This is an interesting paper, healthily interdisciplinary, as it applies techniques from particle (astro)physics to the study of the atmosphere, leading to novel concepts. It is very well written.
Before accepting the paper, however, I recommend that some weaknesses of the data interpretation be addressed, as listed below. Moreover, I would have appreciated the author to elaborate more about the potential relevance of this study for atmospheric science.
My own research is mostly in particle physics, therefore I may be missing a few points that might be more obvious to the typical readers of this journal. However, I assume that this article should aim at being appreciated also outside of the core readership of this journal, therefore I suggest the author to assume that whatever is unclear to me might possibly deserve being elaborated more in the body of the paper.

Main comments:

The entire point of this paper is to extract knowledge about the atmospheric dynamics from measurements (performed at sea level) of observables that implicitly integrate through a very thick amount of atmosphere (from 30 to 0 km a.s.l.). Necessarily, this implies that in order to extract any interpretation from this method we need a reliable simulation of the atmospheric dynamic. This creates a sort of circular logic. Not being an atmospheric scientist, I would like to see the paper elaborate more explicitly on this point: how much can we rely on the models and which free parameters of the models can be pinned down with this new method?

I am not sure how to reconcile the assumptions of this paper with the similar study by Tramontini et al.:
https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2019EA000655
In that paper, the muon flux is used to infer information about the middle-atmosphere dynamics such as SSW. I also note that SSW are mentioned at lines 75-76 of the present paper (but without citing Tramontini et al.).
My understanding is that the two studies are using the same probe to analyze the same medium, but under different assumptions and to infer different free parameters. If that's the case, what could be done to disentangle the middle- and upper- atmosphere dynamics in order to get less model-dependent results?
I suggest to include a citation to that paper (to which I am not related, therefore I might have misunderstood its message), accompanied by a critical discussion of similarity, difference and possible complementarity with the present study.

Similarly, I understand that this work makes also different assumptions with respect to the papers cited at lines 95-98. It is said at lines 99-100 that the difference is that those works focused on barometric and temperature effects. I suppose that the same can be said of the SSW study by Tramontini et al.? Anyway, a few more words would be appreciated here, to also address the previous point.

The paper separates the barometric effect and the isobaric height effect. Usually one has the barometric formula describing the functional dependence between the pressure and the height, in that case, the ground-level pressure should correlate with the height of the muon generation layer. May you please clarify whether by "barometric effect" it is meant here the effect of deviation from the canonical barometric formula?

Lines 151-152: not only the energy loss, but also the probability of meson-nucleus interactions. Can you argue that this has a negligible effect on your conclusions?

One of the crucial and most original developments in this paper is the proposal to use cyclone data for calibration. I find it clever, but again this makes several implicit assumptions: in essence, it is assumed that the barometric dependence of muon flux variations is universal. At the very least, I would like to see a proof that it is universal across different cyclones (and even that, probably, wouldn't be an air-tight proof that the calibration is valid in all contexts). Are there additional cyclone datasets that the author can analyse and compare, to 1) prove that the calibration curve extracted for one particular cyclone is valid for any cyclone, and 2) extract, from the variance of the fit parameters across all datasets, an additional systematic uncertainty to attach to this calibration. And it would be great if the author can find other data (not cyclone-related) where large pressure differences are observed, to be used to further validate the method.

Please perform a test of the sensitivity of the fitted formula of eq.1 to the time window during the Typhoon period, for example, do the coefficients change significantly if only considering the data points between 09/29 and 10/01 ?

How stable is the detector response? Implicit in all this analysis is that any observed variation in muon flux can only be due to factors that are external to the detector. But if e.g. the gain or the noise of the detector depend on some systematic effect, which is usually the case, this should be estimated. For the type of detector considered, for example, one can expect significant dependence on the **local** temperature around it. The statement at line 172 may be too optimistic. If the calibration procedure is such to correct for any temperature-dependent or time-dependent effect at the detector level, additional details must be provided about that.
All this is particularly relevant to support the statement at lines 322-323: one can imagine for example that a dependence of the detector response to the local temperature might have the negative correlation observed here, hence faking the effect sought to be studied. A possible test that shouldn't take too much effort could be a separation of the dataset between day and night.

Provide somewhere the goodness of fit test for the calibration curve of Fig. 2 (right).

It is mentioned at Lines 228-230 that the muon spectrum model is taken from three experimental papers. I think that more details are necessary here. Are these three measurements compatible between them? Were they combined (and if so, how)? Or are they covering disconnected ranges in momentum and zenith angle, such that there is no need of combining but there is a need for interpolating as explained in point B (lines 232-233)? Figure 5 only shows the interpolated model, I suggest to overlay the input data points on this plot (which would probably make my questions unnecessary).

The apparatus comprises slabs of both lead and steel. What is the purpose of having both? I assume that one is more effective against a type of background and the other against another type, but the current text is not explaining that, and it is probably not obvious.

As mentioned in Section 5 the muon flux is normalized to the value observed on Aug. 20, 2017. Are Delta H and Delta N in eq.(2-2) and Delta P and Delta N in eq.1 also the difference to the corresponding values of that day? If so, do the results change if a different reference day is chosen for everything?

In Figure 7 I suggest to overlay also the predicted flux without the barometric correction, to

quantitatively demonstrate that including the barometric correction indeed improves the compatibility between the model prediction and data.

Line 347 mentions limitations from (a) statistics and (b) modeling; it would be important to quantify their relative weight.
It is then mentioned that more statistics will be taken thanks to a larger detector, but at line 349 we are told that the size will be 4.5 m2, which is only twice the detector used for this paper. This means that the purely statistical uncertainty can only decrease by a factor of sqrt(2). Therefore, this is not going to change the game significantly: to reduce the statistical uncertainty by an order of magnitude, the detector area should increase by two orders of magnitude (the lateral size by an order of magnitude), hence it would be better to warn the reader about that. If the purpose of the 4.5 m2 detector is just to verify whether precision really scales by sqrt(2), which would be proof that statistics limitation dominates over modeling uncertainty, then better to state it explicitly.

Line 351 mentions ongoing studies with EAS Monte Carlos. Which MC programs are being used? CORSIKA?

Editorial comments:

Line 14: after “concentration of muons”, in parenthesis, we see an approximate estimate of the amount of muons in the whole muosphere, but not a concentration. That would be muons/km^3. I suggest changing “concentration” into “number”.

It is hinted at the end of the abstract that more case studies will have to be studied in the future, and the Conclusion sections ends with a similar statement. I find these statements a bit vague, I believe that it would be interesting to elaborate a bit, in the Conclusion, about which case studies may be interesting to investigate with this technique.

Lines 29-30 are an exact repetition of the beginning of the abstract.

Line 34: “usually do not”->”rarely”

Line 37: “start to”->”increasingly”; otherwise the reader unfamiliar with particle physics may think that this is an on/off type of process, instead of a continuous one.

Lines 39-40: “do not strongly interact”: I suppose this is a reference to the Strong Nuclear Force, to which muons are insensitive, but this reference will not be obvious to the reader unfamiliar with particle physics. This may be solved by adding a few words to explain that muons only feel electromagnetic and weak nuclear interactions but not strong nuclear ones, or by modifying this sentence to be more specific.

Line 46: “due to its”->”due to their” (the subject is muons, plural)

Line 50 defines the muosphere between +30 and -10 km from sea level but at line 55 the calculation uses a thickness of 35 km, instead of 40.

Line 64: repetition (“and and”).

Line 72: “surface height variations. Studying”->”surface height variations, studying”

Line 75: first occurrence of SSW acronym, must be explained.

Line 85: first occurrence of DoF acronym, must be explained.

Line 88: “decay constant”->”mean lifetime at rest”. In fact, the former is defined as the inverse of the latter. By the way, perhaps not all readers of this journal are familiar with relativistic time dilation, so I would advice to spend a few words to say that while the lifetime at rest is 2.2 microseconds, the actual lifetime as observed from the detector, and therefore the path traveled, is much longer by the relativistic time dilation factor, and that the latter depends on energy. You may also give typical ranges, to be even more informative to non-experts.

Line 98: “recent detailed review”: being from 2011, it is not so recent anymore.

Line 114: “hadronic process”: be more specific, e.g. “hadronic collisions of mesons with nuclei”; “decay process”: add “of the mesons” or rephrase somehow.

Line 127: “detail”->”detailed”

Line 166: Delta P with respect to which reference P value?

Line 174: remove “which”.

Line 196: add “estimated” before “variations”.

Line 197: explain that B is based on Equation 1.

Line 202: I suggest to expand the title: studies of what?

Line 204: after “to monitor” and after “measurement results”, explain which observables are being monitored / measured. Only much later in the section the reader is finally informed about that.

Line 216: I am not sure if “geopotential altitude” is a universally known term, I suggest giving its definition in parenthesis or in a footnote.

Line 269: racker → tracker

Line 281: remove the word “Reference”

Lines 295-296: I would remove the parenthesis around “such as positrons/electrons”

Line 296: “lead block”->”lead slab” (I presume), and also add “slab” after “stainless-steel”

Lines 296-297: remove “with a thickness of 3 cm”, as that information already appeared few words earlier.

Line 330: I suggest to express this number in percent, i.e. 0.987% (as you do at line 342 for its approximation).

Lines 375 and 380: I suppose that LOF here should be replaced by DoF?

Line 453: no better reference? According to Google Scholar, this might be arXiv:1909.01406 and PoS ICRC2019 (2020) 894.
Note by the way that clicking on the link redirects to https://0.0.1.197/

Lines 458-459: I suppose, reading the arXiv entry, that these proceedings have been then published somewhere. Please update the reference.