**EGPGV 2021 Revision Report – Submission 1005**

**coordinator review**  
  
  Summary Rating [Early Submission]  
  
    Major revision: The reviewers request major changes to make this paper acceptable.  
  
  Summary Review Text [Early Submission]  
  
    The reviewers have mixed feelings about this paper.  
  
    On the upside, the paper extends work by Agranovsky et al. that is very useful to  
    the field, and it also extends the aforementioned study with a very thorough  
    evaluation. The idea presented in the paper is simple but still interesting and  
    might also be relevant in settings other than flow-vis.  
  
    The reviewers however conclude that from the evaluation and theoretical  
    considerations alone it is hard to assess how bad the bias introduced by the  
    method can get in practice; it seems quite simple to construct trivial flows (or  
    regions) that are difficult to represent by local flow maps. In particular, this  
    is not only limited to velocity, but to general non-linear flow dynamics such as  
    abruptly changing flow behavior and bifurcations.  
  
    The reviewers also note that publication of the source code (or of a  
    representative sample code) under an open source license would make the  
    contribution far more valuable.  
  
    The authors should prepare a major revision that comprises a visualization that  
    helps to better assess the impact of the error on the flow; in the reviewers'  
    opinion such a visualization could e.g. be based on Finite-time Lyapunov exponent  
    (FTLE) ridges which should be complex enough to judge that error when compared to  
    ground truth generated for example with the method by Agranovsky et al. or with  
    the Eulerian method.

We prepare FTLE visualizations produced by both (Lagrangian\_{Dist} and Lagrangian\_{Local}) techniques for each data set.

    The authors should also publish a sample or prototypical code as part of their  
    revision.

In situ Lagrangian analysis capabilities for integration with a simulation code can be accessed here: <https://github.com/Alpine-DAV/ascent.git>

Single-node shared-memory implementation is available here: https://gitlab.kitware.com/vtk/vtk-m/-/blob/master/vtkm/filter/Lagrangian.h

    For the paper to be acceptable for publication, the authors should also address  
    the other issues brought up by the reviews.

We address issues brought up by reviewers and comment on each such revision in the remainder of the report.

**Reviewer 1**  
  
  Overall Rating [Early Submission]  
  
    Probably accept: I would argue for accepting this paper.  
  
  Expertise [Early Submission]  
  
    Passing Knowledge  
  
  Summary [Early Submission]  
  
    The paper proposes a communication-free model to compute flow maps in situ and  
    compares that to the approach proposed in [ACG\*] that does require communication.  
    The approach is based on computing only local flow maps and just ignoring  
    particles crossing domain boundaries. The respective trajectories are instead  
    interpolated from valid trajectories on adjacent processors during post-hoc  
    analysis. The error from that is analyzed theoretically and in practice by  
    comparing (amongst others) to ground-truth trajectories obtained using the  
    approach from [ACG\*].  
  
  Contribution [Early Submission]  
  
    The authors present a very simple strategy to increase scalability of particle  
    advection with Lagrangian flow maps on distributed memory systems, namely, by just  
    discarding particles that cross domain boundaries. During post hoc analysis, this  
    is accounted for by interpolating adjacent trajectories from neighboring compute  
    ranks. This is obviously a very simple optimization; the authors very thoroughly  
    evaluate the error from that wrt. to storage interval, grid resolution, different  
    data sets, comparisons to both Eulerian and Lagrangian-with-communication ground  
    truth methods, etc. and show that the approach is viable for a wide variety of  
    configurations.  
  
  Clarity of Exposition [Early Submission]  
  
    The paper is well written. The following are only some minor remarks.  
  
    - Figure 1 I found a bit hard to comprehend; specifically, because of the  
    description in the text that the flow maps are denoted FX->Y, and X and Y  
    customarily not used for time / storage intervals. F presumably means snapshot  
    file? The figure seems to be based off of Figure 2 in [ACG\*], where the concepts  
    are explained in more detail. The figure and its description should be edited for  
    clarity.  
    - Figure 3 (a)-(d) are never referenced in the text. (b) and (d) presumably depict  
    the Delaunay triangulations mentioned in Section 3.4, but this is not mentioned in  
    the caption.  
     - In general, please provide more informative figure captions. When devising  
    those, imagine a reader skimming through only the figures and captions; in the  
    best case, the gist of the paper should be comprehensible to them from just the  
    captions.

Thank you for your feedback. We address both the above issues by making the figures and the corresponding captions more self-contained. We hope this modification helps readability and improves clarity.

    - Section 5.3: Typo: Lagrangian\_{D}ist

We have addressed this typo.

  Technical Soundness [Early Submission]  
  
    Apart from the following the methods and the way they are described seem  
    technically sound to me:  
  
    - Figure 3 (b) and (d): do the triangulations really reach across boundaries? The  
    text in Section 5.3 suggests otherwise ("a Delaunay triangulation is per- formed  
    using CGAL [CGA20] on a local cluster"). Please define what you mean by "local  
    cluster" and clarify in the text / caption.

Thank you for pointing this issue out. We believe the confusion arises from the node boundaries in (b) and (d) in the notional example. We address this issue by correcting the figures. Further, we change our “local cluster” to “single-node workstation” for clarity.

  Quality of References [Early Submission]  
  
    I can't identify any missing references, but I only have passing knowledge in this  
    area and am not too familiar with the complete state of the art.  
  
  Reproducibility [Early Submission]  
  
    The methods should be reproducible, possibly by also referring to [ACG\*].  
  
  Evaluation [Early Submission]  
  
    Yes, the method is very thoroughly evaluated. My only critique is that the  
    evaluation only presents statistics, w/o giving a sense of the visual error that  
    can be observed. I understand that the error will not manifest as artifacts in the  
    visualization but rather trajectories just being "wrong" or missing when for  
    example the velocity is very high. Nevertheless, a visual comparison based on a  
    post hoc analysis and pointing out where the visualizations differ (e.g., for a  
    case where the method works well, and for a case where it doesn't, like the Jet  
    Flow data set) would be highly appreciated.  
  
  Explanation of Recommendation [Early Submission]  
  
    The paper presents a very simple optimization to improve scalability of in situ  
    flow map communication, namely just performing no communication whatsoever across  
    ranks assigned to the spatial domain composition.  
  
    Pros:  
    - The authors provide a comprehensive study of the error resulting from that, and  
    the increase in scalability observed; the error and scalability studies are very  
    thorough, take different types of configurations into account.  
    - The evaluation stresses configurations that result in large error (e.g., large  
    storage intervals or data sets with high velocity) and thereby uncover the  
    (obvious) limitations of the method; it is shown that for a wide variety of data  
    sets the error is acceptable.  
  
    Cons:  
    - Figures and captions are a bit hard to read; in the best case, the gist of the  
    paper should be comprehensible from just the figures and captions.  
    - Apart from the measured error, I'm missing a 3D visualization showing the impact  
    in terms of a ground truth comparison and using post hoc analysis.

Thank you for the feedback. We address the first issue by improving our descriptions in the captions. This is valuable feedback that we will carry forward in future works as well. We address the second issue by performing post hoc analysis using the extracted Lagrangian flow maps to produce 3D FTLE visualizations enabling a comparison and visualization of the possible impacts of using a communication-free model.

    I understand that this optimization is really a very simple one; nevertheless I  
    feel that the study is interesting, contributes to the state of the art, and is  
    well suited for EGPGV. Generally, the pros outweigh the cons for me. The latter  
    (figures and captions, 3D visualization to show the error for some assorted data  
    sets and configurations) should be addressable in a revision  
----------------------------------------------------------------

**Reviewer 2  
committee member review**  
  
  Overall Rating [Early Submission]  
  
    Borderline: The strengths and weaknesses balance for this paper.  
  
  Expertise [Early Submission]  
  
    Knowledgeable  
  
  Summary [Early Submission]  
  
    The authors presented a variation of the technique introduced in Agranovsky et al.  
    The more significant change was the elimination of the intercommunication step to  
    calculate flow maps. Removing communication improved the scalability of the  
    technique as the results showed, while some tradeoffs between reconstruction  
    accuracy of discarded particle trajectories and full pathlines, should be  
    considered.  
  
  Contribution [Early Submission]  
  
    Improved scalability while maintaining some acceptable error while reconstruction  
    the flow maps.  
  
  Clarity of Exposition [Early Submission]  
  
    Yes, paper well organized and easy to follow.  
  
  Technical Soundness [Early Submission]  
  
    Yes, the paper also report  some detailed results.  
  
  Quality of References [Early Submission]  
  
    No  
  
  Reproducibility [Early Submission]  
  
    Yes  
  
  Evaluation [Early Submission]  
  
    Yes, it includes four different use cases were the technique is verified in  
    different contexts, i.e. performance evaluation (in-situ) and  post-processing.  
  
  Explanation of Recommendation [Early Submission]  
  
    Removing communication improved the scalability of the technique as the results  
    showed.  
    Still, interested parties in adopting this technique  should be aware of the  
    tradeoffs between reconstruction accuracy of discarded particle trajectories and  
    full pathlines.

Thank you for the feedback. We include this sentiment in the limitations.

----------------------------------------------------------------  
 **committee member review**  
  
  Overall Rating [Early Submission]  
  
    Probably accept: I would argue for accepting this paper.  
  
  Expertise [Early Submission]  
  
    Knowledgeable  
  
  Summary [Early Submission]  
  
    This paper  proposes a communication free algorithm for computing in situ flow  
    representation based  
    on lagrangian flow maps. The original algorithm was introduced by AGRANOVSKY et  
    al. in 2014.  
    The main idea is to compute in situ short flow maps from uniformly spread particles  
    and save these short  
    trajectories. And then build long trajectories post-hoc from the particles chosen  
    by the user using  
    the short trajectories computed in situ. This algorithm enables to significantly  
    reduce the amount  
    of data to save compared to a traditional approach for an equivalent quality of  
    results. This also  
    enables to speed-up the post-hoc computation as the amount of data to handle is  
    reduced (less I/O  
    costs) and part of the computations have already been done in situ (so less  
    computation load  
    post hoc).  
  
  Contribution [Early Submission]  
  
    The original algorithm required communication steps during the in situ phase to  
    move the particles  
    from one cell to another if reaching the cell boundary before the end of the  
    particle transport  
    time. Here the authors propose to simply drop this communication phase and discard  
    the particles that  
    reach cell boundaries before the end time.  This leads to significant performance  
    gains during the in situ  
    phase of course.  The authors present extensive experimental results,  
    comparing the original algorithm with their version, but also with the traditional  
    Eulerian  
    approach. Their results shows that the amount of dropped particles is often  
    reduced to a few percents  
    and that the reconstruction error during the post hoc phase is also small. Notice  
    that they clearly  
    compute the error only for the particles that are concerned with discarded  
    trajectories, as other  
    trajectories are not affected. Obviously as the cell size decreases the percentage  
    of discarded  
    particles tend to increase.  
  
  Clarity of Exposition [Early Submission]  
  
    Could be improved (see below)  
  
  Technical Soundness [Early Submission]  
  
    ok  
  
  Quality of References [Early Submission]  
  
    Probably ok  
  
  Reproducibility [Early Submission]  
  
    Probably but making code available would be much better.  
  
  Evaluation [Early Submission]  
  
    Yes  
  
  Explanation of Recommendation [Early Submission]  
  
    First I remember ready the original AGRANOVSKY et al. in 2014 paper with enthusiasm  
    as seeing there  
    on of the first true in situ algorithm (this paper got LDAV best paper).  And I  
    agree with the authors that this original paper was  
    short on experiments (not even 100% sure they actually ran any real in situ  
    experiment). So the  
    current paper is welcome in the sense that it comes with extensive experimental  
    results of this  
    approach. And the results confirm the quality of AGRANOVSKY algorithm in term of  
    I/O savings and quality  
    of final results. Then the improvement they propose, though simple, is effective.  
  
    Where I would be more critical about this paper is:  
    - Writing:  not well organized/structured, making it difficult to read. Bellow  
    the bar of what we usually expect from a scientific paper.

We hope our revisions – particularly improving the descriptions in the captions and additional editing – helps readability. (I note most other reviewers said the paper is well written)

    - The authors kind of stay in the middle with the evaluation of their approach. Is  
    the proposed algorithm usable as the entailed errors can stay under control or the user can be warned that some of the data may be subject to some errors, or the algorithm requires further work as suggested through the proposed future direction of investigation.

We believe this is an application-specific problem. The errors are dependent on the specific configuration parameters. Thus, for general application we suggest future work that : 1) can improve on linear interpolation for post hoc reconstruction – there are new studies that consider up-sampling of FTLE visualizations from Lagrangian flow maps, or 2) modify in situ extraction to adapt.

    - They should make the code available and present their work as a benchmark suite  
    (but this is till doable for the final version).  
  
    "In this paper, we do not analyze the parallel I/O times, and consider I/O  
    optimization methods to be beyond the scope of this work." I think this is a mistake as one of the main motivation of in situ processing is I/O savings. The saving intervals has clearly an impact on the performance. And one way to limit the amount of discarded particles is to keep this interval small. It would not be complicated to report on the in situ phase exec time, including the I/O time (or this is really not well done and you prefer to "hide" it :-) Does fig 4 includes I/O time or not?

Figure 4 does not include I/O times. We have found for the size of files written by individual ranks in our experiments, write times on Summit are very fast. In fact, they are faster than a single cycle of in situ Lagrangian computation. Thus, for an infrequent operation, the cost is low.

OR

Figure 4 does not include I/O write times. The reviewer is right as our I/O operation was not optimal - we wrote data from VTK-m directly and VTK-m does not support writing binary VTK files. Thus, we have not included these results. We conducted additional studies to measure time to write binary files of the same size concurrently and found for the size of files written by individual ranks in our experiments, write times on Summit are very fast. Overall, for the file sizes we considered, write times are faster than a single cycle of corresponding in situ Lagrangian computation. Thus, for an infrequent operation, we expect the cost to be low. In the implementation provided, I/O write is performed via Ascent (rather than VTK-m).

    I miss in this paper one, even short, clear presentation of the algorithm. Fig 1. gives some elements in the introduction and the next elements are kind of spread in the paper.  I understand that the authors assume that it was already published in AGRANOVSKY et al., but for completeness and as the current algorithm differs from the original one, a clean presentation would help the reader.

I thought the notional example was quite clear. Thoughts?   
  
    I am not convinced by the Theoretical analysis that is detailed for the easy part (eq. 2) while it goes quite fast for the difficult part (3, 4). This can be improved. As it is I would not guarantee this analysis is fully relevant.

Should I ask Roxana if she has any thoughts on this review?  
  
    Fig 5: Put in the caption that left is for the dist algo and the right for the local algo. Would be better to actually have the exec time for both algos and split this exec time in com/advection/I-O

Thank you for your feedback. We improve the description in the caption to distinguish the two more clearly. Our objective with presenting these separately is to highlight the increased cost of weak scaling for particle advection on a single node – where the number of GPUs used increases. In contrast, communication benefits from weak scaling on a single node and shows co-location of more ranks on a single node can benefit from intra-node MPI optimizations.

    Even if the authors made the efforts to detail the error distribution, they do not analyse if there is a spatial impact. I would expect that for the cells where the flow is faster, the percentage of discarded particle be significantly higher than for others, and so the error. But it is unclear if this happens in some critical areas. For instance, plotting a histogram with the amount of dropped trajectories per cell could be one way in that direction.  Something that could probably be easy to do is to attach to post hoc reconstructed trajectories the portions that have been reconstructed from a degraded density of short trajectories. This may be sufficient to warn users that in some areas the data must be taken more carefully.

Thank you for this feedback. We did consider that spatial patterns exist as well. This could be something interesting to do.

    Reproducibility. I did not find mention of code availability in the paper. That a point that could significantly boost the impact of this paper. As the author did a significant work in setting-up these implementations and experiments, that would be very valuable to make them available to the community as it could become a benchmark.  
  
  
    2 typos page 7:  
    - number of CNs increaseS  
    - accurately (under 100 ....)

We address these typos.  
  
  
----------------------------------------------------------------  
  
reviewer 4 review  
  
  Overall Rating [Early Submission]  
  
    Probably reject: I would argue for rejecting this paper.  
  
  Expertise [Early Submission]  
  
    Knowledgeable  
  
  Summary [Early Submission]  
  
    The paper proposes an optimization for the in situ computation of flow maps. To  
    avoid communication and thus to improve scalability, the flow map is limited to a  
    local region. The authors study the introduced error and scalability of the flow  
    map computation.  
  
  Contribution [Early Submission]  
  
    An optimized flow map computation is introduced, which speeds up the in situ  
    analysis performed during a simulation. The scalability of this method is  
    evaluated extensively.  
  
  Clarity of Exposition [Early Submission]  
  
    The paper is well written.  
  
    There are some minor points:  
    - The citation abbreviations are inconsistent. Often, the year is missing, e.g.  
    [ACG\*].  
    - In the references [C\*20], has missing author names.  
    - Table captions should be above the table.  
    - Some figures in  Sec. 6 are barely, if at all, discussed in the paper. Maybe  
    some figures can be moved to a supplementary document or even left out?

I need to figure out how to correct all the references – I’m not sure about the reason for inconsistency. EGPGV allows a maximum of 1 page of references. So I’ll need to reduce the overall length of the document through edits.

I am not seeing where it says table captions should be above the table.

I can remove some images, only retaining those at the extremes for parameter value. I can also improve our discussion of these figures in the captions.

  Technical Soundness [Early Submission]  
  
    The optimization is simple and straightforward. The authors acknowledge some  
    limitations in Sec. 7. However, I feel that the theoretical limitations are far  
    greater than described in the paper. Most concerning, the method is inherently  
    biased since it simply drops sampled trajectories that leave the local region.  
    This implies that it can systematically misrepresent a flow. Indeed, it is quite  
    trivial to construct a theoretical, but not unrealistic, flow that is not well  
    represented by local flow maps. Note that this is not limited to flows/regions of  
    high velocity, but might also be caused by bifurcations and other non-linear flow  
    behavior. Due to the inherent sensitivity to small perturbations in most time-  
    varying flows, this is deeply concerning to me.

Thoughts on systematically misrepresenting the flow --- even considering a data reduction of the form 1:8 introduces error and technically “systematically” introduces error. It is also possible to construct flows that are poorly captured by spatial subsampling. Our reasoning behind discarding trajectories is the following: When using a 1:1 data sampling rate, losing 10% of particles results in 90% as many particles as grid points. That said, the Lagrangian research space is simultaneously exploring 1:8, 1:27 data reduction factors, which use 12.5% as many particles as grid points and 3.7% as many particles as grid points. Of course, as storage interval lengths and data reduction factors increase, the impacts require analysis. We study the impact via our experiments.

While we agree with many of these sentiments, the reality is that large scale simulations often cannot afford to perform time-varying flow analysis. Lagrangian analysis introduces a potential solution, but presents different barrier to overcome in an in situ setting. We believe foundational strategies to address extraction costs can be built upon using more sampling and reconstruction strategies to improve accuracy.

    As shown in Sec. 4, the error is bounded by the time interval, velocities, and  
    spatial extent of the region. Although this is a good result, it also implies that  
    these parameters have to be set carefully and preferably adaptively. This is noted  
    as future work by the authors. In my opinion, this is far from trivial and would  
    ultimately decide how useful local flow maps can be in practice.  
  
  Quality of References [Early Submission]  
  
    To my knowledge, all important references are included and are well discussed. The  
    related work section outlines a clear need for research in this direction.  
  
  Reproducibility [Early Submission]  
  
    The method is straightforward and is discussed in detail. The evaluation is well  
    described.  
  
  Evaluation [Early Submission]  
  
    The quantitative study is well done. Although it does not convince me of the  
    usefulness of the method, it does convey interesting insights into implementing  
    scalable in situ Lagrangian analysis.  
  
  Explanation of Recommendation [Early Submission]  
  
    The paper is well written and the study is insightful. Although I like the method  
    for its simplicity, I have a lot of concerns regarding its correctness and thus  
    usefulness.  
  
    The Cloverleaf3D results are promising, especially the comparison to the Eulerian  
    representation. Results of the ABC flow look good, but the error seems quite  
    extreme for the Nyx and Jet flows. I would like to see more comparisons between  
    the two Lagrangian methods regarding the error on the ABC, Nyx, and Jet flows.  
    Currently, the improved scalability of local flow maps (less than one order of  
    magnitude) does not justify the high error, in my opinion. Although some  
    configurations with a lower error exist, it seems difficult to select such a  
    configuration beforehand and the error can grow quite large. I fear that  
    addressing this will require the adaptive method that has been noted as future  
    work.  
  
    Distributed particle tracing is well studied and not significantly (i.e. orders of  
    magnitude) slower - this makes the method seem impractical to me. I was not  
    convinced otherwise by the paper, but the authors might be able to address this.

Distributed particle tracing in an in situ context is not well studied. The nature of interfacing with the simulation, resource allocation and location, and overall scale impact the overhead. Thus, I would argue that as the scale increases it is possible that it can be a magnitude of order higher.   
  
    The distribution of errors is nicely visualized by the violin plots. However, it  
    is difficult to tell the impact of the error on post hoc flow analysis and  
    visualization methods. Studying visualizations such as the finite-time Lyapunov  
    exponent (FTLE) that operate on the flow map would give a different perspective on  
    the introduced error and how much impact it has on the flow behavior.  
  
    In conclusion, I have strong doubts regarding the correctness and thus usefulness  
    of local flow maps. I could be persuaded to accept the paper on the grounds that  
    this study can serve as a first step for future research. I do believe that this  
    is an interesting research direction.