PEER REVIEW REPORT

PLANETARY RESEARCH

This peer review report contains the editor decisions, reviews, and author rebuttals for each round of peer review. The peer review rounds are ordered chronologically, and the final acceptance decision is not included.

The published manuscript can be found on the journal website: https://planetary-research.org.

TITLE

The theory of cratering statistics: Significant corrections to existing methods

AUTHORS

S. Bell and S. J. Robbins

DOI

https://doi.org/10.53480/46d8n314

EDITOR DECISION
round 1

Editor decision: Round 1

Dear Samuel Bell,

I would like to thank you for submitting your manuscript "The theory of cratering statistics: Significant corrections to existing methods" (#235) to Planetary Research. I have now received the reviews of your submission, which are included at the bottom of this email. Based on the reviewer comments and my own assessment of your work, your manuscript should in principle be suitable for publication after revision.

The reviewers have not identified any major flaws, and they mostly provide a list of detailed comments that should not be too difficult to respond to. In addition to their comments, I have two editorial points that I would like you to address.

First, one of the reviewers points to a problem in tone when referring to previous work. In this regard, I agree that some of the text can be a bit off-putting, and perhaps exaggerated at points. The offending text is mostly found in the earliest portion of the manuscript (summary and abstract), but unfortunately, this might color how a reader sees the remainder of the manuscript.

Second, I note that the text is a bit long, and that this might make it difficult for some readers to appreciate the novelty of this work. I do appreciate the mathematical completeness, and also that the more technical details have been relegated to appendices. It is possible that this manuscript could be split into two separate works, but I do not see a trivial way of doing this. As an example, one could perhaps have a theoretical paper (and move some of the appendices to the main text), and then have a second paper that demonstrated the approach with examples. I will not insist that you do this. Alternatively, it might be possible to just better describe the outline of the paper and add a couple extra sentences between sections to better guide the reader. I would appreciate that you at least reflect on this issue and also ensure that the title, summary, and abstract accurately reflect the totality of the work presented in the manuscript.

Please provide us with a point-by-point response to all comments and identify what changes you have made to the revised manuscript. In addition to the rebuttal, we ask that you also provide a document that tracks all changes that have been made to your manuscript since its initial submission.

When you have completed your revisions, you can upload the revised manuscript, your response and a tracked changes document from your submission dashboard.

If you have any questions about the recommendations in your review, or if you are unsure how to respond or proceed, please do not hesitate to contact me from your submission dashboard.

Sincerely,

Mark Wieczorek
Editor in chief, Planetary Research

The following comments were received from the reviewers.

Reviewer 1:
Recommendation: Revisions Required With Review

All pre-existing data used in the manuscript are publicly available and sufficient information is provided so that they can be found and accessed.

Yes

All significant data products generated as part of the study have been uploaded to a public repository (or will be at the time of acceptance) and are cited using a persistent identifier (such as a DOI).

Not applicable

All significant numerical codes developed as part of the study have been uploaded to a public repository (or will be at the time of acceptance) and are cited using a persistent identifier (such as a DOI or SWHID).

Yes

All pre-existing software used as part of the study are acknowledged, and sufficient information is provided to find, access, and understand the software. If the software is restricted, the conditions under which it can be accessed are publicly available.

No

If you answered no to any of the above questions, please provide supporting information.

The error they identified is from my code: https://github.com/ggmichael/craterstats
Whether that counts as 'used' here is not clear.

I am uploading an annotated version of the manuscript.

No

Comments for the authors

The theory of cratering statistics: Significant corrections to existing methods
Bell and Robbins
Review

The main components of this manuscript are:
1) Correction to implementation of probability calculation from Michael et al (2016)
2) Error modelling: counting and diameter errors, area error, d_min selection error. Other sources of error.
3) Measuring SFD slope
4) SASH model
5) Application of SASH: Saturnian system and lunar farside

There's a lot of material in this manuscript and it takes some time to form a coherent impression of it: not least because of its considerable length, but also because the distinct topics it covers were not fully anticipated from the title, summary and abstract. Here, I first give my general comments on the work, then summarise my impression of it. At the end I make line by line comments.

1. First of all, about the error they found in the code implementation of our 2016 paper: I have verified it, and agree that there should be a dC/dt factor introduced at the change of variable. I thank the authors for pointing this out, and acknowledge them for the correction.

2. They suggest (114-117) that we were unable to construct an error model for lack of the maths. That's neither the case nor an appropriate conjecture to make. The reason I have focussed on the Poisson statistics over several works is because that is where I think the interesting behaviour is. Small number statistics were the bane of the binning/fitting approaches, and a probabilistic approach has given us the possibility to say something real about some actual features of interest. While building an error model is a rational thing to do, I'm not expecting a lot of new insights into surface history from it, although I'll be happy if you show some. Given that we agree that there are possible 100% or greater errors hidden in the μ of the chronology calibration, the mapping from crater densities to absolute ages is already only notional. We know the endpoints of the relationship, and know the function has a behaviour in between, which probably looks somewhat like the chronology we use, but may differ substantially from it. What has been valuable in crater dating is the sequencing of related surfaces, and the technique has proven to work excellently for this purpose. We have the epoch systems for the Moon or Mars or other bodies, and crater dating places surfaces correctly into and within these systems. If the epoch boundaries turn out to be wrongly placed in time – as surely they must to some extent – of course it will change the bigger picture: how changes in surface processes relate to the internal physics of the planets. But I don't expect that understanding to come from crater dating, even with a more descriptive error model. It will come from samples. The established sequencing of the surface features will remain unchanged even in the light of such a revision.

The claim that "the math had yet to be developed" (117) suggests that the error model maths here is novel, but the compounding of PDFs is not new. The authors have written a new piece of code to implement it.

3. An additional point for consideration in connection with an error model is that, within a single author study, many of the modelled measurement errors will have correlations which act to reduce the spread of the error in compounding when several measurements are compared. When your interest is local sequencing, this is an important effect which is not addressed in your text.

Furthermore, if there is any directional bias in the application of a component of the error model, e.g. the d_min error, this needs to be similarly included into the measurements used to derive the chronology model itself, where it would have an opposing effect. If this is not done, the consequence is certainly a degradation of the result compared to doing nothing.

4. What you are calculating is the error on the true age for each measurement (Section 2.5). A reasonable question to ask is: what are you going to do with this? Questions about when you might

expect changes in the surface expression of processes connected to the physics of a planet are never likely to be connected to a single crater dating measurement. Instead, they will be related to the chronology system as a whole – in particular, its epoch boundaries in our present level of understanding - and it's at that point that it is necessary to bring in the reasoning of your error model. For an individual measurement, I don't see what more you will learn from seeing this spread distribution than you do from the µ-function that already represents it: simply, that this uncertainty is present and is large. You seem to recognise this in suggesting showing an additional age without the chronology error (382-384). If you are showing both, it would be pertinent to describe the occasions when you can expect to understand something additional from the chronology-error-inclusive result.

Neukum made some chronology system error estimates as you've said, but I'd suggest that these only hold under the assumption that the connections made between samples and their source regions are broadly correct. What quantification do you give to the possibility that some of them may be wrong? If you discount this possibility, is the error model providing a false confidence that we understand all the uncertainties? Is it still necessary to caveat the result with an acknowledgement that the error could be greater? If so, what do we learn from this exercise?

5. The title does not reflect the content. If the paper covered component 1 only, you could argue for it (in the singular). But the substance of the work is really about the error model and the SASH method. The error model may be considered an addition to existing methods. The SASH method appears to be a valuable approach to assessing the shape of the CSFD. Neither of these are corrections. The summary inexplicably mentions neither the error model nor SASH.

6. The calculation used for the age PDF is exactly that described in Michael et al 2016. The change of variable correction relates to the code implementation of that publication. This work goes to great lengths to present its material as an overturn of previous incorrect work, when in truth it adopts an existing technique in its entirety with a single correction. It has taken you ten years to come to the realisation that my proof was correct (142-145), and you cannot proceed without it; neither had a probabilistic description of the problem been considered in this context before that work. At a minimum, the tone of this manuscript is lacking in respect for the prior work without which the error model could not have been constructed in the way it has; inasmuch as it appears purposefully constructed to lead readers into the mistaken impression that an alternative method for determining age was innovated in this work, it could be considered unethical.

If the work is to remain with its present structure, it should be retitled referring to the error model, the SASH algorithm, and the slope analysis which are the actual innovations here. Certainly describe the change of variable correction, but remove the preachy "let us teach you how random variables are supposed to be" material as well as the claim that random variable maths is a novel development; remove the 'substantial flaws' aspersions repeated throughout the manuscript in reference to the source work that has been reproduced, and describe the new contributions properly in the context of what already exists, including the compounding of random variables.
In summary, the manuscript has severe problems of presentation which make it unacceptable for

publication. The tone of the article, which aims to discredit a previous work while appropriating its advances is not ethical. The choice to use a title and summary not covering the actual content but to promote this aim is not adequate.

The work is structured more like a thesis than a research article. It brings together several chapters on topics which are distinct, although with an over-arching connection. It is not possible to give proper weight of consideration in this review to results on the crater population of the Saturnian system or the lunar farside within what is presented as a work on methods. The SASH algorithm is likewise a distinct contribution from the error model which ought to be evaluated separately: again, there are important details which are beyond what I can evaluate in this review. Thus, the majority of my assessment here relates to the error model.

Nevertheless, an error model of the type presented has not been constructed for crater dating before and is an original contribution. The application of the ASH algorithm is also new, and potentially a valuable tool in the understanding of measured crater size distributions. The work over-claims novelty for the treatment of random variables: the maths is not new, but the authors have produced a new code implementation which they use for their evaluations in the present application.
Greg Michael

116-117 That's not the reason (see above). It is unacceptable to attribute an imagined explanation in a scientific paper.
149-150 "We do not need to adjust for how the production function varies with diameter." I understand what you want to say, but don't think that's good wording – the PF remains implicitly in the expected number of craters.
151-154 Yes, but I already wrote this 10 years ago: "In all cases, the PDF description of the model prediction should be considered superior to results obtained by binning/fitting" (Michael et al 2016).
145-148 I'm glad you are finally convinced.
Section 2.2 – 2.3
Estimating appropriate parameters for new situations and terrains is what makes using an error model difficult in practice (235-236). My opinion is that applying standard parameters (as those you suggest in 231-233, and 247-250) is similar to being generally aware of the magnitude of this type of error, as was established in your 2014 paper. Rather than recommend default assumptions, it would be preferable to recommend a practicable method to measure this fresh for each scenario.
256-263 I agree - this is an interesting observation. You don't say it explicitly, but because of 148-149, it's only the diameter errors which cross the interval boundary which make a difference. And at that, it's d_min which dominates. I realise you recognise this in the discussion, though.
Fig 2. I recall the prominent Chapman talk at the Robbins 2015 workshop "Issues in Crater Studies and the Dating of Planetary Surfaces", which ridiculed the use of three significant figures for ages. Here you are showing four. In the context of the magnitude of errors you discuss elsewhere, you should comment on the utility of this choice.
381-385 I agree with this.
396 Please give an example of a non-contrived 'highly unstable area' and its error measured in

CraterTools. Also mention ArcGIS first in your criticism, since this universally-used tool is the source of the polygon measurements for CraterTools, to avoid giving an impression of bias against particular authors.

Section 2.7

I agree this is worth doing. I would characterise the procedure rather as picking a diameter interval which is consistent with the slope of the PF, rather than just excluding the rollover. That seems to be slightly different from your implementation.

Section 2.8

There's nothing to argue with here until you come to quantify them all together, including unknown unknowns, at 10% (line 475). That's a remarkable insight. Why this number and not any other? I suggest you elaborate on how you chose it. It is better to have a result which we understand may be wrong for various unquantifiable reasons than one which contains arbitrary adjustments. Otherwise, the error model ends up as merely an exercise in formalism.

517-520 It expresses an estimation of the error over the counting statistics. Whether it's a significant improvement depends on what you can interpret from such a result in comparison to what was previously possible. You have not addressed this topic, nor demonstrated a greater interpretive value, so I don't think you can yet make this claim.

633-637 If your wording were consistently applied, this would be described both as an 'incorrect method' and a 'substantial flaw'. Please address this.

722-729 This was already said in Michael et al 2016, p283.

Section 4.3

Use of the ASH algorithm/SASH variant appears to be a promising approach for understanding crater size distributions.

882 …or even marginally exceeds the diameter of the body!

1161-1168 The calibration point was constructed using a counting-error-only measurement, and therefore the 1.70 Ga is showing a substantially different result from the original. There's more to worry about with this count than you acknowledge here: the additional error-smearing doesn't rectify the problem.

1174-1175 This is excessively grandiose. The maths of compound probabilities has always been available. More correctly, you are the first who have tried to apply this to crater dating.

Section 5.4

Although this section is an application of the methods of the paper, it would probably have been better as a follow-up paper: the results are distinct from the remainder of the work, and are given no mention in either the title or abstract. It is difficult to assess them within the same context as the rest but this, of course, is the type of application where the SASH approach could be beneficial.

1414-1416 "we have laid out the fundamental math behind the Age Problem" This is incorrect. You have described an error model making use of known methods for compounding PDFs on top of a derivation that you adopted from Michael et al (2016). Please rectify.

1418 "can be solved without plots" - you still absolutely need the proviso that you gave earlier (152-154), otherwise this is wrong.

Reviewer 2:
Recommendation: Revisions Required Without Review

All pre-existing data used in the manuscript are publicly available and sufficient information is provided so that they can be found and accessed.

Yes

All significant data products generated as part of the study have been uploaded to a public repository (or will be at the time of acceptance) and are cited using a persistent identifier (such as a DOI).

Yes

All significant numerical codes developed as part of the study have been uploaded to a public repository (or will be at the time of acceptance) and are cited using a persistent identifier (such as a DOI or SWHID).

Yes

All pre-existing software used as part of the study are acknowledged, and sufficient information is provided to find, access, and understand the software. If the software is restricted, the conditions under which it can be accessed are publicly available.

Yes

If you answered no to any of the above questions, please provide supporting information.

I am uploading an annotated version of the manuscript.

No

Comments for the authors

The authors develop a set of new tools used to analyze crater statistics with the aim of correcting some of the problems inherent in the standard techniques that have been widely employed to analyze crater counts and infer ages of planetary surfaces. The paper goes into great depth and detail regarding the current methods and their shortcomings, and not only demonstrates how we can do a better job of inferring ages and age uncertainties using more robust statistical methods, but also providing a comprehensive set of tools to do it with the ‘cratersfd’ Python package.

This is a *beautiful* paper. The field of planetary sciences has been desperately in need of a revision to these old techniques for quite some time now. Prior to reviewing this paper, I had been planning to develop a set of crater statistical analysis tools as part of my project, but now I don’t have to, as Bell and Robbins have done a far, far better job than I could ever hope to.

I have no major critiques of the paper, only a handful of very minor suggestions. I do not need to review the manuscript again, and I do not need to be anonymous.
-David Minton, Purdue University

Line 122: The concept that craters could be analyzed as a Poisson point process goes back much further than 2010. The earliest I am aware of is by D.W.G Arthur 1954 in the journal of the British Astronomical Association (cited as the first by A. Marcus in 1966 https://doi.org/10.1016/0019-1035(66)90024-8, a somewhat forgotten paper that had a similar aim as the current authors to challenge the underlying assumptions of statistical models that were in use at the time for answering the "impacts vs. volcanoes" lunar crater origin hypotheses).

[multiple places]: Whenever I review a paper that uses the terms "saturation," or its variants ("saturation equilibrium", "equilibrium saturation", etc.) I am compelled to point out that Gault (1970) gave us two perfectly useful definitions for the terms "saturation" and "equilibrium," which mean two very different things. Saturation is a purely mathematical construct that describes a surface entirely filled with circles. Melosh (1989) prepended the term "geometric" to saturation to drive this point home, and its derivation has recently been revisited by Hirabayashi et al. (2024) (https://doi.org/10.3847/PSJ/ad8883). It is meant to serve as a reference against which to compare crater number densities to. In contrast, equilibrium is the observed maximum crater surface number density seen on planetary surfaces, and it results from the competition between new crater formation and the accompanying topographic degradation that inevitably accompanies the violent formation of each new crater (for more detail, see sec. 1.1 of Minton et al. (2019))

What the authors refer to when they use the term "saturation" is actually equilibrium. For instance, on line 222:
"... Ostrach (2013) found that these effects continue to fade as D increases past the equilibrium saturation rollover into unsaturated diameters."
Should instead be something like:
"... Ostrach (2013) found that these effects continue to fade as D increases past the equilibrium rollover into diameters that have not reached equilibrium."

I would please ask that the authors adopt the definitions of "equilibrium" and "geometric saturation" as I've described here. To me this is quite important to do in this paper, as I foresee this paper having a major influence on how we approach the science of crater statistics.

Line 454:
"Rays of discontinuous ejecta drive erosion of other craters beyond the extent of the continuous ejecta (Minton et al., 2019). Although these effects may not dominate when the production function is steeper than ~-2, they do clearly sometimes happen. However, because these effects have yet to be quantified empirically, it remains unclear precisely how significant they are."

This is a bit of a muddled summary of what we found with regard to rays in our 2019 study. First of all, in that paper we were exclusively focussed on the regime where the production function is steeper than

-2, so our conclusion with regard to rays is about steep-sloped SFDs, not shallow ones. What we found is that the observed magnitude of the equilibrium CSFD of steep-sloped craters (i.e. ~few percent geometric saturation across all sizes) requires that distal ejecta is the dominant driver of topographic degradation of small craters. Furthermore, it is the apparent *stability* of the observed magnitude of the equilibrium CSFD of steep-sloped craters (i.e. that it seems to be the same ~few percent geometric saturation whenever it is measured, all else being equal) that requires that the distal ejecta be spatially heterogeneous, as uniform homogeneous degradation leads to major fluctuations in the observed crater number density in our simulations, and “rays” seem to suppress them.

While it is correct that this effect has yet to be quantified empirically (though we are attempting such quantification right now, actually) it’s emphatically not an issue for shallow-sloped production functions.

Line 1493: There seems to be a repeat of the “Data and code availability” section. The first section discusses code, and provides both a GitHub repository and a working Zenodo link with a DOI. However, the second section provides a Zenodo link for figure files and datasets that does not work. From what I can tell, the GitHub repo (and working Zenodo link) provides what is necessary to reproduce the figures (but not the figures themselves) so perhaps this is just a manuscript versioning error.

AUTHOR REBUTTAL
round 1

Dear Samuel Bell,

I would like to thank you for submitting your manuscript "The theory of cratering statistics: Significant corrections to existing methods" (#235) to Planetary Research. I have now received the reviews of your submission, which are included at the bottom of this email. Based on the reviewer comments and my own assessment of your work, your manuscript should in principle be suitable for publication after revision.

We thank you for this assessment and for your work reviewing our paper.

The reviewers have not identified any major flaws, and they mostly provide a list of detailed comments that should not be too difficult to respond to. In addition to their comments, I have two editorial points that I would like you to address.

First, one of the reviewers points to a problem in tone when referring to previous work. In this regard, I agree that some of the text can be a bit off-putting, and perhaps exaggerated at points. The offending text is mostly found in the earliest portion of the manuscript (summary and abstract), but unfortunately, this might color how a reader sees the remainder of the manuscript.

We have made edits to the title, summary, and abstract to accommodate these concerns. In our edits, we have shifted the emphasis towards the improvements we have developed and away from the problems we have found in previous work.

Second, I note that the text is a bit long, and that this might make it difficult for some readers to appreciate the novelty of this work. I do appreciate the mathematical completeness, and also that the more technical details have been relegated to appendices. It is possible that this manuscript could be split into two separate works, but I do not see a trivial way of doing this. As an example, one could perhaps have a theoretical paper (and move some of the appendices to the main text), and then have a second paper that demonstrated the approach with examples. I will not insist that you do this. Alternatively, it might be possible to just better describe the outline of the paper and add a couple extra sentences between sections to better guide the reader. I would appreciate that you at least reflect on this issue and also ensure that the title, summary, and abstract accurately reflect the totality of the work presented in the manuscript.

We begin by acknowledging the issue. The paper *is* long. We did consider several potential approaches to dividing it into two or more separate publications, but we ultimately concluded that one single publication was the best way forward for the way we hope our work to be used. Were we focused on maximizing our publication count—something we, of course, have professional incentives to do—we could have separated it into as many as four papers: the Age Problem, the Slope Problem, the Plotting Problem, and the worked example of the cratered plains of the moons of Saturn.

However, there were several reasons why we chose a single paper. To begin with, we want this work to provide a comprehensive update to cratering statistics, and we want readers to be able to turn to it to provide a comprehensive reference. We also wished to produce a single comprehensive code, and it would have been awkward to have multiple papers presenting one code. Because each section focuses on the math of probability distributions, each section uses similar fundamental math that would require duplication if split into multiple papers. Furthermore, the sections rely on each other. Additionally, we felt that the theoretical concepts would have been difficult to understand without examples, and each example uses methods from both the Age Problem and Slope Problem sections. Had we first presented a theoretical paper that incorporated material in the appendices, with a follow-up paper for examples, we also worry that it would have, by nature, front-loaded the reader with the meatier mathematical text, making it more difficult for the reader to follow.

For these reasons, we thank you for not requiring us to sever the paper.

In addition to editing the title, summary, and abstract for tone, we added a sentence to the abstract on the worked examples so that it more fully captures the totality of our work. We also added the words "for ages, slopes, and plots" to the title to be more descriptive.

To provide a better outline, we have modified the paragraph at the end of Section 1 outlining the rest of the paper to more clearly state which section discusses the Age Problem, the Slope Problem, and the Plotting Problem. For the Age Problem, Slope Problem, Plotting Problem, and Examples sections, we have added a few sentences to provide an outline for those sections. We also moved the section on applying the inverse chronology function to make more sense in the narrative flow. Finally, we have added transition sentences to improve the narrative flow between the sections.

Ironically, we had excluded much of this language in order to keep the paper shorter, but we agree that including it improves readability.

Please provide us with a point-by-point response to all comments and identify what changes you have made to the revised manuscript. In addition to the rebuttal, we ask that you also provide a document that tracks all changes that have been made to your manuscript since its initial submission.

When you have completed your revisions, you can upload the revised manuscript, your response and a tracked changes document [from your submission dashboard].

If you have any questions about the recommendations in your review, or if you are unsure how to respond or proceed, please do not hesitate to contact me from your submission dashboard.

Sincerely,

Mark Wieczorek
Editor in chief, Planetary Research

The following comments were received from the reviewers.

Reviewer 1:
Recommendation: Revisions Required With Review

All pre-existing data used in the manuscript are publicly available and sufficient information is provided so that they can be found and accessed.

Yes

All significant data products generated as part of the study have been uploaded to a public repository (or will be at the time of acceptance) and are cited using a persistent identifier (such as a DOI).

Not applicable

All significant numerical codes developed as part of the study have been uploaded to a public repository (or will be at the time of acceptance) and are cited using a persistent identifier (such as a DOI or SWHID).

Yes

All pre-existing software used as part of the study are acknowledged, and sufficient information is provided to find, access, and understand the software. If the software is restricted, the conditions under which it can be accessed are publicly available.

No

If you answered no to any of the above questions, please provide supporting information.

The error they identified is from my code: https://github.com/ggmichael/craterstats
Whether that counts as 'used' here is not clear.

As we described below, we now feel that this was less of an error and more of a problematic assumption. This issue was present in the text of Michael et al. (2016) as well as the code. In Figure 3, we show an example of what the age PDF would look like when the *N*(1) PDF is scaled by the inverse chronology function $\mathcal{C}^{-1}$ instead of having $\mathcal{C}^{-1}$ applied to it with random variable math. Figure 3 was not generated using the CraterStats code that implements Michael et al. (2016). We made it using the .scale() method of the

RandomVariable object class in *cratersfd* (the new code released with this manuscript). Here is the code used to generate that part of Figure 3:

```
(lambda_pdf(N) / area).scale(ncf_inv).plot(
    kind='median', label=True,
    pdf_label='Scaled PDF', upshift=-1.1,
    xlim=[1, 3.7], color='red', unit='Ga'
)
```

As with every figure in the paper, we made the Jupyter notebook used to produce it public as part of the GitHub repo. The Jupyter notebook for Figure 3 can be accessed here.

Once we realized that the text of Michael et al. (2016) described a method that scaled the $N(1)$ PDF by $\mathcal{C}^{-1}$ instead of applying $\mathcal{C}^{-1}$ to the $N(1)$ PDF, we checked the CraterStats code to verify that the code contained the same issue as the paper, which it did. However, in the paper, we felt that separately mentioning that the same issue present in the text of Michael et al. (2016) was also present in the CraterStats code would have conveyed an unnecessarily critical tone.

I am uploading an annotated version of the manuscript.

No

Comments for the authors

The theory of cratering statistics: Significant corrections to existing methods
Bell and Robbins
Review

The main components of this manuscript are:
1) Correction to implementation of probability calculation from Michael et al (2016)
2) Error modelling: counting and diameter errors, area error, d_min selection error. Other sources of error.
3) Measuring SFD slope
4) SASH model
5) Application of SASH: Saturnian system and lunar farside

There's a lot of material in this manuscript and it takes some time to form a coherent impression of it: not least because of its considerable length, but also because the distinct topics it covers were not fully anticipated from the title, summary and abstract. Here, I first give my general comments on the work, then summarise my impression of it. At the end I make line by line comments.

1. First of all, about the error they found in the code implementation of our 2016 paper: I

have verified it, and agree that there should be a dC/dt factor introduced at the change of variable. I thank the authors for pointing this out, and acknowledge them for the correction.

We thank the reviewer for this—we appreciate it! As a technical note, we do wish to point out that the dC/dt factor is also missing from Equation 8 of Michael et al. (2016).

However, on deeper reflection, we are less sure it is fair to call this an error per se. As we now describe in the text, Michael et al. (2016) assumed a uniform age prior, and scaling the PDF is the mathematically correct approach for a prior in age space. Now, we think that a uniform prior in age space is a misleading assumption, even if it seems deceptively reasonable on the surface. However, the form used by Michael et al. (2016) must be used when an explicit prior is constructed in age space.

2. They suggest (114-117) that we were unable to construct an error model for lack of the maths. That's neither the case nor an appropriate conjecture to make. The reason I have focussed on the Poisson statistics over several works is because that is where I think the interesting behaviour is. Small number statistics were the bane of the binning/fitting approaches, and a probabilistic approach has given us the possibility to say something real about some actual features of interest. While building an error model is a rational thing to do, I'm not expecting a lot of new insights into surface history from it, although I'll be happy if you show some.

We apologize if the text came across that way. The intent was: 1) to credit the reviewer and their coauthors for acknowledging other errors exist, 2) to point out that an error model like this did not exist, and 3) to make clear that we do not intend to malign all published age estimates for not utilizing a method that had yet to be developed. We were not saying previous work lacked mathematical aptitude; we were saying that the mathematical methods we develop here did not yet exist. However, because it seems to have cause offense when the intent was, in fact, entirely the opposite, we have removed the words saying that a lack of the mathematical methods was the reason. For clarification, we have also replaced "math" with "mathematical methods" in the subsequent sentence.

As we document in our worked examples, counting error does not dominate the age PDF *even when N is low*. When *N* is high—as it is in most published ages—counting error presents a fairly marginal effect. The mathematical methods required to incorporate these sources of error are non-trivial, especially for random diameter measurement error.

We appreciate the reviewer's openness to new insights from this approach, and we hope the worked examples in the manuscript demonstrate its value, as well as some of the listed items later in this response.

Given that we agree that there are possible 100% or greater errors hidden in the μ of the chronology calibration, the mapping from crater densities to absolute ages is already only notional. We know the endpoints of the relationship, and know the function has a behaviour in between, which probably looks somewhat like the chronology we use, but may differ substantially from it. What has been valuable in crater dating is the sequencing of related surfaces, and the technique has proven to work excellently for this purpose.

We agree.

We have the epoch systems for the Moon or Mars or other bodies, and crater dating places surfaces correctly into and within these systems. If the epoch boundaries turn out to be wrongly placed in time – as surely they must to some extent – of course it will change the bigger picture: how changes in surface processes relate to the internal physics of the planets. But I don't expect that understanding to come from crater dating, even with a more descriptive error model. It will come from samples. The established sequencing of the surface features will remain unchanged even in the light of such a revision.

It is true that our work does not dive into the calibration of the chronology system to recalculate it with a more accurate error model. However, as the reviewer notes below, this work does affect the chronology function and production function calibration calculations.

The claim that "the math had yet to be developed" (117) suggests that the error model maths here is novel, but the compounding of PDFs is not new. The authors have written a new piece of code to implement it.

We have removed this line, as described above. We have also reworded the following sentence to say "mathematical methods" to avoid the implication that we have, say, invented the concept of random variable math. We have also added the words "long-established" to the previous part of this paragraph that describes random variable math to reiterate that we did not invent it.

3. An additional point for consideration in connection with an error model is that, within a single author study, many of the modelled measurement errors will have correlations which act to reduce the spread of the error in compounding when several measurements are compared. When your interest is local sequencing, this is an important effect which is not addressed in your text.

We do address this in the text; we refer the reviewer to sections 5.1 and 5.4. We break both human counting error and diameter measurement error into random and systematic components. In the Canala Crater worked example and the comparison of the bright and dark terrain of Iapetus (sections 5.1 and 5.4, respectively), we only include the random error components because the same person did the counts.

However, we did not make clear in the body of Section 2 that the systematic component should not be included when comparing counts by the same person. We have now added that clarification to Sections 2.2 (first paragraph, and last sentence of the last paragraph) and 2.3 (end of the second paragraph).

Furthermore, if there is any directional bias in the application of a component of the error model, e.g. the d_min error, this needs to be similarly included into the measurements used to derive the chronology model itself, where it would have an opposing effect. If this is not done, the consequence is certainly a degradation of the result compared to doing nothing.

We certainly agree that our methods should be applied to a recalibration of the production and chronology functions. However, this falls beyond the scope of this present work. We do, however, make clear that these sources apply to the production and chronology functions, saying in Section 2.5, “all the additional sources of error beyond counting error we describe here apply to the calibration counts (e.g., human counting error also applies to whoever constructed the calibration curves).” We think this would be a prime next step area for future work.

4. What you are calculating is the error on the true age for each measurement (Section 2.5). A reasonable question to ask is: what are you going to do with this? Questions about when you might expect changes in the surface expression of processes connected to the physics of a planet are never likely to be connected to a single crater dating measurement. Instead, they will be related to the chronology system as a whole – in particular, its epoch boundaries in our present level of understanding - and it’s at that point that it is necessary to bring in the reasoning of your error model. For an individual measurement, I don’t see what more you will learn from seeing this spread distribution than you do from the μ-function that already represents it: simply, that this uncertainty is present and is large. You seem to recognise this in suggesting showing an additional age without the chronology error (382-384). If you are showing both, it would be pertinent to describe the occasions when you can expect to understand something additional from the chronology-error-inclusive result.

Let us first restate what we think the reviewer is saying: We are likely correct, but the reviewer does not think it will significantly affect the science from crater population studies; this is effectively a "so what?" question. If we are incorrect in our interpretation, please let us know.

Assuming we are correct, we think that if methods to-date have been approximations of better methods, and now we are presenting the math for those better methods in an easy-to-use interface (i.e., it is not a burden to use the better methods), then the better methods should be used. Our manuscript's entire section 5 of worked examples also demonstrates how the new methods yield different – and what we argue are more correct – results than traditional methods.

As additional examples:

1. Geophysical modeling of a process that uses a feature date as an input. With, say, a heat flow model exploring the end of volcanism on the Moon, it is the true date of late-stage volcanism in Oceanus Procellarum that matters, not the date relative to chronology system error. It is the true age of a permanently shadowed crater that matters for a model of volatile accumulation.
2. Matching a date to another dating method. For instance, when assessing whether a crater could be a source crater of a lunar or Martian meteorite by matching a cosmic ray exposure date from the meteorite with the crater count date for the potential source crater, it is the true age that matters and especially a more full understanding of the errors involved.
3. When a geologic event dated by crater counting affects questions outside geology. For instance, it is the true, not relative, end date of fluvial and lacustrine activity on Mars that matters for the question of how much potential extraterrestrial life would have been able to develop before astrobiological conditions became less favorable.
4. Modeling of impactor sources. For instance, it is the true date of Tycho Crater that matters in assessing how likely it is to have come from, say, the Flora or Baptistina collisional families.

Neukum made some chronology system error estimates as you've said, but I'd suggest that these only hold under the assumption that the connections made between samples and their source regions are broadly correct. What quantification do you give to the possibility that some of them may be wrong? If you discount this possibility, is the error model providing a false confidence that we understand all the uncertainties? Is it still necessary to caveat the result with an acknowledgement that the error could be greater? If so, what do we learn from this exercise?

We have added a sentence noting this additional source of error, but it does not affect our overall estimate of the magnitude of chronology system error of ~2.5 because that factor already includes this consideration. The entire next-to last paragraph of section 2.4 discusses the various factors that lead us to this factor, and if someone disagrees, they would be able to both change this factor in their own work and it is customizable in our code.

In terms of what we learn from this exercise, we think it is always important to quantify uncertainties as much as possible. The μ notation introduced by Michael et al. (2016) serves as a valuable acknowledgment that did not exist in the field that additional uncertainties exist. Our approach attempts to go a step further by quantifying these uncertainties and incorporating them directly into the age PDF, which we think provides a more actionable framework for downstream applications. Further, as we show in the examples in this manuscript, incorporating that uncertainty in the chronology model can significantly increase the uncertainty in ages, something that we think most people in the community do not appreciate.

5. The title does not reflect the content. If the paper covered component 1 only, you could argue for it (in the singular). But the substance of the work is really about the error model and the SASH method. The error model may be considered an addition to existing methods. The SASH method appears to be a valuable approach to assessing the shape of the CSFD. Neither of these are corrections.

To address this, we have replaced "corrections" with "improvements" in the title.

The summary inexplicably mentions neither the error model nor SASH.

We have extensively rewritten the summary to focus on what we develop rather than where we find issues with existing methods.

6. The calculation used for the age PDF is exactly that described in Michael et al 2016. The change of variable correction relates to the code implementation of that publication.

There are several substantive differences between the age PDF calculation in our work and that of Michael et al. (2016). First, equation 8 of that work gives an equation for the relative age PDF that does not include dC/dt; this missing factor is important. Second, the mathematical approach is entirely different. We use random variable math, and Michael et al. (2016) does a variable substitution. Third, we clarify that the λ PDF follows the Gamma Distribution. Finally, and most importantly, we incorporate additional sources of error beyond counting statistics.

This work goes to great lengths to present its material as an overturn of previous incorrect work, when in truth it adopts an existing technique in its entirety with a single correction.

We have reframed the summary and abstract to emphasize what we have developed, rather than the errors and biases we have found in previous methods. However, we disagree with the characterization that we suggest Michael et al. (2016) should be overturned in its entirety. As the reviewer notes in the next sentence, we explicitly state in some detail that Michael et al. (2016)'s core contention that the diameters do not matter is correct.

We also did not adopt the Michael et al. (2016) technique in its entirety with a single correction; rather, as explained elsewhere in the manuscript and in this response document, we have a different technique and mathematical framework. Section 2.1, which confirms a key finding of Michael et al. (2016) but through a completely independent method, is one component of a much larger framework of this manuscript.

It has taken you ten years to come to the realisation that my proof was correct (142-145), and you cannot proceed without it; neither had a probabilistic description of the problem been considered in this context before that work. At a minimum, the tone of this manuscript is lacking in respect for the prior work without which the error model could not have been constructed in the way it has; inasmuch as it appears purposefully constructed to lead readers into the mistaken impression that an alternative method for determining age was innovated in this work, it could be considered unethical.

We take the suggestion of unethical conduct seriously and reject it. We think the manuscript, as detailed below, clearly credits Michael et al. (2016) extensively while presenting substantively different methods. We validated a core conclusion of Michael et al. (2016) and stated "its implications are considerable." That demonstrates a respect for their work.

We did credit Michael et al. (2016) for the probabilistic description, saying, "Michael et al. (2016) replaced the practice of isochron fitting by directly calculating a probability density function (PDF) of age from the observed number of craters and counting statistics."

This work does *not* present the same solution as Michael et al. (2016) for age modeling, so we do not think we are misleading readers. While we use Michael et al. (2016)'s revelation that diameters do not matter for quantifying counting statistics, counting statistics are only one of the *many* sources of error we include. Additionally, we take a different approach to counting statistics, clarifying that the λ PDF follows the Gamma Distribution, explicitly treating λ as a random variable, and deriving a normalized PDF that allows for fractional counts.

Moreover, Michael et al. (2016) did not develop the concept of using a chronology function to calculate ages. The chronology function has been worked on for decades, dating back to the first returned samples from *Apollo 11*. Michael et al. (2016) present a way to do the calculation. We present a different way.

Michael et al. (2016) is probably the work that gets lauded and credited the most in our paper. Just because we found one assumption that does not hold up does not change the fact that we have great respect for it.

Finally, we wish to point out that we do also point out shortcomings in our own previous work. For instance, in the section on the Saturn system data, we highlight problems with Bell (2020)'s methods, and in the Slope Problem section and Appendix D we also document and analytically derive biases in the Robbins et al. (2018) MLE method.

If the work is to remain with its present structure, it should be retitled referring to the error model, the SASH algorithm, and the slope analysis which are the actual innovations here.

We have changed the title to: "The theory of cratering statistics: Significant improvements to existing methods for ages, slopes, and plots."

Certainly describe the change of variable correction, but remove the preachy “let us teach you how random variables are supposed to be” material as well as the claim that random variable maths is a novel development; <snip>

We dispute that Appendix A is “preachy” or unnecessary. Many readers will not know this material. The reviewer may be familiar with it, but he is a luminary in the field, and we cannot expect every reader to have the same background.

We do not claim that we invent random variable math. Instead, we portray it as well-established math, which it is. That is why we simply assert the theorems and do not derive them. However, we added an explicit clarification to this effect at the beginning of Appendix A.

<snip> remove the ‘substantial flaws’ aspersions repeated throughout the manuscript <snip>

We have removed this wording from the summary and abstract, which we note was the *only* place it appeared. Throughout, we have shifted emphasis from the problems we have found with previous methods to the improvements we have made.

<snip> in reference to the source work that has been reproduced, <snip>

As noted above, we strongly dispute the characterization that we have reproduced source work: We validated information from previous work and used that as one component out of *many* for our overall work.

<snip> and describe the new contributions properly in the context of what already exists, including the compounding of random variables.

Although we believe it was clear, we have added an explicit statement in Appendix A clarifying that we did not invent random variable math.

In summary, the manuscript has severe problems of presentation which make it unacceptable for publication. The tone of the article, which aims to discredit a previous work while appropriating its advances is not ethical. The choice to use a title and summary not covering the actual content but to promote this aim is not adequate.

We have shifted the emphasis in the title, summary, and abstract

We strenuously object to the allegation that we have attempted to discredit Michael et al. (2016) or appropriated its advances. We do not think an average reader would reach that conclusion when we explicitly state their work was correct in a fundamental finding.

The work is structured more like a thesis than a research article. It brings together several chapters on topics which are distinct, although with an over-arching connection. It is not possible to give proper weight of consideration in this review to results on the crater population of the Saturnian system or the lunar farside within what is presented as a work on methods. The SASH algorithm is likewise a distinct contribution from the error model which ought to be evaluated separately: again, there are important details which are beyond what I can evaluate in this review. Thus, the majority of my assessment here relates to the error model.

Noted.

Nevertheless, an error model of the type presented has not been constructed for crater dating before and is an original contribution.

We thank the reviewer for this acknowledgement.

The application of the ASH algorithm is also new, and potentially a valuable tool in the understanding of measured crater size distributions.

We also thank the reviewer for this acknowledgement.

The work over-claims novelty for the treatment of random variables: the maths is not new, but the authors have produced a new code implementation which they use for their evaluations in the present application.
Greg Michael

As stated above, we have added a sentence to the top of Appendix A making crystal clear that we did not invent random variable math, and we added a statement where "random variable" first appears to note it is an established branch of mathematics that is simply rarely, if ever, applied to cratering.

116-117 That's not the reason (see above). It is unacceptable to attribute an imagined explanation in a scientific paper.

See the above response to this point.

149-150 "We do not need to adjust for how the production function varies with diameter." I understand what you want to say, but don't think that's good wording – the PF remains implicitly in the expected number of craters.

We have removed this confusingly worded sentence.

151-154 Yes, but I already wrote this 10 years ago: “In all cases, the PDF description of the model prediction should be considered superior to results obtained by binning/fitting” (Michael et al 2016).

We have added the citation there.

145-148 I’m glad you are finally convinced.
Section 2.2 – 2.3
Estimating appropriate parameters for new situations and terrains is what makes using an error model difficult in practice (235-236). My opinion is that applying standard parameters (as those you suggest in 231-233, and 247-250) is similar to being generally aware of the magnitude of this type of error, as was established in your 2014 paper. Rather than recommend default assumptions, it would be preferable to recommend a practicable method to measure this fresh for each scenario.

Error modeling will always inherently involve some estimation. Measuring random error for each case is a good idea, especially with random diameter measurement error. We considered recommending a process to do this. However, in practice these data are often not available, such as when analyzing literature counts. Moreover, we felt that if we recommended recounting a portion of the terrain to do this analysis, realistically most researchers would not do it. Additionally, if the count area is small, the counter is likely to have remembered the crater in question, which might bias the result. Finally, there is the reality that the systematic component of human counting error typically dominates, and it is difficult for an individual researcher to measure their own systematic error without having multiple counters count the same terrain.

256-263 I agree - this is an interesting observation. You don’t say it explicitly, but because of 148-149, it’s only the diameter errors which cross the interval boundary which make a difference. And at that, it’s d_min which dominates. I realise you recognise this in the discussion, though.

We agree.

Fig 2. I recall the prominent Chapman talk at the Robbins 2015 workshop “Issues in Crater Studies and the Dating of Planetary Surfaces”, which ridiculed the use of three significant figures for ages. Here you are showing four. In the context of the magnitude of errors you discuss elsewhere, you should comment on the utility of this choice.

Chapman’s talk was, as the reviewer notes, limited to ages. Our observed N / true N ratios are robustly measured with 10,000 synthetics, so four significant figures are appropriate *insofar as this simulation is showing*. We show four significant figures so that the reader can see subtle effects, like the shift from a mean of 1.008 at 5% random error and 0% systematic error to a mean of 1.009 at 5% random error and 1% systematic error. We do not mean for someone to infer that they should show 4 significant figures, and we have added text to that effect.

381-385 I agree with this.

Thank you.

396 Please give an example of a non-contrived 'highly unstable area' and its error measured in CraterTools. Also mention ArcGIS first in your criticism, since this universally-used tool is the source of the polygon measurements for CraterTools, to avoid giving an impression of bias against particular authors.

We added the example of Yue et al. (2020), and we clarified that CraterTools is based on ArcMap. That being said, it is reasonable for ArcGIS to store polygons by their vertices, and ArcGIS correctly reprojects those vertices. The issue is that the area will not remain the same because the lines between vertices will change, and that is not the fault of ArcGIS. A polygon defined by vertices will not have the same shape when reprojected into an equal area projection, so the area calculation will be unstable.

In terms of the allegation of "bias against particular authors," we remain perplexed. This is the only mention we have made of Kneissl's work here, and we celebrate the seminal work of coauthor Neukum repeatedly throughout the paper. A sole instance of criticism does not create an impression of bias.

Section 2.7
I agree this is worth doing. I would characterise the procedure rather as picking a diameter interval which is consistent with the slope of the PF, rather than just excluding the rollover. That seems to be slightly different from your implementation.

We think this illustrates the point that there are different interpretations of where to draw D_{min}, and the work by Robbins et al. (2014) showed that other researchers had different definitions, too.

Section 2.8
There's nothing to argue with here until you come to quantify them all together, including unknown unknowns, at 10% (line 475). That's a remarkable insight. Why this number and not any other? I suggest you elaborate on how you chose it. It is better to have a result which we understand may be wrong for various unquantifiable reasons than one which

contains arbitrary adjustments. Otherwise, the error model ends up as merely an exercise in formalism.

We have added a justification for this number. However, we disagree with the view that error factors that cannot be precisely quantified should not be included in the error model. In science, it is rare that error factors can be as precisely quantified as Poisson counting error can be. It is nearly always necessary to have some level of intelligent estimation for sources of error that have yet to be precisely quantified. If we limit the error model to only the sources of error that can be precisely quantified, we mislead the reader by implying that the error is smaller than it is. Further, by providing an additional, tunable parameter in our code, a user can see what their own value for this does to the final results. If they think it's 0%, they can see that effect; if they think it's 100%, then they would be able to see the results from that, too, which we think offers insight.

517-520 It expresses an estimation of the error over the counting statistics. Whether it's a significant improvement depends on what you can interpret from such a result in comparison to what was previously possible. You have not addressed this topic, nor demonstrated a greater interpretive value, so I don't think you can yet make this claim.

The advantages of more accurate ages or at the very least a better accounting for the error distributions, we think, are evident: Better accounting for these and a better estimate is, simply, better. For instance, when the correct PDF that incorporates more sources of error than counting error is used, it by definition changes how we would assess the likelihood of a given age hypothesis.

Moreover, after the extensive work we have done to quantify these effects, we think we have clearly demonstrated that we change the age PDFs significantly, leading to much broader error distribution and different median ages. We also provide specific worked examples that demonstrate a greater interpretive value even when *N* is small. The example with the Canala crater counts from HiRISE vs CTX are a prime illustration of how our methods reconcile what otherwise appear to be conflicting results that could change how one interprets ages on the Canala ejecta deposit.

633-637 If your wording were consistently applied, this would be described both as an 'incorrect method' and a 'substantial flaw'. Please address this.

The two instances of the phrase "substantial flaw" have been removed from this manuscript in this revision, so we have not added it here (which would now be the only instance of the phrase). We think the general intent is captured by other words, and we have added to the end of the sentence, "to address this critical issue."

722-729 This was already said in Michael et al 2016, p283.

Michael et al. (2016) did not say what we are saying here: that the absence of zero-crater bins causes the surrounding bins to plot artificially high, leading to an upward bias at the large crater end for differential and R plots.

On page 283, what Michael et al. (2016) did say was “data points representing empty bins do not appear on a logarithmic scale for a differential plot,” and the empty interval between the largest crater and D_{max} must be included in the Age Problem. This is not the same as saying that differential and R plots have an upward bias at the large crater end.

Section 4.3
Use of the ASH algorithm/SASH variant appears to be a promising approach for understanding crater size distributions.

We thank the reviewer for this comment.

882 ...or even marginally exceeds the diameter of the body!

We chose the wording “significantly exceeds” because of the theory that the Borealis Basin on Mars may have been caused by an impact, creating a crater whose diameter slightly exceeds the diameter of the body (Wilhems and Squyres, 1984). Additionally, Rheasilvia on Vesta has a diameter larger than one of the axes of Vesta.

Wilhelms, D. E., and Squyres, S. W. (1984). The Martian hemispheric dichotomy may be due to a giant impact. Nature, 309(5964), 138-140.

1161-1168 The calibration point was constructed using a counting-error-only measurement, and therefore the 1.70 Ga is showing a substantially different result from the original. There’s more to worry about with this count than you acknowledge here: the additional error-smearing doesn’t rectify the problem.

We are unsure what the reviewer's concern is here. The 1.70 Ga is based on completely different data from any original work on this calibration point. The purpose of this example is (a) to provide a worked example, and (b) to illustrate how a seemingly contradictory result (that this site is younger than the radiometric age) can be reconciled using our methods. We do not reference any of "the original" work done on this problem throughout the 1960s–1980s, so we have not modified the text in response to this comment.

1174-1175 This is excessively grandiose. The maths of compound probabilities has always been available. More correctly, you are the first who have tried to apply this to crater dating.

We have changed “math” to “mathematical methods” to avoid the interpretation that we have invented random variable math. As noted above, we do not believe the original text leant itself to this interpretation, but out of an abundance of caution, we have stated explicitly that we did not invent random variable math at the top of Appendix A and in the first place that it appears in the main body of the text.

Section 5.4

Although this section is an application of the methods of the paper, it would probably have been better as a follow-up paper: the results are distinct from the remainder of the work, and are given no mention in either the title or abstract. It is difficult to assess them within the same context as the rest but this, of course, is the type of application where the SASH approach could be beneficial.

We felt that it was necessary to have a worked example demonstrating how to use the SASH model and its synthetic error model. It is true that we chose an example where the results are scientifically meaningful in order to demonstrate the explanatory power of the SASH model. Additionally, the finding is deeply relevant to the rest of our work. We feel that a paper attempting as comprehensive as possible a treatment of the topic of sources of error in crater counting would not be complete without a clear demonstration of the perils that saturation poses.

To avoid scope creep, in this worked example, we limited ourselves to implications directly relevant to this work. There are considerable implications of these results for the fundamental science that we did not discuss because they would be off topic.

Additionally, the first author felt it was important for academic integrity to acknowledge how our improved statistical methods correct his previous work.

1414-1416 “we have laid out the fundamental math behind the Age Problem” This is incorrect. You have described an error model making use of known methods for compounding PDFs on top of a derivation that you adopted from Michael et al (2016). Please rectify.

We carefully chose the wording “laid out” to avoid implying that we developed this math. The rest of the sentence only makes this clearer. The full quote was: “In our investigation into the theory of cratering statistics, we have laid out the fundamental math behind the Age Problem, the Slope Problem, and the Plotting Problem—much of which has never been used in planetary science." The clause after the em dash makes clear that 1) we are not claiming that everything we have laid out is new (only "much of" it) to planetary science; and 2) because we are only saying (much of) it is new to planetary science, it makes especially clear that we are not claiming to have invented random variable math.

1418 "can be solved without plots" - you still absolutely need the proviso that you gave earlier (152-154), otherwise this is wrong.

We have edited the wording to make clear this is specific to the math of the calculation.

Reviewer 2:
Recommendation: Revisions Required Without Review

All pre-existing data used in the manuscript are publicly available and sufficient information is provided so that they can be found and accessed.

Yes

All significant data products generated as part of the study have been uploaded to a public repository (or will be at the time of acceptance) and are cited using a persistent identifier (such as a DOI).

Yes

All significant numerical codes developed as part of the study have been uploaded to a public repository (or will be at the time of acceptance) and are cited using a persistent identifier (such as a DOI or SWHID).

Yes

All pre-existing software used as part of the study are acknowledged, and sufficient information is provided to find, access, and understand the software. If the software is restricted, the conditions under which it can be accessed are publicly available.

Yes

If you answered no to any of the above questions, please provide supporting information.

I am uploading an annotated version of the manuscript.

No

Comments for the authors

The authors develop a set of new tools used to analyze crater statistics with the aim of correcting some of the problems inherent in the standard techniques that have been widely employed to analyze crater counts and infer ages of planetary surfaces. The paper goes into great depth and detail regarding the current methods and their shortcomings, and not only demonstrates how we can do a better job of inferring ages and age uncertainties using more robust statistical methods, but also providing a comprehensive set of tools to do it with the 'cratersfd' Python package.

This is a *beautiful* paper. The field of planetary sciences has been desperately in need of a revision to these old techniques for quite some time now. Prior to reviewing this paper, I had been planning to develop a set of crater statistical analysis tools as part of my project, but now I don't have to, as Bell and Robbins have done a far, far better job than I could ever hope to.

I have no major critiques of the paper, only a handful of very minor suggestions. I do not need to review the manuscript again, and I do not need to be anonymous.
-David Minton, Purdue University

We thank the reviewer for his very kind comments about our work.

Line 122: The concept that craters could be analyzed as a Poisson point process goes back much further than 2010. The earliest I am aware of is by D.W.G Arthur 1954 in the journal of the British Astronomical Association (cited as the first by A. Marcus in 1966 https://doi.org/10.1016/0019-1035(66)90024-8, a somewhat forgotten paper that had a similar aim as the current authors to challenge the underlying assumptions of statistical models that were in use at the time for answering the "impacts vs. volcanoes" lunar crater origin hypotheses).

We thank the reviewer for pointing us to this important citation, which we have added.

[multiple places]: Whenever I review a paper that uses the terms "saturation," or its variants ("saturation equilibrium", "equilibrium saturation", etc.) I am compelled to point out that Gault (1970) gave us two perfectly useful definitions for the terms "saturation" and "equilibrium," which mean two very different things. Saturation is a purely mathematical construct that describes a surface entirely filled with circles. Melosh (1989) prepended the term "geometric" to saturation to drive this point home, and its derivation has recently been revisited by Hirabayashi et al. (2024) (https://doi.org/10.3847/PSJ/ad8883). It is meant to serve as a reference against which to compare crater number densities to. In contrast, equilibrium is the observed maximum crater surface number density seen on planetary surfaces, and it results from the competition between new crater formation and the accompanying topographic degradation that inevitably accompanies the violent formation of each new crater (for more detail, see sec. 1.1 of Minton et al. (2019))

What the authors refer to when they use the term "saturation" is actually equilibrium. For instance, on line 222:
"… Ostrach (2013) found that these effects continue to fade as D increases past the equilibrium saturation rollover into unsaturated diameters."
Should instead be something like:
"… Ostrach (2013) found that these effects continue to fade as D increases past the equilibrium rollover into diameters that have not reached equilibrium."

I would please ask that the authors adopt the definitions of "equilibrium" and "geometric saturation" as I've described here. To me this is quite important to do in this paper, as I foresee this paper having a major influence on how we approach the science of crater statistics.

We are largely adopting the reviewer's suggestions. We have rewritten the text as the reviewer suggests, replacing sections describing portions of the SFD as "saturated" with language describing them as "in equilibrium." We have also added text explicitly defining the terms we use. However, we make two technical notes:

While we agree with the reviewer's definition of "geometric saturation," we do not define this term because we do not use the concept in this paper. Since Gault (1970) defined "saturation" as the maximum density, as we understand it, the term has evolved to refer to the concept of craters erasing preexisting craters in general, including forms like quasi-equilibrium saturation. That is why Melosh (1989) created the term "geometric saturation." We support this redefinition and believe it has become generally accepted. At a minimum, we need some term for the concept in general, and we don't see a good alternative to using "saturation" for this purpose.

We adopt the definition of "equilibrium" that would cause us to amend the examples as the reviewer suggests: we define "equilibrium" as the shape the SFD follows for diameters smaller than the rollover. This is subtly different from the suggested definition of "the observed maximum crater surface number density seen on planetary surfaces." Real SFDs in equilibrium do not necessarily follow the maximum number density; instead, they typically have a slope that is slightly shallower than $\alpha \approx 2$, often rolling over to a slope that is much shallower (Xiao and Werner, 2015). This distinction is critical to our discussion of the Saturn system data. For clarity, we define a new term, "rollover density" for the density where the rollover is reached.

Line 454:

“Rays of discontinuous ejecta drive erosion of other craters beyond the extent of the continuous ejecta (Minton et al., 2019). Although these effects may not dominate when the production function is steeper than ~-2, they do clearly sometimes happen. However, because these effects have yet to be quantified empirically, it remains unclear precisely how significant they are.”

This is a bit of a muddled summary of what we found with regard to rays in our 2019 study. First of all, in that paper we were exclusively focussed on the regime where the production function is steeper than -2, so our conclusion with regard to rays is about steep-sloped SFDs, not shallow ones. What we found is that the observed magnitude of the equilibrium CSFD of steep-sloped craters (i.e. ~few percent geometric saturation across all sizes) requires that distal ejecta is the dominant driver of topographic degradation of small craters. Furthermore, it is the apparent *stability* of the observed magnitude of the equilibrium CSFD of steep-sloped craters (i.e. that it seems to be the same ~few percent geometric saturation whenever it is measured, all else being equal) that requires that the distal ejecta be spatially heterogeneous, as uniform homogeneous degradation leads to major fluctuations in the observed crater number density in our simulations, and “rays” seem to suppress them.

While it is correct that this effect has yet to be quantified empirically (though we are attempting such quantification right now, actually) it’s emphatically not an issue for shallow-sloped production functions.

The original section was poorly worded and did not clearly convey what we meant to say. We have rewritten it for clarity. What we meant to say is that erosion by rays, like overprinting by large craters, clearly sometimes happens when the surface is not yet at equilibrium. Even in the equilibrium domain, we cannot entirely ignore saturation.

Line 1493: There seems to be a repeat of the “Data and code availability” section. The first section discusses code, and provides both a GitHub repository and a working Zenodo link with a DOI. However, the second section provides a Zenodo link for figure files and datasets that does not work. From what I can tell, the GitHub repo (and working Zenodo link) provides what is necessary to reproduce the figures (but not the figures themselves) so perhaps this is just a manuscript versioning error.

The Planetary Research template calls for separate data and code availability sections. We copied part of the text because it was applicable to both cases. However, we neglected to change the section titles—an error we have now corrected.

There are two separate Zenodo records. The code record linked to a published code version so that the code as it was when the paper was published can be preserved with a DOI. We have made minor edits to the code since we submitted it, so we have updated the link with a new DOI for a new version.

There is also a Zenodo record for the data files associated with the code that we describe in the data availability section. We have added a clearer explanation of its purpose. That record we left in draft form in case edits might need to be made during the revisions. We did make a minor edit. During the revisions stage, we discovered that a *cratersfd* update subsequent to writing the code for Figure 6 accidentally caused Figures 6l and 6m to appear without zero-crater bins. With this error remedied, we have now submitted the DOI for publication in anticipation of publication of this article.

EDITOR DECISION
round 2

Editor Decision: Round 2

Dear Samuel Bell,

I would like to thank you for submitting your manuscript "The theory of cratering statistics: Significant corrections to existing methods" (#235) to Planetary Research. I have now received a review from the previous reviewer, which is included at the bottom of this email. Based on my evaluation of your submission, your manuscript will be suitable for publication after addressing the remaining (minor) comments.

We ask that you please provide a point-by-point response to all comments and identify what changes you have made to the revised manuscript. In addition to the rebuttal, we ask that you also provide a document that tracks all changes that have been made to your manuscript since its initial submission.

When you have completed your revisions, you can upload the revised manuscript, your response and a tracked changes document from your submission dashboard.

If you have any questions about the recommendations in your review, or if you are unsure how to respond or proceed, please feel free to contact me from your submission dashboard.

Sincerely,

Mark Wieczorek
Editor in chief, Planetary Research

The following comments were received from the reviewers.

Reviewer 1:
Recommendation: Revisions Required Without Review

All pre-existing data used in the manuscript are publicly available and sufficient information is provided so that they can be found and accessed.

Yes

All significant data products generated as part of the study have been uploaded to a public repository (or will be at the time of acceptance) and are cited using a persistent identifier (such as a DOI).

Yes

All significant numerical codes developed as part of the study have been uploaded to a public repository (or will be at the time of acceptance) and are cited using a persistent identifier (such as a DOI or SWHID).

Yes

All pre-existing software used as part of the study are acknowledged, and sufficient information is provided to find, access, and understand the software. If the software is restricted, the conditions under which it can be accessed are publicly available.

Yes

If you answered no to any of the above questions, please provide supporting information.

I am uploading an annotated version of the manuscript.

No

Comments for the authors

see attachment

The theory of cratering statistics: Significant improvements to existing methods for ages, slopes, and plots

Bell and Robbins

Review2

I appreciate the authors consideration of my previous comments, and acknowledge their improvements to the tone of the revised manuscript.

I have some final comments which should be addressed (2 & 3) and may be addressed (1 & 4), after which the manuscript can be published.

Greg Michael

Section 2.4 Area Error

1. The wording with ArcMap is fine now, and the example is a fair illustration of this type of imprecision. It's already an improvement that you removed the 'highly' from 'highly unstable': I would also suggest that 'imprecise' is a more appropriate word than 'unstable' unless you have seen published measured areas tending to infinity.

Section 2.8

2. I do not agree with your interpretation here. Please replace "The reason why Michael et al. (2016) did not include" -> "Our interpretation of the reason why Michael et al. (2016) did not include…". As before, it is not appropriate to attribute a supposed motivation.

Comments on responses:

> Furthermore, if there is any directional bias in the application of a component of the error model, e.g. the d_min error, this needs to be similarly included into the measurements used to derive the chronology model itself, where it would have an opposing effect. If this is not done, the consequence is certainly a degradation of the result compared to doing nothing.

> We certainly agree that our methods should be applied to a recalibration of the production and chronology functions. However, this falls beyond the scope of this present work. We do, however, make clear that these sources apply to the production and chronology functions, saying in Section 2.5, "all the additional sources of error beyond counting error we describe here apply to the calibration counts (e.g., human counting error also applies to whoever constructed the calibration curves)." We think this would be a prime next step area for future work.

3. The point is that it is wrong to apply such an adjustment without also doing so at the calibration stage: it would be worse than doing nothing. The recommendations should make this clear.

1. What you are calculating is the error on the true age for each measurement (Section 2.5). A reasonable question to ask is: what are you going to do with this? Questions about when you might expect changes in the surface expression of processes connected to the physics of a planet are never likely to be connected to a single crater dating measurement. Instead, they will be related to the chronology system as a whole – in particular, its epoch boundaries in our present level of understanding - and it's at that point that it is necessary to bring in the reasoning of your error model. For an individual measurement, I don't see what more you will learn from seeing this spread distribution than you do from the μ- function that already represents it: simply, that this uncertainty is present and is large. You seem to recognise this in suggesting showing an additional age without the chronology error (382-384). If you are showing both, it would be pertinent to describe the occasions when you can expect to understand something additional from the chronology-error-inclusive result.

Let us first restate what we think the reviewer is saying: We are likely correct, but the reviewer does not think it will significantly affect the science from crater population studies; this is effectively a "so what?" question. If we are incorrect in our interpretation, please let us know.

Assuming we are correct, we think that if methods to-date have been approximations of better methods, and now we are presenting the math for those better methods in an easy- to-use interface (i.e., it is not a burden to use the better methods), then the better methods should be used. Our manuscript's entire section 5 of worked examples also demonstrates how the new methods yield different – and what we argue are more correct – results than traditional methods.

As additional examples:

1. Geophysical modeling of a process that uses a feature date as an input. With, say, a heat flow model exploring the end of volcanism on the Moon, it is the true date of late-stage volcanism in Oceanus Procellarum that matters, not the date relative to chronology system error. It is the true age of a permanently shadowed crater that matters for a model of volatile accumulation.
2. Matching a date to another dating method. For instance, when assessing whether a crater could be a source crater of a lunar or Martian meteorite by matching a cosmic ray exposure date from the meteorite with the crater count date for the potential source crater, it is the true age that matters and especially a more full understanding of the errors involved.
3. When a geologic event dated by crater counting affects questions outside geology. For instance, it is the true, not relative, end date of fluvial and lacustrine activity on Mars that matters for the question of how much potential extraterrestrial life would have been able to develop before astrobiological conditions became less favorable.
4. Modeling of impactor sources. For instance, it is the true date of Tycho Crater that matters in assessing how likely it is to have come from, say, the Flora or Baptistina collisional families.

4. Having an evaluation of the shape of the chronology uncertainty for a specific measurement does not make the uncertainty go away. We don't need to know exactly how a factor 2.5 uncertainty convolves with the other uncertainties to understand how it constrains meaningful answers to questions of the type you describe above. The questions are good, but are very hard to address satisfactorily with the current crater chronology calibration uncertainty. The error model should hammer this point, which is beneficial, but my comment is simply that for routine measurements – as you are recommending – it will generally not be possible to interpret anything additional from doing this. You suggest simultaneous presentation of the result without chronology error – I agree with this, because it lets you understand sequencing. I'm asking what you expect to interpret from the routine chronology-error-inclusive results? Maybe you just answer "it doesn't matter - it's just better": then I think we understand each other.

AUTHOR REBUTTAL
round 2

The theory of cratering statistics: Significant improvements to existing methods for ages, slopes, and plots

Bell and Robbins

Review2

I appreciate the authors consideration of my previous comments, and acknowledge their improvements to the tone of the revised manuscript.

I have some final comments which should be addressed (2 & 3) and may be addressed (1 & 4), after which the manuscript can be published.

Greg Michael

Section 2.4 Area Error

1. The wording with ArcMap is fine now, and the example is a fair illustration of this type of imprecision. It's already an improvement that you removed the 'highly' from 'highly unstable': I would also suggest that 'imprecise' is a more appropriate word than 'unstable' unless you have seen published measured areas tending to infinity.

Change made.

Section 2.8
2. I do not agree with your interpretation here. Please replace "The reason why Michael et al. (2016) did not include" -> "Our interpretation of the reason why Michael et al. (2016) did not include…". As before, it is not appropriate to attribute a supposed motivation.

Change made.

Comments on responses:
Furthermore, if there is any directional bias in the application of a component of the error model, e.g. the d_min error, this needs to be similarly included into the measurements used to derive the chronology model itself, where it would have an opposing effect. If this is not done, the consequence is certainly a degradation of the result compared to doing nothing.
We certainly agree that our methods should be applied to a recalibration of the production and chronology functions. However, this falls beyond the scope of this present work. We do, however, make clear that these sources apply to the production and chronology functions, saying in Section 2.5, "all the additional sources of error beyond counting error we describe here apply to the calibration counts (e.g., human counting error also applies to whoever constructed the calibration curves)." We think this would be a prime next step area for future work.
3. The point is that it is wrong to apply such an adjustment without also doing so at the calibration stage: it would be worse than doing nothing. The recommendations should make this clear.

We have added the following text acknowledging this effect in Section 2.4: "Although the primary effect of our error model is to increase uncertainty, it also corrects bias, and there is likely a small average bias correction. (We are unsure of the direction of the average bias correction because of the uncertainty of the average effect of the D_{min} PDF.) Because this average bias is not removed from the chronology function calibration data, it adds to the effective chronology function uncertainty."

We disagree that the average bias correction effect renders the whole effort worse than doing nothing. The average bias correction is small compared to the average magnitude of the bias correction and very small compared to the overall uncertainty. It does not outweigh the benefits of correcting the bias, and it does not outweigh the benefits of acknowledging the uncertainty. Additionally, we note that this only matters for the true age calculation. For a relative age calculation—which the reviewer argues below is more important—the average bias correction effect does not matter.

1. What you are calculating is the error on the true age for each measurement (Section 2.5). A reasonable question to ask is: what are you going to do with this? Questions about when you might expect changes in the surface expression of processes connected to the physics of a planet are never likely to be connected to a single crater dating measurement. Instead, they will be related to the chronology system as a whole – in particular, its epoch boundaries in our present level of understanding - and it's at that point that it is necessary to bring in the reasoning of your error model. For an individual measurement, I don't see what more you will learn from seeing this spread distribution than you do from the μ- function that already represents it: simply, that this uncertainty is present and is large. You seem to recognise this in suggesting showing an additional age without the chronology error (382-384). If you are showing both, it would be pertinent to describe the occasions when you can expect to understand something additional from the chronology-error-inclusive result.

Let us first restate what we think the reviewer is saying: We are likely correct, but the reviewer does not think it will significantly affect the science from crater population studies; this is effectively a "so what?" question. If we are incorrect in our interpretation, please let us know. Assuming we are correct, we think that if methods to-date have been approximations of better methods, and now we are presenting the math for those better methods in an easy- to-use interface (i.e., it is not a burden to use the better methods), then the better methods should be used. Our manuscript's entire section 5 of worked examples also demonstrates how the new methods yield different – and what we argue are more correct – results than traditional methods.

As additional examples:

1. Geophysical modeling of a process that uses a feature date as an input. With, say, a heat flow model exploring the end of volcanism on the Moon, it is the true date of late-stage volcanism in Oceanus Procellarum that matters, not the date relative to chronology system error. It is the true age of a permanently shadowed crater that matters for a model of volatile accumulation.
2. Matching a date to another dating method. For instance, when assessing whether a crater could be a source crater of a lunar or Martian meteorite by matching a cosmic ray exposure date from the meteorite with the crater count date for the potential source crater, it is the true age that matters and especially a more full understanding of the errors involved.
3. When a geologic event dated by crater counting affects questions outside geology. For

instance, it is the true, not relative, end date of fluvial and lacustrine activity on Mars that matters for the question of how much potential extraterrestrial life would have been able to develop before astrobiological conditions became less favorable.
4. Modeling of impactor sources. For instance, it is the true date of Tycho Crater that matters in assessing how likely it is to have come from, say, the Flora or Baptistina collisional families.

4. Having an evaluation of the shape of the chronology uncertainty for a specific measurement does not make the uncertainty go away. We don't need to know exactly how a factor 2.5 uncertainty convolves with the other uncertainties to understand how it constrains meaningful answers to questions of the type you describe above. The questions are good, but are very hard to address satisfactorily with the current crater chronology calibration uncertainty. The error model should hammer this point, which is beneficial, but my comment is simply that for routine measurements – as you are recommending – it will generally not be possible to interpret anything additional from doing this. You suggest simultaneous presentation of the result without chronology error – I agree with this, because it lets you understand sequencing. I'm asking what you expect to interpret from the routine chronology-error-inclusive results?
Maybe you just answer "it doesn't matter - it's just better": then I think we understand each other.

We will meet the reviewer half way. "It's just better" is a major motivation behind our desire for true ages. We are, in fact, motivated by the intangible benefits of true ages being the true ages.

However, there are specific examples where this matters. For instance, if a crater counting date of a lunar crater comes back at 10Ma, the factor of ~2.5 kernel tells us that it could match a lunar meteorite with a cosmic ray exposure age of 6Ma but probably does not match one with a cosmic ray exposure age of 1.5Ma.

In recognition of these issues, we have added this language to the end of Section 2.4: "For many geologic questions, ages without chronology system error still have deep explanatory power, but the true ages are the true ages. For questions where the true age matters, it is important to acknowledge the reality of these wide error distributions."