We thank the referee for his comments and respond in the following. Below, bold texts are the original referee comments.

The paper discusses the problem of determining of the injection height of smoke from large-scale landscape fires, which is both interesting and urgent. The author group includes well-known names in the fire research area, so I started the reading with high hopes. Unfortunately, the paper appeared disappointing in several senses described below. This is highly surprising since I am familiar with many papers of some of the manuscript authors. After spending a lot of time digging in this one, I had to accept that they did not pay enough attention to it.

We were slightly taken aback by the abruptness of the first paragraph of this review, but we did find some constructive remarks in this review and here provide an answer to all the points raised.

The paper is astonishingly difficult to read. This is a huge manuscript (totally over 80 pages of the ACPD format) with long sentences, whose meaning is by no means easy to comprehend. The reader gets this already in the abstract, which is long and chaotic. I had to read it several times to finally decipher the simple message: the paper is extending an unnamed (why unnamed?) model with a mass conservation equation and a new entrainment algorithm and compares it with remote-sensing observations to find the model coefficients. The authors spent two pages to say it. I was particularly baffled by the last sentence, which took 9 lines and contained no message whatsoever. The paper continues this style throughout the whole manuscript, which is chaotic and combines several iterations of the same with absence of vital information. For instance, the "steady-state fire requirement" is discussed at least thrice, whereas the basic information on meteorological data used in the analysis is missing, except for a remark that it came from ECMWF. Poor style and organization of the paper are enough to suggest rewriting the whole manuscript in a reader-friendly way.

The methodology selected by the authors seems to have several large issues. The exercise itself is straightforward: the authors put a few new equations in the existing model, which brought about six new parameters that have to be identified through the fitting to the observation data. The only dataset of sufficient volume is the MINX database derived from MISR retrievals, which has now about 13,000 fire plumes (10,000 if "poor" retrievals are removed). The authors took that one and removed 99.7% of the data following arbitrary and vaguely explained criteria, finally ending up with 38 cases! of which they further removed two cases just because the model failed to meet them. Well, having 36 points and six dimensions of freedom in the system is a red light, which means that the approach is wrong. The data mis-management had numerous far-reaching implications. The main ones are: (i) evident over-fitting of the model to the data noise since RMSE of the model is much less than the uncertainty of the MISR data themselves, (ii) complete loss of large classes of fires, i.e. the model is not even deemed to work for them, (iii) fantastic steps in the analysis, such as a correlation coefficient made for three dots, (iv) statistical significance neither calculated nor considered, (v) absence of an evaluation dataset turned the whole evaluation section into a hand-waving, (vi) fire plume climatology got no ground: extrapolation of 36 fires to a few tens/hundreds of thousands observed by MODIS is going much too far.

We restructured the paper to make sections on the optimisation and the evaluation of the model clearer. Section 5 now only deals with the model optimization, i.e. estimation of the input parameters defined in Section 2.2, while Section 6 is re-named "PRMv2 performance evaluation" and presents the model evaluation.

We have improved the presentation and the discussion of the model parameter optimisation approach we have deployed here. Sections 5.1 and 5.2 were re-written, and in addition we updated and re-ran our MCMC routines with an increased focus on realistic uncertainty values (in particular the precision of the input data). The new results are shown in an updated Figure 9 and in Table 1.

In the updated analysis we:

1/ used (an updated version of) the MCMC algorithm to estimate the precision of the fit (as in Braswell et al., 2005). Previously some aspects of the approach were set too arbitrarily in order to ensure that the MCMC algorithm converged properly, which in hindsight was an inadequate strategy. The new root mean square error value estimated by the algorithm (~800 m) is close to previous literature estimates of the precision (of the MISR-derived data). The new strategy led however to a broadly similar set of results to those obtained previously, with a few differences (e.g. shape of Cd distribution) which are now discussed.

2/ We re-ran the MCMC algorithm for 60,000 iterations. The increased number of iterations helps to ensure that the algorithm properly and fully explored the parameter space.

3/ We have re-written the optimisation Section to increase its clarity and moved non-essential details to an Appendix. We now discuss the results of the parameter uncertainty analysis in more detail and describe which parameters are well constrained by the data, and which are not. We also removed the likelihood (Equation 22), due to the statistical implications of the inclusion of the penalty term in this equation (i.e. non-normal).

4/ We use the SA (sensitivity analysis) parameter estimates as our final estimates for the parameters of our plume rise model. We used these values as they were representative of the posterior distributions (they were found within the 68 % credible interval). This gave us confidence that our SA approach had found a minimum (error) region of the parameter space.

5/ We ran the MCMC starting from the above parameter values to assess the uncertainty around these points, whereas in the previous analysis we started at arbitrary points. Our new strategy was designed to eliminate the need for 'burn-in'. Burn-in is when the MCMC algorithm spends initial (and subsequently discarded) time/iterations searching for the optimal region of the parameter space. In the new situation we already had the optimal starting point from the forerunning sensitivity analysis .

Responses to major comments:

Abstract: The abstract is written for a public which is not familiar with the subject. That's why there is a bit more information than the simple picture drawn by the reviewer. However we agree that he was long, and therefore re-word it to make it shorter.

Why the plume rise model is unnamed?: The plume rise model of Freitas et al. (2007; 2010) does not have any specific name. To avoid using a paper reference in the abstract the model is therefore unnamed there - though we could use a reference to a paper if that is preferred by the journal (we know this is usually avoided in paper abstracts).

Missing basic information on meteorological data: A paragraph dedicated to this topic is in Section 4.1.

Steady state discussion: The steady state criteria used in the selection of the fires to be included in our analysis, selected from the complete list of fires present in the MISR plume rise dataset, is an important aspect of our approach. It is first introduced in Section 1.2, and then further developed in Section 4.2.

The approach is wrong: We do not agree that removing data points from the training dataset is an inherently wrong approach. Indeed, it many areas of classification and machine learning, the accuracy of the training dataset is key. Therefore we believe that we are right to focus in on the components of the complete MISR plume rise dataset that have the most value for inclusion in our training data. We believe that the MISR plume rise dataset cannot be used blindly for "model training", and that we will obtain a better and more optimised result by focusing on the use of fires are carefully selected - in particular to avoid including fires where the scenarios of fire activity and the plume height are not coupled. The whole of Section 4.2 is dedicated to this task and all criteria used in the selection are introduced and discussed therein. We appreciate that this approach is different to some existing work on the subject of plume rise estimation, which have used the complete MISR dataset and may themselves also be a sensible choice, but we feel our approach to the careful selection of the training data is also very well justified.

"Mismanagement" of the data: The different points identified by the referee are listed below

(i) *Over fitting*: This is a typo for which we apologise for including erroneously. For more details, see the Section below on specific comments.

(ii) *Complete loss of large class of fire*: the only class of fire which is systematically and completely removed from the training data extracted from the MISR database is agricultural fires (see Section 4.2). We feel this is valid as the smoke from these types of fire typically do not rise to large plume heights, and perhaps more importantly the fires are under careful control of human influence which the plume rise model makes no account of. Section 6 describes the operational set up of PRMv2. There it is specified that PRMv2 is not run for small fires (such as agricultural fires) as the Dozier algorithm used to derive the fire size and this has been show to have limitations for fire smaller than 1 ha (See Appendix B2; and also much of the literature about the Dozier algorithm robustness for smaller fire events). The smoke from fires smaller than 1 ha in active fire area (including the agricultural fires) are assumed to be trapped in the boundary layer (this was also made clear in Section 6 covering "evaluation").

(iii) *fantastic steps such as correlation made of three points*: Of course we would prefer more points on this plot, but we emphasise that no major conclusion is made from this plot. We just note that, the selection of "good" fire clusters exhibits the same linear relationship between FRP and InjH for fires occurring in unstable atmosphere as the one observed by Val Martin et al. 2010, and Amidiris et al. 2010.

(iv) *statistical significance neither considered*: This is considered in the MCMC uncertainty discussion in Section 5.

(v) *absence of an evaluation dataset*: Section 6 directly relates to our evaluation of the optimized PRMv2. Rio et al. (2010) used a rather similar approach to evaluate their plume rise parameterization, with few direct comparisons to modelled and measured plume InjH. In particular, they studied the sensitivity of their parameterization when run over large areas (in their case southern Africa). Their parameterization is plugged into the LMDZ model and CO_2 concentration is considered. In our case, PRMv2 is run off line, and sensitivity of model outputs to atmospheric stability and season is discussed in our manuscript.

(vi) *fire plume climatology got no ground*: we are not presenting a completed "plume rise climatology", but rather showing how our efforts are working towards a climatology. Section 6 is essentially used as a test of the optimized PRMv2. This was made cleared in the reorganised manuscript (see above details on our reorganisation).

Response to specific comments:

P.9819, L.18-19. There is always a space for improvement but simply wiping out all existing models and parameterizations as "unsatisfactory" without even formulating the criterion to be satisfied is not an acceptable style:

As shown in Val Martin et al. (2012) (see Fig.2), and the review of Paugam et al. (2015) (see Fig. 7), none of the current plume rise model parameterizations compare well with the MISR-derived plume rise dataset. In particular, the fact that for all fires in the MISR dataset the modelled plume height derived using the parameterization of Sofiev et al (2012) remains below 4km appears to be a significant limitation.

P.9819, L.18-25. Is that all? I recall about a dozen of papers discussing the impact of injection height, either directly working with it or mentioning the issue in connection with sensitivity studies and modelling efforts. Some of these studies are later even quoted by the authors – and yet not included here. Two pages later, the authors make a U-turn and mark-out three approaches, which should be mentioned here.

Here, we are specifically referring to work that has demonstrated current limitations of plume rise models - for example Val Martin et al 2012 who ran an extensive analysis of the Freitas model we base our adapted model on, and Paugam et al. 2015 who present an relatively exhaustive review of fire plume rise models and where most existing works are referenced.

To make our referencing clearer however, we have remove mention of Val martin et al 2012 here and only reference the Paugam et al. 2015 review which gathers and presents all the information that the reviewer is expecting at this stage in one place.

P.9819, L.25. What is "improving large scale transport relatively locally"?
This sentence was modified to make it clearer and we apologise if it was unclear before. It refers to the discussion of Elguindi et al 2010, which shows that modifying constant InjH predictions only improves tracer transport at certain locations. The new sentence reads: "Conversely current efforts on InjH predictions have only been able to substantially improve large scale plume transport in some scenarios (Elguindi et al., 2010)."

P.9819, L.27. Freitas et al, 2007 is a 1D model, not a parameterization.
Thank you for this comment, we agree and we have modified the paper accordingly.

P.9822, 9837 and in other places. The whole concept of steady-state fire sounds ill- fated to me. None of fires is ever in such condition: changing wind, fuel type and density, evolution of the fire front position and shape, its interaction with landscape topography, etc, all these parameters are never stable. The task of back-tracing of the fire intensity is interesting and challenging but I have a feeling that it should be addressed up-front rather than pushed under-the-carpet by assuming that some fires are more "steady" than others with little reason to do so.
This we believe is a rather confusing comment. Our approach is trying to sort out the issue of the synchronisation of the activity observations between the plume and the causal fire. This is required because with only one measurement time (i.e. at the Terra satellite overpass time that carries MISR and MODIS sensors that respectively make our plume height and FRP observations) we need to select fires where the fire activity and plume dynamics are clearly related to one another at this moment (e.g. avoiding fires where the activity to too close to the start or end of the fire and where, for example, the plume has reached a height that is unrepresentative of the actual thermal energy emission being produced at the time of the Terra satellite overpass). This is inherently because the height to which a smoke has risen at time X will depend not on the fires heat release rate at time X, but on the heat release rate somewhat earlier on when that smoke was originally formed. We believe that the use of "unsynchronised" plume and FRP situations is best avoided by focusing on fires that have reached a type of "steady state" condition, usually near the peak intensity of their thermal energy release rate, and where for example there is a period of semi-constant thermal energy emission rate. The heat release rate (FRP) measured at the Terra satellite overpass time can thus be related to the plume rise altitude measured at the same time, since the observed FRP is likely to be representative of the FRP when the smoke was formed quite some minutes before. The end of Section 4.2 discusses this topic, and to try to make our point even clearer we have defined what we mean by "steady state" in Section 1.2. Our new text thus reads:

„Ideally there should be a focus on remotely sensed observations that best represent a 'steady state' situation or situation where both the causal fire and the resultant plume are least variable, avoiding situations where (i) the plume is potentially masking parts of the driving fire, or (ii) where the fire and plume are too close to their initialisation times and thus still changing their nature, or too near to their end point where plume and atmospheric dynamical processes can be too highly coupled. Although mathematically not fully correct, we use the term steady state fires for simplicity here. The current work takes account of potential time delay between fire activity and plume development, making careful use of simultaneously recorded fire and plume observations to try to minimise the effect of this issue, whilst...“

P.9834, L.12-13. Why? If the fire is not related to the plume then why to include it?
We cannot fully understand this comment here - perhaps the page or line number is incorrect?

P.9834, L.18-22. I did not understand the value of re-extracting the MODIS FRP data instead of using these very data already picked in the MINX dataset. MINX project has this extraction done with much more care than the approach suggested by the authors: 20km around the plume reference point easily picks the fires not associated with the plume but also can miss the needed fires if the plume long enough. This step is worsened the quality of the dataset, not improved it.
Our method was used to avoid relying on the manual delineation of the "plume" seen by MISR, which is used as an input in the MINX algorithm as drawn manually by the MINX-operator. In some cases we have checked, the selected "plume" dataset did not encompass all the active fire pixels producing FRP at the base of the plume, which is what is driving the plume lofting. We also removed scenarios with too many fire clusters near the plume origin (see Section 4.2) as we could not tell which fires were contributing to which plumes. Our selected plumes are isolated from other neighbouring fires and are therefore likely to be linked with the most intense fire located in a perimeter around the plume origin. The plume length has no impact here, as we only consider fire pixels at the origin of the plume where the vertical updraft is instigated.

P.9837, L.21-26. A very strange move. About 90% of fires are ignited by humans, either on purpose or accidentally. Having removed the agriculture-areas fires, what to do with the deforestation fires? Or with those in urban-rural interface? Conversely, I would suggest that these fires are the ones to be used rather than ignored: MISR makes its observations before but close to midday local-time, i.e. regularly. So, the regular deliberate fires constitute a dataset with “known features”, which can be taken into account and correlated with MISR. Why do the authors think that randomly ignited natural fires (even assuming that they are deciphered out of MINX dataset) are any better?
We are only focusing on North America in this analysis, and it is clear that "deforestation fires" are a minor issue at such latitudes. We agree that when applied to a geographic location such as South America we would have to re-optimise the plume rise model using a set of completely new data.
Fires in agricultural or urban-rural interface landscapes are much more likely to be influenced by humans, including fire-fighter units, and this includes North America (Lee et al. 2013 record a successful "initial attack" period requiring a response time of 30 to 60 min in California). If the fire activity is perturbed by such anthropic activity, our requirement to identify fire events with fire and plume activities "synchronised" becomes more difficult. Therefore we believe that we can find more "good quality" training data by removing the agricultural areas, and thus focusing our screening on "natural landscapes" in which fires are more likely following their diurnal cycle.

P.9838, L.5-8. I did not understand the procedure of derivation: the MINX dataset does not have any wind profile. The whole procedure described in this paragraph is unclear, both from scientific and technical points of view. Since the authors removed three quarters of the observations using this criterion, a much better ground and description is needed. In fact, this is one of corner stones of the problems of the paper: the authors formulate a vague criterion, which has severe consequences, but do not make any attempt to justify the choice. In this case, the problem is difficult because the MISR wind and height retrievals are correlated, and careful analysis is needed to understand the data with and without wind correction, to estimate the related uncertainty of the plume height, may be, to filter out some retrievals with evident problems (but not 75%, of course), etc. An important aspect is the quality (or, rather, existence) of the reference point: modelled global meteorological fields form a shaky ground for any sharp action with the data. The representativeness issues may be overwhelming.
As explained in our manuscript, the objective of the criterion commented on by the Referee is to filter out fire events where the "MISR-derived" and "ECMWF-modelled" wind profiles differ signficantly. This screening removes small fires, and for our analysis we require plumes that extend at least over 300 m, and events where the modelled wind profile failed. MISR derived wind profile is used as “reference point”, not the modelled wind profile. A large number of fires from the full MISR dataset are removed by this criterion, and it would be better if we were able to keep a larger number of these in our training data, but it is difficult to optimize a physically based model when the ambient atmospheric input profiles are unrepresentative of reality.

P.9838, L22. Another strange criterion. The authors have previously rejected the MINX fire clustering, which uses the actual plume edges to associate the fires, replacing this procedure with a 20km circle, which has no physical ground. Here is another arbitrary requirement claiming 30km area to be free from fires. Plumes cannot interact at such scales and resolution of MODIS and MISR are both an order of magnitude higher, so retrievals of those clusters cannot interact with each other either.

We disagree with this comment. This criterion is used to remove events where multiple fires occur in the same area and it is difficult to match fires with their resulting plumes. For example, see the image below.

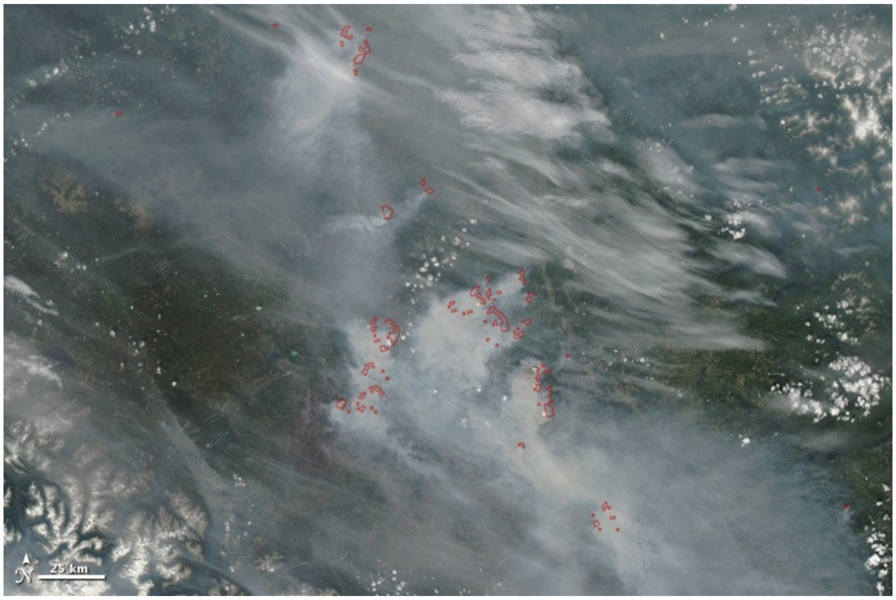

Image extracted from http://earthobservatory.nasa.gov/NaturalHazards/view.php?id=45056

In this area of British Columbia (Canada) many fires occurred at the same date and time (August 4 2010). In the MISR dataset, several fires were selected from such large fire events as these. But to avoid effects from surrounding plumes (i.e. changes in atmospheric stability), we only selected in the “good” cluster training dataset isolated fire events, based on a 30 km radius that we consider not conservative since large plumes can easily spread downwind over longer distances - reaching locations where other fires start (see middle of the figure). However tests show that this criterion for screening out such events was sufficient to remove most of the dense fire events where confusion between plumes and their causal fires existed. Furthermore, we are interested here to remove downwind fires, not lateral interactions.

The 20 km radius used in the derivation of the FRP is set to ensure that we are considering all fire in the vicinity of the isolated selected plume. A rather large radius is used (i.e. 20 km) to correct for misplacement of the plume reference point. This point is set manually by the MISR operator as the point where the contour is started, and is not always located at the base of the plume updraft.

P.9839. This is baffling! Having initially over 13,000 plumes (the MINX project database today), the authors left 39 for the model development and evaluation. Should we really accept that the whole dataset is unusable? The remaining 39 cases are not sufficient for any feasible application because both statistical significance and extrapolation possibility of the results will be negligible.

The current MISR dataset for North America includes 3820 fires, among which 3313 are classified having a "good" quality flag. Of these, only 38 fires came out of our training data selection process. Previous works where the MISR data set were used extensively in the derivation of a method for plume rise estimation have employed the entire dataset (e.g. Sofiev et al. 2012) - but we wished to focus on the optimum set of training data. The 38 fires we selected for inclusion in the training dataset were therefore used for the optimisation of the model, but the evaluation of the model performance was performed based on a statistical analysis of model runs made on the complete 2003 MODIS active fire archive. This has been made clearer in the manuscript via a change the Section title and by specifying clearly in Sections 4 and 5 that the "good" cluster dataset is not used for evaluation, but only for the optimisation.

P.9842, L.15. This must be a joke!! Do the authors really think that correlation can be meaningfully computed for 3 points??

As mentioned towards the start of our responses, this plat was not used as a main validation but rather as a side comment.

P.9847, L.2. Being within 4.5 m from the MISR observations (error of 20m2) means that there is a huge over-fitting to the noise in the data: declared accuracy of MISRis 200m, if I recall correctly. The independent studies show even larger uncertainties. Whatever is much better than this is nothing but over-fitting. Not surprising though: 36 points for 6 dimensions to catch is a guaranteed over-fitting.

There was a typo in the manuscript, for which we apologise as stated above. The sum square residual (SSR) error is in km^2. It is clear in Fig. 8, which shows the best instance of the 32 simulating annealing runs, that our optimised model is not giving an SRR error of 20 m^2 - but rather the units are km^2. The section has been modified and rms values introduced to make the discussion clearer (see Section 5.2).

P.9847. And here is another problem: one cannot kick out the data just because the model does not fit them. This is not acceptable.

We used the available MODIS observations at each of our 38 fires in our "good" fire cluster dataset to evaluate the evolution of the plume during each fire event. The two fires we removed show massive changes in their plume shape just after the MISR observation, and so we removed these two fires because they are considered not to meet our "steady state" criteria.

Section 5.2 has little about the actual performance evaluation. Physical reasoning can be used in discussion but evaluation requires directly comparable quantities: one observed, one predicted. The section does not have them because the authors have disqualified 99.7% of MISR data following arbitrarily picked criteria criticized above. Out of curiosity, how does the scatter-plot for the whole MINX dataset looks like? The authors heavily refer to Sofiev et al works – there such plots are presented, with consideration of "poor" MISR retrievals in one case and having them removed in another.

As mentioned previously, the scatter plots in Sofiev et al 2012 does not show any fires with a plume higher than 4 km. Are those plots more valuable than the sensitivity test we perform on PRMv2 for a year of MODIS data? In our approach we questioned the validity of the MISR data set when used in the optimization of plume rise parameterizations, in particular when used with simultaneous measures of FRP and injection height. We show that we can have potentially a better behaviour when considering PRMv2 responses to unstable atmosphere and season variability. We mention also that there is still *a space for improvement* (see first paragraph of Sec 6 and Conclusion) and that the validation of PRMv2 could be improved, but would "rely on high spatial and temporal resolution information, including ideally simultaneous field measurements of the key parameters driving smoke plume rise".

References

Yohan Lee, Jeremy S. Fried, Heidi J. Albers, Robert G. Haight: Deploying initial attack resources for wildfire suppression: spatial coordination, budget constraints, and capacity constraints. Canadian Journal of Forest Research, 2013, 43:56-65, 10.1139/cjfr-2011-0433