Dear Dr. Galmarini:

Below you will find the colleague reviews to this paper, along with our response. Reviewer comments are in black. Responses are provided in *blue italics*. Together with the comments and responses we attach a copy of the manuscript text with additional changes tracked. The revised manuscript with figures is provided as a separate file. We hope you find these responses adequate to merit publication.

Sincerely,

Àlex Martí.

Reviewer #1

General comments:

1. p. 2, line 30-31
There are on-line operational systems, e.g., for dust transport.

The text has been modified to include "However, ... , the experiences from other fields (e.g. on-line models for air quality, dust, etc.)..."

2. p. 10, line 15 and elsewhere in the manuscript The phrase "... decreases with coupling frequency..." may be misunderstood to mean the opposite of what is actually meant, i.e. "... decreases as the coupling frequency increases ..." Therefore, I would recommend to say "... decreases with decreasing coupling frequency ..." instead.

We agree with the reviewer. We have modified the manuscript with: "decreases with longer coupling intervals..."

Suggestions for technical corrections

1. p.1,line20 Replace "credited to" by "due to"

Corrected.

2. p.1,line20 Replace "that" by "as"

Corrected.

3. p. 2, line 14 and elsewhere in the text Leave out "of" in "require of"

Corrected.

Reviewer #2

General comments:

1. How does the offline version of your model handle meteorology at the times between the coupling times? On p. 5, line 16, you suggest that meteorological parameters are set to constant in between coupling times. But many offline models linearly interpolate. Could higher-order interpolation schemes reduce error?

That is a good observation. Thank you. The objective of this paper was to compare the effects of the on-line and off-line coupling strategies between the MetM and VATDM. However, it is true that most off-line systems employed at operational level, perform a linear interpolation in time to attenuate the off-line coupling effects. This is not possible in our off-line strategy because of the concurrent solution of both meteorology and dispersal. Higher-order interpolation schemes to drive the meteorological input would, indeed, reduce the error associated to off-line forecasts.

2. Section 2.3.1. It was difficult for me to grasp the physical significance of some of the quantities used to compare output from the offline and online models. For example, the Structure component S is said (line 3, page 8) to capture information about the size and shape of cloud objects. But all of the terms in S refer to mass; of a node, column or nodes, or cloud object. How do differences in S reflect variations in size and/or shape?

Thanks for the comment. The basic idea of the structure (S) component in SAL is to compare the volume of the normalized ash column load (ACL) objects. An object is a group of adjacent grid cells that have an ash cloud loading value above a given threshold. Scaled masses (V_n) are calculated separately for each object between the off-line and on-line forecasts. The scaled mass provides the ACL area-integrated (vertical integration for all z nodes) of each forecast, and therefore offers insights of the size and shape of the ACL for the on-line and off-line forecasts.

The exact meaning of some other parameters, such as R_xy and D were unclear.

D corresponds to the domain area of the simulation, while R_xy is the area-integrated ACL in grid cell xy within each (on vs off-line) simulation. The text has been updated to clarify this section.

And it was unclear to me how one could get a value less than zero for L_1.

That is correct. We apologize for the typo. The values of L_1 are in the range of [-1,1], with L_1=0 suggesting identical centers of mass for both forecast. Corrected.

Also, I didn't see a definition of the parameter B in eq. 13

The definition of the parameter B was missing. Thank you! We have added them in the manuscript.

3. Several source parameter terms in Table 2 are not adequately explained. Details are in comments in that table. Also, I don't see sources cited for the observations used to constrain the eruption source parameters. Other key observations, like the arrival time of the Cordon Caulle ash cloud in Buenos Aires on 4 June 2011 (p. 20, line 3) do not cite sources.

Thank you. We have included additional information to most parameters in Table 2 and added the corresponding references regarding the arrival of ash to Buenos Aires airports (Collini et al., 2013)

Collini, E., Osores, M. S., Folch, A., Viramonte, J., Villarosa, G. and Salmuni, G.: Volcanic ash forecast during the June 2011 Cordón Caulle eruption, Nat. Hazards, 66, 389–412, doi:10.1007/s11069-012-0492-y, 2013.

4. Some of the figures need more description. For example, the methods of estimating mass eruption rate plotted in Fig. 5b. And the various lines in Fig. 10b representing mass eruption rate with time.

We have added additional information to each figure's description.

Figure 5b. Resulting MER over time considering different plume parameterizations (FPLUME - Folch et al. (2016); Woodhouse - Woodhouse et al. (2013); Mastin - Mastin et al. (2009); Degruyter - Degruyter and Bonadonna (2012));

Figure 10b. Resulting MER for each coupling strategy (meteorology coupled on-line or with intervals of time of 1h, 3h, 6h and 12h) with Degruyter option only.

5. Section 4.1.2: In examining the Eyjafjallajökull eruption, why do you think the misses were in the south and the false alarms were in the north? Dacre et al. (2011) suggested there was an error in the modeled wind speed over England. Was the south-oriented wind just accelerating, and the acceleration was not being caught when the coupling intervals were too infrequent?

That was an excellent observation. While the errors suggested by Dacre et al (2011, 2016) could be a factor to the errors shown in Fig. 8, the errors in the plume position shown by the off-line forecast are probably caused by the cumulative effect of errors in the infrequent coupling of driving meteorology en route. The synoptic meteorological situation over South Iceland and the North Sea increases (see Fig1. Folch et al., 2012) suggests that the south-oriented wind speed (>30m/s) increased during the 15 and 16 of April (purple area if Fig. 1 from Folch et al., 2012). This scenario, together with employing infrequent intervals to drive the meteorological conditions, could explain why our results showed misses in the south and false *alarms in the north of the plume. The manuscript has been updated to clarify this section.*

Folch, A., Costa, A., Basart, S., 2012. Validation of the FALL3D ash dispersion model using observations of the 2010 Eyjafjallajökull volcanic ash clouds. Atmospheric Environment 48, 165e183.

Dacre, H. F., et al. (2011), Evaluating the structure and magnitude of the ash plume during the initial phase of the 2010 Eyjafjallajökull eruption using lidar observations and NAME simulations, Journal of Geophysical Research: Atmospheres, 116(D20),

Dacre, H. F., N. J. Harvey, P. W. Webley, and D. Morton (2016), How accurate are volcanic ash simulations of the 2010 Eyjafjallajökull eruption?, J. Geophys. Res. Atmos., 121, 3534–3547, doi:10.1002/2015JD024265.

6. Section 4.1.3: You say that bias scores suggest that offline forecasts tend to systematically underestimate ash column loading? Is there a physical explanation for this? If the model is conserving mass, does this imply that offline models also systematically overestimate cloud area? And how does this statement square with your statement on p. 19, line 6, that all off-line forecasts OVER-estimate ash column loading?

Thank you for the observation. First of all, we have noticed that statements claiming overestimation of ACL were incorrect. In most cases, off-line forecast tend to UNDER-estimate ACL compared to the on-line forecast. For the Amplitude component of SAL and the Bias score, off-line forecasts underestimate the domain-averaged ACL when A or Bias are < 0. However, in the case of the Frequency Bias (FBI) score, off-line forecasts underestimate the domain-averaged ACL for values of FBI < 1. The original text was incorrect in the description of some of the FBI and A scores and it has been updated accordantly.

The fact that most off-line forecasts show a FBI < 1 at the end of the simulation indicates that the forecast system has a tendency to underestimate ACL events. However, it does not measure how well the off-line forecast corresponds to the on-line simulation, only measures relative frequencies.

Amplitude scores, on the other end, provide a simple measure of the quantitative accuracy of the total concentration of airborne ash in the domain ignoring the field's subregional structure. According to Wernili et al. (2008) Amplitude scores range between [-2,2] with 0 denoting no difference between off-line and on-line forecasts. An amplitude score of +1/-1 indicates that off-line forecasts overestimate/underestimate the domain-averaged ACL by a factor of 3. Scores of A=0.4 and 0.67 correspond to factors of 1.5 and 2, respectively. Our results indicate that most off-line forecast have a small tendency (i.e. about 0.2 times) to underestimate ACL objects from the on-line forecasts. The infrequent coupling of the meteorology in the off-line forecasts could be the cause of this difference. In some cases, the infrequent coupling interval (e.g. 3h, 6h, 12h) of an off-line forecast *might imply that some of the ash is deposited on the ground or leaves the computational domain. We must highlight that Amplitude scores only compare airborne ACL events within the computational.*

Wernli, H., Paulat, M., Hagen, M. and Frei, C.: SAL—A Novel Quality Measure for the Verification of Quantitative Precipitation Forecasts, Mon. Weather Rev., 136(11), 4470–4487, doi:10.1175/2008MWR2415.1, 2008.

Additional minor and specific comments are included in the attached pdf. I look forward to seeing the final version of this paper.

We appreciate the additional constructive comments, which helped to improve the clarity and the quality of the manuscript.