

RC1: Andreas Stohl

This study compares different Lagrangian models in their capability to decompose different terms in the heat budget, at the example of a specific event. I have several concerns regarding this paper, which prevent me from recommending it for publication.

Thank you very much for your careful review, detailed comments and constructive suggestions, which have greatly helped to improve the quality of our paper.

Most importantly, it is a “mechanical” intercomparison exercise that does not offer any physical interpretation. Differences are found but they are not explained. A scientific paper should go beyond simply comparing and should also offer explanations. In fact, I am wondering why this study is needed at all. It is clear that trajectory models do not account for turbulence. However, all trajectories end up in the boundary layer, where turbulence is extremely important and where trajectory models fail in capturing the real transport history of an air mass. So, in my opinion, they are simply not fit for the purpose, and I would avoid using them, unless you are driving them with the output of a large eddy simulation, where turbulent motions are captured in the input wind fields.

We thank the reviewer for this critical and insightful comment. We fully agree with you about the important limitation highlighted here: pure-trajectory models do not account for boundary layer turbulence, and, as trajectories enter the boundary layer at the end, where turbulence is extremely important, trajectory models may fail to capture the transport history of the air mass. We agree with your understanding of the limitations of pure-trajectory models in this context.

However, we respectfully take a somewhat different view regarding the claim that this study is not needed at all. In this sense, the physical limitations you highlight further motivate the present comparative study.

Currently, there is a rapidly growing body of research utilising the Lagrangian temperature anomaly decomposition approach to examine the physical drivers of heatwaves (e.g., Röthlisberger and Papritz, 2023; Oppong et al., 2025; Fernández-Alvarez et al., 2025, 2026; Lemburg et al., 2026; Pérez-Alacón et al., 2025). Overall, some studies use pure-trajectory models (e.g., LAGRANTO), while others use dispersion models (such as FLEXPART) that include turbulence parameterisations. As our manuscript shows, applying these two model classes to the same heatwave event yields different physical interpretations regarding the relative contributions of adiabatic and diabatic processes. This is a critical issue for the community, as the choice of model approach may influence the resulting attribution and, consequently, the interpretation of the thermodynamic drivers of a given meteorological event.

Therefore, the main goal of our study is not to offer a novel physical interpretation of a specific heatwave, but rather to provide a mechanistic methodological comparison of the Lagrangian temperature anomaly decomposition approach itself. By conducting this mechanistic intercomparison, we are able to explicitly trace differences in the diagnosed adiabatic and diabatic contributions to the presence or absence of turbulence parameterisation in the models. Given the widespread and continued use of pure-trajectory models in heatwave research, we believe that this comparison provides useful information for the community by highlighting how the choice of model approach can affect the resulting physical interpretation. Importantly, this study does not seek to establish which model class is most suitable, but rather represents a first step towards systematically investigating these discrepancies and their implications for heatwave attribution.

We also agree that the results can not be generalised to all heatwave events, as our intercomparison is based on a single case study. As we noted in the Final Remarks and Conclusions section, assessing the generality of these findings would require a comprehensive analysis of multiple heatwaves across different regions and synoptic conditions. Finally, we hope this work can provide a useful basis for such future studies, helping to further investigate the origin of model-dependent differences and to improve the robustness of Lagrangian heatwave attribution.

Should the editor consider that resubmission of the manuscript is appropriate, we will revise the last paragraph of the Introduction to clarify the aim of the study as follows:

On this basis, there is a rapidly growing body of research using both pure-trajectory and dispersion models to examine the relative contribution of physical processes to HWs (e.g., Röthlisberger and Papritz, 2023; Oppong et al., 2025; Fernández-Alvarez et al., 2025, 2026; Lemburg et al., 2026; Pérez-Alarcón et al., 2025). However, discrepancies in the reported findings among these studies warrant further examination to accurately quantify the role of each process in HW development. Therefore, this study compares the temperature anomaly decomposition results using LAGRANTO, FLEXPART, the Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model (Stein et al., 2015), and the Hybrid Parcel Tracker (HPT; Shukla et al., 2026) for the PNW21 HW event. In particular, the main goal is to provide an initial assessment of how model-specific characteristics influence the quantification of adiabatic, advective, and diabatic contributions to temperature anomalies, and to lay the groundwork for interpreting Lagrangian HW analyses across different modelling frameworks.

For the particular problem studied in this paper, the difference between trajectory models and dispersion models seems largely a result of the neglect of entrainment of free-tropospheric air into the boundary layer in the trajectory models (which is related to the lack of turbulence in the trajectory models). Warming of air in the boundary layer is the result of diabatic warming at the surface but, equally or even more importantly, entrainment of warm free tropospheric air at the boundary-layer top. Since the trajectory models do not capture the entrainment, it is not surprising that they attribute a larger fraction of the warming to diabatic processes, whereas FLEXPART attributes more to adiabatic warming (i.e., entrainment of descending free tropospheric air into the boundary layer). I am very surprised that entrainment is not even mentioned in the paper, although it seems like a logical explanation that should be explored specifically. It seems also plausible that trajectory models ignoring turbulence will result in a wrong process attribution. For dispersion models, their reliability will be controlled to a large extent by how realistically they capture boundary layer dynamics, in particular it will depend on the daily cycle of boundary-layer depth.

We thank the reviewer for this observation. We completely agree that the entrainment of free-tropospheric air into the boundary layer could be a fundamental driver of the discrepancies between pure-trajectory and dispersion models.

As you correctly point out, the entrainment of descending, adiabatically warmed air at the top of the boundary layer is a critical mechanism for boundary-layer warming. Because pure-trajectory models omit turbulence, they could fail to capture this downward mixing of free-tropospheric air, which inevitably forces them to erroneously attribute a larger fraction of the warming to surface diabatic processes. Conversely, because dispersion models simulate turbulent motions, they capture this entrainment and attribute the warming to adiabatic descent.

We sincerely thank the reviewer for highlighting this omission. Should the editor consider that resubmission of the manuscript is appropriate, we will explicitly incorporate this mechanism into our analysis in the last paragraph of Section 3.3, as follows:

Furthermore, the representation of free-tropospheric air entrainment into the boundary layer is another factor that modulates differences between the two model classes. While pure-trajectory models fail to capture this entrainment and therefore emphasise diabatic heating, dispersion models better represent the exchange with descending free-tropospheric air, retaining the adiabatic-warming signature.

Another problem of the paper is the rather random choice of models. While different trajectory models are compared, all dispersion model calculations are based only on different versions of FLEXPART. Why did you add all these different versions of FLEXPART? What do you learn from this? The discussion of the results only focusses on the differences between trajectory models (as a group) and the FLEXPART ensemble. Differences between the trajectory models are not discussed at all, so why do you even need them all? To make the points the paper wants to make, it would have been completely sufficient to compare FLEXPART with and without parameterizations. I can't see any discussions/conclusions that go beyond that, and what the different models (and model versions) used do add.

We thank the reviewer for raising this point regarding our experimental design. We acknowledge that using FLEXPART with and without parameterisations would effectively show the impact of the parameterisations. Indeed, we performed this analysis in Section 3.3. However, focusing only on FLEXPART with and without parameterisations would constitute a pure "model sensitivity study," which is not the goal of this paper. Our goal is to provide an initial assessment of how model-specific characteristics (pure-trajectory and dispersion models) influence the quantification of adiabatic, advective, and diabatic contributions to temperature anomalies, and to lay the groundwork for interpreting Lagrangian HW analyses across different modelling frameworks.

Our model selection was designed to mirror current methodological choices in the literature. Specifically, we included both pure-trajectory and dispersion models because previous studies rely on both frameworks for Lagrangian temperature anomaly decomposition.

To address your specific questions regarding our model choices:

1. We intentionally selected the Pacific Northwest 2021 (PNW21) heatwave because it served as the seminal example case in Röthlisberger and Papritz (2023) when presenting the decomposition method to the community. Consequently, it was essential to include LAGRANTO and the trajectories provided by Röthlisberger and Papritz (2023) as our baseline. We also included another widely used pure-trajectory model (such as HYSPLIT for trajectories) and the recently launched HPT to verify that the results were not merely a consequence of LAGRANTO's numerical scheme and were a consistent characteristic of the entire pure-trajectory model class. On this basis, we also include a version of FLEXPART trajectories derived by switching OFF all parameterisations.

As you noted, we do not discuss the internal differences between these pure-trajectory models. This is because, despite some discrepancies, they fundamentally agree on the dominant physical process (diabatic heating) driving the heatwave. Exploring their inter-model differences falls outside the scope of our goal, which is the inter-class (trajectory versus dispersion) comparison.

2. While FLEXPART is the only dispersion model included, it is the dispersion model used in the Lagrangian methodology in the current literature. Furthermore, we used three distinct, long-term Lagrangian datasets that are currently publicly available and actively used in the literature: the LARA reanalysis (Bakels et al., 2025), and the datasets from the EPhysLab, University of Vigo (FLEXPART-10.4; Vázquez et al., 2024) and University of Ghent (FLEXPART-11; Deman et al., 2026). By including these three datasets, we show that the discrepancy with pure-trajectory models is not related to the model setups used in EPhysLab simulations (our primary FLEXPART data). Instead, the discrepancies are consistent across all FLEXPART (dispersion) datasets.

In summary, the models and datasets chosen are not random. They represent the exact tools driving the current divergence in the literature.

I have a few other specific points:

Lines 135-158: The domain-filling FLEXPART runs all use too few particles to obtain statistically robust results. This is mentioned for the LARA data set on line 200 but I think this is true for all data sets. The global domain-filling set-up (e.g. LARA) has been developed for climatologies but it is definitely not ideal (or even suitable) for case studies, unless you use hundreds of millions of particles. A better option here would be to run the model backward in time from the target domain, such that all particles contribute to the results.

We completely agree that the global domain-filling setup is better for climatological studies, and running the model backwards in time from the target domain is the best option for analysing specific case studies, as it ensures all simulated particles contribute to the event. However, our rationale for utilising the LARA reanalysis and the University of Vigo (UVigo) and University of Ghent (UGhent) datasets relies on their public availability and community use. A core objective of this paper is to evaluate the actual datasets and frameworks researchers are actively applying to Lagrangian temperature anomaly decompositions. Evaluating these specific, readily available long-term datasets provides practical value to the community using them.

Furthermore, despite the theoretical limitations of using domain-filling simulations for single events, recent literature and our results demonstrate their viability. For example, a recent intercomparison of moisture-tracking approaches shows that moisture sources for specific precipitation events identified using global domain-filling FLEXPART simulations mostly agree with the sources identified using LAGRANTO and Eulerian approaches ([Benedict et al., 2026](#)). This consistency is strongly mirrored in our current findings. Except for LARA (due to its small number of parcels), the resulting temperature anomaly footprint derived from the UVigo and UGhent datasets largely agrees with the footprints from all other models. Crucially, they show strong spatial agreement with LAGRANTO and HYSPLIT in the temperature anomaly footprint, both of which were run backwards in time from the target domain.

Equation (1): The most interesting term is the last one (diabatic term). However, without splitting it further up into the different processes, it is not extremely useful, since you then don't know which process is actually responsible (e.g., latent versus sensible heat flux, radiation).

We agree that the diabatic term is highly interesting and that partitioning it into its constituent physical processes (e.g., surface sensible heat flux, latent heat, and radiation terms) would make the

decomposition significantly more useful for understanding the exact drivers of the temperature anomaly. Indeed, we fully recognise this limitation of the Lagrangian temperature anomaly decomposition. We explicitly highlight this constraint in the last paragraph of the 'Final Remarks and Conclusions' section, noting that the inability to isolate specific diabatic processes remains a key limitation of the Lagrangian temperature anomaly decomposition.

While we agree that extending the Lagrangian temperature anomaly decomposition approach to separate and quantify individual diabatic terms would be a valuable methodological advancement for the field, its development and validation is outside of the scope of the current manuscript. Separating the specific diabatic terms should be addressed in future work.

Figure 1: The last column (Budget) is pretty useless, as it cannot be distinguished from the first column. Showing the differences would be much more meaningful.

We thank the reviewer for this observation. We agree that the last column (Budget) appears visually indistinguishable from the first column (Tanom). However, this visual similarity is exactly what we intended to convey, and we deliberately chose to show the budget term rather than the differences.

Our rationale for this choice is to provide rapid, visual evidence of the agreement between the temperature anomaly and the sum of all decomposition terms (adiabatic, advective, diabatic and seasonality). By showing that the budget is perfectly closed in every model, we visually assure the reader that the fundamental differences we report between pure-trajectory and dispersion models are not budget closure errors, numerical loss, or computational inconsistencies within the decomposition method itself.

While plotting the differences highlights the residuals, displaying the full budget acts as a quality-control check. It shows that the decomposition method works correctly across all frameworks, allowing the reader to focus on the physical differences that form the core of this paper.

Figure 4: I don't think the cluster analysis provides any further insight. To me, it is not clear what it is supposed to show.

We thank the reviewer for this comment; however, we believe the cluster analysis provides valuable information necessary to fully understand the differences between the model classes. Specifically, this analysis shows that while the geographical origin of the air masses is similar across all models, differences in temperature anomaly decomposition are directly linked to different vertical transport histories. For instance, the cluster analysis provides evidence that the dominant air mass in dispersion models undergoes a steep descent. Meanwhile, in pure-trajectory models, the same air mass remains at lower levels.

Lines 385-390: I do not understand the comments about closed boundaries with respect to turbulent exchanges. Of course, diabatic surface heating is only happening at the surface. That's exactly how it should be!

We thank the reviewer for pointing this out. We agree that the phrase 'closed boundaries' caused unnecessary confusion. Our intention was to describe how the treatment of vertical motion in each model class influences diabatic heating. Should the editor consider that resubmission of the manuscript is appropriate, we will remove this terminology from the revised manuscript to avoid any ambiguity.