“The Landscape evolution modeling, on which the authors’ interpretations rely on heavily, is wrong. Details are given further on in this review.”

In the Discussion, we relied on the landscape evolution model to:

- Support inferences from other observations
- Suggest that box canyon formation is best represented by a linear diffusive model
- Estimate the ratio of volume of water to the volume of eroded material

The landscape evolution modelling was a useful addition to the study, but the study’s objectives could be addressed without relying on landscape evolution modelling.

We would like to point out that re-running the landscape evolution model simulations as described below would take ~30 days.

“The authors use two models to simulate the erosion of the Canterbury Coast. The first is the stream power model and the second is a linear diffusion model. Neither of these is likely appropriate to simulate the erosion here, nor do the authors provide the necessary analysis to justify their use. First, the stream power model is really for detachment-limited landscapes. However, the authors do not make a case that the erosion and sediment transport conditions here justify the use of the stream power model. Particularly, as the presence of major slope failures and alluvium in the channels suggests that a significant sediment cover effect may be happening here.”

In the paper we were testing which model, between stream power and linear diffusion, better fits the observations to infer the geomorphic process responsible for canyon formation. We agree with the reviewer that the stream power model is not appropriate to represent the evolution of the box canyons. We can make this clearer in the methodology and discussion.

“Next, the authors use a linear diffusion model as another endmember. However, with the presence of major slope failures, which are significantly non-local and thus not diffusive, it is hard to justify its use and the authors do not provide this justification in the manuscript. The discussion section completely disregards this fact and overall is not convincing.”

The mass wasting flux is taken to be linearly proportional to the topographic slope at low slopes. A non-linear component can be incorporated to take into account episodic landsliding processes that occur on steeper slopes where the traditional linear formulation may break down.

The diffusion model (Howard, 1994, 2007; Forsberg-Taylor et al., 2004; Luo and Howard, 2008; Barnhart et al., 2009, Boatwright and Head, 2019) can be expressed as:

$$\frac{\partial \eta}{\partial t} = D_l S + D_n \left(\frac{1}{1 - (\frac{S}{S_t})^2} - 1\right)$$

where t is time [T], D_l and D_n are the diffusion coefficients [L 1-3m T m-1], and S_t is the threshold value or critical slope gradient. S is:

$$S = \nabla^2 \eta$$

Where S is slope (dimensionless) and η is the topographic elevation [L].

If S << S_t, then the non-linear part of the equation can be considered negligible (i.e. a very small value compared to the linear part), converting the equation above into a linear diffusion model (equation 9).

$$\frac{\partial \eta}{\partial t} = D_l S$$

The slope of the Canterbury Plains is close to 0°, and we assume that the threshold value or critical slope gradient is much higher than the 70° slope of the cliff.

"Another problem in the methodology occurs with the values used in the model. One problem is when the authors equate the values for K and D in the stream power model and the linear diffusion model. These values represent very physically processes and are likely very different, by orders of magnitude in some cases. Their use of also assuming that K/D is proportional to the seepage flux and surface water shear stress, while I am open to the idea, isn't backed up with a reference or proof of concept."

We used similar values for K and D for the sake of simplicity and to easily compare the results. We will use different values (based on sensitivity analyses – see below) in the revised simulations.
The assumption that K/D is proportional to the seepage flux and surface water shear stress is based on Howard (1988) and Luo and Howard (2008). This will be clarified in the text.

"Next the authors state that they obtain the values for M and Tau_t "by trial and error." What does this mean? Did they just pick values until they obtained behavior they wanted?
Because the authors picked values by trial and error, rather than by constraining them from some means, it is not clear if the modeling results presented are meaningful.
Normally, when one does not know what input parameters should be, a sensitivity analysis is needed to evaluate the outcomes from a full range of possible values for a specific parameter.
Finally, there is no sensitivity analysis given on how model outputs change with change in the input parameters. This is needed to interpret the results from landscape evolution modeling."

We will carry out a sensitivity analyses that we will include in the revised version of the manuscript.

"On line 297, it says that the authors are testing the models, but these methods are not really a formal test of these LEMs as currently written."

This should have been written to indicate that we were using the models to test which erosional process forms the canyons, rather than testing the models themselves. This will be corrected in the revised version of the manuscript.

"One line 337: Actually, there probably is a way in landlab to incorporate sand lenses, but this would essentially involve developing a new landlab component to do so."

This is correct. We should have written:
"Since Landlab is a 2-D modeling environment, it was not possible to include 3-D sand lenses in its current version. For this reason, we simplified the geology by considering two types of sedimentary units: sandy gravels with a NW-SE strip of sand."

"Why is Equation 10 the same as Equation 7?"

This was a mistake. Equation 10 will be removed.