**Atmospheric
Chemistry
and Physics
Discussions**

Interactive comment on "A new real-time Lagrangian diagnostic system for stratosphere-troposphere exchange: evaluation during a balloon sonde campaign in eastern Canada" *by* M. S. Bourqui et al.

M. S. Bourqui et al.

michel.bourqui@mcgill.ca

Author's reply to Reviewers comments

We thank both Reviewers for their excellent comments that have helped improve substantially the clarity of the manuscript. We provide hereafter our answers to each of the comments raised by the Reviewers. The Reviewers comments are reproduced in italic font. In addition, we have made minor changes to help further clarify the text. These are listed at the end of this report. Line and page numbers refer to the on-line

C15145

discussion paper.

Reply to Reviewer 1 (M. Sprenger)

We thank M. Sprenger for his encouraging and relevant comments that have contributed to improve the clarity of the revised manuscript. In his general comments, he suggests a change in the structure of sections 4.3, 4.4 and 5. We have considered this possibility, have weighed the pros and cons, but have decided not to change the structure in the revised manuscript. A detailed answer to this general comment is provided below. We have answered each minor comment after that. We have integrated most of them into the revised manuscript. In our answers, we use the page and line numbers that appear in the on-line version of the discussion paper.

GENERAL COMMENTS:

One major point concerning the overall structure of the manuscript might be the following: Section 4.4 is a rather detailed discussion of the internal structure of the stratospheric intrusions and how well it is represented in the Lagrangian STE diagnostic. It was a little 'too detailed' for fluent reading! I wonder whether this part can be somewhat shortened, and possibly be combined with section 4.3. On the other hand, section 5 is a rather interesting and, I think, enthusiastic discussion of STE forecast errors. I like this section, and also like very much the cluster analysis of missed events! It would deserve a more prominent place in the manuscript - I would 'vote' for a place before the much more 'technical' section 4.3. But of course, there are also good reasons for the present structure...
In summary, considering the already well-written state of the manuscript I suggest that it can be published with minor revisions, which are enlisted below. Most of them are requests to clarify certain statements.

The main goal of this manuscript is to provide an objective evaluation of the new Lagrangian STE data set. For this reason, we think that the comparison against observations using objective criteria (categorical scores) must appear before, and be detached

C15146

from, the more subjective discussion provided in section 5. Furthermore, section 5 makes the link between the (more technical) objective evaluation and the overall picture, and thereby smoothly brings the reader to the Conclusions section. We believe this sequence provides a better overall clarity than otherwise.

Sections 4.3 and 4.4 offer two different levels of evaluation of the Lagrangian STE data, one on the overall depth of intrusions and one on the detailed vertical structure. We think that merging the two sections may make these two levels of evaluation less clear to the reader. We have carefully reread the text and could not identify paragraphs that could be removed from these sections without changing substantially the content. But we would be happy to reconsider this, following a more specific suggestion from M. Sprenger.

MINOR COMMENTS

L37: The listing of physical processes related to STE ends with a numeric process, numerical diffusion. This is somewhat unhappy, because numerical aspects should be separated from true physics.
We agree of course that numerical diffusion is not a real process. To avoid confusion with those real processes listed in the manuscript, we have removed this term.

L40-43: The term 'chemical gradients' is readily understood, but not very fortunate. More correctly, it should be 'the gradient of the chemical constituents'. Furthermore, I think the same statement (L40-43) would also be true if the gradients are not so large.
Well spotted. We have changed "chemical gradients" as requested and have rephrased the sentence as: "However, because the gradients in chemical constituents across the tropopause are large, it is the separate stratosphere-to-troposphere (S→T) and troposphere-to-stratosphere (T→S) mass fluxes that control the transport of chemical species across the tropopause, not the net mass flux." In the case of a zero (or small) gradient, it is the transport of the mass of a tracer that is controlled by the net mass

C15147

flux, not its distribution.

L45: Is a resolution of 0.5 deg sufficient to resolve all of the above mentionend physical processes (L36-37). Probably not, but the text suggests that this is the case.
Thanks for spotting this. It is not the processes, but their contributions to STE that are expected to be represented adequately. We have changed the sentence to: "Bourqui (2006) suggests that consideration of hourly meteorological fields with horizontal resolution $0.5^o \times 0.5^o$ is necessary to resolve the most important contributions from these processes on STE. "

L52: It might be helpful for the readers not familiar with 'residence time' to explain it already at this place with one sentence.
Thanks for the suggestion. We have added the following sentence: "In such a case, the parcel is required to reside on either side of the tropopause for a time interval larger than a given threshold."

L67-69: Here it is suggested that Wernli and Bourqui (2002) and Sprenger and Wernli (2003) significantly underestimate the frequency of deep STE events. But in these studies a different residence time (96 h) was applied compared to the present study, which enforces a 12 h threshold. I guess that the discrepancy can be explained by this different choice of residence times. Please comment I
We agree that the choice of residence time may play a central role in this discrepancy. However, other factors may also play roles and more research is necessary to reconcile these different estimates. Such a discussion appears to be beyond the scope of the manuscript. Throughout the manuscript, we refer to Sprenger and Wernli (2003) as it provides a good and comparable reference. We understand that the following sentence may be misleading and we have removed it: "This suggests that Wernli and Bourqui (2002) and Sprenger and Wernli (2003) may significantly underestimate the

C15148

frequency of deep STE events, possibly due to the long residence time used (96 h)."
We have also added "in the 15-year climatology" in the following sentence in order to make sure that the difference on the periods considered in the different studies cited are clear: "The mass flux associated with this deep S→T transport activity was found to be about one order of magnitude larger than the 15-year climatological estimate from Sprenger and Wernli (2003)."

L104: 'within five succesive 24 h time windows with a 12 h residence time' -> at first reading the meaning of the sentence is somewhat difficult to grasp. Please reformulate I

This passage has been reworded as follows: "Here, we introduce the first real-time Lagrangian STE data set based on global weather forecasts. These data have been calculated daily since July 2010 at Environment Canada (EC) following the methodology introduced in Bourqui (2006). They consist of global, five-day STE forecasts calculated daily using the 10-day global weather forecast initiated at 00 h UTC. The five-day STE forecasts are based on six-day trajectories started at $00 + I$ h UTC, I = 0, 24, 48, 72, 96, respectively, and selected as follows: they must cross either the ± 2 PVU dynamic tropopause or the 380 K isentrope with a residence time of 12 h within the time window $[12 + I; 36 + I[$ h UTC."

L114: What is the 'spatial density of initial trajectories'? More precisely, it would be the spatial density of the initial trajectory points. To put it otherwise: by definition, a trajectory is a whole path in space and time, and cannot be used to refer to a single point/time step along the trajectory.
Agreed. It has been changed as suggested.

L145: Low-level PV anomalies with PV > 2 PVU might mimic a stratosphere, but are actually of tropospheric, diabatic origin. How are such tropospheric PV anomalies

handled? Furthermore, the definition of the of STE needs a clarification. What if both, the 2-PVU and the 380-K isosurface, are crossed? Obviously, what is meant is that the lower crossing counts. Right?
The identification (and subsequent removal) of events related to low-level PV generation is a difficult task and there is no known optimal way to do it. They are most likely associated with relatively high mountains, such as the Rockies, Himalayas, Alps, etc. In order to avoid introducing *a priori* a bias of some sort in the data set, we have decided to keep all trajectories following the criteria spelled in the manuscript. It is then up to the user of the data set to remove such false events if they are significant. In the context of this manuscript, such false events are not expected since there are no high mountains in the area around the balloon sites. We are currently working on this issue in the context of a global climatology.
The crossing of the +2PVU, -2PVU and the 380K are considered independently in the selection procedure, and the decision to include trajectories crossing the 2PVU surface, the 380K surface, or a combination of them belongs to the user of the data set.

L160, 172-177: Some reference is made here to the T->S calculations. But the whole study relies only on the S->T trajectories. Indeed, the restriction is motivated in L174-176, and is pertectly ok. I simply wonder whether any reference to non-used products of the methodology must already be described here. I would prefer to skip these parts.
The description of the data set provides the context to this study. We think that providing a full description helps the reader understand the motivation of study and its consequences/limitations. We have tried to keep the description of the data set to its minimum (e.g. no example, illustrating figure, etc). Reducing the description of the data set to the portion used in this study would lead us to remove mentions about the T→S, but also about the crossing of 380K, about the fact that it is a five-day forecast, that it has been delivering data every day since July 2010, etc... We think that this would lead to a serious lack of completeness on the context of the study and

its motivations.

L236-243: Here, the objective STE identification is described. I wonder a little why ozone gradients are used in the criterion, but absolute ozone concentrations are not. Is there a specific reason for omitting [03]?
Significant vertical gradients in ozone mixing ratios, when conjugated with vertical gradients in humidity, suggest that air masses of different origins are found on either part of the gradient. This is why the vertical gradients occupy the central role in this algorithm. The criterion on the absolute value of RH inside and outside of the stratospheric intrusion provides a way to select air masses that have a not too old origin. As seen in Fig 1, RH allows the removal of very moist regions that are clearly not of stratospheric origin. Adding a criterion on the ozone mixing ratio did not change significantly the results. Of course, some degree of subjectivity is unavoidable in the choice of these criteria. This issue is discussed in sections 5 and 6.

L251: At this place place it becomes not clear why the comparison with the Lagrangian STE data is simplified by the choice of 50-h Pa bins
This sentence has been corrected as: "This vertical grid is also used with the Lagrangian STE data set (see Section 4.1)."

Figure 1: The thermal tropopause is shown, most likely because it can be derived on the measured temperature alone. On the other hand, the Lagrangian STE diagnosis uses the 2 PVU/380 K tropopause. Would it make sense to include this dynamical tropopause also in Figure 1?!
We tried this but decided not to include it for the following three reasons: (1) The thermal and 2PVU tropopauses are fairly close to each other, and the figure would be less clear with both overlapping surfaces. (2) This figure only shows results for Montreal and most discrepancies between the two tropopauses are at the two other

sites. (3) This figure focusses on the observational results and we would prefer as much as possible to keep it independent from the modelling results.
The difference between the thermal and 2PVU tropopauses is shown on Figure 4: Black line for the thermal tropopause, and top of blue squares for the 2PVU tropopause. A comment on this issue is available on p.27984, l.24-27.

Figure 2: The upper-most intrusion might more reasonably be called 'the stratosphere' I How is this handled in the validation of the method? Is the highest high-PV reservoir taken as an intrusion, or is it excluded from the validation because it is the lower part of the stratosphere and hence deserves special treatment?
The detection of the thermal tropopause and the detection of stratospheric intrusions are independent. Intrusions detected just below the tropopause are kept as intrusions. These likely represent fine filamentary structures found below the thermal tropopause. Our detection and categorisation algorithms move and smear these layers onto a 50hPa vertical grid. The Lagrangian STE data are then considered to capture this event if at least one trajectory is found within the same 50hPa layer. We have introduced the following comment on line 13 of p.27980: "It shows two detected stratospheric intrusions and the accompanying categorisation of each 50 hPa layer. The bottom level of the lower intrusion marks a transition between a tropospheric region below and a dry layer with slightly enhanced ozone mixing ratios above. Ozone and RH show a visible anticorrelation in the vertical from 800 hPa up to 500 hPa, a typical fingerprint of air of stratospheric origin. The upper part of this intrusion is more ambiguous and part of this ambiguity is absorbed in the Intermediate (Top) category. The second intrusion's bottom is detected around 180 hPa, but its top is not detected by the algorithm because it is thin and close to the thermal tropopause. This example illustrates the difficulty faced when reducing an observed profile into binary information on the presence / absence of stratospheric intrusion. Here, the definition of an intermediate region between the inside and the outside of the intrusion provides a useful palliative for the lower, deep intrusion but not for the upper, shallow one."

L182-283: The number is about two orders of magnitude larger than estimates in Sprenger and Wernli (2003). But note that the latter study is climatological in nature, and hence considerably lower values must be expected. It would be interesting to compare the value in the manuscript with Sprenger and Wernli (2003) on an event basis. Furthermore, note again the different residence times which are applied. In short, a comparison with Sprenger and Wernli (2003) is difficult, and it should be stated so.

As mentioned above, we use Sprenger and Wernli (2003) as a reference throughout the manuscript. The reasons for the discrepancy certainly include the different residence times but may also include other causes. It is beyond the scope of this paper to elaborate on this issue and we have tried to clarify any ambiguous statements. We have added "15-year climatological" in the sentence "Yet, it can be estimated from the total black shaded area below 700 hPa in Fig. **??** that around ten percent of the air below this level originates in the stratosphere, a number which is about two orders of magnitude larger than the 15-year climatological estimates from Sprenger and Wernli (2003)."

Figure 3: This is an important figure of the manuscript; and it presents two quite different pieces of information - if I correctly grasp it: (i) the frequency of bins in the different categories; and (ii) the distribution of RH, Q and 03 within the categories. Aspect (i) is clearly discernible, but could easily be shown in an extra row - note that the frequency is the same for all three rows. On the other hand, aspect (ii) is partly rather dificult to see; e.g. RH in the intermediate (top) bin. As a remedy: If (i) is shown in an extra row, (ii) could be shown in the next three lines, but now each bar would be equally high and represent 100 %.

We agree that the comparison of the histograms through the different categories would be simpler in the suggested graphical representation. However, we are hesitant to use

it because it would expand the bars that have only very few events and for which the histogram is not well defined, to the same size as the ones that include many events and for which histograms are well defined.

L323-324: Here the initial grid of trajectory starting points (55 km, 5 hPa, 24 h) is compared to the mapping grid of STE mass fluxes (2x2 deg, 50 hPa, 24 h). Are the numbers for the 'mapping grid' subjective, based on your experience, or are there some objective criteria which define the mapping grid in terms of the initial grid?

This is an example based on the analysis provided in Bourqui (2001) [M. Bourqui; Analysis and Quantification of STE: A Novel Approach; Swiss Federal Institute of Technology (ETH); GCA-Verlag; 2001], Appendix B, Figure B.6. We have added the reference and have removed the "x50hPa" in "(2x2 deg x 24 h)", as this was an error (mass fluxes across the tropopause surface lose their vertical dimension).

L371: 'In general, the PC score artificially increases in cases of low event frequency.'->A eference might be made to the pitfalls of skill scores based on contincency tables. See for instance, "The Finley affair: A signal event in the history of forecast verification. Murphy- Weather and Forecasting, 1996"

Reference added. Thanks for pointing out this interesting article!

Figure 4, L405-406: Within several time periods the thermal tropopause is rather high (e.g. 15 July)? I wonder where the dynamical tropopause based on 2 PVU/380 K is situated for these periods. My guess is that the dynamical tropopause behaves quite differently, i.e. that it is even found at lower-than-average heights during these periods. If so, I wonder whether it is even worthwhile to show the thermal tropopause? Please comment!

In fact, the thermal and dynamical tropopauses do agree fairly well. The dynamical tropopause coincides with the top of the blue columns in Fig. 4. (see remark l. 24-27,

p.27984). We have added the following sentence in Figure 4's caption to make it more obvious: "The dynamical tropopause coincides with the top of the blue shaded columns."

L417: '(3, left column)'-> confusingI Most likely, you mean (Table 3, left column)?
Yes. Corrected.

Figure 6: Above 300 hPa (dark blue), the number of occurrences drops from 24 at 48 h quite dramatically to 7 at 60 h. Is this sharp drop due to the small sample size, or is there a good physical reason for it?
We think this is due to the relatively small sample size, and a climatological study will be necessary to study these distributions.

Reply to Reviewer 2 (Anonymous)

We thank the Reviewer 2 for his/her very relevant comments that have contributed to improve the clarity of the revised manuscript. The remarks are answered hereafter with details on the associated changes made to the manuscript.

However, I think that the short time duration of the experimental campaign prevents this global system to be fully evaluated and validated.
It is not the ambition of the manuscript to provide a general validation of the data set. Instead, it is clearly stated that this study is a first attempt to validate the data set in a limited spatio-temporal framework. This limitation is explicit in the title and throughout the manuscript. However, we have modified several statements in the Abstract, Introduction, and Conclusions sections to try to make this limitation as clear as possible:
Abstract l.4-6 p.27970: "**Within the limits of this study,** this allows us to expect a negligible bias throughout the troposphere in the spatially-averaged STE frequency

derived from this data set, for example in climatological maps of STE mass fluxes. **This first evaluation is limited to eastern Canada in one summer month with a high frequency of stratospheric intrusions, and further work is needed to evaluate this STE data set in other months and locations**".
Introduction: l. 25-29 p.27972: "This evaluation covers only one summer season and is restrained spatially. **Since the skill of this STE data set may vary** in space and time, the characterisation of errors made here may not be automatically generalised to other seasons and regions. Nevertheless, this is a first step towards understanding the capabilities and limitations of this new data set."
Conclusion: l. 25-27, p. 27995: "This study represents the first evaluation of this new Lagrangian STE data set. It is however limited to eastern Canada in one summer season with a high frequency of stratospheric intrusions, and therefore the skill characterised here may not be automatically generalised to other seasons and regions. Further evaluations in different seasons and locations around the world will be useful in order to characterise its errors in different parts of the world."

Moreover, as the "core" of the paper is the presentation of the real-time system and the "regional" comparison with observations in the eastern Canada, I'm wondering if the paper can be more profitably presented as a "Technical note"?
STE is of general interest. Past studies aiming to provide global estimates of STE flux using a Lagrangian perspective have not been evaluated against observations. This study is an attempt to fill this gap and to discuss the underlying scientific challenges. For these reasons, we think that this manuscript is a scientific contribution and not a technical note.

Principal remarks

1.The comparison with the measurements only cover about 20 consecutive days during summer season. Thus, no information about the skill of this system are provided

for other seasons characterised by different meteorology which can also affect STE features (both in term of frequency and modality of occurrence). Probably, the authors should provide more information about the meteorological patterns observed over East Canada during the measurement campaigns to assess the capacity of their system in "catch" STE under different modality of stratosphere-to-troposphere transport.

It is beyond the scope of this manuscript to provide a detailed description of the meteorological patterns associated with the STE during this campaign. Overall, however, baroclinic waves can be seen throughout the campaign above Canada moving eastwards, similar to those described in detail in Bourqui and Trépanier (2010). We have added the following sentence on l. 23 p.27972: "Meteorological conditions prevailing during the campaign are similar to those described in Bourqui and Trépanier (2010), with the presence of baroclinic waves over Canada moving eastwards."

2. As also reported in the abstract the authors claimed that "the predictive skill for the overall intrusion depth is excellent for intrusions penetrating down to 300 and 500 hPa". As reported in the conclusions: "the statistical bias was found to be slightly positive in the upper troposphere". Since the authors indicated that 89% (79%) of days showed signature of stratospheric intrusions below 300 (500) hPa, I suspect that the "excellent" predictive skill in the upper troposphere can be simply due to the fact that almost for the entire measurement periods STE signatures were present over East Canada. Please comment on that!

It is true that this period is "saturated" with upper-level stratospheric intrusions. The fact that the Lagrangian data set captures this "saturation" of intrusions is positive. It can be argued that the skill might degrade in a non-saturated situation, such as when intrusions are present only once in a few days. However, since the individual dynamical/physical processes leading to intrusions in both saturated and non-saturated conditions are the same, there is no *a priori* reason for such degradation. The repetition of a similar evaluation during a period with fewer events would however be necessary to confirm this. As mentioned above, it is stated throughout in the

C15157

manuscript that the scope of this study is limited to the period of the campaign. In order to clarify this, we have added an explicit statement in our conclusion l.26 p.27995: "It is however limited to eastern Canada in one summer season with a high frequency of stratospheric intrusions, and therefore the skill characterised here may not be automatically generalised to other seasons and regions." and in the abstract l.6 p.27969: "This first evaluation is limited to eastern Canada in one summer month with a high frequency of stratospheric intrusions, and further work is needed to evaluate this STE data set in other months and locations"

Also for better clarifying this point I think that a more extended validation exercise should be done before claiming "excellent" skill" (to my knowledge ozone-sondes are lunched routinely at Egbert with a coarser time frequency, is there any possibility of comparison even only for a specific event during a winter month?) or at least conclusions should be presented in a more and more cautious way.

There are weekly balloon soundings available at Egbert and at other sites in Canada that we plan to use in a future study to evaluate the data set further. This manuscript is however limited to this measurement campaign. As mentioned above, we have tried to make the limits of the study more visible in the text. As a response to this point, we have also decided to use the terminology "very good skill" instead of "excellent skill" in the abstract and conclusions in an attempt to be more cautious in qualifying the error: l. 23 p.27968: "We find that the predictive skill for the overall intrusion depth is very good for intrusions penetrating down to 300 and 500 hPa,"; l. 19 p. 27994: "Evaluation of the STE data set at representing the overall depth of stratospheric intrusions identified in the observed profiles shows very good predictive skill for intrusions penetrating below 300 hPa and 500 hPa, respectively."

Finally, as reported on page 27974, the evaluation was restricted only to the trajectories started at 00:00 UTC and representing exchange occurring within [12h, 36h[. This

C15158

should be better stated both in the abstract and in the conclusions.

We have added a sentence in the Introduction and Conclusions sections to clarify this point: l.25 p.27972: "We use the first day forecast (i.e. $I = 0\,\mathrm{h}$), since it is expected to be associated with the smallest possible weather forecast errors."; l. 22 p.27994: "The Lagrangian STE data set is evaluated with respect to its capacity to capture stratospheric intrusions identified in the observations. Here, the evaluation is restrained to the first-day forecasts, since they are expected to be associated with the lowest possible weather forecast errors."

However, we felt that the abstract was not the appropriate place to mention it since the temporal component of the STE forecasts is not explained in sufficient detail for this to be clearly understood by the reader (the abstract should be self-explanatory). We think that this is not possible without introducing too many technical details in the abstract.

3. Is there a possibility that the methodology for identifying intrusions from ozone sondes could overestimate the actual occurrence of events. Did you perform a sensitivity study by changing the threshold values described in the paragraph 3.2 (for instance decreasing the RH threshold value)? Moreover, can you compare your results with pre-existing studies about STE climatology over Canada/North America?

We compared these results with those of Bourqui and Trépanier (2010) and they showed a good consistency. The results also compare coherently with those of Lefohn et al. (2011). We have played a lot with the parameters given in section 3.2. These parameters are tuned to follow a subjective, experience-based identification of intrusions (see l. 4-6 p. 27977). Small changes in these parameters only change the results slightly. The more fundamental difficulties with this detection algorithm are explained in section 5 item 1. For instance, point (i) may lead to an overestimate of the frequency of detected stratospheric intrusions, defined as stratospheric air irreversibly exchanged through the tropopause. Here, the question becomes: where do we place the tropopause? Nevertheless, most of the stratospheric intrusions show

C15159

a clear, unambiguous stratospheric signature and the independent STE estimates from this Lagrangian data set, using the 2PVU tropopause, lead to a similar frequency, suggesting a good degree of robustness. More investigations are needed on this aspect, though, as noted in l.21-29, p.27988.

Specific remarks

Abstract, line 28: "A significant low statistical bias...is found in the layer..". In respect to what?

With respect to the intrusions detected from observations. We are not sure if we understand the question here. The fact that the STE data are evaluated against the intrusions detected from observations should be clear in this context.

Introduction, pag 27969, line 14: numerical diffusion is more a modelling issue that a process of STE.

We agree of course that numerical diffusion is not a real process. To avoid confusion with those real processes listed in the manuscript, we have removed this term.

Pag 27971, line 5: please specify the scenario.

It is their S5-S2 scenario. A note was added on l. 5 p.27971: "S→T fluxes were predicted to increase by an average of 8% by 2030 under a climate change scenario (their S5-S2 scenario)."

Pag 27972, line 6: Actually is not the first time that a Lagrangian STE data-set on global forecast has been used. See please" Trickl, T, Feldmann, H., Kanter, H.-J., Scheel, H.-E., Sprenger, M., Stohl, A., and Wernli, H.: Forecasted deep stratospheric intrusions over Central Europe: case studies and climatologies, Atmos. Chem. Phys.,10,499-524, doi:10.5194/acp-10-499-2010, 2010"

C15160

Thank you for pointing out to this manuscript, which we did not know about. Indeed it represents another interesting initiative using weather forecasts. However it does not provide global STE forecasts. We have added the following statement: l.5 p.27972: "Trickl et al. (2010) used routine trajectory calculations based on global weather forecasts performed at ETH Zürich and covering the Atlantic Ocean / Western European sector and showed a satisfactory consistency with observations from ozone lidar over the period 2001-2005." We have corrected the following statement l.6 p.27972: "Here, we introduce the first *global* real-time Lagrangian STE data set based on global weather forecasts."

Page 27972, line 11: "...within five successive 24h time windows with a 12h residence time: It is not clear to me. Please rephrase.
This statement has been rephrased as follows: "They consist of global, five-day STE forecasts calculated daily using the 10-day global weather forecast initiated at 00 h UTC. The five-day STE forecasts are based on six-day forward trajectories started at $00 + I$ h UTC, $I = 0, 24, 48, 72, 96$, respectively, and selected as follows: they must cross either the ± 2 PVU dynamic tropopause or the 380 K isentrope with a residence time of 12 h within the time window $[12 + I; 36 + I[$ h UTC."

Page 27973, line 8: please specify "EC"
EC=Environment Canada is introduced on l.8 p.27972.

Page 27973, line 23: it is not clear to me why, along the 6-day trajectory, only the time window [12h, 36h] was analysed for STE occurrence.
The succession of such 24 h time windows estimated from daily forecasts form a continuous temporal grid with a 24 h resolution. This is important in order to avoid counting twice the same event. A clarifying sentence has been added on l.23 p. 27973: "The combination of the $[12$ h, 36 h$[$ UTC periods from the successive forecasts

C15161

forms a temporal grid with a 24 h resolution."

Pag 27976, line 15: "three large deep intrusion events"
Correction made as suggested.

Pag 27981, line 14: I wouldn't indicate 03 as "passive "tracer"l
Over the time scale of an intrusion, this is what Fig. 3 suggests. The sentence has been clarified as follows: "This similarity suggests that ozone behaves approximately as a passive tracer over the time scale of an intrusion."

Pag 27986, line 5: "Above 300 hPa, the predictive skill is excellent". According to Fig. 1, basing on the selection methodology, almost all the days presented an intrusion. Thus, are your sure that the forecast system is really excellent or simply it always "see" STE?
The question concerning the use of the "excellent skill" terminology and the problem of saturation in the frequency of events and possible overestimate of skills with respect to a non-saturated period have been answered above (see replies to Principal Remarks). We are currently working on a one-year climatology of STE that should be submitted soon, and our results show that the STE data does not "always see STE" (fortunately!). The period examined here appears to have a higher frequency than average, but its use was dictated by the availability of the field campaign measurements.

Pag 27986, line 26: I would say simply that:" the STE data not provide useful predictive skill"
We think that the fact that the STE data provides the right frequency of intrusions below 700hPa is a non-negligible result and should not be ignored in this discussion.

Pag 27987, line 23: "...which is well captured with about 5% overtorecasts". The

C15162

forecast system strongly underestimate intrusions in the lower troposphere below 700hPa (which are quite rare events). Thus, this "good" result can be simply related to the high number of non-intrusion day.

We agree that this sentence was confusing. We have rephrased it as follows: "Finally, in the region below 700 hPa, the only significant category is the "Below Intrusion" category, which is captured with about 5% overforecasts, a number strongly constrained by the large number of non-intrusion days in both observations and STE data."

Pag 27990, line 27: "it is likely that these errors cancel on climatological averages". Please explain why.

The reasoning here was that the grid-scale winds are corrected with subgrid-scale parameterisations and in the lower troposphere are expected to be mostly un-biased as a climatological average. However, the implications for trajectories might be more complicated (especially in terms of dispersion). The phrase has been removed.

Pag 27993, line 22: Also for the reasons you explained later in the text (Pag 27994, line 11), I'm not convinced that only using RH and 03 an intrusion can be identified with accuracy. Please, rephrase

We have moved and reworded the sentence from line 11 p. 27994 to l. 22 p.27993: "It should be noted however, that the identification of stratospheric intrusions based solely on individual profiles of ozone and RH has inherent flaws. In particular, it is not able to distinguish descents of dry upper tropospheric air from descents of air from above the 2 PVU tropopause. Ambiguous layers exist in the observed profiles. Summertime low-level in-situ ozone production and upward vertical transport may add further errors."

Pag 27994, line 17: You should add a comment (and a sentence in the abstract also) about the fact that the evaluation can be influenced by the large occurrence of STE diagnosed by the ozone-sonde algorithm.

C15163

We have discussed this point above. Most intrusions identified from the profiles show marked, unambiguous stratospheric signatures. The fact that the STE data independently capture these is positive and this is what we state in the manuscript. However, it is necessary to continue evaluating the data set in different conditions (including over periods with fewer events). Accordingly, we have revised the sentence defining the limits in the conclusion l.26 p.27995 as follows: "It is however limited to eastern Canada in one summer season with a high frequency of stratospheric intrusions," and have revised the last phrase of the abstract as follows: "Within the limits of this study, this allows us to expect a negligible bias throughout the troposphere in the spatially averaged STE frequency derived from this data set, for example in climatological maps of STE mass fluxes. This first evaluation is limited to eastern Canada in one summer month with a high frequency of stratospheric intrusions, and further work is needed to evaluate this STE data set in other months and locations."

Additional Minor Changes

Title: We have added the word "global" in the title as: "A new global real-time Lagrangian diagnostic system for stratosphere-troposphere exchange: Evaluation during a balloon sonde campaign in eastern Canada"
l.2 p.27968: "A new real-time" changed into "A new global real-time"
l.4 p.27968: "performed globally following" changed into "performed following"
l.7 p.27968: "are calculated for six days" changed into "are calculated forward in time for six days"
l.13 p.27974: Changed "This allows the analysis of (rapid) upward transport as well." into "This allows the analysis of (rapid) upward transport using six-day trajectories as well."
l.18 p.27972: "of meteorological data" changed into "of the meteorological data"
l.2 p.27973: "campaign, provides" changed into "campaign and provides"
l.8 p.27973: "multi-day" changed into "10-day"

C15164

l.11 p.27973: "Multi-scale" changed into "Multiscale"
l.15 p.27973" "upon a Eulerian iterative" changed into "upon an iterative Euler"
l.6 p.27974: "for the same forecast" changed into "for the same weather forecast"
l.9 p.27974: "offer a forecasting" changed into "offer an STE forecasting"
l.26 p.27974: "In this evaluation paper" changed into "In this first evaluation paper"
l.17 p.27977: "multiple intrusions layers" changed into "multiple intrusion layers"
l.11 p.27979: "intrusions, with still one day" changed into "intrusions, though still with one day"
l.10 p.27981: "identified here as from stratospheric" changed into "identified here as of stratospheric"
l.20 p.27981: "with instant, point-observed" changed into "with instantaneous, point-observed"
l.1 p.27982: "spacings from" changed into "spacing from"
l.29 p.27982: "forecasts" changed into "forecast"
l.27 p.27983: "trustable" changed into "trustworthy"
l.28 p.27983: "false alarm" changed into "false alarms"
l.12 p.27984: "intrusions seem also" changed into "intrusions also seem"
l.24 p.27985: "STE data is therefore" changed into "STE data are therefore"
l.10 p.27986: "STE data has" changed into "STE data have"
l.27 p.27986: "STE data shows" changed into "STE data show"
l.28 p.27986: "but does" changed into "but do"
l.20 p.27986: "through the entire" changed into "throughout the entire"
l.23 p.27986: "(FB=87)" changed into "(FB=0.87)"
l.25 p.27986: "This is artificial, and due" changed into "This score is artificial, and is due"
l.1 p.27989: "trajectory's starting grid" changed into "trajectory starting grid"
l.2 p.27989: "evalutation" changed into "evaluation"
l.27 p.27989 "before that any" changed into "before any"
l.21 p.27991: "distinct candidates clusters" changed into "distinct candidate clusters"

l.9 p.27992: "cluster passes at a few" changed into "cluster passes a few"
l.22 p.27992: "either shifted or too" changed into "either shifted geographically or are too"
l.2 p.27993: "new real-time" changed into "new global real-time"
l.26 p.27993: "every second measured" changed into "half of the measured"
l.17 p.27994: "Evaluation of the STE data set at representing" changed into "Evaluation of the ability of the STE data set to represent"
l.16 p.27995 "factor three at least" changed into "factor of at least three"
l.8 p.27996: "We thank Seok-Woo Son" changed into "We also thank Balbir Pabla for his assistance with the GEM data transfer and Seok-Woo Son"
Table 3 caption: "bracket" "no false alarm" changed into "brackets" "no false alarms", respectively
Table 4 caption: "for the four pressure" "no false alarm" changed into "for four pressure" "no false alarms", respectively
Table 5: "16(sh)" was changed in normal font, instead of bold font, as it is not used in Fig. 7
Figure 6 caption: changed "caused by the limitation to 6 days of the trajectory length." into "caused by the limitation of the trajectory length to 6 days."