I thank the reviewers, Michael Prather and Joyce Penner for their comments which helped to streamline this paper and to clarify its content and purpose. The new version has been drastically shortened and goes directly to the mathematical derivation of the trade-off inequality plus its application to GWA. The other application to the equilibrium temperature changes has been retained, but shifted to an Appendix. The first section of the original discussion has been deleted and replaced by another one discussing whether ΔERF can be determined with higher precision than ERF itself (c.f. comment 3 of RC1). The other points in the discussion (choice of uncertainty distribution and consideration of all gaseous emissions in a contrail-avoidance scenario) are retained. The conclusion has been completely rewritten.

Since the discussion of the differences of the two notions "exchange rate" and "trade-off" is no longer present in the paper, the word "Exchange rate" has been removed from the title as well.

With these major changes I hope that the paper is now much clearer in its scope and up to the point.

Replies are printed in light blue.

RC1

Summary

This paper engages with an important and genuinely underexplored problem: how to rigorously assess the risk that a proposed aviation climate mitigation measure will fail to deliver a net climate benefit, given the substantial uncertainties in aviation climate metrics. The motivation for critically examining Prather et al. (2025) (hereafter PGP) is reasonable, that paper does leave a number of methodological questions open. Unfortunately, the present manuscript does not succeed in either of the two goals it appears to pursue: it neither constitutes a sufficiently focused or complete critique of PGP, nor does it develop an alternative framework that is demonstrably more sound. For these reasons, and as elaborated below, I do not think the manuscript is suitable for publication in its current form. I do not believe these issues can be resolved through revision without what would amount to a substantially new paper, and I therefore recommend rejection, while encouraging the author to consider the paths forward suggested at the end of this review.

Major Comments

1. Unclear scope and contribution

The most fundamental problem with this manuscript is that it pursues two distinct goals simultaneously, a critique of PGP and the presentation of an alternative framework, without fully achieving either. This ambiguity undermines the paper throughout. A focused critical response to PGP, limited to identifying specific mathematical or conceptual errors, would more naturally take the form of a correspondence or comment submitted to Nature, where PGP was published. The present venue and format imply that the paper is making a positive contribution beyond critique, but the alternative framework presented here is not developed to a standard that would justify publication on its own terms. The author should decide which goal is primary and restructure accordingly. As it stands, the two goals pull against each other: the critique sections are sometimes too detailed for a paper primarily presenting new methodology, while the alternative framework is not developed rigorously enough to stand as an independent contribution.

I accept this critique. It is my fault that the scope of this paper was not made sufficiently clear. It was not even clear to me when I started to get engaged in this topic. My main motivation was simply to

understand what PGP had done, why the ratio of two uncertainty distributions allows to assess a risk and what the x-axis in PGPs fig. 3 actually means. This required first to understand all the notions PGP used and then to derive the equations from scratch, that is, I had to start from a blank sheet of paper. All other points of critique are secondary, they are rather a reaction to the strong statement made in the paper title. My motivation was not to provide an alternative to PGPs risk analysis, although the misleading word "alternative" was used here and there.

Now, after a complete revision of the paper and after deleting all the marginal issues that I needed initially to get accustomed with the topic (and not to state anything wrong), the paper's scope should be clear: It is simply the mathematical derivation of PGPs method, that I (and others) missed in the original work and the check whether it works as intended (and so, indirectly, whether claims made by PGP and others are tenable).

2. **The most important weakness of PGP is not addressed**

The paper raises several legitimate concerns about PGP's mathematical presentation, but conspicuously omits what is arguably the most fundamental weakness of that paper and of the present one: the premise that a GWP-like metric provides a reliable basis for determining whether a mitigation action delivers a genuine climate benefit. GWP and GWA are time-horizon-dependent, globally and temporally averaged quantities that aggregate very different physical processes into a single number. The scientific literature contains extensive debate about whether such metrics are appropriate for decision-making, particularly when comparing long-lived forcing agents like CO_2 with short-lived ones like contrail-induced cirrus, see for example Shine et al. (2005, Climatic Change, 68, 281–302), who proposed alternatives to GWP precisely because of its limitations in comparing gases with very different atmospheric lifetimes, and Shine (2009, Climatic Change, 96, 467–472), who argued explicitly for a fundamental reappraisal of GWP as a policy metric. Fuglestvedt et al. (2010) provide a thorough overview of the metric landscape in the transport context, and Lee et al. (2021, 2023) discuss at length the difficulties of applying such metrics to aviation non-CO_2 effects specifically. By accepting the GWA framework as a valid foundation and confining the critique to mathematical implementation, this paper implicitly endorses a premise that deserves much more scrutiny. A paper positioning itself as a rigorous examination of PGP's methodology should engage seriously with this deeper issue.

I don't agree. If the mathematics is wrong, everything is wrong! The derivation is not a special one for GWA or any other metric, it is general. It only requires that the chosen metrics is additive and an extensive quantity.

I acknowledge that the choice of GWA can be criticized. Obviously, even the authors themselves are open for such critique (see the comment by M. Prather and J. Penner). I have now at least mentioned that there are these doubts, and have quoted one of the papers mentioned above, but this must suffice. My knowledge of metrics is too limited to venture an opinion on the choice PGP had made. I prefer to leave this to experts who have the insight that I don't have. Engaging in the metrics discussion (if I would be able at all to do this) would again open a new large topic aside from the mathematics, and would again lead to an unclear scope.

3. **The assumption that ERF changes are free of uncertainty is not justified**

The central methodological claim of the alternative framework is that the right-hand side of the risk inequality, the ratio of ERF changes due to a mitigation measure, can be treated as a known, certain quantity because it is determined by a concrete operational or technical decision (lines 200–202). This assumption is very difficult to justify in practice and the paper does not seriously defend it. ERF changes are not directly controlled by operators; they are emergent outcomes of operational

decisions mediated by complex atmospheric and microphysical processes. The change in contrail ERF resulting from a rerouting decision, for example, depends on ambient humidity, temperature, aircraft type, contrail microphysics, and radiative conditions — all of which carry substantial uncertainty. Further, ERF varies over the lifetime of a contrail, suggesting that even absent uncertainty, there wouldn't be a one-to-one mapping between a re-routing decision and an ERF change, although this issue could be avoided by instead considering something like lifetime-integrated ERF. Treating ΔERFs as essentially known while locating all relevant uncertainty in the impulse response function parameters is not a physically motivated choice. It is an assumption made for mathematical convenience, and one that is arguably less realistic than PGP's treatment.

This is certainly a valid point, but I stress that this is no assumption, it is rather a requirement. In response to your question – and to clarify the matter – I have written some nec comments in the section where this is derived and a new subsection in the Discussion section (S. 3.1). There I give arguments that Delta ERF should be calculable with much higher precision than ERF itself. This is because errors are more or less cancelled once ERF and ERF' are computed with the same model. An example of this is given in the paper by Burkhardt et al. (2018) on the effect on RF of a reduction of the soot number concentration in aircraft emissions.

In any case, if for any reason ∆ERF cannot be determined with sufficient precision (that is with higher precision than the ratio of the sensitivities), then there is no possibility at all to apply the initial postulate that the total metric value must get reduced in a mitigation measure for a climatological risk assessment, neither in the sensitivity form of the inequalitiy nor in the form proposed by PGP. Probably this was not clearly stated in the initial version of the paper, but now it is.

4. The contrail lifetime argument exposes an internal inconsistency

This issue is made concrete by examining what the left-hand side of equation (17) actually contains. The author claims that the relevant uncertainties reside in the parameters of the impulse response functions K_C(H), specifically the atmospheric lifetimes and decay timescales. For CO_2 this is reasonable; the multi-timescale decay of atmospheric CO_2 perturbations is a genuine and well-studied source of uncertainty. For contrail-induced cirrus, the impulse response function does include a lifetime parameter, and there is some genuine uncertainty in mean contrail lifetime. However, contrail lifetime is short enough (hours to at most a few days) that the integral K_CiC(H) is relatively insensitive to the precise value of that lifetime for any climatologically relevant time horizon H, the uncertainty in K_CiC(H) is small compared to the overall uncertainty in contrail climate forcing. This means that while the left-hand side of equation (17) carries some contrail-related uncertainty, it captures only a narrow slice of the total uncertainty associated with contrail forcing. The much larger sources of contrail climate uncertainty, including radiative properties, coverage, microphysics, and the relationship between operational decisions and actual contrail ERF, are not parameters of the impulse response function and therefore do not appear on the left-hand side at all. They must instead reside on the right-hand side, precisely where this paper claims everything is known. This is not a minor technical quibble; it is a structural problem that undermines the paper's central claim about where the relevant uncertainties lie.

Eq. 17 is the analogue of eq, 12 (original numbering), when the method is applied to GWA. If in this and analogue inequalities (in "sensitivity form" as I now call it) the sensitivity factors are precisely known, no risk assessment is necessary (as stated in the paper). Then one needs only to assure that the ratio of ERF changes is larger than the sensitivity ratio. It is another question how the ERF changes can be computed and with which precision (c.f. Q 3). If it can be assured that the ratio of ERF changes is larger than the sensitivity ratio, one may perform the mitigation, but if not, one should forego with it.

5. The conclusion undermines the case for publication

The author states in the conclusions that "the method is not mature enough to base practical decisions on" and that "the stochastic model for the uncertainties is close to arbitrary." While this intellectual honesty is appreciated, it raises a question the paper does not adequately answer: what is the positive contribution being made? If the proposed framework is acknowledged to be immature and not actionable, and if its mathematical relationship to PGP is closer than claimed, then the manuscript needs to articulate much more clearly what a reader should take away from it. Scientific papers presenting methodological frameworks that are not yet fit for use need to make a compelling case for why the framework is nonetheless a meaningful step forward. That case is not made here.

This problem is related to the one raised in comment 1, namely the unclear scope. As the paper is now completely revised and as marginal topics have been deleted, also the conclusions have been completely rewritten. Still, the paper does not offer a new method, but it demonstrates that another method that has been praised by some scientists in the field does not work as intended and that a claim made in the title of that paper is untenable. I think, it is important that people know that and know the reasons. It is good scientific practice to check the work of colleagues.

Minor Comments

- Section 2.1 (exchange rate vs. trade-off ratio): The conceptual distinction drawn here is interesting and potentially valuable, but it is developed at greater length than its ultimate role in the paper justifies. If the paper were reframed as a pure conceptual/theoretical contribution, this section could anchor it more effectively.

  I agree. But this was necessary for myself to clarify the notions. This section has been completely deleted.

- Section 3.4 (equilibrium temperature change example): This section is one of the more self-contained and readable parts of the paper. The finding that the result is good for climate as soon as the contrail ERF reduction exceeds approximately 2.63 times the CO_2 ERF increase (under the best-estimate sensitivity factors) is useful and concrete. However, the subsequent risk analysis here suffers from the same problems identified in Major Comment 4. The uncertainty in λ_ERF for contrails is very large relative to that for CO_2 precisely because contrail forcing is so uncertain, which again suggests the real uncertainty is not where the framework places it.

  Please note that this equation is a simple consequence of the mathematical derivation. It is stated several times that if Delta ERF cannot be determined with some precision (with higher precision than the sensitivity factors), the method cannot be applied. The “sensitivity form” of this trade-off and risk analysis is the only one that might work at all in contrast to the PGP method, which evidently does not work. Note that this section has been shifted to the Appendix in order not to appear as a alternative that is claimed to work better than general method.

- Section 4.3: The demonstration that NO_x emissions must be included in the trade-off calculation for contrail avoidance is a valid and useful point. This would be worth preserving in any future revised work.

  Thanks. This is one section that is retained in the revised version.

- The suggestion in the conclusion that ensemble weather forecast methods (Borella et al., 2026; Hanst et al., 2025) represent a more promising direction is well-taken, but feels

somewhat tangential to the rest of the paper. If this is the author's preferred path forward, it might be worth a separate short perspective piece rather than a brief mention at the end of a paper focused on a different methodology.

Yes, a good point and correct observation. It was indeed appended in the end to give the paper a less negative touch. It does no longer appear in the revised version since I think it is not necessary and again it blurs the focus.

Recommendation and Paths Forward

For the reasons outlined above, I recommend rejection of the manuscript in its current form. The issues identified are not, in my view, resolvable through revision without producing what would effectively be a new paper. I want to be clear that this recommendation reflects the mismatch between the paper's scope and its execution, not a judgment that the underlying questions are unimportant. Indeed, they are important, and the aviation climate community would benefit from rigorous work on them.

I would suggest the author consider two potential paths forward. First, the specific mathematical criticisms of PGP, particularly the point about uncertain quantities appearing on both sides of the risk inequality, could be developed into a focused comment or correspondence submitted to Nature. This way the comment would reach the audience most directly engaged with PGP. Second, if the goal is to develop a genuinely alternative framework, the author should begin from a more critical examination of whether GWP-like metrics are an appropriate foundation at all, resolve the ERF uncertainty problem identified above, and ensure that the uncertainty structure of the proposed framework reflects the actual physical sources of uncertainty in aviation climate science rather than mathematical convenience. That would be a substantial and valuable contribution.

The revised version is effectively a new paper. All the marginal considerations, which initially were necessary for me to understand basic notions and their interpretation and application, have been removed. Only the derivation of the trade-off inequality and its application on GWA has been retained. To my opinion, this is not something that can appear in a simple comment to Nature, since it is too long. Some detail is necessary for a thorough understanding, and the complete derivation is not just a comment, it is rather a supplement to the original paper.

I stress that the goal was never to develop an alternative framework and I am sorry that some formulations in the original version were not clear enough in this respect. In the revision I have replaced words like “alternative” with “other” or similar.

M. Prather and J. Penner

We are impressed by the effort Dr. Gierens put into his treatise dissecting the novel method we proposed to evaluate tradeoff mitigation options in a quantifiable risk framework (denoted PGP: M. *Prather, A. Gettelman & J.E. Penner (2025) Trade-offs in aviation impacts on climate favour non-CO2 mitigation, Nature, 643, 988–995, 2025)*. we are also delighted that he took up the challenge of expanding the PGP approach and questioning where it might work, or where it might go wrong. It is good to have an open discussion here. Our paper closes with an open ended discussion of the problems and possible opportunities for our quantifiable risk application. We will go through the paper linearly, responding to Dr. Gierens' interpretation of our paper and comparing that with what we had intended.

I realise that my analysis must feel to the authors like a dissection of their method, but it is, first and foremost, a derivation. The fact that this derivation ultimately discloses a serious problem for the method can be seen as a dissection; that is simply how it turned out.

L32ff: We cannot understand how PGP is ad hoc, without 'derivation', and with unjustified uncertainties. We selected what was clearly available from the recent literature and admitted that it maybe was not good enough to make decisions yet. What we provided was a pathway and a decision risk curve that was clearly defined. Others (e.g*., C. Voigt, 2025, Solving aviation's climate-action conundrum, Nature 643 (8073), 921-923*) have been able to follow the derivation and reproduce the base case.

Your paper states: "In going from a climate warming metric (GWA) to a decision tool that measures the likelihood of a positive outcome, we needed not just to provide out best estimate but a statistical view of the risk. From the individual GWA probability distributions, we built a probability distribution of the GWA(non-CO2) to GWA(CO2) ratio." Unfortunately, it was not as clear for me as for the reviewers of PGP that this is the method one usually applies in such cases. There is no derivation of this method, neither an explanation, nor a reference in your paper. If there is indeed recent literature, why didn't you quote it? It was particularly this statement that I wanted to understand.

L40ff: This discussion of tradeoff or exchange rate is semantics. GWPs are used for trading emissions, hence tradeoffs. I do not disagree with some of your approach but it is only words.

Yes, these are only words, I agree. This is a minor problem. My experience is however, that in physics and mathematics it is beneficial to be strict with the wording in order to avoid misunderstandings. I needed this discussion for my own understanding. For the revised version the corresponding section has been deleted.

L51ff: Example 2 does not really work if we accept a "social cost" of carbon (which some do, but we might not) then the cosmogenic health problem can be readily costed and considered in tradoffs with CO2.

This is just an example for a trade-off without a firm physical basis. Certainly, one can form a connection by introducing "social costs" etc., but these are subjective measures that cannot be derived from physical principles. But anyway, it has been deleted.

L84ff: "PGP assume … PDFs are log-normal". This is an incorrect reading of the paper, In PGP, the PDFs of uncertainty can be in any form as long as it can be defined. They can even change sign if that is logical. We happened to select log-normal for the example shown. This is not a problem with the methodology.

I accept that the word “assume” may indicate something incorrect. I can instead write “use”. The caption of table 1 says “Uncertainties are expressed as log-normal distributions”. I admit that “express” and “assume” are not the same words, but the semantic difference is hard to see.

I also agree that the choice of a distribution does not change the validity or invalidity of the method. I think, I have not stated that. Rather I have shown an example using normal distributions. The point is however, that in concrete applications there is an element of arbitrariness in selecting a pdf, such that the result cannot be fixed. This is the point raised in the paper.

The algebraic derivations beginning here are fine, but not applicable to the PGP general methodology. Gierens’ derivations focus entirely on carbon, which is core to the example in PGP, but the methodology can be applied to any climate tradeoff (or trading exchange rate).

I dispute that. The derivation is general, and the only requirement is that it starts from a pair of ERFs and their changes due to a mitigation measure.

L109ff: Yes, this is a good explanation of the PGP method, but veers off at L120.

Again, this was initially needed for me to understand the method. But all this has now been deleted.

L128: Yes, this is a very good point as to how these tradeoffs could be achieved in time, but that is so far outside the PGP paper. We did not go through the technology to define what CO2:contrail tradeoff was feasible.

This consideration has been deleted.

L135: The complexity of what a tradeoff includes is indeed an interesting topic. For example, with the dual-annular combustors used for CO2:NOx tradeoffs, how are contrails impacted? This is an important study, but peripheral to the PGP work. Same with L154, the tradeoff will depend on specific situations with concrete numbers, but the PGP methodology can then provide a risk curve to describe the likelihood of a successful climate outcome.

The examples are just examples, insofar peripheral, agreed. And all this has been deleted since it is peripheral and no longer needed.

L174: Sorry, NOx has two time scales (short O3 = ~1 month, and CH4 =~11 yr). But this whole discussion is irrelevant as long as we know the times scales (and their uncertainties) that are involved.

Yes, the text was incorrect. This means, Π(NOx) is a vector, not a scalar. Corrected.

L213ff: This criticism is obscure and confusing. The community has generally accepted that a change in ERF can be transformed into a change in global mean temperature. That requires a great deal of faith in how we calculate ERF from a change in the atmosphere. There really is no absolute ERF value here, only a change in its value. It is a valid question of whether /how the change in ERF (for any constituent) from a fleet operating over several years can be scaled down to one airline or even one flight. Is this linear? Maybe for CO2, but I doubt it for NOx or contrails. This is a major scientific question for which we really do not yet have an answer. Parts of this are noted in PGP in the final discussion, but it is certainly outside its scope.

Let me repeat what you write here: “The community has generally accepted that a change in ERF can be transformed into a change in global mean temperature.” Indeed, it is the *change in ERF* that counts, not the ERF itself. This is my point. I agree that it may be as difficult to determine ΔERF as it is to determine ERF (cf. RC1), but this is a problem for the practical application, not for the mathematical derivation. But I have now added in the new section 3.1 an argument why ΔERF can in

principle be determined with much smaller uncertainty than ERF itself. My point is also (mentioned in the paper): if one is not able to at least estimate the ΔERF that a mitigation action would involve, one should probably not do it.

L231ff: Gierens' math here is technically correct but introduces new variables (absolute ERF) that are unnecessary to the PGP tradeoff ratios. The problem with expanding ratios is that one can add terms that would otherwise cancel out, and these added terms induce an artificial uncertainty that should not be included. For example, the uncertainties in abs-ERF and ERF are highly correlated, and any uncertainty in abs-ERF will/should be propagated into the tradeoff ERF. The extra uncertainty in the example here, I believe is double counting.

Either the math is correct or not. ERF (or absolute ERF) is not a new variable. It is the ERF you begin with when calculating GWA or any other metric. The new variable that is absolutely necessary is rather ΔERF, and this is a variable I miss in the description of your method. Furthermore, they are not *added* but *multiplied (or divided)*. There is nothing extra or wrong in the inequality.

Unfortunately, it is not clear to me what you mean with "absolute ERF". You cannot find such an expression in my paper.

Concerning the double counting: This is exactly what you arrive with once the risk inequality is brought into a form where a ratio of metrics stands on the lhs. Then the ERF ratio appears automatically on the rhs and the problem with uncertain quantities on both sides is there. As I have shown, to avoid this double counting one needs to bring the inequality into another form where only sensitivities and ΔERF appears. I call this now the "sensitivity form" of the trade-off inequality.

This fundamental problem with uncertainty in ratios is why PGP specifically avoided any use of GWPs (e.g., GPW-CH4 = AGWP-CH4/AGWP-CO2) and stopped at the calculation of AGWP. If one wanted to do a mitigation tradeoff between CH4 and N2O, using the ratios of GWP-CH4:GWP-N2O with their individual uncertainties greatly exaggerates the uncertainty in the tradeoff because the uncertainty in AGWP-CO2 propagates into the GWP ratio. Instead we should start with AGWP-CH4 and AGWP-N2O. This problem gets even worse with contrails where a clear definition of causal emissions is not available.

I understand this argument. In the paper I add that AGWP is not additive and can therefore not be applied. The remainder of this comment refer probably to parts that are now deleted.

L333ff: The Discussion wanders off into objections of using PGP method for a single flight, or the PGP choice of uncertainty PDFs, neither of which are core to the methodology. PGP is giving an example, and only speculates into specific cases where it might be applied.

This part of the discussion has been removed completely.

The Conclusion wraps up with a long list of objections to the PGP methodology, many of which are based on a misreading of our conclusions and path forward, and others of which are based on the incorrect addition of uncertainties.

The conclusion section has been rewritten completely and does no longer touch side aspects.

Gierens has included, here and throughout the paper, a number of important, very useful insights into how the PGP methodology could be used or misused. These ideas are valuable and need to be highlighted as they begin the scientific discussion and research on the way forward.

I thank you for this positive comment and I look forward to the scientific discussion.

We support the comments of Anonymous Referee #1 (10 Jun 2026), which we just read before posting our comments, including the weakness of the PGP paper in adopting GWA as the climate metric. The methodology can and should be applied to alternative metrics to see if the risk curve between tradeoffs changes.

As the derivation demonstrates, the method does only work (or might work) if it is given the "sensitivity form" and only if the changes of ERF can be determined with sufficient precision. This may be possible if the required global model is treated carefully, with nudging of important weather variables. This is something one could try in future work.

RC2

Aviation has much discussed effects on climate through its CO2 and non-CO2 emissions. The non-CO2 effects have inherently much larger uncertainties than the CO2 effects (in terms of effective radiative forcing, ERF). Mitigation measures tend to focus on technology, operations, and fuels, and often have so-called 'tradeoffs', where the non-CO2 effects may be reduced at the expense of an increase in CO2 effect. Prather et al address this. The present paper is a critique of the Prather et al approach, and as such is an important contribution. However, the present draft has some shortcomings which could be successfully addressed.

This is an important paper in the debate over operational mitigation approaches for aviation. Thus, some extra effort bringing some clarity, particularly to the style of presentation is worthwhile.

Abstract

The tone is quite critical of the original work and could be phased more equitably and factually. For example, "…leaves, to this author's opinion, many issues open for discussion" could be " leaves some issues open for discussion' – indeed the original authors (Prather et al; hereafter PGP25) highlighted this. "Vague" could be "less clear". The sentence "Unfortunately…" could be "The method proposed by PGP25 may have some shortcomings that are addressed with derivations presented here" (omit the next sentence). Remove "unfounded" "…suffer from a degree of arbitrariness" could be "..are highly dependent on the assumptions made, some of which are subjective or 'user choices'" The last sentence needs to be put in context of what the present paper found in a more positive way (rather than the abstract reading as a list of criticisms of PGP25). At the moment, the last sentence is uninterpretable. Indeed, what this relates to is a rather passing comment at the end of the paper, which seems incomplete.

The abstract is currently letting a good paper down significantly and needs a complete re-think and re-write. Instead of focusing on the PGP25 problems and rather critical statements, the author should focus on some of the significant findings and interpretation that are present in the discussion.

I agree and admit that I was indeed quite critical. As the paper has been revised completely, also the abstract has been reformulated. The paper now just gives a "supplement" to PGPs one, namely with the necessary derivation of the trade-off inequality and its application to GWA. Side aspects have been deleted from the paper.

Introduction

L10, the referencing is extensive but the outline non-specific. It would be helpful so say something like "a number of mitigation strategies have been proposed that require developments in technology, operations, and fuel type".

Yes indeed, good idea! Thanks for the suggestion.

L14 refers to tradeoffs, but is non-explicit. Many mitigation strategies do indeed imply compromises or tradeoffs, but the main problem lies in tradeoffs between short-lived climate forcers and CO2. This is the nub of the problem. There are other SLCF-SLCF tradeoffs but these are usually technological and of lesser consequence. It would be helpful to the reader to get to the core of the problem that is being addressed by PGP25 and the present paper. L15-21 identify the problem but not the critical issue (SLCF vs CO2) but omits to mention that the outcome of 'positive for climate' is potentially one that could even change sign over time.

I thank you for this clarification. I have added a few sentences into the introductory text such that these difficulties are now explicitly mentioned.

L27 introduces the PGP25 metric but does not illuminate. Is it an AGWP per activity? In which case it is worthwhile to explain what the "Global Warming" is, since AGWP itself is a misnomer – it is simply an integrated RF over a user-chosen time horizon.

I agree. A short explanation of its definition has been added following the explanation of PGP.

L32 "ad hoc, without derivation" is overly harsh. Please rephrase. Similarly, for the next two sentences, "not convincingly justified", "vague" could be rephrased.

I agree. The sentence has been reformulated.

Critical review of the method of PGP

L41 I find the critique of PGP25 here entirely justified and the PGP25 sentence quoted rather muddles the history and usage of CO2 equivalence metrics. A tradeoff is an extension of the original intended usage of GWPs, as an exchange rate mechanism but has rather different implications. L:48 Example 1 is rather strange, going to the length of equating an individual source of CH4, and is not really illustrating successfully the equivalence/exchange rate mechanism originally intended. As originally conceived, GWP is a reporting convenience that recognizes different GHGs (and balances of a country, or sector) that by necessity requires an exchange rate mechanism, since the end effect of releasing (e.g.) 1 kg of CH4 is not the same as 1 kg of CO2. Also, by necessity, a time horizon (TH) needs to be specified, a compromise and subjective user choice that has been extensively debated in the literature since the concept of GWP was introduced in the late 1980s. Example 2 is correct, but seems unnecessarily off topic, and could be restated as the SLCF vs CO2 tradeoff, which has already been stated for NOx/fuel (and thus CO2), earlier. Example 3, is more to the point but identical in principle to Example 2.

The purpose of the examples was foremost to clarify my understanding of these notions, and to clarify that exchange rates and trade-offs are different things. This section has been deleted, since it blurs the focus of the paper, as RC1 rightly stated.

L60, this is an oversimplification in that it omits the aspect of TH. TH has the potential to change the 'outcome', and should at least be acknowledged here.

This may be, but as the text has been deleted nothing needs to be added here.

It should also be noted that three are, as PGP25 noted, other CO2 equivalence metrics, notably the Global Temperature change Potential (GTP, Shine et al., 2005) and an alternative derivation of the GWP, the GWP* (Allen et al., 2015). These at least need to be mentioned as context somewhere (not necessarily in detail), earlier in the introduction of the concept and usage of CO2 equivalence emissions.

I am not sure whether an excursion to other metrics would not again blur the focus of the paper. As you can see, the derivation is a general one for any metric that is additive and extensive. And if the final inequality is written in the "sensitivity form", the metric itself does no longer appear.

L67 "impossible", at this point in the reading of the m/s, I would judge this to be incorrect (although the reader is directed forwards). Surely, the risk-analysis method will contain intrinsic uncertainties, and required some degree of user discretion and caution, but "impossible"?

I agree, but this sentence has been removed together with the whole section.

2.2 Tradeoff risk

L78 – 80. This is a poor characterization of a particular GWP, in this case CH4. It is not a “best guess”, it is the best estimate of a ratio of an integrated RF from a pulse emission of CH4 divided by the same mass pulse emission of CO2. Both nominator and denominator have uncertainties, and the calculation will have assumptions of, e.g. background CO2 concentrations and TH selected. Dismissing it as a “best guess” is incorrect. The last part of L80 is correct in that the intrinsic uncertainties will make the decisions uncertain (in the sense of the efficacy of those decisions).

As a non-expert in the topic of metrics I accept this critique. It does not apply to the revised version since this section has been deleted.

L98 “..if the GWA ratio is large…” I do not understand this. Surely, the success of the outcome is not simply dependent on the size of the ratio but the size of the uncertainties, particularly, in the nominator?

The problem you encountered here shows that this method is not as easy to understand as the original Nature paper seems to assume. But my expression is correct. The GWA ratio is an uncertain quantity and its uncertainty or our lack of knowledge can be described using a statistical distribution, in this case a ratio distribution (F(x), which of course depends strongly on the uncertainty distribution of the nominator). The expression “…if the GWA ratio is large…” simply means: you assume a certain value x (in this case a large one) and ask for the probability, that the GWA ratio is equal to or exceeds x. Although there is an uncertainty distribution, in concrete situations one asks whether the uncertain quantity has a value x±dx or whether it is equal to or exceeds x.

Please note that this section has been deleted.

Finally, I note the PGP have themselves provided comments which are generally very helpful and I would recommend that the author consider these comments.

Agreed, but as far as this recommendation refers to the original section 2.2, it does no longer apply.

L156, for clarity and accuracy, please refer to CO2 equivalent emission metrics, not ‘climate metrics’ (see IPCC 1r6 glossary for what a ‘climate metric’ is generally agreed to be).

Thank you for this clarification. I have changed the text as suggested.

L213 I don’t understand this assertion. Yes, it is the ‘deltas’ that matter, but to some degree these will be dependent on the size of the absolutes, surely? For example, if the global ERF of contrail cirrus were 5 mW/m2, rather than 50 mW/m2, then that would affect the decision as to whether a mitigation measure was worthwhile. Hence, at this point, I disagree with the assertion and agree with that of PGP25 on “pressing need”.

No. Even if the Deltas would somehow be correlated to the basic ERFs, it is only the Deltas that appear in the trade-off inequality. The question whether the mentioned correlation exists and what consequences this might have, is another one. Probably a large basic ERF would allow a larger lever to change it, while a small ERF would perhaps rather lead to the decision that there is not much to gain from a change. I think also, that my last sentence remains correct even if a kind of correlation between Delta ERF and ERF is present. The statement of PGP mentions “most pressing need” and “firmly establish”. These are strong words, and to my opinion this is not justified here, even if there is some of the mentioned correlation. Nevertheless, I have reformulated the sentence: In light of this result, PGPs postulate …… appears as an exaggeration

L350 Bickel et al. (2025) are quite clear that an ERF for a single flight cannot be computed from a global computation, nor a local climate sensitivity parameter.

Yes, rightly. But this part of the discussion has been deleted. It leads too far astray from the central theme, the derivation and its consequences.

L364-367. Absolutely. This is the core finding. This has been potentially misrepresented or misunderstood by Voigt companion commentary to the PGP25 paper.

Certainly, and I thank you for this comment. However, for the sake of a clear scope of the paper, I had to give up this section.

L404 Commendable – a widely ignored point. In general, there is a likelihood that an upwards vertical displacement for the purposes of avoidance will increase aviation induced ozone, and vice versa. Again, but for different reasons, the ERF of an individual flight is virtually impossible to compute robustly with chemistry-climate model. Similarly, the direct forcing of water vapour may operate similarly for altitudinal shifts.

Yes, I agree.

L441 – L444. While agreeing completely, the author has not at any point dug into the practical difficulties that ERFs are only correctly computed with GCMs. Global integrations of plume calculations to derive a global RF, which is then corrected to an independent ERF/RF ratio are (as pointed out here and by Bickel et al. 2025) not valid. Avoidance calculations are carried out with plume calculations, at present, with mostly the single CoCIP model. The RF change of a flight is therefore not necessarily consistent with the global ERF (actually two different metrics), and an ERF of a single flight with a plume model without a coupled water vapour/energy budget, using a globalized adjustment is invalid.

I agree that ERF changes cannot be determined for single flights. Neither Δ ERF can be determined for a single flight. But the question of whether this method is applicable for single flights is no longer touched upon in the paper, and the conclusion section has been completely rewritten. This was done to achieve a clear scope of the paper, which referee RC1 had quite rightly pointed out.

L472 “hardly” could be replaces with “not” on the basis of the arguments presented.

Yes, indeed. But, as mentioned, this discussion does no longer appear.

L479 – L486 I understand what the author is alluding to, but I think that this is confusing issues. Presumably “weather forecasts” refers to accurate prediction of ice-supersaturation (for the formation of the individual persistent contrail in itself). PGP25’s work, seems to assume perfect prediction of contrail formation and persistence and deals with the tradeoffs of delta ERFs. I think that this point would be better dropped into the argument earlier, when summarizing the assumptions of the PGP25 work, rather than alluding to alternative methods later.

This final allusion was included since the paper otherwise is overly negative. I wanted to at least show that positive developments are possible and are in progress. However, yes, the addition is a bit unconnected to the rest of the paper. It has been deleted together with the large majority of the former conclusions.

Final comments

I think in revising the paper, a better summary of what PGP25 are saying could be provided (see their comments on the present paper), rather than the sole focus on the mathematics of their risk calculations. They do indeed outline reservations over the usefulness and make some

recommendations. Unfortunately, Nature-stable publications like “headlines” and the well-intentioned caveats of authors can get lost.

You are right, the strong headline of the paper triggered me to be strong as well. This might explain some harsh statements. Since the revised version no longer touches upon topics that I had to think through in the first version, many of my critical comments have been removed. The paper is now actually merely a supplement to PGPs paper that delivers the mathematical derivation which one needs to judge whether the method is applicable or not as it has been presented. In the original form, it is not, but it might work in the “sensitivity form”. This is an open issue and can trigger interesting discussions and studies, something that M. Prather and J. Penner have suggested in their comments.

Style – the paper is extensively written in the first person, e.g. “to my feeling” L426, “To my view” L479, which I think is better phrased more passively voiced

The passive style is now used in the revised version.