This is a well-written manuscript. I would recommend it to be published with some minor revisions.

Major issues:

1) I agree with RC4 by Wolfang Bangerth that the main message of the manuscript should be clarified. If the target audience is geodynamicists who are testing different approximations of governing equations or the performance of different solver algorithms, this manuscript is an excellent introduction and demonstration and can be published with some minor revisions in the introduction and conclusion. If the target audience is geodynamicsists who are working on real-world geophysical problems, this manuscript falls short in several aspects. It is probably beyond the scope of this manuscript to address these aspects, but I would list them below nonetheless:

a) Since solving the governing equation is only a small part of a mantle convection code, some part of the generated C kernel must be extended. I would like to know how easy is it to modify and extend the generated C kernel? Can it be done in python or must be done in C? For example, to add dynamic time stepping, one has to calculate min(v/h) over all elements, where h is the element size. Can the contents of v and h be accessed in python, or only in C?

b) Most legacy mantle convection codes are using domain decomposition to parallelize. But Firedrake seems to parallelize by distributing the matrices and vectors. How will this affect the extension of the code, for example, adding markers?
Some minor suggestions:

2) In the code listings, it will be better if some function arguments are passed as keywords, instead of just as numbers. For example, line 8 in listing 1 becomes: \( W = \text{FunctionSpace}(\text{mesh}, \text{family}="CG", \text{degree}=1) \)

3) Line 40-41 listing 1, the boundary tags 1, 2, 3, 4 are better to named as left_id, right_id, bottom_id, top_id, respectively. Related, Listing 2, bottom_id and top_id are used but not defined.

4) The solver parameters for energy solver in Listing 1 and as described in line 405 appear to be inconsistent with those described in line 300. Also, the solver parameters for Stokes solver in line 336 are inconsistent with those in later code listings.

5) In the code, the buoyancy term is using \( T_{\theta} \), rather than \( T_{old} \) or \( T_{new} \). Why?