Reviewer 2 (anonymous):

I wish to thank this reviewer for his thoughtful comments (I will use the male pronoun since the AE identified this reviewer as male). Although I do not agree with some of his comments, I agree with many and the manuscript has been significantly improved based on his review. I greatly appreciate the time he took to engage with the manuscript, both in this round of review and in a prior round for a different journal.

<u>Comment:</u> (1) First, I concur with reviewers Heimsath and Whipple on the fact that it is nonstandard at best to assign a separate P0 value to each measured P value using exponential scaling relationship in Eq 1. It would be equivalent to predicting different values of a y intercept in a linear regression of y on x when you know the slope of the regression and the value of y and x for each data point. There is only one y intercept per regression through a cloud of data. This business of inferring one y-intercept per data point strikes me as – at best – a kooky way (i.e., that differs from established norms) of quantifying the uncertainty in the y intercept. Based on what I can see in their reviewer comments, Heimsath and Whipple had the same reaction. And the author's response to their comment – i.e., "On a more practical level, I don't understand how we, as a community, could make significant progress on understanding the controls on P0 values if we accept the logic of Heimsath and Whipple that only two P0 values can be reliably determined from 57 CRN analyses" – is not compelling. The alternative logic of Pelletier seems to be that we should suspend conventions of statistics and stretch data farther than they can be stretched just to support some as yet non mechanistic formulation that he has presented here. I prefer the less radical option of recognizing the limits of data and working to overcome them in more traditionally acceptable ways – i.e., with new measurements and perhaps a more clever analysis approach. For example, as an alternative to the methods presented here, the author might think of ways to model P rather than P0 using some sort of multiple regression analysis that includes h explicitly in a model of rock damage and microclimatic effects. This business of calculating a P0 effectively corrects for the exponential-withdepth variation in P before the modeling begins. In a true multiple nonlinear regression, one could account for h, D, A and everything else simultaneously, and as an outcome of the approach also quantify the relative importance (leverage) of each variable in the regression. If the outcome is that h dominates while D and A add little to the predictive power of the model, then the author would be forced to confess that D and A are not strong predictors of P and thereby P0. And as I point out below, there is good reason to suspect that that is precisely what he would learn.

<u>Response:</u> I am not asking the reader to suspend the conventions of statistics. A soil production function is the outcome of a regression analysis. A regression analysis yields two types of outputs: the coefficients of the regression equation and a set of residuals. Computing residuals and testing for additional controls is a recommended step in regression analysis.

I defined P_0 in the paper as the maximum soil production rate at each point on Earth's surface. To estimate P_0 values defined in this way, one begins by accounting for the effects of soil cover (which has the effect of decreasing the soil production rate below its maximum or potential value) by regressing log-transformed P values to soil thickness. Following regression, P values are divided by the regression equation (which is equivalent to subtracting the regression of the log-transformed data) to obtain a set of residuals that can and should be interrogated for additional controls. That is all I have done to estimate P_0 values. If the regression of P data to soil thickness yields no statistically significant trend, then there is no statistically significant regression to soil thickness and hence no residuals (P_0 values) to study. That is not the case here, as Heimsath et al. (2012) clearly demonstrated that gently and steeply sloping portions of the landscape fit exponential soil production functions with nearly identical decay constants.

I think it is reasonable to ask reviewers to at least consider my definition of P_0. However, they simply define P_0 differently (as the y-intercept of the soil production function) and then criticize me on the basis of that alternative definition. I think the two definitions are complementary. I don't think see any reason why the residuals of this particular regression should be ignored when the output of this or any other regression is a set of regression coefficients and a set of residuals, both of which contain important information.

Reviewer 2 joins Heimsath and Whipple in criticizing me for not providing a new suite of CRN-based soil production rate data. I think there is broad agreement in the scientific community that it is appropriate for some studies to focus on measurements and data analysis (e.g., Heimsath et al., 2012) and others to focus on analyses of existing data and modeling/process-based interpretation (this paper). Science would move forward more slowly and with a less diverse range of perspectives if, for example, every study of soil production required a new *in situ* CRN dataset. In this specific case I think it is clear that what is most needed is a process-based understanding of trends in the existing data that can be used to guide additional targeted 10Be analysis. I agree with the reviewers that my *ESurfD* paper did not provide such an analysis, in part because it did not consider the potentially important process of topographically induced stress fracture opening. However, I believe that my proposed revision does provide a process-based understanding that is both consistent with trends in the data and well-grounded in theory.

The reviewer also calls for a multivariate regression to all of the potential controlling variables including soil thickness. I think it is more appropriate to honor the work of Heimsath et al. (2012) (as the reviewer recommends in many of his comments) by using the residuals of their regression as a starting point. The combination of stepwise regression and cluster analysis I use in the revision is based on standard statistical methods. The reviewer may not agree with every step of my revised analysis, but I respectfully ask that he consider it.

Comment: (2) Second – and this is a bigger concern in my view – is the degree to which the predicted values of P0 ***diverge*** from the observed values in the dataset. It seems like a key goal in this paper is to use indices of rock damage and aspect to predict P0 and ultimately map the variations in P0 and some additional offshoots of it (E and E*L) onto the landscape. Starting at section 2 and continuing through to the end of this paper, this is actually ***the*** central focus of results and discussion. The trouble is, one must be willing to believe that the model in Eq 2 is a good predictor of P0 in order to confidently follow the author in this vital leap of faith. Personally, after reading this paper, I am not willing to make that leap. Nor should any self-respecting data analyst, once he/she realizes that the predicted values are not actually a very good match to the observations. Sure – as the reviewer points out – there is a highly significant correlation between the measured and predicted values P0, but the existence of such a correlation is not sufficient in and of itself to demonstrate that the predictions are good enough to explain the variations in P0. Recognizing this challenge nearly fifty years ago, hydrologists Nash and Sutcliffe (1970) developed their own measure of model efficiency in their quest to objectively evaluate whether their models of river discharge were good predictors of observations. Though this Nash- Sutcliffe (N-S) statistic – as it later became known – was developed for models of river flow, it has also been widely used to assess model efficiency for other natural variables, including erosion rates and nitrogen and phosphorus loading. My quick calculation of a N-S efficiency statistic for the model of predicted P0 yields a value of 0.18 based on data provided in the supplemental table. For context, realize that the maximum value of the coefficient, which is 1, would indicate that the model explains ***all*** of the observed variation in the data. By contrast a value of 0 would indicate that the model is ***just as good as*** the average value of the data at explaining observed variations across the data set. Values less than 0 imply that the model ***is worse than*** the average. In this case, my estimated value of 0.18 indicates that ***the model in Eq. 2 is a little bit better than the average P0*** at predicting the distribution of measured P0 across the dataset. For this reason, I think the machinations of the predictive modeling exercise (i.e., most of the paper) are not really warranted, irrespective of the significance of any inferred correlations between P0 and D and between P0 and A. Importantly the reader should only commit to believing those correlations to the extent that he/she can overlook the suspect exercise of calculating P0 for each measured value of P. Ultimately, it is not clear to me that the author understands that there is a vital difference between documenting a statistically significant correlation between a measured and predicted value and demonstrating that a model is good at doing what it is supposed to do. If he does, he is hiding it at the top of page 7, where he seems to suggest that statistical significance in classical regression metrics is sufficient. I will not deny that the correlation coefficient and thus the coefficient of determination by themselves provide a very loose first approximation of model fitness. But even then, this is true only to the extent that high coefficients of determination (close to one) imply better correlations and low coefficients imply poor correlations (irrespective of whether they are statistically significant). To understand the problem with using Rˆ2 in the way the author seems to want to use it here, consider the toy example in which P0 observed is exactly 0.2 times the value of P0 predicted for each inferred value of P0; in that case, the coefficient of determination of P0 predicted and P0 observed would be 1.0 with a very very low p value even though the predicted value of P0 is 5 times higher than the observed value at each site. P0 predicted is dead wrong but the coefficient of determination is fantastically good. This illustrates how simple correlation indices for predicted versus observed data sets can (and probably often do) fall short on gauging the predictive power of a model.

Response: I did not include the Nash-Sutcliffe efficiency for the simple reason that it does not apply to regression models. The definition of the Nash-Sutcliffe efficiency is

$$E = 1 - \frac{\sum_{n=1}^{N}(X_{n,obs}-X_{n,sim})^2}{\sum_{n=1}^{N}(X_{n,obs}-X_{mean,obs})^2}$$

where $X_{\mathrm{n,sim}}$ are the values predicted by a simulation or other type of model that is *not based on regression*. In cases where the predicted values are based on a regression (as is the case here), the closest analog of the Nash-Sutcliffe efficiency is the coefficient of determination, R^2, defined as

$$R^2 = 1 - \frac{\sum_{n=1}^{N}(X_{n,obs}-X_{n,reg})^2}{\sum_{n=1}^{N}(X_{n,obs}-X_{mean,obs})^2}$$

where $X_{\mathrm{n,reg}}$ are the predicted values based on the regression. If these two equations look almost identical it is because they are. When using a regression model, one uses R^2, which I did. When not using a regression model, one uses E.

The reviewer criticizes me for not reporting a Nash-Sutcliffe efficiency despite his request for one in a review of a prior version of the paper for the journal *Geology*. I am not going to include a statistic that, by definition, does not apply to the method I am using.

I don't want to antagonize this reviewer, but I would like to point out that his discussion of the *Geology* review violates GSA's ethical guidelines for publication, which require reviewer confidentiality. I am, of course, glad that the paper was rejected by *Geology* and that my *ESurfD* paper was also negatively reviewed, because this has prompted me to take a fresh look at the problem and redouble my efforts to understand the process basis for the trends in P_0 values in the SGM. This type of rethinking and major revision is often a positive outcome of a negative review. That said, I still think it is reasonable to ask that the review process follow established ethical guidelines. I don't think it is fair or accurate that the literature now suggests that I am not a careful scientist who carefully considers reviewer comments. There are simple reasons, identified here, why I did not explicitly address some of his concerns from the prior review.

The 0.18 value computed by the reviewer is inappropriate because it is weighted towards the errors associated with higher P_0 values since the reviewer did not log-transform the data. Such weighting is appropriate for many applications, such as modeling discharges of water, sediment, or contaminants, in which the performance of the model must be judged on its ability to predict both individual data points and the integrated value of the quantity under study. Since the integrated value is dominated by the largest values in the dataset, it is appropriate to weigh the errors associated with larger values more heavily in such cases. That is not the case here. There is no reason to weigh samples from areas with larger potential soil production rates more heavily than samples from areas with low potential soil production rates in judging the model fitness. Because the data have a positive skew, it is more appropriate to log-transform the data.

The reviewer poses the case of an independent variable regressed to a dependent variable offset by a factor of 5. A regression of the logarithms of the independent variable to the dependent variable yields a model with no offset (the unique result of the regression to the hypothetical data posed by the reviewer is ln y = ln(0.2) + ln x, R^2 = 1). Therefore, the supposed counterexample suggested by the reviewer is impossible using the method I am using (regression of log-transformed data).

I understand very well that there is value in having a low value of p and a high value of R^2 (or the Nash-Sutcliffe efficiency, if one is evaluating a simulation model). However, there are many geomorphology papers that are based on regressions with R^2 values lower than the ones I obtained (one example: R^2 = 0.17 in Nature Geosciences, v. 8, p. 462-465, 2015, Fig. 3a). In my revised paper I obtained R^2 = 0.50 (R^2 = 0.87 when data with the same S_{av} value are averaged to minimize local variability). I don't know whether this will satisfy the reviewer since I don't know what he considers an acceptable value of E or R^2.

Comment: On a side note, when I plotted the P0 measured and P0 predicted values in the supplemental table against each other, I get a pattern that looks slightly different than the one shown in the figure. The differences are not big enough to explain away the problem of low Nash-Sutcliffe statistics (Fig 3D and 4C), but it made me worry that the author has some version inconsistencies between his figures and the data he provided in the table. Not sure which version is "correct."

Response: I could not reproduce this error. I am as certain as I can be that the data presented in the table and plotted in the figures of the proposed revision are the same.

Comment: (3) Like Heimsath and Whipple, I am unimpressed with the theoretical basis of Eq 2, and moreover, I am not compelled by the author's response – i.e., "It is far beyond the scope of the paper to develop a comprehensive theory for how microclimate relates to vegetation cover, wildfire frequency and severity, and soil production rates, assuming such a theory is even possible." However, whereas Heimsath and Whipple rightly seem very worried about how D might connect to rock damage at landscape scales, and how those variations in D would actually connect to P0 in a mechanistic way, I am stuck on the fact that the authors never actually showed me that aspect should matter at SGM. The references cited on page 3 have nothing to do with the effect of aspect on vegetation or the effect of either vegetation or aspect on fire intensity or severity in the SGM. Where is the proof that vegetation, fire frequency, and slope steepness vary with aspect in the SGM? It seems it would be crucial to demonstrate this is the case before motivating the paper and the formulation of equation 2 more specifically. The aspect story fits with some of the author's work in other landscapes but not here – at least not according to the references cited here. If anything, the Keeley and Zedler study seems to suggest that the current regime – in which the landscape is prone to large fires that sweep through the landscape with indifference to aspect – has been the norm for a long time. Additionally, this study seems to hang a lot of its motivation on the idea that fire promotes weathering. But - despite the good investigative work cited on page 3 - I am not sure I concur that the connection has been well documented at SGM. All of the studies cited here are fascinating but ultimately just anecdotal investigations of weathering of boulders – not weathering of rock under soil, which presumably is important here since much of the SGM area is covered by soil. Moreover, they do not report faster weathering rates on fire-prone versus not-fire-prone slopes. In fact, none of the studies actually report rates (focusing instead on processes) and none compare fire-prone versus not fire-prone slopes. Shtober-Zisu et al. comes close to reporting a rate but ultimately says it is hard to say how the boulder spalls in carbonate outcrops influence denuda tion rates across the landscape. And again, there is no comparison to a landscape that is not fire prone, so there is no control in the experiment – and importantly no support for the author's claim here that weathering is faster in fire-prone versus not-fire-prone landscapes.

However strong the correlation between P and A may be, I think it is very important for the author to step back from this generic claim that aspect-driven differences in wildfire are driving the show and more precisely drill in on how anecdotal studies from the SGM in particular support the slope aspect idea. Bottom line is there needs to be some stronger motivation here – hopefully shored up some sound mechanistic explanations for why both D (measured in the S&B 2011 approach) and A should matter. I do NOT think it is "beyond the scope" of this paper to justify the formulations that it presumes to impose broadly on the landscape.

<u>Response:</u> I agree with the criticism that the processes I invoked in my previous model were not necessarily the best or only controls on P_0 values, so I have thought hard about this issue, tested the topographically induced stress fracture opening process, and found it to be a better explanation of trends in the data. Once this control is accounted for, a climatic control on P_0 values becomes apparent at the highest elevations of the SGM.

I did not state or imply that it was beyond the scope of the paper to justify the formulations I am invoking. Rather, the point I made was that there is value in documenting statistically significant correlations between P_0 values and controls that are based on reasonable process-based models given that the literature has only identified one control (average slope) on P_0 in this dataset and no process-based understanding for even that trend. My proposed revision is an advance because it identifies topographically induced stress fracture opening as the process most likely responsible for the average slope control on P_0 values in the SGM.

My reanalysis of the data shows that, contrary to my *ESurfD* paper, the null hypothesis that P_0 values are independent of slope aspect cannot be rejected. Ten of the sample locations are ridgetops where the local slope is zero and slope aspect should be undefined. However, my initial extraction routine did not account for this fact. Instead, the routine returned values that in several cases indicated that the slopes faced nearly directly south or north (which was correct given the location data, which in some cases is 10-30 m from the ridge due to roundoff error in the sample location). When the data are reanalyzed to include only areas that are not ridgetops (47 of 57 samples), P_0 values are slightly higher, on average, on south-facing slopes, but the null hypothesis that P_0 values are independent of slope aspect cannot be rejected.

Note: The AE has instructed me to respond to reviewer 2 prior to drafting a revision. However, I don't think it is possible to fully respond to their concerns without drafting a revision, since the requested changes were so extensive and fundamental.

<u>Comment:</u> (4) The statistical analyses are nonstandard. My discomfort with them is very high. My discomfort started with the first indication – I think on page 5 – that the author thinks of statistical significance as the logical and quantitative complement to a calculated p value. This is not the case, of course. Rather "significance" is commonly reserved referring to the threshold false positive rate that is allowed in a statistical hypothesis test. So the idea that the author thinks that a calculated p = 0.001 corresponds to a "statistical significance" of 99.9% set me on edge. This misappropriation of terminology was repeated many times throughout the text. But that was just the start. The author also evidently thinks it is ok to calculate a y-intercept for each measured value in a dataset using an overall regression slope that was calculated from the entire data set – and which also yields an overall regression intercept. To be honest,

this seems akin to data fabrication to me, but I can settle on the gentler view of Heimsath and Whipple that it is really just of a crude way to estimate the uncertainty in the intercept. Next, the author follows a rather stilted approach to quantifying the relationship between P0 and D and A. I personally think it should be P versus D, A and h, thus recognizing h as a factor regulating P and avoiding the problem of getting just two P0 values from 57 values of P. In addition, I think the author missed an opportunity to perform a very standard multiple regression analysis on log-transformed variables and instead opted or a multi stage approach that undoubtedly underestimates errors and fails to produce vital outputs like leverage plots and partial regression coefficients which would help the audience gauge the relative importance of the different factors in the regression. In addition, there is no attempt to propagate uncertainties through any of this. This is a major oversight that needs to be fixed. Last and not least, the author also thinks it is ok to use the significance of Rˆ2 for the relationship between predicted and observed values to judge the performance of his model. In the hydrology community that idea has been rejected for nearly half a century. I am very concerned about the strength of the analyses for these reasons.

Response: The reviewer is correct that it is more accurate to define a threshold false positive rate (typically 0.05) and then compare the p value to this threshold to determine whether the null hypothesis is accepted or rejected. I have rephrased my discussion of statistical significance accordingly in the revision. An example from the proposed revision is as follows: “Assuming a significance level of 0.05, the null hypothesis that the cluster of blue points has a mean that is indistinguishable from that of the remaining points with Sav > 30° can be rejected based on the standard t test with unequal variances (t = 0.021).”

The reviewer’s claim that computing residuals is “akin to data fabrication” is troubling. Given that even a whiff of fabrication can ruin a scientist’s career, this is language that, if taken out of context, could be very damaging. I am stunned that anyone would invoke this charge on a fellow scientist in an open review without any evidence of actual fabrication.

Comment: 2.10. I see that Heimsath and Whipple have provided a review of the manuscript and will defer to them as experts on evaluating this paragraph as a motivating theme for the paper. They did not call attention to any problems here. However, as I read line 21 on this page, I guess I have to say that this was not the take home message I got from Heimsath et al., 2012. Higher frequency of disturbance?

Response: “… a greater frequency of disturbance for a given soil thickness” is a defining phrase in the concluding paragraph of Heimsath et al. (2012). As such, I think it is appropriate to include it in a review of the relevant literature. However, I have rephrased this text as follows: “Heimsath et al. (2012) concluded that high erosion rates, triggered by high tectonic uplift rates and the resulting steep topography, cause potential soil production rates to increase above any limit set by climate and bedrock characteristics. Their results challenge the traditional view that P_0 values are controlled solely by climate and rock characteristics.”

Comment: 3.6-3.10. This study seems to hang a lot of its motivation on the unsupported idea that aspect promotes differences in vegetation which in turn promote differences in fire that promote differences in weathering in the SGM area. See general comment above.

Response: Text removed.

Comment: 4.11. I think I understand what the author is trying to do here (correct the measured P0 for the hump in the SPF), but on reading this, I am confused. You used 1.78P for P0? Not 1.78P0? The way I want to read it is the author is correcting the "measured" P0 – which is inferred from the exponential function to the data – by some correction factor. But again, I am confused by this statement.

Response: A humped production function means that the maximum or potential soil production rate is higher than the P value measured on bare ground. Hence P_0, defined as the maximum soil production rate at a point, has to be higher than P for these four cases. As explained in the paper, the data suggest that the factor increase is 1.78. Hence P_0 = 1.78P. I don't see how P0=1.78P0 could possibly be an alternative way of estimating P0, as the reviewer suggests.

Comment: 4.12. "This modification of equation (1) affects 4 of the 57 data points." This would only be comforting if there was actually a very strong trend across all the data. Instead, it seems that the data form really loose clouds of correlations that are hinged entirely on a few points. So the fact that this affects 4 of the points is actually troubling – not comforting – to me.

Response: Text removed.

Comment: 5.3. This equation does not include the fault specific constant of Savage and Brodsky. So I think this assumes that the constant is the same across the study area. Is this justified? Also, to make D dimensionless wouldn't delta x need to be raised to the 0.8 power too?

Response: Savage and Brodsky found no relationship between the fault specific constant and fault displacement (which correlates strongly with fault length). That is, there was variation from fault to fault in terms of their effect on fracture density in nearby rocks, but no systematic variations that one could use in a predictive equation. They stated "When we plot the entire data set shown in Figure 5, there is no clear relationship between c (the fault-specific constant) and displacement"). Savage and Brodsky did propose a weak pattern for faults in siliciclastic rocks, which is clearly not relevant for SGM.

Regarding the units of D, I have thought about this more and run some tests to determine how D should be defined so that the results are most nearly independent of grid resolution. I have found that D should not be dimensionless but should have units of length since it represents the total length of fault segments in a region (albeit weighted by an inverse power-law function of distance). The proposed revision addresses this point as follows: "I define the bedrock damage index D (Fig. 5A) as the sum of the inverse distances, raised to an exponent 0.8, from the point where the D value is being computed to every pixel in the study area were a fault is located:

$$D = \sum_{\mathbf{x}'} \Delta x \left(\Delta x / |\mathbf{x} - \mathbf{x}'|\right)^{0.8} \qquad (6)$$

where Δx is the pixel width, $\mathbf{x}$ is the map location where bedrock damage is being computed, and $\mathbf{x}'$ is the location of each mapped pixel in SGM where a fault exists. *D* has units of length since it is the sum of all fault lengths in the vicinity of a point, weighted by a power-law function of inverse distance."

Comment: 5.10. ***This is very important.*** The line plotted in Fig. 3A is a log-log regression that ignores the cluster of five data points circled in the figure. There is NO justification for ignoring these points!!! He says in line 5.20 that they occur in an area of unusually dense landslides. I do not see this in figure 1!!! Even if I did, it would not justify excluding them from the analysis. Heimsath and Whipple seem to agree. I think it is complete nonsense. Makes the line look steeper than is should be. Sweeping these points under the rug does not make them go away. Including them in the regression would undermines his story that D plays a "subequal" role with tectonics. It not only looks suspicious. It is suspicious. Author needs to HONOR the data in this study and in his other work and not try to sweep data points away like this.

Response: The line plotted in Figure 3A was the linear trend predicted by a simultaneous multivariate regression of P_0 to D and A that included all of the data points. My discussion of these 5 points was limited to a thought exercise in which I reported p values of the relationship between P_0 and D with and without these points included. That thought exercise did not extend to the multivariate regression or any other part of the paper. I made it clear in my Sept 25 response to Heimsath and Whipple that any mention of the cluster of five points would be removed from the proposed revision.

Comment: 6.10. I do not understand why the correlation would shut off on north-facing slopes. Is there a mechanistic/theoretical basis for this? If not than the relationship is purely empirical.

Response: My reanalysis of the data shows that, contrary to my *ESurfD* paper, the null hypothesis that P_0 values are independent of slope aspect cannot be rejected. In my earlier analysis I extracted slope aspect using the location data provided by Heimsath et al. (2012), which identify sample locations to an accuracy of approximately 10-30 m. Ten of the sample locations were ridgetops where the local slope is zero and slope aspect should be undefined. However, my initial extraction routine did not account for the local slope, hence my routine returned a slope aspect close to directly south- or directly north-facing for some of these ridgetop samples. When the data are reanalyzed to include only areas that are not ridgetops, P_0 values are slightly higher, on average, on south-facing slopes, but the null hypothesis that P_0 values are independent of slope aspect cannot be rejected. All of the discussion of aspect has therefore been removed from the proposed revision.

Comment: 6.20. Some more non-standard statistical machinations. The author does a regression that suggests that the power law exponents of A and D are 1.1 +/- some error. Then he reanalyzes things assuming that they are 1 to determine the value of c – the constant in front of A and D in Eq. 2. I am at a loss here. I know the author to be very bright and competent quantitatively. Yet here he invoking using some unnecessary, non-standard, and potentially misleading steps to avoid what would be a fairly straightforward multiple regression analysis of all of the parameters (slopes and intercepts) implied by a power law formulation of Equation 2. Doing this in a more standard way would yield some very useful metrics like partial correlation coefficients and leverage plots. Perhaps his approach seemed easier to explain at the time he wrote it. But I would argue that the community deserves and expects more.

Response: When a power-law relationship has an exponent of 1.1 +/- 0.3, I think it is appropriate to assume a linear relationship for simplicity (since 1.1 and 1.0 are indistinguishable, within uncertainty). However, this text has been removed as it is no longer included in the revised analysis.

Comment: 7.1-7.2. This is actually not a very good correlation for predicted versus observed – especially since it is strangely for a log-log plot. To understand this, look at the plot. There is almost an order of magnitude of variation in predicted P0 at any given value of P measured. To evaluate this model, rather than see an Rˆ2 for a log-log observed versus predicted plot, I think we need to see something like a Nash-Sutcliffe statistic, which would tell us how good the model is compared to simply assuming that we could use the average P measured to estimate P everywhere.

Response: As I have already noted, the Nash-Sutcliffe statistic does not apply to regression models, and I have provided the closest analogous statistic (R^2). I don't know what the reviewer means by a "strange" log-log plot. When plotting data that have a large positive skew, it is common to plot log-log simply so that the points that would otherwise cluster in the lower left corner of a linear-linear scale can be resolved in the graph.

Comment: 7.5 What are the assumptions inherent in simplifying the equations in this way? Simply citing off to previous work here is not sufficient. What are the assumptions inherent in doing this? For equation 6 you assume slopes are planar, right? Is that reasonable here? What are the limitations of removing the higher order terms of Roering et al.?

Response: I have clarified the assumption as follows: "Equation (8) assumes that the mean slope gradient at the base of hillslopes (where sediment leaves the slope) can be approximated by the average slope, S_{av}." Roering et al. proposed that sediment flux is proportional to slope with a one minus slope squared term in the denominator. If the divergence of the flux is computed, the result is a complex expression with higher-order terms, but I am using the same equation Roering et al. proposed for flux. I am happy to clarify further but I would need more information from the reviewer to do so.

Comment: 7.19. Why 0.03? Just because this is the minimum finite thickness measured? But the whole point is they have no thickness!!! The mathematical inconvenience of having a value of 0 on what you want to plot on a log scale does not justify making up a value that ***drives*** a regression that you then plot through the data. Importantly it is very true that these points have a lot of leverage on the regression. Since calculating understanding the relationship between h and S is vital to calculating E from topography, this ends up being key to the paper. And I really do not think it is well justified.

Response: These locations have no thickness *today* but must have episodically had soil in the past or else they would never erode (absent landsliding in bedrock or intact regolith, which can certainly occur but are not widespread in granitic rocks). It is common practice to add a small constant (comparable to the uncertainty of the data) prior to performing a linear regression of log-transformed data. I don't think the alternative (leaving out these values entirely from the analysis, thereby biasing the results to those with finite soil thickness) is a better choice. If the reviewer would please provide a suggestion as to how these data could be included in a way that would satisfy him, I would be willing to try whatever alternative he proposes.

Comment: 9.8. If this is the key result, then you need to demonstrate it using more conventional statistical approaches. A multiple linear regression of the log of P versus log D and h and log A would be a good place to start. This would avoid the strange – and thus hard-to-justify – correction of P to P0 that you have employed here. It would also avoid the strange practice of finding a 1.1 +/- error power slope and then redoing the regression assuming the slopes are 1 to find the best fit intercept term. This whole analysis seemed like a contorted and potentially error-prone way of doing what could have been a textbook application of multiple linear regression analysis on transformed variables.

Response: In the proposed revision I have used conventional statistical approaches throughout.

Comment: 10.12. This is misleading at best. I see a factor of 2 to 3 in either direction, so a factor of 4 to 6 overall. For example, in Fig 3D, at a value of P0 observed of _150 m/My I see a range of predicted values running from 85 to 450 m/My. That's a factor of nearly 6 range in predictions for a single value of P observed. That is NOT a good prediction in my book and my assertion is asserted by the very low N-S statistic for this modeling exercise.

Response: The sentence is correct as stated. When saying that a prediction is correct to within a factor of 2 from the observed value for 72% of the data points, that includes differences both above and below the prediction (resulting in a factor of 4 difference between the max and min predictions at a given observed value of P_0). However, I have removed the sentence because it is not central to the argument.

Proposed revision to Section 2.1:

2.1 Controls on potential soil production rates in the SGM

P_0 values for the SGM can be estimated using the residuals obtained from the regression of soil production rates to soil thicknesses reported by Heimsath et al. (2012) (their Fig. 3). The exponential form of the soil production function quantifies the decrease in soil production rates with increasing soil thickness:

$$P = P_0 e^{-h/h_0}, \tag{1}$$

where h is soil thickness and h_0 is a length scale quantifying the relative decrease in soil production rates for each unit increase in soil thickness. Heimsath et al. (2012) obtained $h_0 = 0.32$ m for locations with an average slope, S_{av}, of less than or equal to 30° and $h_0 = 0.37$ m for locations with $S_{av} > 30°$. S_{av} is defined by Heimsath et al. (2012) as the average slope over hillslopes adjacent to each sample location. P_0 values (Supplementary Table 1) can be estimated as the residuals obtained by dividing P values by the exponential term in equation (1):

$$P_{0,\mathrm{resid}} = \begin{matrix} Pe^{h/0.32\ \mathrm{m}} & \text{if } S_{av} \leq 30° \\ Pe^{h/0.37\ \mathrm{m}} & \text{if } S_{av} > 30° \end{matrix}. \tag{2}$$

where $P_{0,\mathrm{resid}}$ denotes P_0 values estimated using the residuals of the regression. Note that equation (2) is equivalent to subtracting the logarithms of the exponential term from the logarithms of P values, since division is equivalent to subtraction under log transformation. Log transformation is appropriate in this case because P values are positive and positively skewed (i.e., there are many P values in the range of 50-200 m/Myr and a smaller number of values in the range of 200-600 m/Myr that would be heavily weighted in the analysis if the data were not log-transformed). $P_{0,\mathrm{resid}}$ values estimated from equation (2) increase, on average, with increasing S_{av} (Fig. 2A). $P_{0,\mathrm{resid}}$ values exhibit an abrupt increase at an S_{av} of approximately 30°.

Heimsath et al. (2012) did not include data points from locations without soil cover in their regressions because these data points appear (especially for areas with $S_{av} > 30°$) to fit below the trend of equation (1). This implies that a humped production function may be at work in some portions of the SGM. The mean value of P from areas with $S_{av} \leq 30°$ that lack soil cover is 183 m/Myr, i.e., slightly higher than, but within 2σ uncertainty of, the 170 ± 10 m/Myr value expected based on the exponential soil production function fit by Heimsath et al. (2012). As such, the evidence indicates that for areas with $S_{av} \leq 30°$, data from locations with and without soil cover are both consistent with an exponential soil production function. The mean value of P from areas with $S_{av} > 30°$ that lack soil cover is 207 m/Myr, i.e., significantly lower than the 370 ± 40 m/Myr expected based on the exponential soil production function. This suggests that a hump may exist in the soil production function for steep ($S_{av} > 30°$) slopes as they transition to a bare (no soil cover) condition. To account for this, I estimated P_0 to be equal to 1.78P (i.e., the ratio of 370 to 207) at locations with S_{av} > 30° that lack soil cover.

The SGM has horizontal compressive stresses of ~10 MPa in an approximately N-S direction at depths of less than a few hundred meters (e.g., Sbar et al., 1979; Zoback et al., 1980; Yang and Hauksson, 2013). The development of rugged topography can lead to topographically induced fracturing of bedrock and/or opening of pre-existing fractures in compressive-stress environments (e.g., Miller and Dunne, 1996; Martel, 2006; Slim et al., 2014; St. Clair et al., 2015). Given the pervasively fractured nature of bedrock in the SGM (e.g., Dibiase et al., 2015), I assume that changes in the stress state of bedrock or intact regolith beneath hillslopes leads to the opening of pre-existing fractures (i.e., an increase in the bulk porosity of bedrock or intact regolith) rather than the fracturing of intact rock. I adopt the analytic solutions of Savage and Swolfs (1986), who solved for the topographic modification of regional compressive stresses beneath ridges and valleys oriented perpendicular to the most compressive stress direction. Savage and Swolfs (1986) demonstrated that the horizontal stress (σ_{xx}) in bedrock or intact regolith becomes less compressive under ridges as the slope increases (Fig. 3). In landscapes with a maximum slope larger than 45˚ (equivalent to an average slope of approximately 27˚ or atan(0.5) in the mathematical framework of Savage and Swolfs, 1986), bedrock or intact regolith that would otherwise be in compression develops tensile stresses close to the surface beneath hillslopes (Fig. 3A). An average slope of 27˚ is close to the threshold value of 30˚ that Heimsath et al. (2012) identified as representing the transition from low to high P_0 values in the SGM. Therefore, the abrupt increase in $P_{0,resid}$ values at approximately 30˚ is consistent with a transition from compression to tension in bedrock or intact regolith beneath hillslopes of the SGM. In addition to this sign change in the horizontal stress state in the rocks beneath hillslopes of the SGM, the Savage and Swolfs (1986) model predicts a gradual decline in horizontal compressive stress as S_{av} increases between 0 and approximately 27˚ (Fig. 3B):

$$\frac{\sigma_{xx}}{N_1} = \frac{2-4S_{av}}{(2+4S_{av})(1+4S_{av})} \quad (3)$$

where N_1 is the regional maximum compressive stress and S_{av} has units of m/m in equation (3). Equation (3) is simply equation (36) of Savage and Swolfs (1986) expressed in terms of the average slope from the drainage divide to the location of maximum slope rather than the shape parameter b/a used by Savage and Swolfs (1986). Note that the tangent of the slope angle (units of m/m) is averaged to obtain S_{av} in all cases in this paper. However, after this averaging S_{av} is reported in degrees in some cases to facilitate comparison with the results of Heimsath et al. (2012).

Figure 3 illustrates the effects of topography on tectonic stresses only. Gravitational stresses can be included in the model by superposing the analytic solutions of Savage and Swolfs (1986) (their equations (34) and (35)) with the solutions of Savage et al. (1985) for the effects of topography on gravitational stresses (their equations (39) and (40)). The result is a three-dimensional phase space of solutions corresponding to different values of the regional tectonic stress N_1, the characteristic gravitational stress $\rho g b$ (where ρ is the density of rock, g is the acceleration due to gravity, and b is the ridge height), and the Poisson ratio μ. The effects of including gravitational stresses are (1) to increase the compression at depth via the lithostatic term (at soil depths this corresponds to an addition of ~10 kPa, which is negligible compared to the regional compressive stress of ~10 MPa in the SGM), and 2) to increase the compressive stresses near the point of inflection on hillslopes (e.g., Fig. 2a of Savage et al., 1985). These modifications do not alter the first-order behavior illustrated in Figure 3 for rocks close to the surface that are not close to hollows or other points of inflection. Section 3 provides additional discussion of the assumptions and alternative approaches to modeling topographically induced stresses.

The fit of the solid curve in Figure 2A to $P_{0,\mathrm{resid}}$ values is based on equation (3), together with an assumption that the transition from compressive to tensile stresses triggers an step increase in $P_{0,\mathrm{resid}}$ values over a small range of S_{av} values in the vicinity of the transition from compression to tension:

$$P_{0,S} = \begin{cases} P_{0,l}\left(1-\frac{\sigma_{xx}}{N_1}\right) \text{ if } S_{\mathrm{av}} \le S_l \\ P_{0,h}\left(1-\frac{\sigma_{xx}}{N_1}\right) \text{ if } S_{\mathrm{av}} > S_h \\ \left(P_{0,l}+\left(P_{0,h}-P_{0,l}\right)\frac{S_{\mathrm{av}}-S_{\mathrm{l}}}{S_{\mathrm{h}}-S_{\mathrm{l}}}\right)\left(1-\frac{\sigma_{xx}}{N_1}\right) \text{ if } S_l \le S_{\mathrm{av}} < S_h \end{cases} \tag{4}$$

where $P_{0,S}$ denotes the model for the dependence of P_0 values on S_{av}, $P_{0,l}$ and $P_{0,h}$ are coefficients defining the low and high values of P_0, and S_l and S_h are the average slopes defining the range over which P_0 values increase from low to high values across the transition from compression to tension. $P_{0,l}$ and $P_{0,h}$ were determined to be 170 m/Myr and 500 m/Myr based on least-squares minimization to the data (data from elevations above 2300 m were excluded because of the climatic influence described below). S_l and S_h were chosen to be 30˚ and 32˚, respectively, to characterize the abrupt increase in P_0 values in the vicinity of 30˚.

In addition to the average slope control associated with the topographically induced stress fracture opening process, a climatic control on P_0 values can be identified using cluster analysis. This type of analysis involves identifying clusters in the data defined by distinctive values of the independent variables that also have different mean values of the dependent variable. The four points colored in blue in Figure 2A are the four highest elevation samples in the dataset, with elevations ≥ 2300 m a.s.l. The logarithms (base 10) of this cluster have a mean value of -0.40 after subtracting the logarithms of $P_{0,S}$ to account for the average slope control on $P_{0,\mathrm{resid}}$ values, compared with a mean of 0.00 for the logarithms of the remaining data points with $S_{\mathrm{av}} > 30°$ (also with the logarithms of $P_{0,S}$ subtracted). Assuming a significance level of 0.05, the null hypothesis that the cluster of blue points has a mean that is indistinguishable from that of the remaining points with $S_{\mathrm{av}} > 30°$ can be rejected based on the standard t test with unequal variances (t = 0.021).

Figures 4A-4C illustrate the mean annual temperature (MAT), mean annual precipitation (MAP), and existing vegetation height (EVH) for the central portion of the SGM. Above elevations of approximately 1800 m a.s.l., vegetation height decreases systematically with increasing elevation (Fig. 4D). This limitation is likely to be primarily a result of temperature limitations on vegetation growth because MAP increases with elevation up to and including the highest elevations of the range. This result is consistent with the hypothesis that vegetation is a key driver of soil production. The decrease in P_0 values with elevation is likely to be gradual rather than abrupt, and indeed there is evidence of a peak in the climatic control of P_0 values. Figure 4E plots the ratio of $P_{0,\mathrm{resid}}$ to $P_{0,\mathrm{S}}$ as a function of elevation. The closed circles are binned averages of the data (each bin equals 100 m in elevation). The ratio of $P_{0,\mathrm{resid}}$ to $P_{0,\mathrm{S}}$ (equivalent to the residuals under log transformation after the effects of average slope are removed) increases, on average, and then decreases within the range of elevations between 1500 and 2600 m, broadly similar to the trend of EVH (Fig. 4D).

Local variability in P_0 estimates due to variations in soil thickness, mineralogical variations within a given lithology, spatial variations in fracture density, etc. can be minimized by averaging P_0 values (not including the four highest-elevation points because of the climatic control) from locations that have the same average slope (Fig. 2C). This process tends to average data from the same local cluster since local clusters often have average slopes that are both equal within the cluster and different from other clusters. Figure 2C demonstrates that the predictions of the topographically induced stress fracture opening hypothesis are consistent with the observed dependence of $P_{0,\mathrm{resid}}$ values on S_{av} values.

The average slope and climatic controls on P_0 values can be combined into a single predictive equation for P_0 values:

$$P_{0,\mathrm{pred}} = P_{0,s}C \tag{5}$$

where $P_{0,pred}$ denotes predicted values for P_0, C is a climatic index defined as 1 for $z < 2300$ m and 0.4 (i.e., the ratio of the geometric mean of the data for $z > 2300$ m to the remaining data points with $S_{av} > 30°$) for $z > 2300$ m. A regression of $P_{0,pred}$ values to $P_{0,resid}$ values yields an R^2 of 0.50 (Fig. 2D). When data with equal S_{av} values are averaged (i.e., the filled circles in Fig. 2D), the resulting R^2 value is 0.87.

The results of this section demonstrate that average slope and climate exert controls on P_0 values in the SGM. Although I did not find additional controls that were clearly distinct from these, it is worth discussing additional controls that I tested for. The data points colored in gray in Figure 2B are from the three rock types most resistant to weathering as determined by Spotila et al. (2002): granite, anorthosite, and the Mount Lowe intrusive suite. Spotila et al. (2002) also identified gabbro as a relatively resistant rock in the SGM, but no soil production rates are available from this rock type. Figure 2B suggests that lithology might exert some control on P_0 values. Specifically, 7 samples from the more resistant lithologies sit above the least-squares fit of equation (4) to the data, while 13 (including the 7 lowest P_0 values) sit below the least-squares fit. However, the null hypothesis that the residuals of the gray cluster after the effects of average slope are removed has a mean that is indistinguishable from the residuals of the remaining points (colored black in Figure 2B) cannot be rejected ($t = 0.21$).

Many studies have proposed a relationship between fracture density and bedrock weatherability on the basis that fractures provide additional surface area for chemical weathering and pathways for physical weathering agents to penetrate into the bedrock or intact regolith (e.g., Molnar, 2004; Molnar et al., 2007; Goodfellow et al., 2014; Roy et al., 2016a,b). The difference in erosion rates between the SGM and adjacent San Bernadino Mountains, for example, has been attributed in part to differences in fracture density between these ranges (Lifton and Chase, 1992; Spotila et al., 2002). As such, it is reasonable to hypothesize that differences in P_0 values might result from spatial variations in fracture density within each range. I computed a bedrock damage index D based on the concept that P_0 values increase in bedrock that is more pervasively fractured, together with the fact that bedrock fracture densities are correlated with local fault density in the SGM (Chester et al., 2005; Savage and Brodsky, 2011). Savage and Brodsky (2011) documented that bedrock fracture density decreases as a power-law function of distance from small isolated faults, i.e. as $r^{-0.8}$ where r is the distance from the fault. Fracture densities around larger faults and faults surrounded by secondary fault networks can be modeled as a superposition of $r^{-0.8}$ decays from all fault strands (Savage and Brodsky, 2011). Chester et al. (2005) documented similar power-law relationships between bedrock fracture density and local fault density in the SGM specifically. I define the bedrock damage index D (Fig. 5A) as the sum of the inverse distances, raised to an exponent 0.8, from the point where the D value is being computed to every pixel in the study area were a fault is located:

$$D = \sum_{\mathbf{x}'} \Delta x \left(\Delta x / \left| \mathbf{x} - \mathbf{x}' \right| \right)^{0.8} \tag{6}$$

where Δx is the pixel width, **x** is the map location where bedrock damage is being computed, and **x'** is the location of each mapped pixel in SGM where a fault exists. D has units of length since it is the sum of all fault lengths in the vicinity of a point, weighted by a power function of inverse distance. Equation (6) honors the roles of both the distance to and the local density of faults documented by Savage and Brodsky (2011) because longer faults and/or more mature fault zones with many secondary faults have more pixels that contribute to the summation. The fact that a relationship exists between $P_{0,\mathrm{resid}}$ values and D (Fig. 5B, $p = 0.035$) and between D and S_{av} (Fig. 5C, $p = 0.015$) suggests that some of the control by average slope that I have attributed to the topographically induced stress fracture opening process may reflect differences in the density of pre-existing fractures related to local fault density. However, the much higher R^2 value of the relationship between $P_{0,\mathrm{resid}}$ and $P_{0,\mathrm{pred}}$ ($R^2 = 0.50$) compared to that for the relationship between $P_{0,\mathrm{resid}}$ and D ($R^2 = 0.08$) suggests that the topographically induced stress fracture opening process is the dominant mechanism controlling P_0 values in the SGM. In addition, this process has a stronger theoretical foundation.

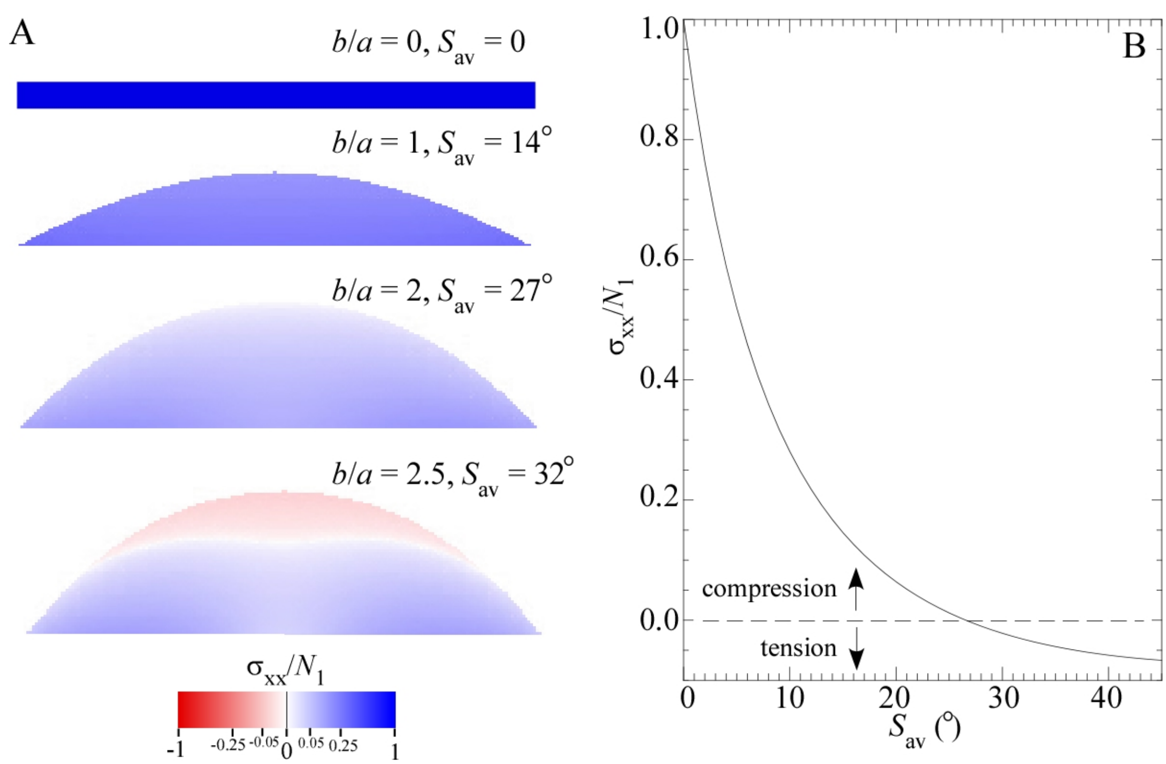

Figure 2. Analytic solutions illustrating the perturbation of a regional compressive stress field by topography. (A) Color maps of the horizontal normal stress, σ_{xx} (normalized to the regional stress, N_1), as a function of ridge steepness (defined by the shape factor b/a of Savage and Swolfs (1986) and the average slope S_{av}) using equations (34) and (35) of Savage and Swolfs (1986). The hillslopes are plotted with no vertical exaggeration. (B) Plot of σ_{xx} directly beneath the ridge as a function of S_{av} using equation (36) of Savage and Swolfs (1986). The plot illustrates the decrease in compressive stress with increasing average slope and the transition to tensile stresses at a S_{av} value of approximately 27°.

,

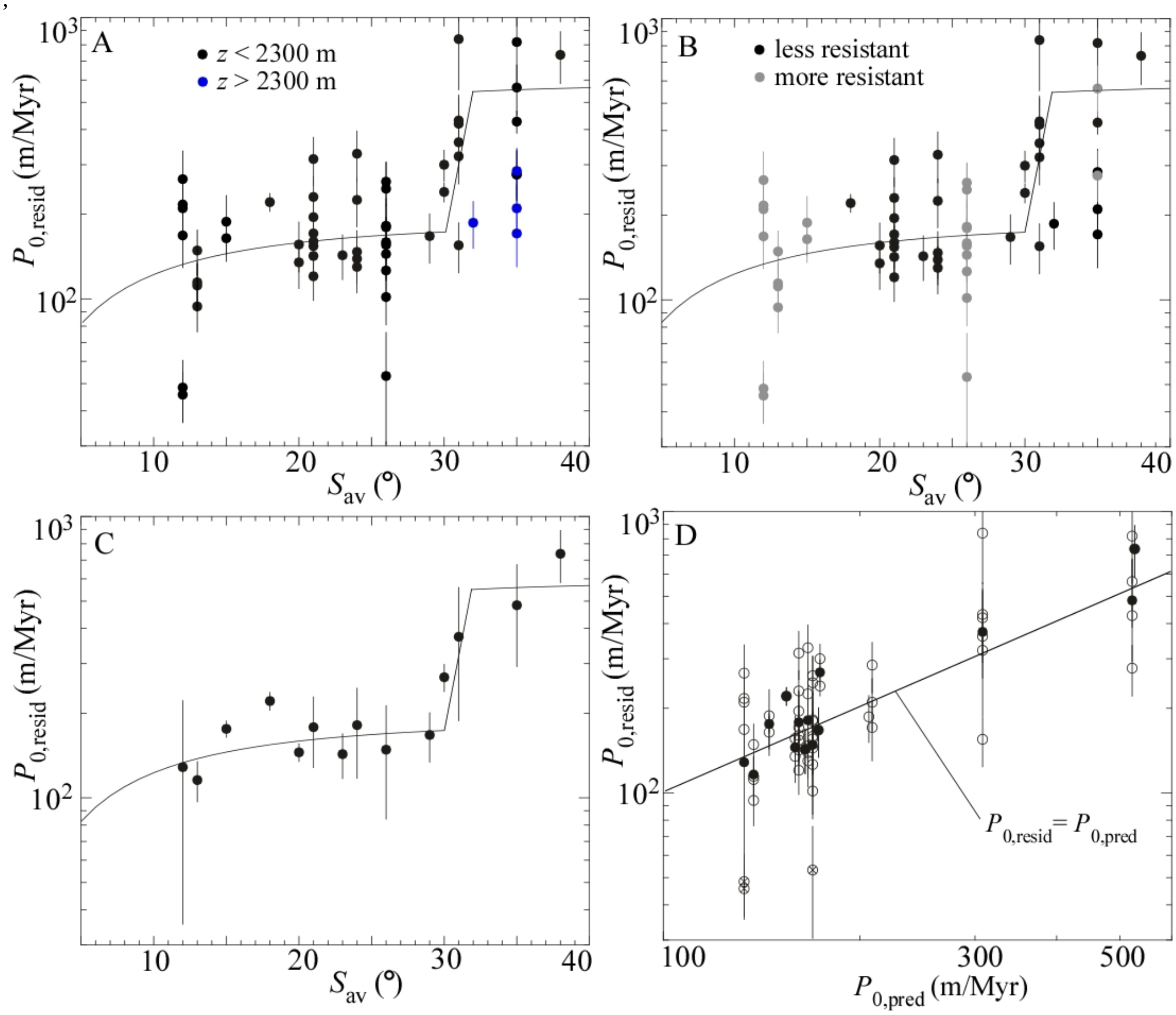

Figure 3. Plots of $P_{0,resid}$ and their relationship to average slope, S_{av}, and other potential controlling factors. (A) Plot of $P_{0,resid}$ values versus S_{av}. Data points colored blue are from the highest elevations of the range (z > 2300 m). (B) The same plot as (A), except that data points are colored according to whether they from rocks that are relatively more resistant (gray) or less resistant (black) to weathering. (C) Plot of $P_{0,resid}$ values averaged for each value of S_{av}. In (A) and (B), error bars represent the uncertainty of each data point, while in (C) the error bar represents the standard deviation of the data points averaged for each S_{av} value. (D) Plot of $P_{0,resid}$ versus values predicted from equation (5). Unfilled circles show individual data points, while filled circles represent the averaged data plotted in (C).

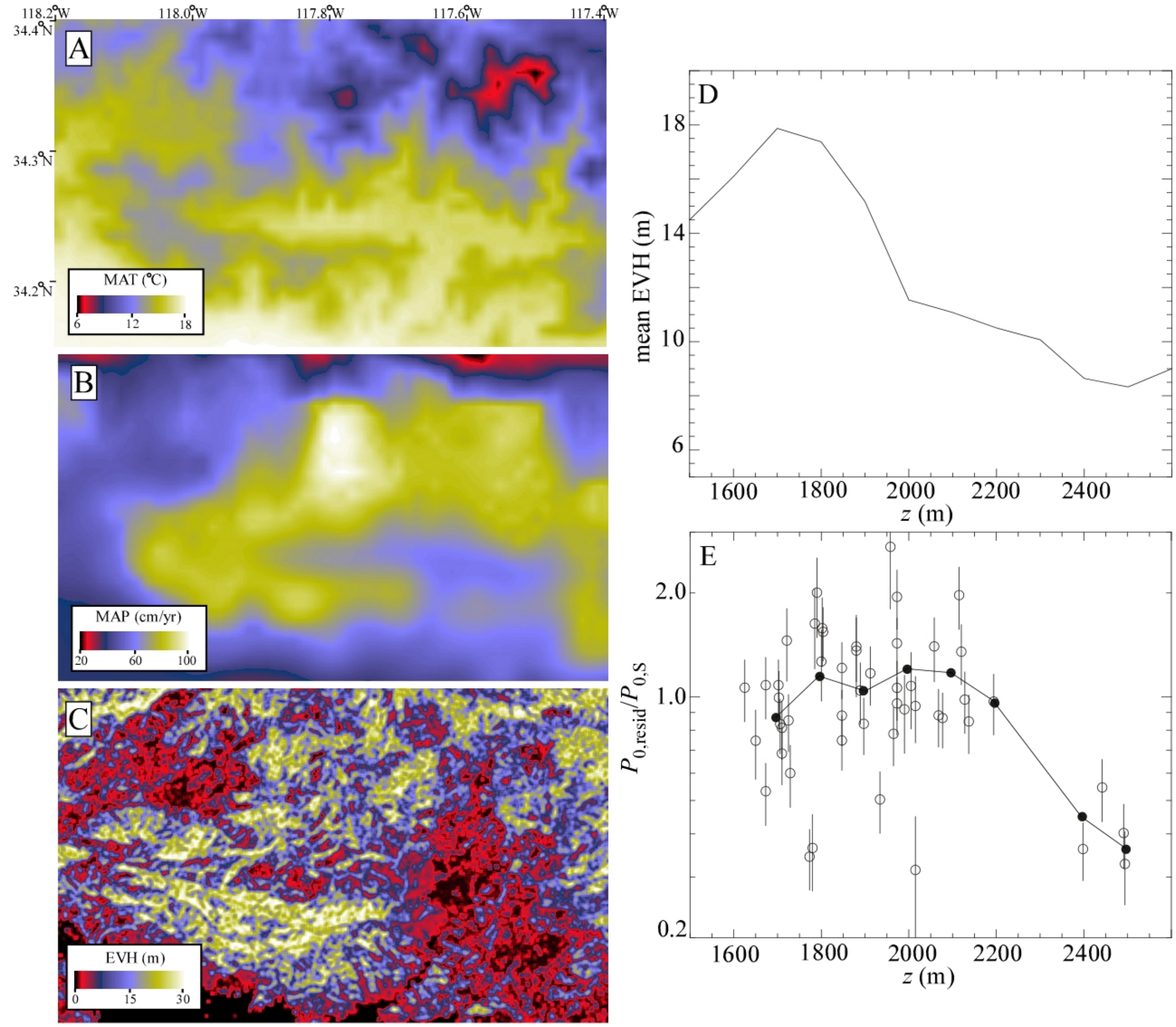

Figure 4. Climate and vegetation cover of the central San Gabriel Mountains. Color maps of (A) mean annual temperature (MAT) and (B) mean annual precipitation (MAP) from the PRISM dataset (Daly et al., 2001). (C) Color map of mean existing vegetation height (EVH) from the U.S. Geological Survey LANDFIRE database (U.S.G.S., 2016). (D) Plot of mean EVH versus elevation above sea level, *z*, using the data illustrated in (C). (E) Plot of the ratio of $P_{0,resid}$ to $P_{0,S}$ as a function of elevation. Filled circles are binned averages of the data (each bin equals 100 m in elevation).

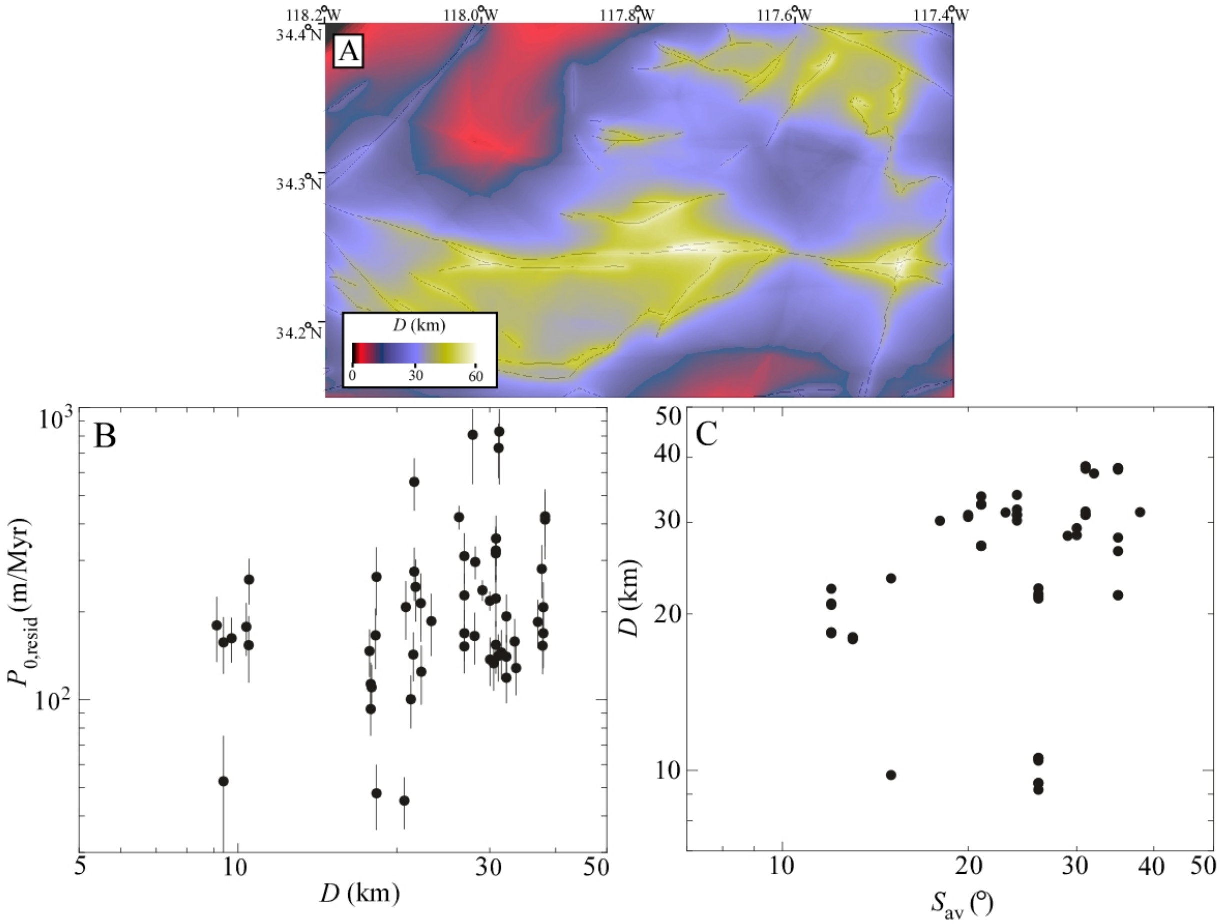

Figure 5. Map of the bedrock damage index, *D*, and its correlation with S_{av}. (A) Color map of spatial variations *D*. (B) Plot of *D* versus S_{av} for the 57 sample locations of Heimsath et al. (2012).