Review of "Islet: Interpolation semi-Lagrangian element-based transport" by Andrew Bradley, Peter Bosler, and Oksana Guba.

The paper presents a new set of basis functions for semi-Lagrangian advection method in spectral element models. The presented method is novel and its performance in atmospheric test cases is encouraging. The paper therefore warrants to be published. The main drawback of the paper is that it is quite tedious to read. This paper essentially presents two things: 1) a novel optimized basis functions for 1D semi-Lagrangian advection schemes, and 2) its implementation in an atmospheric model using three different (sub-element) grids. Both of these are quite complex topics, and their discussion is intertwined (e.g. in section 2) and the reader is easily lost in details. The overall presentation should be improved before the paper can be accepted for publication. Considering the amount of work, I recommend a major review.

Major comments:

For clarity, I suggest that you clearly define the two considered problems, 1D SL advection method, and the implementation in 3-grid atmospheric models, already in the introduction. The introduction is now quite short and actually does not mention many important concepts relevant for the paper.

Section 2 is rather difficult to follow, consider revising. I suggest to start by defining the 1D discretization with N elements, the interpolant functions within each element, and the properties of the interpolant functions (e.g. basis functions, continuity and symmetry). Presently, the interpolant functions first appear only in section 3.1. A figure could clarify the concepts, including the source and target elements/nodes. I would also define a symbol for the basis functions themselves, instead of using L from section 3.1 (L is the interpolant function itself). The discussion of stability becomes comprehensible only after the discretization has been introduced.

As I understand, the 3 axioms of advection methods are: 1) global conservation, 2) preservation of constant tracers (sometimes called local conservation, or tracer consistency), and 3) monotonicity (i.e. no spurious overshoots appear). These properties apply to both the SL tracer advection scheme and the remap operators between grids. In section 1.2 these concepts seem to be mixed and referred to by different names (plausibly
due to historical reasons; the so called "property preservation" is just a combination of 1, 2, and 3). Consider revising.

Figures 20 and 21 shows larger diffusion for shorter time steps. Could you elaborate on this? Is the traditional SL property that longer time steps reduce diffusion (while the solution can degrade in other metrics)? How would one choose the right time step in practice? In the case of Eulerian transport it is easy: take the maximum stable one, and you are guaranteed to satisfy all the necessary properties.

Minor comments:

Throughout the manuscript the authors use the terms "mixing ratio" and "tracer" for the advected quantity \( q_i \), seemingly interchangeably. For the sake of clarity I would prefer just to use "tracer".

I35: what is the definition of "local conservation" here?

I155: "This structure arises as follows. Consider a continuous discretization using a nodal \( n_p \)-basis, \( n_p = d+1 \), with \( n_p \) the number of nodes. The grid has \( N \) elements. Each row of the space-time matrix corresponds to a target node."

This description is too brief to be understandable, please elaborate. This is my interpretation of the discretization: The 1D domain is divided into \( N \) elements. The solution in each element is approximated by a continuous function, defined by \( n_p \) basis functions. Thus a function \( f \) in an element \( e \) can be written as \( f(x) = \sum_{i=1}^{n_p} f_i \psi_i(x) \) where \( \psi_i \) and \( f_i \) denote the \( i \)-th basis function and its corresponding coefficient. Each basis function is associated with a node \( x_i \) within the element; The basis is Lagrangian (a.k.a. nodal), i.e. \( f_i(x_i) = 1 \). Furthermore, the discretization of the function is continuous across element interfaces, implying that the neighboring elements share a (exactly one?) basis function. Furthermore, the basis is assumed to be symmetric about the center point of the element.

I156: "Each row of the space-time matrix corresponds to a target node." You should define the space-time matrix for this statement to be comprehensible.

I177: Have you defined the basis to be symmetric somewhere?

I195: "L provides a basis for degree-d polynomials." I would say that the basis functions are the \( \Pi_i \) functions defined in the equation of \( L \); \( L \) itself is the interpolant function defined by the basis and the specific nodal values \( y(i) \).

I195: "These are supported by \( n = d + 1 \) points, each an element in the \( n \)-vector \( x_n \)." To be consistent with the literature I would use the term "node" instead of "point".

I204: "Given a departure point \( x \)" These properties define the interpolant functions, thus there's no need to say that \( x \) is a departure point, it can be any point within the element.

I208: I think this constraint is equivalent to saying that the basis must Lagrangian or nodal?

I246: only here you define a d-degree polynomial. This would be useful already in section 2.

I257: what is Runge's phenomenon? what is Lebesque constant? help the reader to
understand the rationale behind your work.

Section 3.6: the description of the search algorithm is quite technical and could perhaps be moved to the appendix; it is not necessary to follow the main storyline of the paper.

Section 4: mention TTPR already in the introduction as it seems to be relevant for the entire Islet method.

l439: earlier tracer tendencies were denoted by $f_i \Delta t$

l442: "immediate element neighbors" Are these neighbors that share an edge or vertex?

l464: "In contrast, ..." Meaning unclear, please revise.

l507: What is "tracer density"? Is it just $\rho$? Then, for clarity, I'd call it "density"

section 4.3: while computational efficiency is important, I would move this section to the appendix, as it

l570: "is proportional to the number of grid points" Should read: "proportional to the square of number of grid points"

l572: what is a naive h-halo exchange and how does it differ from what is proposed here?

section 5, l606: please mention the problem domain (full sphere?) and the equations that are being solved (pure advection on the tracer grid?)

l631: why can you not use the dynamical core to compute the density? the chosen approach seems rather ad-hoc; can you guarantee that density values are realistic?

scaling figures 6-8, 11-16: for easier readability do not use red line twice for np=6 and np=12. x axis label is missing.

l703: "has even more accuracy" Can you quantify this? Is the difference significant?

l711: "For each value ..." unclear sentence, please revise.

l741: 0 in $m(0)$ stands for time $t=0$?

l780: what is a terminator?

l897: Here you define the basis functions $\phi_i$. This notation would be useful throughout the manuscript, already in Sect 2 and 3.1.

abstract and intro: you define an abbreviation "dycore". I would omit it as it does not really save space, and it is not used frequently in the paper.

Typos:

l745: "shows the results"

l815: SYPD numbers printed above ...