Answer to Referee 2

We wish to thank to this referee for his/her very useful comments that have helped us to improve the manuscript and have been addressed as follows:

General comments:

1. *This paper addresses the issue of Lagrangian transport in the Stratospheric Polar Vortex (SPV). The first part of the paper analyzes SPV data from the ECMWF using the technique of Lagrangian Descriptors (LDs, developed over the years by some of the authors of this paper and their collaborators) for a specific time period in September 2002. A three-mode kinematic model which possesses the gross characteristics of the data is then developed, and there is some discussion on how it is possible by adjusting its parameters to mimic certain behaviors of the observational data. The paper is well-written and readable. However, I believe that some more work is needed to show that LDs are relevant to this situation, and that the kinematic model provides useful information. I have expanded on this in my specific comments below. My feeling is therefore that a major revision would be required before being acceptable for publication.*

We have clarified in a new version of the Introduction, the major goals of the article as maybe they were not sufficiently elaborated in original manuscript. The major goal is to gain new insights into the fundamental mechanisms responsible for complex fluid parcel evolution by providing a simple model (a kinematic model). The model allows in a controlled manner to recognize the physical mechanism responsible for the key observed transport features of SPV. In order to highlight the Lagrangian skeleton responsible for transport features both in the stratosphere and in the model, we use a Lagrangian tool, the function M, which has been extensively used in the literature. We consider that the references we provide in Section 2.2 provide a sufficient basis to use this tool, and we do not focus on justifying again in this new paper the efficiency of M in highlighting Lagrangian features, we just use it.

Specific comments:

1. *It seems that the major focus is on modeling the SPV breakdown in September 2002. If trying to use Figure 4 as evidence that LDs provides an excellent way to explain this, then I feel that there must be some comparison to other studies which show this. Beyond a few brief references (page 2, line 27-28), the authors do not seem to do much in this direction. After all, how good are the results of Figure 4? What are the other symptoms of the SPV breakdownwhat other observations showed that this indeed did break down? (Using Figure 3 is a startbut this is using an Eulerian observation to predict something Lagrangianor is it?) And is Figure 4 consistent with any other observations? Several references which might help are: Nishii & Nakamura (Geophys. Res. Lett., 2004), Kruger et al (J. Atmos. Sci., 2005), Taguchi (J. Atmos. Sci., 2014), Fisher et al (Atmos. Chem. Phys., 2008), Esler & Scott (J. Atmos. Sci., 2005), Konopka et al (J. Atmos. Sci., 2005), Varotsos (Environ. Sci. Pollution Res., 2002, 2003, 2004) and Allen et al (Geophys. Res. Lett., 2003). In addition to these, I feel that it is imperative that there be comparisons (or relevant discussions) with the paper by Santitissadeekorn et al (Phys. Rev. E, 2010) which provides a Lagrangian analysis and provides pictures very similar to Figure 4. ?*

The SPV breakdown in September 2002 has been extensively studied in the literature using ERA-Interim data and these references are now quoted in the manuscript. A novelty of our study is trying to understand the breakdown and its previous stages in a simple model that shows that the breakdown is related to wave propagation phenomena. The Lagrangian analysis of the breakdown exhibits what are the transport implications of the breaking, showing that the splitting leads to no mass transfer between the two vortices.

The paper by Santitissadeekorn et al (Phys. Rev. E, 2010) presents an interesting approach to estimating the three-dimensional location of the vortex. the promise of this approach is demonstrated by examination of the period from August 1 to September 31 in 1999. (The similarity of pictures during different final warming events can be expected from the similarity in evolution reported by Mechoso et al. (1988). Our paper focuses on a different year (2002) and our concerns are not on the precise location of the polar vortex edge. Therefore, we will keep the paper the paper by Santitissadeekorn et al (Phys. Rev. E, 2010) in mind for future studies, but shall not include a reference in the text.

2. *The term Hyperbolic Trajectories (HTs) is used often in this paper, and described briefly in the introduction. The ideas and intuition given in the third paragraph of the introduction are however only valid in infinite-time flows. There are sometimes additional limitations of steadines: the cats-eyes structures in these models de- pends on drawing streamfunction contours (either in the steady frame or in a moving frame), and so are associated with steady situations. While the remain- der of the discussion does not necessarily confine itself to steadiness, as far as I am aware, hyperbolic trajectories can only unambiguously be defined for infinite- time situations, using the ideas of exponential dichotomies. The paper by Ide et al (Nonlin. Proc. Geophys., 2002), for example, cites the exponential dichotomy definitionbut this cannot be adequate for finite-time flows since the variational equation associated with any trajectory will obey the exponential decay requirement by choosing a suitably large prefactor. There have been attempts to fix this: by choosing a prefactor of 1 (Doan et al (J. Differential Equations, 2012), Karrasch (J. Differential Equations, 2013), Duc & Seigmind (Int. J. Bifurc. Chaos, 2008)), or by extending to infinite-times in some fashion (Balasuriya, (J. Nonlin. Sci., 2016)). In general, it seems that HTs are ill-defined for finite-time flows. Throughout the paper, however, the authors seem to be using saddle-like locations of the LD field as their method of identifying HTs. I understand why such locations can be called hyperbolic, but there does not seem to be any justification in calling them trajectories since it is not at all clear if by following these in a time-varying way by computing LDs over a range of t0 values, an actual trajectory of the system (5) arises. If the flow is nearly steady, it seems that it might be possible to establish the existence of time-varying saddle-points which are close to an actual (infinite-time) hyperbolic trajectory in some instances (Ide et al (Nonlin. Proc. Geophys., 2002), Balasuriya, (J. Nonlin. Sci., 2016)). But is this necessarily so for this situation, viz. using finite-time data, with moderate unsteadiness, and specifically using LD fields to identify saddle points? If the actual term hyperbolic trajectories is not important to what the authors are doing, then perhaps they should simply call them saddle points of the LD field? But even so, claiming a direct relationship to stable and unstable manifoldswhich are undefined for finite-time flowsseems problematic.*

We have extended the explanations on HTs in the Introduction and in Section 2.2. We provide references that compute and justify the use of HT in finite time data sets and also briefly summarize their content. In section 2.2 we have provided also references and arguments that allow us to refer to the the "saddle-like locations of the LD field" as HTs. We have provided references in Section 2.2 that provide a constructive definition for finite time stable and unstable manifolds. We have also briefly summarized the content of these references in the text.

3. *I have some concern about the centered nature of the definition for M in (6). If requiring to find information on the skeleton of transport at time t0 using FTLEs/FSLEs/.../variational LCSs, the basic approach is to seed initial values at t0. If looking for the analog of the stable manifold at t0 (i.e., repelling LCSs, ridges of forward-time FTLEs), these needs to be advected in forward time. Similarly, the advection is in backward time if looking for analogs of the unstable manifold. It is this information which tells us about the skeleton at time t0. For example, Gaultier et al (J. Marine Sci., 2013; J. Geophys. Res. Oceans, 2014) do this advection in backwards time in order to compare with sea-surface temperature fields at the time t0. This is also because the advected scalar field (temperature in their case, whereas in this case it could be temperature, ozone concentration, etc, depending on the specific observable of interest in the SPV) at time t0 would depend on the advection occurring until the time t0. Future times surely cannot have an impact. Therefore, why is the integral in (6) being taken from times t0-tau to t0 + tau? This seems inconsistent with all other Lagrangian approaches. Moreover, its hard to argue that the SPV knows the future! The pinch-off on September 24 in Figure 4(b), for example, uses velocity data into October.*

In Section 2.2 we have included an explanation about the forward and backward integration time used for M, its relation with FTLE and the convenience of this choice for our study. Our approach is completely consistent with all other Lagrangian approaches, found in the literature.

4. *The authors state that M reveal[s]/highlights Lagrangian coherent structures (page 5, lines 12 and 15). Is there a rigorous justification for this - that M specifically reveals coherent structures which move in a Lagrangian way according to the flow? If so, in what way? I am not able to find it directly in the cited references, though I am unable to get access to the latest article (Loposito et al, 2017) that is still in press. To my knowledge and judgment, a relationship has only been established in heuristic senses (and this is also so for other Lagrangian methods used and advocated by others), and in incredibly simplified test cases. Moreover, the authors talk of stable and unstable manifolds here, but of course these things do not have a proper definition in finite-time flows. I believe that the description here needs to be watered down. The LD field is being used as a heuristic, and there is some evidence that it provides the right understanding.*

There are rigorous justifications that invariant manifolds are aligned with singular features of LDs only for specific examples discussed in Lopesino et al 2015 for discrete dynamical systems and Lopesino et al. 2017 for continuous time dynamical systems. Also, the ability of LDs to highlight invariant sets has been explained, and the tool has been linked to the ergodic decomposition theory.

For geophysical flows Mendoza and Mancho (2010, 2012) have compared and found that numerically computed invariant manifolds systematically are aligned with singular features of M, but in these cases there is not any theorem supporting these facts, just numerical evidence. de la Cámara et al. (2013) show that for similar ERA-Interim fields, singular features of M are aligned with numerically computed stable and unstable manifolds (see their Fig. 2).

These issues are explained now in Section 2.2

5. *The kinematic model requires more justification. Why do the amplitudes of the Fourier modes in the kinematic model have these particular r-dependencies? ?The r(r-a) in vr is understandable, but why exp(- r)? And why the specific forms chosen for PHI1 and PHI2? And why these particular forms of time-dependencies for eps1 and eps2? Certain parameter values are used in the simulationswhy were these chosen? In what way are they consistent with parameter values of the SPV? Since the flow for the kinematic model is unsteady, the pictures of Figure 6 must be drawn at a particular time value t0, I guess. What is it? I also have a much more general question regarding the kinematical model: What particular understanding does it give to this situation? It is probably possible to have the LD field display all sorts of crazy behavior by choosing the s in various ways, and so what does this particular model do? Now, if it was possible to argue, for example, that a particular instability arising from this kinematic model led to the SPV breakdown, then that might be interesting.*

Section 4 has been extensively revised to address the issues raised by the referee. In particular, the choice of free parameters in the kinematic model is explained in more detail. Further, the SPV breaking is reproduced by the kinematic model (see figure 8). The times at which specific patterns are achieved are also reported.

6. *I am confused by what the authors are trying to achieve in Section 5. Are they trying to say (page 15, line 11) that their kinematic model can be made dynamically- consistent but inserting their PHI into (14) and (15) but then treating h as unknown, and thereby getting an expression for h? This can possibly be done (though h will satisfy a PDE which may not be easy to solve), but this is highly artificial. This would be demanding that the topography adjusts to the kinematic model that we insist is a solution. One possibility in which this part of the paper might have value is if the s in the kinematic model were somehow chosen as modes associated with the conservation equation (14)this would be similar to the work of Pierrehumbert (Geophys. Astrophys. Fluid Dyn.,1991). The discussion of the earlier parts of this section also appears to lack relevance. If Q were constant in patches, then complicated dynamics are possible subject to Qs conservation but this simply amounts to nullifying the dynamical constraint, and adds the extra condition (not talked about here) that the streamfunction needs to be chosen such that (15), for a constant Q, is satisfied. Basically, it is true that the potential vorticity distribution imposes constraints on the Lagrangian motion, which may be an aspect the authors are trying to highlight here. For these, the papers by Brown & Samelson (Phys. Fluids, 1994) and Balasuriya (Nonlin. Proc. Geophys., 2001), which deal with both constant and nonconstant Q, may be relevant. In general, Im not sure I understand the goals this section, and so it requires some attention.*

Section 5 has been rewritten and an explicit calculation of the forcing h is reported, that achieves the conservation of potential vorticity Q for one of the proposed Ψ. The calculation is illustrated for a simple Q choice but it could be repeated for more realistic Q distributions as far as they are defined as piecewise constant functions.