12h, 1d, 2d, 3d, 5d), we carry out a set of experiments using the RK 4th-order scheme with an integration time step of 3 hours, and, by setting the minimum number of particles to $N = 1, N = 4, N = 16$, for different spatial subsamplings. For compatibility with [1], we also run experiments for the different temporal samplings using the Euler integration method (only with $N = 1$).

B. Pathline Error Estimation

For the difference analysis between the different time resolutions and the two integrations methods, we used $N = 1$ particles in each cell. The initial particles have the same start position and the same starting time thanks to our emulation modifications. We propagate all particles forward in time to the end or until they leave the domain. Looking for a value, which describes the deviation from trajectories computed with a low temporal sampling ($\Delta t = 6h, 12h, 1d, 2d, 3d, 5d$) to the path lines calculated with the finest temporal sampling ($\Delta t = 3h$), we calculate the mean absolute distance (MAD), the mean squared distance (MSD), and the mean line difference (MLD).

\[
MAD(m, t) = \frac{1}{n} \sum_{i=1}^{n} ||pos_{3h}^t(i) - pos_{m}^t(i)||
\]

\[
MSD(m, t) = \frac{1}{n} \sum_{i=1}^{n} (pos_{3h}^t(i))^2 - (pos_{m}^t(i))^2
\]

\[
MLD(m, t) = \frac{1}{n} \sum_{i=1}^{n} ||pos_{3h}^t(i) - pos_{m}^t(i)|| - ||pos_{3h}^t(i) - pos_{m}^t(i)||
\]

Here $pos_{3h}^t(i)$ is the position of particle $i$ at time $t$ for the 3h data while $pos_{m}^t(i)$ is the position of particle $i$ (i.e. a particle started at the very same position and time) at time $t$ integrated using the data with $\Delta t = m$. MAD and MLD compare the particle position with the position in the 3-hour mean trajectory set directly. In contrast, MLD checks the distances between initial position and current position.

![Fig. 2. This plot shows the mean absolute distance (MAD) of pathlines integrated on the 6h, 12h, and 1d dataset compared to the 3h data. As integration methods we use modified Euler and Runge Kutta 4th-order schemes.](image312x397 to 563x516)

C. Trajectory Based Tracer Field

The idea of the trajectory based tracer field is to reproduce a passive tracer field based on information contained in our trajectory sets. While the classical passive tracers and their source regions have to be defined ahead of the simulation run, pathline predicates allow a post-processing analysis which is much faster and much more flexible. Advection and diffusion of the passive tracers, which distribute them in the model domain, is calculated within the simulation using time steps of 20 minutes. During the simulation, the passive tracer is continuously released at a defined source plane or region from a defined time onward till the end of the simulation.

For emulation of the tracer concentration, we imitate the source area by marking all cells intersecting the tracer source region as tracer cells. All particles running through these cells are marked as tracer particles. By evaluating the according pathline predicates, the trajectory set is partitioned into a set of non-tracer-trajectories and a set of tracer-trajectories (Figure 1). As a generalization of this UPR technique proposed by Nardini et al. [1], we define the \textit{predicate particle ratio (PPR)} between marked particles (here: particles marked with the tracer predicate) and unmarked particles in each cell at
each time step, which is our approximation of the local tracer concentration. For the visualization of the passive tracer concentration and PPR, we use volume rendering based on simple ray casting.

![Image of distribution of passive tracer and PPR](image)

**V. RESULTS**

**A. Pathline Error Analysis**

We compare pathlines starting at the same spatial position and time stamp for different temporal resolutions. We use the best available temporal resolution, i.e. \( \Delta t = 3h \), with a standard RK integration (stepsize of 3h) as reference. In Figure 2, we show the mean average distance (MAD) for the 6h, 12h, and 1d resolutions. Besides comparing pathlines computed for the different temporal resolutions, we also compare a modified Euler scheme with the RK integrator. The graphs show that the error grows with time. As expected, the lowest MAD is achieved by using the RK integration using the 6h data. While using a time step of 6h creates an average error of 41 km within one year, we see 70 km for 12h and 100 km for daily data. We conclude that the temporal resolution has a severe impact on the accuracy of individual particle pathlines also for temporal resolutions of much less than one day. It seems likely that we would also see some differences between the given 3h data and the original simulation timestep of 20 minutes, which is chosen to keep the model numerically stable for a given spatial resolution.

Interestingly, the increase in accuracy by using a higher order integration method seem to be smaller for samplings above 12h, as can also be seen on Figure 2 by comparing the curves for Euler and RK. Nevertheless, there is a substantial difference that justifies the use of the computationally more expensive RK integration in this case. We have also analyzed MSD and MLD, but they show very similar effects and do not lead to generally other findings.

**B. Comparison of Tracer and Emulated Tracer**

Nardini et al. [1] introduced the upwelling particle ratio (UPR) as the ratio between upwelling particles and other particles for each cell at a given time. This post-hoc derived scalar field was introduced to enable application of scalar field visualization methods, in particular volume rendering, to filtered trajectory data. We further investigate the potential of PPR to emulate passive tracers with respect to shape and position of the concentration fields as well as to absolute concentrations. Figure 3 shows six volume rendering visualizations. The upper left image shows the concentration of the passive tracer as directly simulated by the ocean model (A), while the remaining five (B-F) show the post-hoc derived PPR-fields for different temporal samplings and different numbers of particles per cell.

Images B-D show PPR based on \( \Delta t = 3h, \Delta t = 1d \), and \( \Delta t = 5d \) data with the minimum number of particles per cell set to \( N = 4 \). The remaining two images E-F show the results for \( \Delta t = 3h \), but with \( N = 1 \) resp. \( N = 16 \). It is evident at first glance that all emulated tracer fields overestimate the concentration compared to that of the original passive tracer - as can be seen from the color code. Interestingly, the quantitatively best emulation is F, the one with a 4-particle sampling and 5d time interval.

**Fig. 3.** Distribution of the passive tracer (Figure A) and PPR distributions (B-D) 69 time steps after tracer initialization (about 1 year) with a varying temporal sampling (\( \Delta t \)) and minimum number of particles per cell (\( N \)). Parameters: B with \( \Delta t = 3h, N = 1 \); C with \( \Delta t = 3h, N = 4 \); D with \( \Delta t = 3h, N = 16 \); E with \( \Delta t = 1d, N = 4 \); F with \( \Delta t = 5d, N = 4 \); G with \( \Delta t = 5d, N = 16 \).

![Image of average tracer difference](image)

**Fig. 4.** The plot shows the average difference between PPR and the concentration of the passive tracer. Cases of a zero value in both fields receive no consideration.

However, the shape of the area with noticeable tracer concentration above 0 is reproduced rather well by PPR. To quantify this finding, we calculate the average difference between the values of the simulated tracer fields and the original one. Figure 4 shows a plot with the average difference of the same variants as before. In the case of \( \Delta t = 5d \) we could improve the results by using more than one particle per cell. Apart from the \( \Delta t = 5d \) dataset, the results are comparable, similarly to Figure 3.
VI. DISCUSSION AND CONCLUSION

We evaluated the precision of the pathline predicates approach proposed by Nardini et al. [1] for temporal samplings between 3h and 5d. Similarly to [14] for temporal samplings between 3d and one month, we found that the temporal resolution of the data has a rather strong effect on the pathlines also for much shorter time intervals. If one uses a sampling of 6h instead of 3h, the mean absolute distance is already of the order of the grid interval after a few days. Since typical analyses of climatic processes span months to years, this positional error has to be taken into account. This finding also raises the issue of the difference between the available 3h resolution and the 20 minute simulation time step used within the model. As expected, it is shown that Runge-Kutta integration with a step size of about the temporal resolution of the data works much better than the modified Euler scheme. Here, additional research is necessary, including methods like LPDM as described by Stohl [12].

Our evaluation of PPR as a replacement for passive tracer concentration shows two major results. First, even the highest tested number of particles (a minimum of $N = 16$ particles in each cell at any timestep) and the highest available temporal resolution $\Delta t = 3h$ shows a substantial deviation of PPR from tracer density. The differences are much larger than the differences to the PPR values derived from lower temporal resolutions except for the variant with $\Delta t = 5d$ and $N = 5$. With this variant, the big step sizes in combination with a small number of particles lead to rather small values. In contrast, we found that PPR indicates quite well where the tracer would move, i.e. which cells are reached. With some caution, we can therefore state that the emulated tracer ratio allows for qualitative studies, but may not directly be used for quantitative mass transport analyses within the upwelling system. This somewhat surprising combination of simulated particle transport to the correct locations and, at the same time, a clear overestimation of the resulting emulated tracer density is a pressing question for further research.

A possible explanation might be the lack of diffusion that passive tracers undergo while the particles in our implementation do not. However, for sparse temporal samplings, Qin et al. [14] couldn’t reduce the error by adding diffusion to their trajectories. The work of Wolfram et al. [18] about diffusivity could be a reasonable enhancement of our method in order to evaluate the impact of diffusion on pathlines for dense temporal sampling. Also, the passive tracers in ocean models are designed to accomplish exact mass conservation, while the new particles inserted by our method to ensure dense sampling may compromise mass conservation.

As future work we envisage to further enhance our methodology with respect to a quantitatively correct reproduction of passive tracer concentrations. To do this, more experiments and more detailed analyses will be needed. Furthermore, we would like to improve the visualization of filtered pathlines fulfilling researcher defined requirements. The work of Festl et al. [19] or Whitaker et al. [20] could be a promising approach.

ACKNOWLEDGEMENTS

This work was funded by the German Federal Ministry of Education and Research within the project Competence Center for Scalable Data Services and Solutions (ScaDS) Dresden/Leipzig (BMBF 01IS14014B).

REFERENCES


