7th August 2020

RESPONSE TO REVIEWERS LETTER

Dear Editor:

Attached please find the revised version of the manuscript 'Box canyon erosion along the Canterbury coast (New Zealand): A rapid and episodic process controlled by rainfall intensity and substrate variability' (esurf-2020-29). My co-authors and I would like to thank you and the two reviewers for the thorough review and constructive comments. We think that by taking your suggestions into consideration, the paper has been greatly improved.

What follows is a point-to-point discussion of the reviewers' comments and how these have been addressed.

Comments made by Reviewer 1:

1. *The Landscape evolution modeling, on which the authors' interpretations rely on heavily, is wrong. Details are given further on in this review. Because the evidence provided from the landscape evolution modeling is not robust, I fear the authors made need to redo their entire LEM analysis including re-performing model runs, including calibrating their parameter values and conducting sensitivity analyses. I fear that the authors will need more time than is offered during Esurf's review process. As of now, I find the authors statements in the discussion unconvincing.*

   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.

   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.

   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. 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. 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.

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.

Why is Equation 10 is the same as Equation 7?

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 an interesting addition to the study, but the study's objectives could be addressed without relying on it. In view of the above comments, I have decided to remove the sections related to landscape evolution modelling from the paper. The ratio of volume of water to the volume of eroded material is now estimated using the slope stability model. The inference on the linear diffusive model has been removed.

2. *The data presented are not fully synthesized into a clear and complete story or analysis. The challenge here is that it is hard to understand how each of the data the authors present ties into the overall picture of 'box canyon' erosion. As such, it is not clear as the paper is currently written what the contribution of the paper is towards understanding Earth surface dynamics. After reading the paper, the only things I got out of it are that the authors did a lot of work collecting a lot of data, but I can't say I understand what the contribution is. Two good examples are the luminescence ages and the geophysical resistivity surveying. The ages are used mostly in a passing way and the geophysical resistivity doesn't seem like it was very successful, so it not clear why it is included?*

In our contribution we address three points: (i) how coastal gullies are formed by groundwater erosion, (ii) the role that lithology and permeability play in gully initiation and evolution, and (iii) the temporal scale of gully formation. Because of the reliance on experiments and numerical models, and the paucity of process-based observations and instrumental analyses, previous studies had not thoroughly addressed these issues in the past. By integrating field data with modelling, we show that gully formation is an episodic process associated to groundwater flow that occurs when a threshold rainfall intensity is exceeded. The gullies in our study area are actively eroding, and erosion rates can be as high as 30 m per day. Hydraulically-conductive zones (e.g. relict braided rivers channels, tunnels) and sand lenses exposed at the cliff face control the location of the gullies. We believe that these outcomes are important because: (i) there are only a few places where the mechanisms by which seepage erosion occurs have been clearly defined, and (ii) the results can be used to test and quantify models of gully formation and improve the reconstruction/prediction of landscape evolution by groundwater-related processes.

In the revised version of the manuscript we have updated the Discussion by:

- Subdividing the section into 3 sub-sections that directly address each individual objective
- Referring to the observations, figures and tables in the Results section to support the inferences made in each section

The OSL ages are used to demonstrate that the largest gullies have been inactive in the last 2 ka. This contrasts with the inference by Schumm and Phillips (1986) that they were formed by spillage of water from swamps behind the cliffs in the 19th century. It also suggests that they may have formed by processes other than groundwater flow, and may explain why the longer gullies have different morphometrics in comparison to the smaller (younger) gullies.

The comment about the geophysical resistivity is addressed in point 9.

3. *The term Box Canyon doesn't bring to mind the features that the study focuses on. I understand box canyons to be canyons with vertical walls, flat bottoms, and generally ingress and egress is only possible through the mouth of the canyon. The classic box canyon are the narrow canyons typically associated with the arid western United States. I think what the authors are studying are actually gullies that progress into canyons when they collect enough drainage area. As the authors note in Lines 383-385, the younger and lower-discharge/drainage areas/stream orders of these features have V and U shaped cross sectional profiles. Whereas the typical box canyons I have seen in the field continue to have flat bottoms and vertical walls even at low stream orders. I think this will be a problem for the authors in that researchers looking for information on box canyons will not consider this work relevant, and those studying coastal gully erosion will miss this paper due to the title and mislabeling of the features of interest. Instead, I think the authors should rewrite this paper while avoiding the use of the term box canyon. Either referring to these as coastal cliff gullies, headward cliff erosion, or some new term the author's come up with.*

The reviewer's point is valid. We now refer to these landforms as coastal gullies in the revised version of the manuscript. We have also revised the Introduction to include a section on the state of the art and gaps in knowledge related to coastal gully research.

4. *These values don't seem all that great to me. I generally try changing my protocol if the dose recovery ratio comes out worse than 1.10 or 0.9. What could be causes of this not-great dose recovery? Is this typical for feldspar from New Zealand?*

So far, several studies reported poor dose recovery results for $pIRIR_{290}$ (e.g. Stevens et al., 2011; Thiel et al., 2011, 2014; Lowick et al., 2012; Roberts et al., 2012; Murray et al., 2014, Veres et al., 2019; Avram et al., 2020). Usually, these poor dose recovery ratios were greater than unity. On the other hand, Sohbati et al. (2016) reported good dose recovery ratios for K-feldspars using $pIRIR_{290}$ extracted from palaeorockfall boulders from South Island, New Zealand.

Qin and Zhou (2012) suggested that the dose recovery ratio is dependent on the test dose size. Yi et al. (2016) concluded that the best dose recovery ratios are found when the test dose ranges between 15 and 80% of the total dose. Here, following this suggestion, dose recovery measurements were carried out using a test dose ranging between 20 and 60% of the total

dose. On the other hand, a recent study by Avram et al. (2020) reported poor dose recovery ratio using a test dose of 41% of the total dose on polymineral fine grains extracted from Serbian loess.

As the reviewer suggested, two protocols are actually being used, namely $pIRIR_{290}$ and $pIRIR_{225}$. As presented in the manuscript "the ratios between the recovered and given dose for polymineral fine grains are 1.11±0.03 (NZ13A) and 1.01±0.03 (NZ14A) using $piRIR_{225}$ protocol and 1.14±0.05 (NZ13A) and 1.03±0.06 (NZ14A) using $pIRIR_{290}$ protocol. In the case of coarse K-feldspars, the dose recovery ratios are 0.96±0.01 (NZ13A) and 0.87±0.03 (NZ14A) using $pIRIR_{225}$ protocol, while those obtained by applying $pIRIR_{290}$ protocol are 1.17±0.06 (NZ13A) and 1.16±0.04 (NZ14A)." Please note that in the case of the $pIRIR_{225}$ protocol, the results are generally satisfactory – they are within 1σ uncertainty and they are consistent with the 0.9-1.1 interval. We have stated that "We conclude that the overall behaviour of the $pIRIR_{225}$ protocol is satisfactory whereas the $pIRIR_{290}$ protocol overestimates the given dose by ~ 17%."

Some authors (e.g Stevens et al., 2011; Alexanderson and Murray, 2012) concluded that poor dose recovery ratios can be the result of the incorrect residual dose estimation.

In order to avoid potential complications related to laboratory bleaching of the natural samples, the dose recovery test could be performed by adding a beta dose on top of the natural dose. Then, the dose recovery ratios are obtained by dividing the measured dose by the sum of the natural and the given dose.

Such a dose recovery test was performed in the case of $pIRIR_{290}$ protocol by adding a beta dose of 100 Gy on top of natural dose. The results are presented in column 4 of the table below and compared with the previously reported values stated above and presented in the manuscript (column 3).

Sample code	Grain size	Recovered/given ratio (residual dose subtraction)	Recovered/given ratio (100 Gy on top of natural)
NZ13A	4-11 μm polymineral	1.14±0.05	1.12±0.05
NZ14A	4-11 μm polymineral	1.03±0.06	1.15±0.03
NZ13A	63-90 μm K-feldspar	1.17±0.06	1.03±0.03
NZ14A	63-90 μm K-feldspar	1.16±0.04	1.02±0.02

For coarse K-feldspars, the dose recovery ratio improved from 1.17±0.06 to 1.03±0.03 for sample NZ13A and from 1.16±0.04 to 1.02±0.02 for sample NZ14A. In the case of polymineral fine grains, the dose recovery ratio for sample NZ13A was still slightly overestimated (1.12±0.05), while for sample NZ14A, the dose recovery ratio increased from 1.03±0.06 to 1.15±0.03. As such there is no clear trend. We can only conclude that while the results are not ideal, they are acceptable, and the used protocols are providing the best attainable results for these samples given state-of-the-art available methods (see also answers below regarding changing the preheat temperature).

5. *Those g-values might not be insignificant. If you perform the fading correction on the k-feldspar for NZ14A, do the ages agree with the polymineral?*

The g-values are not significant in our view. We have discussed this in detail in Avram et al. (2020), where we are presenting more extended fading measurements, including on calibration quartz. Consequently, as we do not consider these short-term measurements to be reliable and we do not consider the values to be significant, we are not presenting corrected ages. For the sake of the exercise, following the reviewer's suggestion, we have calculated the corrected K-feldspars ages ($pIRIR_{225}$) for both samples NZ13A and NZ14A. The results are presented in the next table:

Sample code	g-value (%/decade)	Uncorrected Age (ka)				Corrected Age (ka)
	K-feldspar ($pIRIR_{225}$)	pfg ($pIRIR_{225}$)	K-feldspar ($pIRIR_{225}$)	pfg ($pIRIR_{290}$)	K-feldspar ($pIRIR_{290}$)	K-feldspar ($pIRIR_{225}$)
NZ 13A	0.60±0.51	16±0.1	20.1±1.5	20.9±2.0	26.2±2.1	20.8±1.5
NZ 14A	0.85±0.09	4.6±0.4	1.9±0.1	6.0±0.7	3.1±0.3	2.0±0.1

The corrected pIRIR ages were calculated using the R Luminescence-package (Dietze et al., 2013) according to Huntley and Lamothe (2001) method.

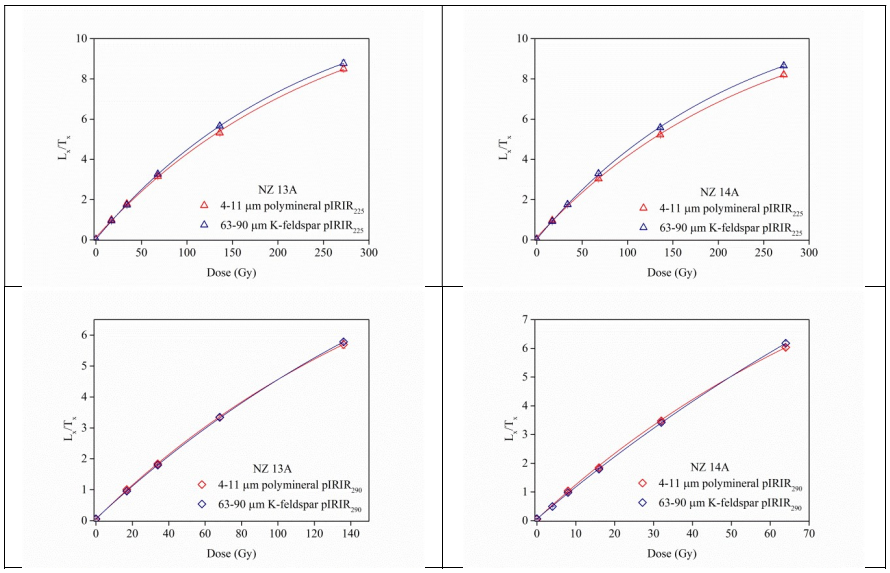
As can be seen, there is no significant difference in the ages obtained.

6. *Perhaps the discrepancy in age between the polymineral fine grain and the k-feldspar (for NZ14A) is due to the presence of unknown minerals in the polymineral aliquots? It is possible that there are minerals in there with unknown pIRIR properties that might be transferring charge in unexpected ways?*

This is indeed a reasonable assumption that is worth further study. The polymineral fraction consists of a mixture of different kinds of minerals but it is assumed that the measured emission in the violet-blue band is predominantly caused by feldspar minerals (Tsukamoto et al., 2012; Kreutzer et al., 2014). Li and Wintle (1992) investigated the stability of luminescence signal of polymineral fine grains in comparison with those from K- and Na- rich feldspars, all extracted from loess. They found that the thermal stability of the IRSL signals from polymineral fine grains extracted from different areas around the world is similar, and the IRSL signal from fine grains is less stable than the signal of K-feldspars. Tsukamoto et al. (2012) found that IRSL and pIRIR signals from polymineral fine grains are less stable than K-feldspar sample when a lower preheat temperature (260–300 °C) is used, but the stability among different types of feldspars becomes similar when a higher preheat temperature (>320 °C) is included. From a dating point of view, they suggested that the IRSL in polymineral fine grains behaves similarly to the Na-feldspars signal, rather than to the K-feldspar's signal. To our knowledge, there are few, if any, published luminescence dating studies that documented both polymineral and K-feldspars ages obtained by using pIRIR methods. Rahimzadeh et al., (2019 - conference abstract) reported polymineral fine grains and K-feldspars ages using $pIRIR_{225}$ protocol on Bavarian loess. They have found that the fading corrected ages for the

two different grain sizes agree within errors for most of the samples, observing a larger discrepancy for samples with an age above 100 ka. These kinds of studies are worth pursuing, however; in our view it would be better to attempt such studies on sites where independent age control or at least well-behaved quartz chronologies are available.

Here, we have further investigated the growth curves of polymineral fine grains and K-feldspars (see figure below). As can be seen, the growth curves overlap over the dose range investigated.



7. *One other thing that might be useful is if the authors could report the distribution of the residual dose. If there is large variance, maybe this could explain the discrepancy of the polymineral and feldspar ages.*

Residual doses measured for each individual aliquot are reported in the table below. There is no significant scatter in between the measured residual doses for each aliquot, and certainly not a difference that could account for the reported age discrepancy. Therefore, we conclude that this scenario cannot explain the observed discrepancy between polymineral and K-feldspars ages.

Sample code	Grain size	protocol	Residual dose	Average residual dose
NZ 13A	4-11 μm polymineral	$pIRIR_{225}$	4.2±0.2	3.3±0.30
			3.1±0.1	
			3.9±0.2	
			2.5±0.1	
			3.0±0.1	
NZ 14A	4-11 μm polymineral	$pIRIR_{225}$	2.7±0.2	3.0±0.2
			2.8±0.1	
			3.0±0.2	
			2.6±0.2	
			3.7±0.2	
NZ 13A	63-90 μm K-feldspar	$pIRIR_{225}$	2.7±0.1	2.9±0.1
			2.9±0.1	
			3.0±0.1	
NZ 14A	63-90 μm K-feldspar	$pIRIR_{225}$	1.1±0.05	1.3±0.1
			1.5±0.04	
			1.3±0.03	
NZ 13A	4-11 μm polymineral	$pIRIR_{290}$	7.6±0.3	10.0±1.0
			11.4±0.4	
			10.6±0.4	
			12.5±0.5	
			7.6±0.2	
NZ 14A	4-11 μm polymineral	$pIRIR_{290}$	10.4±0.4	10.8±0.9
			12.3±0.5	
			12.4±0.5	
			7.4±0.3	
			11.4±0.5	
NZ 13A	63-90 μm K-feldspar	$pIRIR_{290}$	9.7±0.4	9.0±0.3
			8.9±0.3	
			8.6±0.2	
NZ 14A	63-90 μm K-feldspar	$pIRIR_{290}$	5.2±0.2	4.9±0.2
			4.6±0.3	
			5.0±0.2	

8. *One fairly big issue is that there is limited reporting of what landform the luminescence samples were collected from. The locations are given on the aerial photo, but I cannot determine if they were collected from sediment in the channel or from the walls of the canyon, etc. Because of this, it is hard to interpret the interpretations of partial bleaching fully. As the pIR290 ages overestimate the pIR225 ages, there is likely to be partial bleaching, but how much? Are the ages themselves suspect or not? How do you know? Do you need a lower temperature pIR signal to get a better bleached age?*

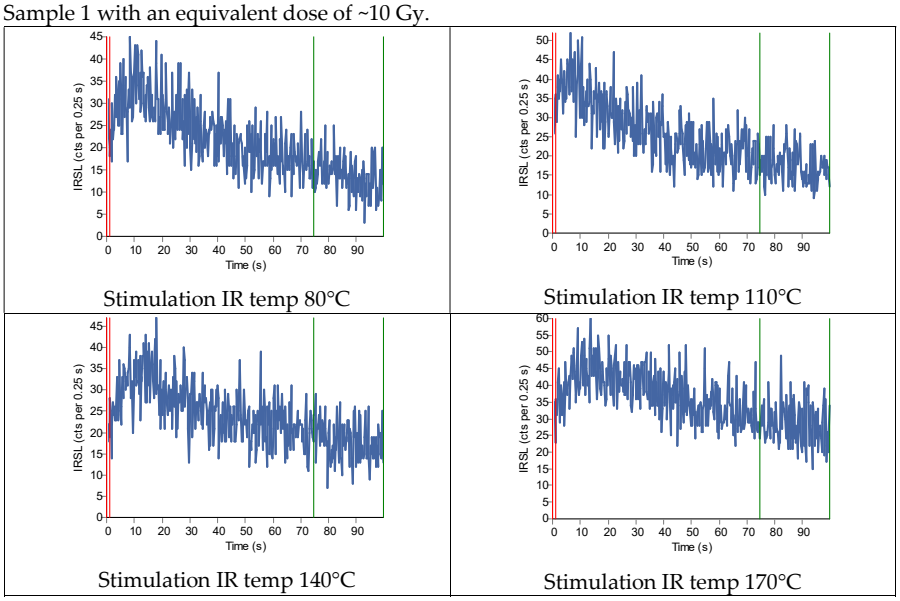
The samples were collected from the flanks of the two largest gullies, just above the boundary between the gravels below and loess above (see revised Fig. 1).

Indeed, partial bleaching is an important concern especially as pIRIR signals are known to be more difficult to bleach. Existing studies have shown that the pIRIR residual doses can range between a few Gray (<2 Gy) to 10-20 Gy or even more (Thiel et al., 2011; Stevens et al., 2011; Buylaert et al., 2011, 2012; Murray et al., 2012, 2014; Yi et al., 2016, 2018; Veres et al., 2018; Avram et al., 2020). Several studies reported a slight dependency between the natural and residual dose, the latter increasing with equivalent dose (Sohbati et al., 2012; Buylaert et al., 2012; Murray et al., 2014). Long term laboratory bleaching experiments have been conducted on loess samples from northern (Yi et al., 2016) and south-eastern (Yi et al., 2018) China. They investigated the degree to which the $pIRIR_{290}$ signal was bleachable by exposing several samples in a Honle SOL2 simulator over various periods. For the loess samples from northern China, they have shown that a constant (or highly difficult to bleach) residual $pIRIR_{290}$ signal (corresponding to a dose of 6.2±0.7 Gy) is reached after 300 h, while for the loess samples from south-eastern China a constant residual dose of 4±1 Gy was obtained after 300 h bleaching in a solar simulator.

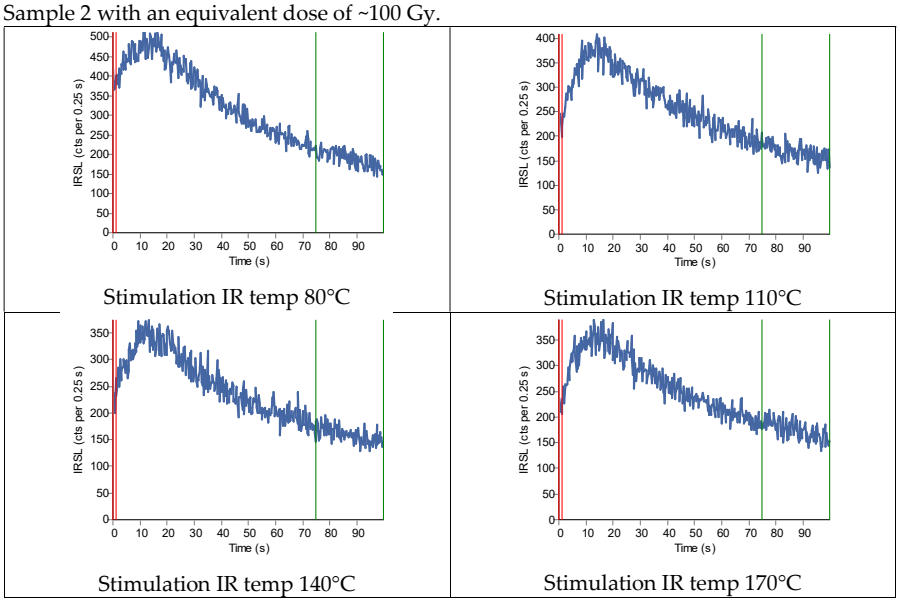
A similar experiment was conducted by our team for a sample collected from a nearby site from South Island, New Zealand (unpublished data). Groups of five polymineral fine grains aliquots were exposed to sunlight (to window – only during daylight) over different periods of time. The residual doses were measured using both pIRIR protocols. In the case of the $pIRIR_{225}$ protocol, a constant residual dose was obtained after 48 h exposure to sunlight while in the case of $pIRIR_{290}$ protocol, the residual dose appeared consistent with the constant dose after a bleaching period of 96 h. Also, we have found that the residual dose decrease to 15% of equivalent dose for $pIRIR_{225}$ after 20 minutes exposure to sunlight. We state that these results are not reported yet. However, this is in our view evidence that $pIRIR_{225}$ signals are bleached.

Also, we have tried to measure pIRIR signals using lower temperature on a nearby site from New Zealand, South Island (unpublished data). We have applied the MET protocol (Fu and Li, 2013) on polymineral fine grains extracted from two samples with no success. In the next figures we show the natural pIRIR signals measured using different IR stimulation temperatures (80°C, 110°C, 140°C and 170°C).

Sample 1 with an equivalent dose of ~10 Gy.



Sample 2 with an equivalent dose of ~100 Gy.



Moreover, the corrected IRSL signal does not grow properly with the magnitude of the irradiation dose; therefore, the construction of the dose response curve cannot be achieved.

Given these poor results on samples collected from a nearby site, the authors decided that such measurements will bring no added value for dating the samples investigated in this study.

9. *The geophysical resistivity doesn't seem like it was very successful, so it not clear why it is included?"*

The inability to fit a 1-D resistivity-vs-depth model to the observed slingram responses, coupled with the observation that the first-time-gate response profiles show considerable along-transect variability (original Figs. 9a-c), suggest that the geo-electrical subsurface structure is strongly heterogeneous within the footprint of the G-TEM transmitter. The geoelectrical heterogeneity suggested by the G-TEM responses could be caused by the resistivity signature of discrete subsurface groundwater conduits, perhaps in the form of cylindrical tubes of irregular cross-section, carrying groundwater seaward across the Canterbury Plains and discharging either at the cliffs or the beach.

This is the key information provided by G-TEM in support of the hypothesis of discrete bodies of groundwater flowing along preferential seepage pathways and supporting the growth and evolution of gullies along the coastal cliffs.

Work is currently underway, as part of a different study, on 2-D forward modelling of the slingram responses; our preliminary results (see new section in Supplementary Materials) indicate that the G-TEM data can be explained by the aforementioned electrically conductive groundwater conduits. It suffices to say here that, to depths of several tens of metres, the subsurface landward of the gullies is certainly not one-dimensional from the perspective of electrical resistivity.

Please see also response to comments 12-24.

Comments made by Reviewer 2:

10. *The paper is interesting but the geophysical approach needs more attention and some more information. Therefore, even if the authors did a nice job, they should rewrite the geophysical part or eliminate it, in order to work more on the data elaboration.*

11. *Line 153: Why the authors made the survey in these dates? I suggest to introduce if there were some climatic phenomena or some motivation about these specific dates.*

The UAV surveys were carried out after rainfall events. This is now clarified in section 3.1.2.

12. *Line 165: the investigation depth depends to the electrical characteristic of the subsoil. How you estimated this depth of investigation?*

The sentence "*All soundings were collected in the 20-gate mode with an acquisition interval of 6 × 10^{-6} s to 8 × 10^{-4} s (after ramp-off), corresponding to investigation depths of ~80 m*" was removed and replaced with: "The maximum depth of investigation of the G-TEM system is given approximately by the formula d=8.94$L^{0.4}\rho^{0.25}$ where L (m) is the TX loop size and ρ (Ωm) is the

upper layer resistivity (Geonics 2016). Setting ρ=100 Ωm yields a depth of investigation of *d*=71 m, whereas ρ=1000 Ωm yields *d*=126 m. Our investigation depth in New Zealand may be slightly greater than these values since the Geonics formula assumes a 1-turn TX loop carrying current 3 A whereas we used a more powerful combination of 4 turns at 1.5 A. "

13. *Line 167: I suggest to add some references about*

The sentence "*...using the IXG-TEM software from Interpex Limited"* has been replaced with "using IXG-TEM commercial software (Interpex 2012). This is a standard 1-D TDEM inversion code that has been successfully used for coastal hydrogeophysical studies (Pondthai et al., 2020)."

14. *Figure 8b: Which is the graph of a model with lower resistive layers? In example, did you tried to use the same of the 1D inversion result?*

After the sentence on line 464, *"As a result, we cannot trust 1-D inversions of the slingram-mode data in such a 3-D geological environment"*, the following sentence was added: "We did not try to use the 1-D inversion software to further analyse and interpret the G-TEM data. However, even though the individual slingram-mode responses cannot be fit reliably by a 1-D model, we can still analyse lateral changes in the observed response curves along the slingram profiles to reveal information about subsurface heterogeneity; this is elaborated below."

15. *Figure 8b: How have you produced these models? I think it is important to describe them or to add some references about.*

After the sentence on line 476*: "Figure 8b displays a 1-D model (A) that contains a conductive layer of 100 Ωm between 10-20 m depths in a homogeneous 1000 Ωm background"*, the following sentences were included: "The 1-D model A in Fig. 8b is motivated by the inversion results of deep-penetrating 40 × 40 m TX loop TDEM soundings carried out on top of the cliffs several tens of meters inland (Weymer et al, submitted manuscript), which revealed such a conductive zone at these depths. Unlike the slingram profiles, the deeper-penetrating, larger-loop sounding curves are readily fit by a 1-D model."

16. *Line 473: Which are the advantages to use this approach? I suggest to describe it in detail or add some references about.*

After the sentence on line 473*: "Instead of performing 1-D inversions, we present time-gate plots for all three transects"*, the following sentences were added: "A time-gate plot is defined as a graph of the observed G-TEM voltage response, evaluated at a particular time-gate, as a function of position along a profile. Time-gate plots are a useful alternative to explore the lateral variability of the G-TEM response along a profile in the event that the sounding curves at individual stations cannot be fit with 1-D models. It is presumed that variability in a time-gate plot is correlated with lateral heterogeneity in the subsurface geoelectrical structure, since a 1-D Earth structure would yield no spatial variability in a time-gate plot."

17. *Line 477: Larger amplitude? What do you mean with larger?*

The sentence on line 477: "*This model generates a G-TEM slingram response that has a substantially larger amplitude at all time gates than does the model (B) without the conductive layer*" was changed to "This model generates a G-TEM slingram response that has a substantially larger ramp-off voltage amplitude at all time gates than does the model (B) without the conductive layer, as shown in Fig. 7b."

18. *Line 478: Before to write this sentence I suggest to add a paragraph on your approach, in order to explain this results. I suggest to add some references about this kind of results.*

After the sentence on line 478: "*Thus we regard an enhancement of response at the first time gate as indicative of a conductive zone at depth beneath the slingram station*", we added the following sentences: "The spatial analysis of time-gate plots is not a conventional approach in time-domain electromagnetics but it is somewhat analogous to the spatial analysis of apparent resistivity profiles in frequency-domain electromagnetics using terrain conductivity meters (e.g. Weymer et al. 2016), based on the idea that the G-TEM response at a fixed time-gate carries information similar to that of a terrain conductivity meter response at a fixed frequency."

19. *Line 481: The description is not clear. The used plots are associated to "first-time-gate", what do you mean? In the model you use a different plot than for the acquired data, I suggest to use the same approach in order to understand the used methodologies, which is not common.*

After the sentence on line 478: "*The first-time-gate profile of transect May15-1 is located upslope of small but recently eroded box canyons (Fig. 9a)*" we added the following sentence: "In this figure, the 'first-time-gate profile' is a plot as a function of distance along the transect of the G-TEM ramp-off voltage at time gate number 1 - as shown, for example, in Fig. 7b - at the first sampled point of the transient response immediately after the TX current has been switched off."

20. *Line 489: With all this "not shown" results, it is hard to understand your approach and the final results.*

The sentence on line 488 from: "*The fluctuating signals remain similar in shape for time-gates 2 to 6 (not shown)*" was changed to "The time-gate plots for gates 2 to 7 remain similar in shape to that of the time-gate-1 plot and hence are not shown. After time-gate 7, the time-gate plots start to lose coherence due to the low signal-to-noise ratio of the decaying RX voltage at late times after TX ramp-off."

21. *Line 493: From my point of view, this observation is not clear taking in account the GTEM results. I suggest to improve the elaboration of them. May be these results are not so consistent and compatible with the geological ones.*

The G-TEM results are now more thoroughly explained, as shown in our response to the previous comments. We also include the following sentences in the Discussion (section 5.2): "With regards to the G-TEM slingram time-gate plots (Fig. 8), we interpret the higher-amplitude responses on the time-gate-1 plots that are preferentially located upslope of recently active gullies as zones of relatively high electrical conductivity in the subsurface at

depths of ~10 m. These zones are suggestive of buried groundwater conduits made up of gravel and/or sandy units (e.g. Weymer et al., in review), or tunnels formed by sub-surface groundwater flow in sand units. Further analysis of the G-TEM data, including 2-D modelling and inversion, is required to ascertain the sub-surface hydraulic geometry responsible for the along-profile amplitude variations. This is elaborated further in the Supplementary Materials... If seaward-directed groundwater conduits are responsible for the location of gullies, the G-TEM results predict that, along the Canterbury coast, we should generally observe active gully development downslope of peaks in slingram time-gate plots. If this is the case, G-TEM could be used to identify locations of incipient and even future gully development."

22. *Figure 9: I suggest to define the x legend...If I think it is meters.*

The sentence "A yellow line marks a slingram transect, the length of which can be determined from the scale bar" was added to the caption of the original Fig. 9.

23. *Line 654: From my point of view, this geophysical results are not so clear. What do you mean with "higher amplitude"? I think the paper needs more detail on the geophysical approach, in order to write this sentence.*

The sentence on line 654: "*We interpret the higher amplitude response in G-TEM profiles located upslope of recently active box canyons as buried groundwater conduits...*" was changed to " With regards to the G-TEM slingram time-gate plots (Fig. 8), we interpret the higher-amplitude responses on the time-gate-1 plots that are preferentially located upslope of recently active gullies as zones of relatively high electrical conductivity in the subsurface at depths of ~10 m. These zones are suggestive of buried groundwater conduits..."

24. *Line 656: I agree with you. Therefore, I suggest to obtain these results and after to add the geophysical data in this paper.*

After the sentence on line 656: "*Further analysis of the G-TEM data, including 2-D modelling and inversion, is required to ascertain the sub-surface hydraulic geometry responsible for the along-profile amplitude variations"*, we added the sentence "This is elaborated further in the Supplementary Materials." A new related section has been included in the Supplementary Materials that describes the G-TEM slingram modelling in detail.

We hope that we have managed to address all your concerns and that you are satisfied with the above changes. I also kindly ask the editor to update the title of the manuscript to:

Groundwater erosion of coastal gullies along the Canterbury coast (New Zealand): A rapid and episodic process controlled by rainfall intensity and substrate variability

and to remove Roger Clavera-Gispert from the author list and include Phillipe Wernette.

With best regards also on behalf of my co-authors,

Aaron Micallef

References:

Alexanderson, H. & Murray, A. S. 2012. Luminescence signals from modern sediments in a glaciated bay, NW Svalbard. Quaternary Geochronology 10, 250–256.

Avram, A., Constantin, D., Veres, D., Kelemen, S., Obreht, I., Hambach, U., Marković, S.B., Timar-Gabor, A., 2020. Testing polymineral post-IR IRSL and quartz SAR-OSL protocols on Middle to Late Pleistocene loess at Batajnica, Serbia. Boreas, https://doi.org/10.1111/bor.12442. ISSN 0300-9483.

Buylaert, J.P., Thiel, C., Murray, A.S., Vandenberghe, D., Yi, S.W., Lu, H.Y., 2011. IRSL and post IR-IRSL residual doses recorded in modern dust samples from the Chinese loess plateau. Geochronometria 38, 432–440.

Buylaert, J.P., Jain, M., Murray, A.S., Thomsen, K.J., Thiel, C., Sohbati, R., 2012. A robust feldspar luminescence dating method for Middle and Late Pleistocene sediments. Boreas 41, 435–451.

Dietze, M., Kreutzer, S., Fuchs, M.C., Burrow, C., Fischer, M., Schmidt, C., 2013. A practical guide to the R package luminescence. Ancient TL 31, 11-18.

Fu, X. & Li, S. H. 2013: A modified multi-elevated-temperature post-IR IRSL protocol for dating Holocene sediments using K-feldspar. Quaternary Geochronology 17, 44–54.

Huntley, D. J. & Lamothe, M. 2001. Ubiquity of anomalous fading in K-feldspars and the measurement and correction for it in optical dating. Canadian Journal of Earth Sciences 38, 1093–1106.

Kreutzer, S., Schmidt, C., DeWitt, R., Fuchs, M., 2014. The a-value of polymineral fine grain samples measured with the post-IR IRSL protocol. Radiation Measurements 69, 18-29.

Lowick, S. E., Trauerstein, M. & Preusser, F. 2012. Testing the application of post IR-IRSL dating to fine grain waterlain sediments. Quaternary Geochronology 8, 33–40.

Li, S.-H., Wintle, A.G., 1992. A global view of the stability of luminescence signal from loess. Quaternary Science Reviews 11, 133-137

Murray, A. S., Schmidt, E. D., Stevens, T., Buylaert, J.-P., Markovi_c, S. B., Tsukamoto, S. & Frechen, M. 2014. Dating Middle Pleistocene loess from Stari Slankamen (Vojvodina, Serbia) — Limitations imposed by the saturation behaviour of an elevated temperature IRSL signal. Catena 117, 34–42.

Murray, A.S., Thomsen, K.J., Masuda, N., Buylaert, J.P., Jain, M., 2012. Identifying wellbleached quartz using the different bleaching rates of quartz and feldspar luminescence signals. Radiation Measurements 47, 688–695.

Qin, J. T. & Zhou, L. P. 2012. Effects of thermally transferred signals in the post-IR IRSL SAR protocol. Radiation Measurements 47, 710–715.

Roberts, H. M. 2012. Testing Post-IR IRSL protocols for minimising fading in feldspars, using Alaskan loess with independent chronological control. Radiation Measurements 47, 716–724.

Rahimzadeh, N., Thiel, C., Sprafke, T., Frechen, M. Conference abstract. DLED 2019, Bingen, Germany.

Stevens, T., Marković, S.B., Zech, M., Sümegi, P., 2011. Dust deposition and climate in the Carpathian Basin over an independently dated last glacial-interglacial cycle. Quaternary Science Reviews 30, 662–681.

Sohbati, R., Borella, J., Murray, A., Quigley, M., Buylaert, J.-P., 2016. Optical dating of loessic hillslope sediments constrains timing of prehistoric rockfalls, Christchurch, New Zealand. Journal of Quaternary Science 31, 678-690.

Thiel, C., Buylaert, J.-P., Murray, A. S. & Tsukamoto, S. 2011a. On the applicability of post-IR IRSL dating to Japanese loess. Geochronometria 38, 369–378.

Thiel, C., Buylaert, J.P., Murray, A.S., Terhorst, B., Hofer, I., Tsukamoto, S., Frechen, M., 2011b. Luminescence dating of the Stratzing loess profile (Austria) – testing the potential of an elevated temperature post-IR IRSL protocol. Quaternary International 234, 23–31.

Thiel, C., Horv_ath, E. & Frechen, M. 2014. Revisiting the loess/ palaeosol sequence in Paks, Hungary: a post-IR IRSL based chronology for the 'Young Loess Series'. Quaternary International 319, 88–98.

Tsukamoto, S., Jain, M., Murray, A., Thiel, C., Schmidt, E., Wacha, L., Dohrmann, R., Frechen, M., 2012. A comparative study of the luminescence characteristics of polymineral fine grains and coarse-grained K- and Na- rich feldspars. Radiat. Meas. 47, 903-908.

Veres, D., Tecsa, V., Gerasimenko, N., Zeeden, C., Hambach, U. & Timar-Gabor, A. 2018. Short-term soil formation events in last glacial east European loess, evidence from multi-method luminescence dating. Quaternary Science Reviews 200, 34–51.

Yi,S., Buylaert, J.-P.,Murray, A.S., Lu, H., Thiel,C.&Zeng, L. 2016. A detailed post-IR IRSL dating study of the Niuyangzigou loess site in northeastern China. Boreas 45, 644–657.

Yi, S.W., Buylaert, J.P., Murray, A.S., Lu, H.Y., Thiel, C., Zeng, L., 2016. A detailed post-IR IRSL dating study of the Niuyangzigou loess site in northeastern China. Boreas 45, 644–657.

570 Yi, S., Li, X., Han, Z., Lu, H., Liu, J, Wu, J., 2018. High resolution luminescence chronology for Xiashu Loess deposits of Southeastern China. Journal of Asian Earth Sciences 155, 188-197.

Groundwater erosion of coastal gullies along the Canterbury coast (New Zealand): A rapid and episodic process controlled by rainfall intensity and substrate variability

Aaron Micallef[1,2], Remus Marchis[3], Nader Saadatkhah[1], Potpreecha Pondthai[4], Mark E. Everett[4], Anca Avram[5,6], Alida Timar-Gabor[5,6], Denis Cohen[2], Rachel Preca Trapani[2], Bradley A. Weymer[1], Phillipe Wernette[7]

[1]Helmholtz Centre for Ocean Research, GEOMAR, Kiel, Germany
[2]Marine Geology & Seafloor Surveying, Department of Geosciences, University of Malta, Malta
[3]Department of Geological Sciences, University of Canterbury, Christchurch, New Zealand
[4]Department of Geology and Geophysics, Texas A&M University, Texas, USA
[5]Faculty of Environmental Science and Engineering, Babes-Bolyai University, Cluj-Napoca, Romania
[6]Interdisciplinary Research Institute on Bio-Nano-Sciences, Babeş-Bolyai University, Cluj-Napoca, Romania
[7]School of the Environment, University of Windsor, Ontario, Canada

Correspondence to: Aaron Micallef (amicallef@geomar.de)

Abstract: Gully formation has been associated to groundwater seepage in unconsolidated sand to gravel sized sediments. Our understanding of gully evolution by groundwater seepage mostly relies on experiments and numerical simulations, and these rarely take into consideration contrasts in lithology and permeability. In addition, process-based observations and detailed instrumental analyses are rare. As a result, we have a poor understanding of the temporal scale of gully formation by groundwater seepage and the influence of geological heterogeneity on their formation. This is particularly the case for coastal gullies, where the role of groundwater in their formation and evolution has rarely been assessed. We address these knowledge gaps along the Canterbury coast of the South Island (New Zealand) by integrating field observations, optically stimulated luminescence dating, multi-temporal Unoccupied Aerial Vehicle and satellite data, time-domain electromagnetic data, and slope stability modelling. We show that gully formation is a key process shaping the sandy gravel cliffs of the Canterbury coastline. It is an episodic process associated to groundwater flow that occurs once every 227 days on average, when rainfall intensities exceed 40 mm per day. The majority of the gullies in a study area SE of Ashburton has undergone erosion, predominantly by elongation, during the last 11 years, with the most recent episode occurring 3 years ago. Gullies longer than 200 m are relict features formed by higher groundwater flow and surface erosion >2 ka ago. Gullies can form at rates of up to 30 m per day via two processes: the formation of alcoves and tunnels by groundwater seepage, followed by retrogressive slope failure due to undermining and a decrease in shear strength driven by excess pore pressure development. The location of gullies is determined by the occurrence of hydraulically-conductive zones, such as relict braided river channels and possibly tunnels, and of sand lenses exposed across sandy gravel cliff. We also show that gully planform shape is generally geometrically similar at consecutive stages of evolution. These outcomes will facilitate reconstruction and prediction of a prevalent erosive process and overlooked geohazard along the Canterbury coastline.

Deleted: Box canyon
Deleted: erosion
Deleted: Roger Clavera-Gispert[1],
Formatted: Font: 10 pt
Deleted: Box canyon
Deleted: box canyon
Deleted: box canyon
Formatted: Not Highlight
Formatted: Not Highlight
Formatted: Not Highlight
Deleted: issues
Deleted: Unmanned
Formatted: Not Highlight
Deleted: and landscape evolution
Deleted: box canyon
Deleted: box canyons
Deleted: The two largest box canyons
Formatted: Not Highlight
Formatted: Not Highlight
Formatted: Not Highlight
Deleted: have not been eroded in the last 2 ka, however. Canyons
Deleted: box canyons
Deleted: box canyon formation is best represented by a linear diffusive model and geometrical scaling

1 Introduction

1.1 Coastal gullies

Gullies can be incised into coastal cliffs and bluffs in a variety of geologic settings around the world, owing their formation to a complex interaction of hydrologic, lithospheric, tectonic, and atmospheric processes. While much research has focused on gully formation and evolution in non-coastal settings in response to changes such as land cover and use, natural hazards, and/or changes in precipitation, relatively little work has focused on gully geomorphology and morphodynamics in coastal cliffs and bluffs. The most commonly accepted mechanism for coastal gully formation is through concentrated overland flow and knickpoint migration (Leyland and Darby, 2008; Leyland and Darby, 2009; Limber and Barnard, 2018; Mackey et al., 2014; Ye et al., 2013). Changes in land cover and use due to agriculture, logging, forest fire, and other factors can decrease surface roughness and increase concentrated overland flow, which, given sufficient energy and/or time, can erode a narrow section of coastal cliff and form a knickpoint. Depending on the resistive forces (e.g. geology, uplift) relative to the erosive force of the overland flow, this knickpoint will migrate inland over time, incising a gully into the cliff or bluff. Recent work has focused on modelling coastal gully formation and evolution as knickpoint migration (Limber and Barnard, 2018).

Coastal cliff stability and gully incision can be affected by processes of concentrated overland flow, quarrying by waves at the base of the cliff, and groundwater discharge (Kline et al., 2014; Limber and Barnard, 2018), although it is unclear when and where each of these factors is important (Collins and Sitar, 2009; Collins and Sitar, 2011). While overland flow is a common formation mechanism, it is possible to have coastal gullies form where the cliff is affronted by a beach, which limits the basal quarrying or notching by waves, and where there is no outward sign of overland flow. Relatively little attention has been paid to the potentially important role of groundwater as a driver of coastal gully formation and evolution, despite the potential for groundwater to affect the geotechnical properties of coastal cliffs (Collins and Sitar, 2009; Collins and Sitar, 2011).

1.2 Channel erosion by groundwater seepage

Groundwater has been implicated as an important geomorphic agent in channel network development, both on Earth and on Mars (Abotalib et al., 2016; Dunne, 1990; Harrison and Grimm, 2005; Higgins, 1984; Kochel and Piper, 1986; Malin and Carr, 1999; Salese et al., 2019). The classic model entails a channel headwall that lowers the local hydraulic head and focuses groundwater flow to a seepage face. This leads to upstream erosion by undercutting, the rate of which is limited by the capacity of seepage water to transport sediment from the seepage face (Abrams et al., 2009; Dunne, 1990; Howard and McLane, 1988). Groundwater seepage has been shown to unambiguously lead to channel formation in unconsolidated sand to gravel sized sediments (Dunne, 1990; Lapotre and Lamb, 2018) (Table 1). In sediments finer than sands, erosion is typically limited by detachment of the grains at the seepage face. In silts and clays, the permeability is so low that the groundwater discharge is often less than that required to overcome the cohesive forces of the grains (Dunne, 1990). The role of groundwater seepage and

Deleted: Whereas erosion by overland flow has been extensively studied and associated geomorphic rate laws have been developed (Sklar and Dietrich, 2001; Whipple et al., 2000), erosion by groundwater flow has received considerably less attention. Groundwater can mechanically erode or deform sediment and bedrock by removal of mass from a seepage face through flowage, frictional slipping and particle dislodgement (Dunne, 1990; Iverson and Major, 1986). Groundwater seepage can also lead to slope instability by undercutting, reduction of soil shear strength and shear surface development (Carey and Petley, 2014; Chu-Agor et al., 2008). Groundwater has been implicated as an important geomorphic agent in valley network development, both on Earth and on Mars (Abotalib et al., 2016; Dunne, 1990; Harrison and Grimm, 2005; Higgins, 1984; Kochel and Piper, 1986; Malin and Carr, 1999; Salese et al., 2019). For over one hundred years, a characteristic morphology - comprising theatre-shaped heads, steep and high valley walls, constant valley width, flat floors, low drainage densities, and short tributaries with large junction angles - has been cited as diagnostic of groundwater activity (Abrams et al., 2009; Dunne, 1990; Higgins, 1984; Russel, 1902; Schumm and Phillips, 1986). We refer to valleys with these characteristics as box canyons. ¶
¶
The classic model explaining the development of box canyons entails a canyon headwall that lowers the local hydraulic head and focuses groundwater flow to a seepage face. This leads to upstream erosion by undercutting, the rate of which is limited by the capacity of seepage water to transport sediment from the seepage face (Abrams et al., 2009; Dunne, 1990; Howard and McLane, 1988).

Deleted: box canyon

channel formation in bedrock, on the other hand, remains controversial (Lamb et al., 2006; Pelletier and Baker, 2011).

Table 1: Examples of studies documenting channels attributed to groundwater seepage erosion in unconsolidated sediments (sand- to gravel-sized).

Location	Substrate type	Reference
Alaska, USA	Gravel braided river deposits	(Sunderlin et al., 2014)
Canterbury Plains, New Zealand	Gravel braided river deposits	(Schumm and Phillips, 1986)
Florida Panhandle, USA	Non-marine quartz sands that contain discontinuous layers of clay or gravel	(Schumm et al., 1995)
Kalahari, southern Africa	Conglomerates, marls, duricrusts and unconsolidated sands	(Nash et al., 1994)
Martha's Vineyard and Nantucket Island, USA	Outwash gravelly sand	(Uchupi and Oldale, 1994)
Obara area, Japan	Granodiorite regolith	(Onda, 1994)
South Taranaki, New Zealand	Terrestrial dune sand and tephra overlying marine sands and gravels	(Pillans, 1985)
Vocorocas, Brazil	Alluvial sands	(Coelho Netto et al., 1988)

Our understanding of channel evolution by groundwater seepage is predominantly derived from theoretical, experimental and numerical models (Chu-Agor et al., 2008; Higgins, 1982; Howard, 1995; Lobkovsky et al., 2004; Pelletier and Baker, 2011; Petroff et al., 2011; Wilson et al., 2007). Such studies suggest that the velocity at which channel heads advance is a function of groundwater flux and the capacity of seepage water to transport sediment from the seepage face (Abrams et al., 2009; Fox et al., 2006; Howard, 1988; Howard and McLane, 1988), and that channel head erosion occurs by episodic headwall slumping (Howard, 1990; Kochel et al., 1985). Channels incised by groundwater seepage have been shown to branch at a characteristic angle of 72° at stream tips, which increases to 120° near stream junctions (Devauchelle et al., 2012; Yi et al., 2017), whereas growing indentations competing for draining groundwater result in periodically-spaced channels (Dunne, 1990; Schorghofer et al., 2004). Channel network geometry appears to be determined by the external groundwater flow field rather than flow within the channels themselves (Devauchelle et al., 2012).

A number of fundamental questions related to the evolution of channels by groundwater seepage in unconsolidated sediments remain unanswered. Firstly, the temporal scale at which channels form is poorly quantified due to a paucity of process-based observations and detailed instrumental analysis. Field observations of groundwater processes are rare (e.g. Onda, 1994), primarily due to the difficulty with accessing the headwalls of active channels, the potential long timescales involved, and the complexity of the erosive process (Chu-Agor et al., 2008; Dunne, 1990). Quantitative assessments of channel evolution have relied on experimental and numerical analyses, but these tend to be based on simplistic assumptions about flow processes and hydraulic characteristics. Experimental approaches are based on a range of different methods, which limits comparison of their outcomes (Nash, 1996). Published erosion rates vary between 2-5 cm per century (Abrams et al., 2009; Schumm et al., 1995) and 450-1600 m^3 per year (Coelho Netto et al., 1988). Secondly, the influence of geologic heterogeneities on channel evolution is also poorly understood. Lithological strength and permeability contrasts are rarely simulated by experimental and numerical analyses. Thirdly, there only a few places where the mechanisms by which seepage

Deleted: box canyon

Deleted: box canyons

Deleted: ¶

Deleted: box canyon

Deleted: , which have reproduced the characteristic box canyon morphologies in unconsolidated sediments

Deleted: box canyon

Deleted: canyon

Deleted: Streams

Deleted: valleys

Deleted: Canyon

Deleted: canyons

Deleted: box canyons

Deleted: box canyons

Deleted: canyons

Deleted: box canyon

Deleted: box canyon

erosion occurs have been clearly defined (e.g. Abrams et al., 2009). Basic observations and measurements of channel erosion rates and substrate geologic heterogeneities are needed to test and quantify models for channel formation and improve our ability to reconstruct and predict landscape evolution by groundwater-related processes.

1.3 Objectives

In this study we revisited the Canterbury Plains study site (Schumm and Phillips, 1986) and carried out field observations, geochronological analyses, repeated remote sensing surveys, near-surface geophysical surveying and slope stability modelling of coastal gullies to: (i) identify the processes by which groundwater erodes gullies along the coast, (ii) assess the influence of geological/permeability heterogeneity on the gully formation process, and (iii) quantify the timing of gully erosion and its key controls.

2 Regional setting

The flat to gently inclined Canterbury Plains, located in the eastern South Island of New Zealand, extend from sea level up to 400 m above sea level, and cover an area of 185 km by 75 km (Fig. 1a). A series of high energy braided rivers emerge from the >3500 m high Southern Alps and flow south-eastwards to the shoreline (Kirk, 1991). The plains were formed by coalescence of several alluvial fans sourced from the these rivers (Browne and Naish, 2003; Leckie, 2003). The Quaternary sedimentary sequence comprises a >600 m thick succession of cyclically stacked fluvio-deltaic gravel, sand and mud with associated aeolian deposits and palaeosols (Bal, 1996; Browne and Naish, 2003). The gravels consist of greywacke and represent a variety of channel fill beds and bar forms, whereas the isolated bodies of sand are relict bars and abandoned channels. The interglacial sediments are better sorted and have higher permeability than the glacial outwash, resulting in a wide range of hydraulic conductivities (Scott, 1980). New Zealand's largest groundwater resource is hosted in the gravels down to at least 150 m depth (Davey, 2006). The upper Quaternary sediments are exposed along a 70 km long coastline south-west of the Banks Peninsula (Moreton et al., 2002). This coastline is retrograding at approximately 0.5–1 m per year and consists of cliffs fringed by mixed gravel and sand beaches (Gibb, 1978). The study area is a 2.5 km long stretch of cultivated coastline located 16 km to the south-east of Ashburton (Fig. 1b). The coastline within the study area consists of a 15-20 m thick exposure of poorly-sorted and uncemented matrix-supported outwash gravel, which is capped by up to 1 m of post-glacial loess and modern soil (Berger et al., 1996). The cliff face is punctuated by ~0.5 m thick lenses of sand or clean gravel.

Deleted: box
Deleted: canyon
Deleted: canyon
Deleted: ¶
Deleted: numerical
Deleted: box canyons
Deleted: box canyons
Deleted: box canyon
Deleted: , and (iii) assess the influence of geological/permeability heterogeneity on the box canyon formation process
Deleted: long and
Deleted: wide

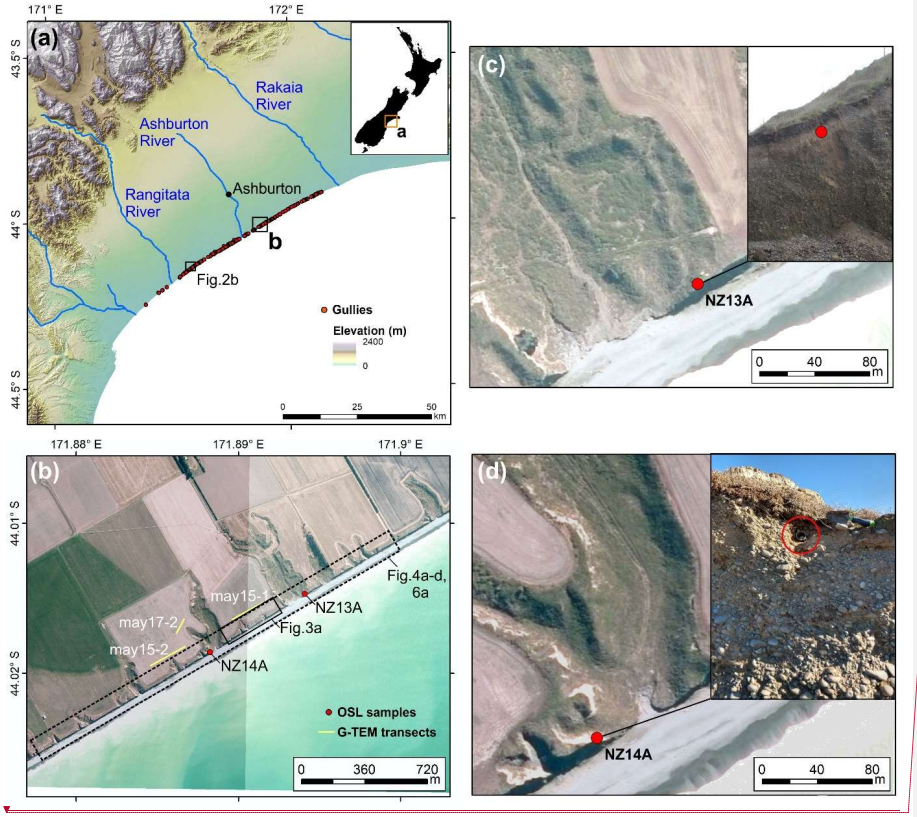


Figure 1: (a) Digital elevation model of the Canterbury Plains (source: Environment Canterbury), located along the eastern coast of the South Island of New Zealand, showing the location of mapped gullies. Location of figure is shown in inset. (b) Mosaic of aerial photographs of the study area (source: Environment Canterbury). Location in a. Location of Optically Stimulated Luminescence (OSL) samples, G-TEM transects, and other figures is shown. (c-d) Zoomed sections of the aerial photograph and site photographs of the OSL sampling sites NZ13A and NZ14A.

3 Materials and Methods

3.1 Data

3.1.1 Field visits

Site visits were carried out in May 2017 and 2019. During these visits, geomorphic features of interest were noted and photographed, and samples collected. The latter included outcropping sediment layers across the cliff face, sediments with coating, and loess sediments for geochronological analysis (NZ13A, NZ14A; Fig. 1). The latter

Deleted:

Deleted: box canyon

Deleted: draping the two largest box canyons

Deleted: b

were collected from the base of the loess draping the flanks of the two largest gullies, above the boundary with the underlying gravels, by hammering stainless steel tubes into the sediment and ensuring that the material was not exposed to light.

3.1.2 Unoccupied aerial vehicle (UAV) surveys

UAV surveys were carried out using DJI Phantom 4 Pro and DJI Mavic Pro drones. The surveys were carried out after rainfall events and on the following dates: 11th May, 19th June, 30th June, 11th July, 15th July, 23rd July, 29th July, 4th August, 26th August, 11th September, 23rd September, 6th October, 13th October and 30th October 2017. The drones were flown at an altitude of 40-55 m, speed of 5 m s^{-1}, and side lap of 65-70%. Eight ground control points were selected and their location and elevation were determined by differential GPS with centimetric horizontal and vertical accuracy. Orthophotos and digital elevation models with a horizontal resolution of 10 cm/pixel were generated from the UAV data using Drone Deploy. The mean distance between the ground control points and the generated orthophoto and model grid cell centres was 0.03 m. Root-mean square and bias were used to estimate the vertical accuracy of the digital elevation models (equations in Laporte-Fauret et al., 2019). The root-mean square error and bias were 0.05 m and 0.03 m, respectively. The model elevations were slightly underestimated (0.1 m).

3.1.3 Near-surface geophysics

Time-domain electromagnetic (TEM) measurements were carried out in May 2019 using the Geonics (Canada) G-TEM system (Fig. 1b). The operating principles of the inductive TEM technique are described in Nabighian and Macnae (1991) and Fitterman (2015). The survey parameters included 4 turns, a 10 × 10 m^2 square TX loop, and a TX current output of 1 A. The G-TEM was operated in a fixed offset-sounding configuration, which is termed "Slingram" mode, in which the RX coil was placed 30 m from the centre of the TX loop and the TX-RX pair moved together along a linear transect at 5 m station spacing, maintaining the 30 m offset. The maximum depth of investigation of the G-TEM system is given approximately by the formula:

$$d = 8.94 L^{0.4} \rho^{0.25} \quad (1)$$

where L (m) is the TX loop size and ρ (Ωm) is the upper layer resistivity (Geonics, 2016). Setting ρ=100 Ωm yields a depth of investigation of d=71 m, whereas ρ=1000 Ωm yields d=126 m. Our investigation depth in New Zealand may be slightly greater than these values since the Geonics formula assumes a 1-turn TX loop carrying current 3 A, whereas we used a more powerful combination of 4 turns at 1.5 A. At each station, a consistent 1-D smooth model of electrical resistivity vs. depth was performed based on the iterative Occam-regularised inversion method (Constable et al., 1987) and using IXG-TEM commercial software (Interpex, 2012). This is a standard 1-D TDEM inversion code that has been successfully used for coastal hydrogeophysical studies (e.g. Pondthai et al., 2020).

Deleted: Unmanned

Deleted: The surveys were carried out on the following dates: 11th May, 19th June, 30th June, 11th July, 15th July, 23rd July, 29th July, 4th August, 26th August, 11th September, 23rd September, 6th October, 13th October and 30th October 2017.¶

Deleted: All soundings were collected in the 20-gate mode with an acquisition interval of 6 × 10^{-6} s to 8 × 10^{-4} s (after ramp-off), corresponding to investigation depths of ~80 m.

Formatted: Font color: Auto

Formatted: Font: (Default) Times New Roman, 10 pt, Font color: Auto

Formatted: Font: (Default) Times New Roman, 10 pt, Font color: Auto

Formatted: Font: (Default) Times New Roman, 10 pt, Font color: Auto

Formatted: Right

Formatted: Font: (Default) Times New Roman, 10 pt, Font color: Auto

Formatted: Font: (Default) Times New Roman, 10 pt, Font color: Auto

Formatted: Font: (Default) Times New Roman, 10 pt, Font color: Auto

Formatted: Font: (Default) Times New Roman, 10 pt, Font color: Auto

Formatted: Font: (Default) Times New Roman, 10 pt, Font color: Auto

Formatted: Font: (Default) Times New Roman, 10 pt, Font color: Auto

Formatted: Font: (Default) Times New Roman, 10 pt, Font color: Auto

Formatted: Font: 10 pt

Formatted: Font: (Default) Times New Roman, 10 pt, Font color: Auto

Formatted: Font: (Default) Times New Roman, 10 pt, Font color: Auto

Formatted: Font: (Default) Times New Roman, 10 pt, Font color: Auto

Formatted: Font: (Default) Times New Roman, 10 pt, Font color: Auto

Formatted: Font: (Default) Times New Roman, 10 pt, Font color: Auto

Formatted ... [1]

Deleted: {Geonics, 2016 #3145}

Formatted ... [2]

Formatted: Font color: Auto

Deleted:

Deleted: using the IXG-TEM software from Interpex ... [3]

3.1.4 Other data

Satellite images with a horizontal resolution of 1 m/pixel and dating back to 2004 were obtained from Google Earth. Precipitation records dating back to 1927 were provided by Environment Canterbury. The latter also provided a time series of water level data since 2015 from a 30 m deep well located 10 km to the north-east of the study area.

3.2 Methods

3.2.1 Sample analyses

Sediment samples were analysed for grain size distribution using sieves following the ASTM D0422 protocol. The composition of the coating on selected sediment outcrops within the gullies was determined using X-ray fluorescence.

3.2.2 Optically Stimulated Luminescence (OSL) dating

Luminescence dating is numerical-age technique that uses optically and thermally sensitive signals measured in the form of light emissions in the constituent minerals that form sediment deposits. Quartz and feldspars are among the most often used minerals. Sediment ages obtained via luminescence dating reflect the last exposure of the analysed mineral grains to daylight, when the resetting (called bleaching) of the previously incorporated luminescence signal occurs.

In order to obtain luminescence ages, two types of measurements were performed. The dose accrued by the crystal from natural radioactivity since its last exposure to daylight (called the palaeodose) was determined as an equivalent dose (D_e). This was done by measuring the light emission of the crystal upon optical stimulation, and matching this emission to signals generated by the exposure to a known dose of radiation given in the laboratory. This is expressed as the amount of absorbed energy per mass of mineral (1 J kg^{-1} = 1 Gy (Gray)). Radioactivity measurements were carried out on each sample in order to determine the annual dose (D_a), which represents the rate at which the environmental dose was delivered to the sample (Gy ka^{-1}). The age was obtained by dividing the two determined parameters. As low luminescence sensitivity and poor dosimetric characteristics were reported for quartz from sediments in New Zealand (see Preusser et al. (2009) and the references cited therein) we have used signals from feldspars by the application of infrared stimulation based on the post IR-$IRSL_{225}$ (Buylaert et al., 2009) and post IR-$IRSL_{290}$ (Thiel et al., 2011) protocols on polymineral fine (4-11µm) grains, as well as coarse (63-90 µm) potassium feldspars.

A detailed description of luminescence dating methodology, including sample preparation, equivalent dose determination, annual dose determination, luminescence properties (including residual doses, dose recovery tests and fading tests), is presented as Supplementary Materials.

Formatted: Justified

Deleted: box canyons

Formatted: Space After: 0 pt, Line spacing: 1.5 lines

Deleted: ¶
¶
Page Break

3.2.3 Morphological change detection

The method used to measure gully aerial erosion in between surveys entailed the manual delineation of shapefiles around gully boundaries for each survey (using orthophotos, digital elevation models and slope gradient maps in the case of the UAV data, and satellite images in the case of the Google Earth data), the estimation of their areas, and the comparison of the latter in between surveys. The uncertainty inherent in this approach is related to the digitisation of the gully boundaries. We made sure that a vertex was added at least every 5 pixels for both the UAV (0.5 m) and Google Earth data (5 m). This ensures that a minimum erosion of 0.25 m^2 (in the case of the UAV data) and 25 m^2 (in the case of the Google Earth data) was detected.

3.2.4 Slope stability modelling

We developed a slope stability model based on the limit equilibrium and segmentation strategy of the Bishop method, where a soil mass is discretised into vertical slices. The factor of safety F_f is calculated using the following (Fredlund and Krahn, 1977; Fredlund et al., 1981):

$$F_f = \frac{\sum (c'\beta \cos\alpha + (N - u\beta)\tan\phi' \cos\alpha)}{\sum N \sin\alpha - \sum D \cos\omega} \quad (2)$$

where c' (in kPa) is effective cohesion, ϕ' (in °) is effective angle of friction, u (in kPa) is pore-water pressure, D (in kN) is concentrated point load, β (in m) represents the slice base length, ω (in °) is the angle between the top part of the slope and surface forces, and α (in °) is inclination of the slice base. N is the normal force acting on the slice base and can be computed by:

$$N = W \cos\alpha - kW \sin\alpha + [D\cos(\omega + \alpha - 90)] \quad (3)$$

where W (in kN) is slice weight (unit weight γ_s (in kN m^{-3}) × volume (in m^3)) and k is hydraulic conductivity (in m s^{-1}).

We also modelled the water flow and pore pressure distribution within the soil using the Poisson equation, which is the generalised form of the Laplace equation (Whitaker, 1986):

$$k_x \frac{\partial^2 h}{\partial x^2} + k_y \frac{\partial^2 h}{\partial y^2} = q \quad (4)$$

where q is the total discharge (m^3 s^{-1}), k_x and k_y are equal to the hydraulic conductivity (m s^{-1}) in the horizontal and vertical directions, respectively, and h is the hydraulic head (m).

Deleted: box canyon
Deleted: box canyon
Deleted: canyon
Deleted: M
Deleted: *Slope stability modelling* ¶ ¶
Deleted: 1
Deleted: geometric parameters,
Deleted: geometric parameter
Deleted: d
Deleted: 2
Deleted: famous
Deleted: 3

Equation (4) applies to water flow under steady-state and homogeneous conditions, whereas the following equation is applicable to dynamic and inhomogeneous conditions:

$$\frac{\partial}{\partial_x}(k_x\frac{\partial h}{\partial x})+\frac{\partial}{\partial_y}(k_y\frac{\partial h}{\partial y})=q+\frac{\partial\theta}{\partial t} \quad (5)$$

where $\partial\theta/\partial t$ describes how the volumetric water content changes over the time.

The water transfer theory accounts for transient behaviour, which can be defined by the following equation (Domenico and Schwartz, 1997):

$$M_{st}=\frac{dM_{st}}{dt}=m_{in}-m_{out}+M_s \quad (6)$$

where m_{in} is the cumulative mass of water that enters the porous medium, m_{out} is equal to the mass of water that leaves the porous medium, and M_s is the mass source within the representative elementary volume. The rate of increase in the mass of water stored within the representative elementary volume is:

$$M_{st}=M_w+M_v \quad (7)$$

where Mw and Mv represent the rate of change of liquid water and water vapour, respectively.

The mechanical and hydraulic soil properties employed in this model are listed in Table 2 and were obtained from Dann et al. (2009) and Aqualinc Research Limited (2007). We modelled two scenarios, based on the available rainfall data (see Sect. 4.4). The first is a 3-day long intense rainfall event ($(I\text{-}D)_3$) covering the period 20th July – 22nd July 2017. The second is a 14-day period with occasional, low intensity rain ($(I\text{-}D)_{14}$) between the 21st June and 4th July 2017. Each scenario is modelled for two sandy gravel slopes, one with a 0.5 m thick sand lens and the other with a 0.5 m thick gravel lens. Both lenses are located at a height of 5 m above sea-level. Lateral water inflow and surface water infiltration were estimated from the hydrological model in Micallef et al. (2020). Slope stability modelling was carried out using the Slide2 software package by Rocscience.

Deleted: 3

Deleted: 4

Deleted: 5

Deleted: 6

Table 2: Mechanical and hydraulic soil properties used in slope stability modelling.

Soil type	Unit weight (kN m^{-3})	Cohesion (kPa)	Friction angle (°)	Saturated hydraulic conductivity (m/day)	Residual water content (m^3 m^{-3})	Saturated water content (m^3 m^{-3})
	γ_s	c	φ	k	θr	θs
Sand	20.5	7	34.5	3.216	0.01	0.078
Sandy gravel	23	8	37	0.64	0.01	0.128
Gravel	24	4	36.5	7.376	0.016	0.142

4 Results

4.1 Gullies along the Canterbury coast - distribution and morphology

We have mapped 315 gullies (locally also known as "dongas") along 70 km of the Canterbury coastline (mean of 4.5 gullies per km of coastline). The distribution of the gullies is clustered (nearest neighbour ratio of 0.33 with a z-score of -22.67 and a p-value of 0); the majority of the gullies are located between Rakaia and Rangitata Rivers (Fig. 1a), particularly in the vicinity of Ashburton River. The heads of many gullies connect to shallow, relict meandering channels (Fig. 2a). Some of these channels are visible in aerial photographs, in spite of the terrain having been worked by farmers (Fig. 2b).

Formatted: Left

Deleted: ¶
Landscape evolution modelling¶
¶
We also built a landscape evolution model (LEM) using the Python modelling environment Landlab (Barnhardt et al., 2020; Hobley et al., 2017). The model allowed us to simulate two main processes: groundwater flow and associated erosion (Fig. 2). ¶
¶

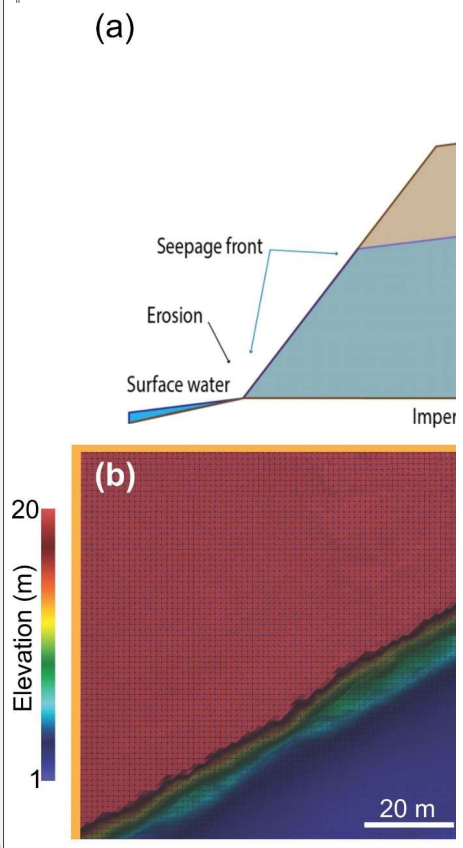


Figure 2: (a) Conceptual model of the LEM. (b) Initial topography used in the LEM. Orange line denotes cl ... [4]

Deleted: Box canyons

Deleted: box canyons

Deleted: canyons

Deleted: box canyons

Deleted: canyons

Deleted: box canyons

Deleted: 3

Deleted: 3b

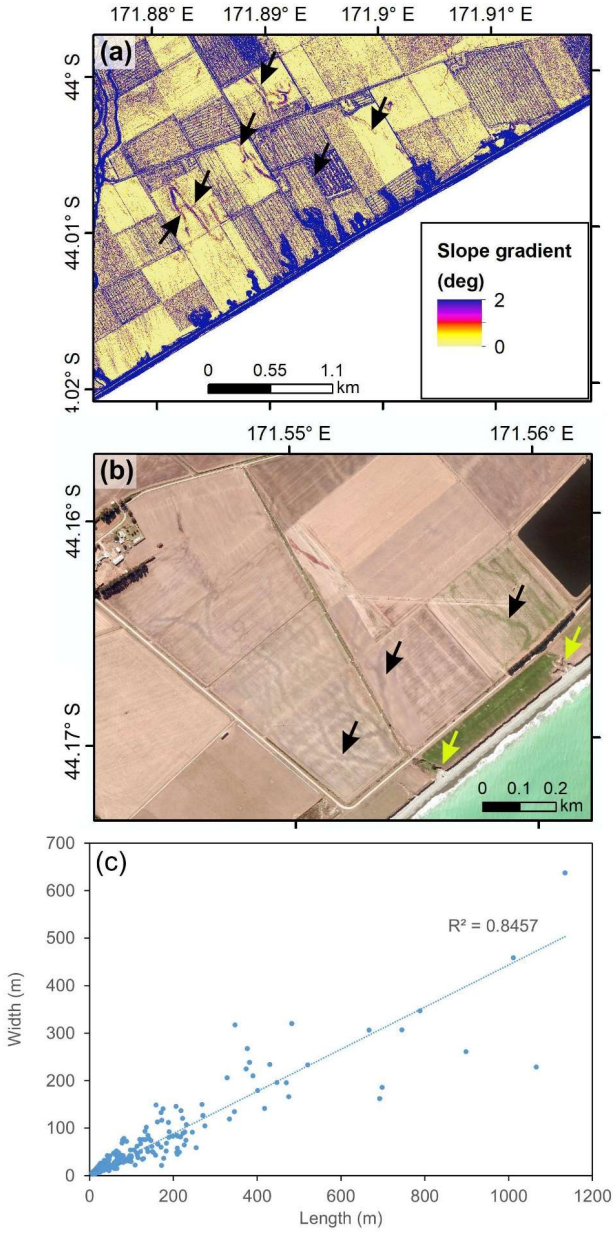


Figure 2: (a) Slope gradient map of the study area. Black arrows indicate relict, infilled channels. (b) Aerial photograph of Coldstream, Canterbury coast (source: Environment Canterbury). Black arrows indicate relict, infilled channels. Yellow arrows indicate gullies. Location in Fig. 1a. (c) Plot of length vs. width for gullies mapped along the Canterbury coastline.

Deleted: 3

Deleted: box canyons

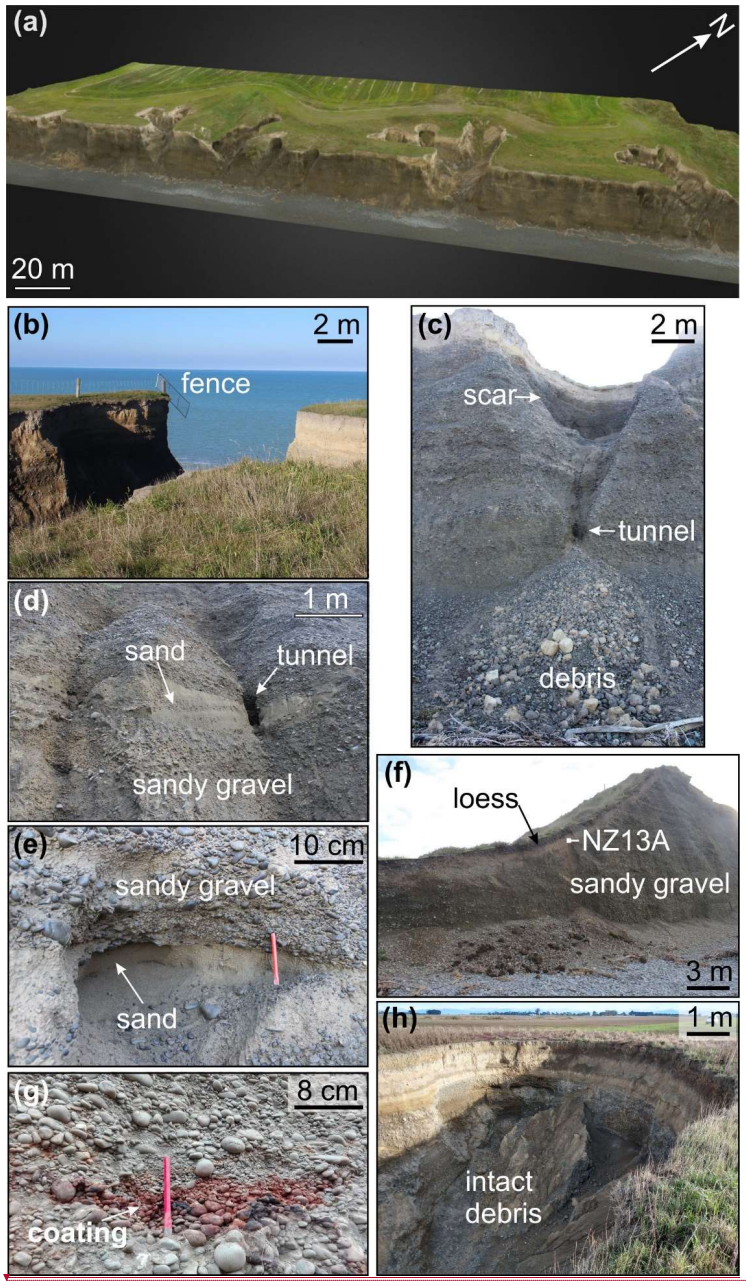
Deleted: box canyons

In plan, the gullies are predominantly linear to slightly sinuous (sinuosity of 1-1.3) and characterised by a concave head. In profile, the gullies have linear, gently sloping (2-10°) axes, with a concave break of slope separating the axis from a steep (up to 70°) head. In cross-section, the gullies are U-shaped with walls up to 70° in slope gradient. The gullies are between 5 and 1134 m long (mean of 116 m) and between 3 and 637 m wide (mean of 56 m). Gullies generally exhibit a constant width with distance upslope. They have a length to width ratio that varies between 1 and 7.9, with a mean of 2 (standard deviation of 0.89) (Fig. 2c).

4.2 Field site observations

In May 2017, our study area hosted 33 gullies that vary between 15 m and 600 m in length (Figs. 1b; 3a). During the site visits we did not encounter evidence of surface flow. However, the middle to lower sections of the gully walls and cliffs were consistently wet. These sections were also characterised by failure scars and alcoves, particularly above the sandy lenses. Alcoves were also encountered at the base of gully heads, where they were wet and up to 1 m deep (Fig. 3e). Some sandy layers outcropping across the cliff face hosted tunnels (Fig. 3c-d). Above these tunnels, theatre-shaped scars with a shallow and narrow gully at their base were observed (Fig. 3c). At the base of the scars, the gully heads and some gully mouths, we encountered mass movement debris that was occasionally intact and that predominantly consisted of gravel, sandy gravel and loess (Fig. 3c, h). Gullies have gravel-covered irregular floors. Whereas the smaller gullies have a U-shaped cross-section, the three longest gullies have gently sloping V-shaped cross-sections, with loess draping their walls (Fig. 3f). Sandy and clean gravel layers outcropping within the gullies were wet; the former appeared weathered, whereas the latter were coated by Fe and Mn (Fig. 3g). Fences were locally seen suspended across a number of gullies (Fig. 3b).

Deleted: box canyons
Deleted: canyon
Deleted: canyons
Deleted: canyons
Deleted: box canyons
Deleted: Canyons
Deleted: valley
Deleted: 3c
Deleted: box canyons
Deleted: 4a
Deleted: canyon
Deleted: canyon
Deleted: 4e
Deleted: 4c
Deleted: 4c
Deleted: box canyon
Deleted: box canyon
Deleted: 4c
Deleted: Box canyons
Deleted: box canyon
Deleted: larger
Deleted: box canyons
Deleted: 4f
Deleted: canyons
Deleted: 4g
Deleted: box canyons
Deleted: 4b



Deleted:

(a)
20 m
(b)
2 m
fence
(d)
1 m
sand
tunnel
sandy gravel
(e)
10 cm
sandy gravel
sand
(g)
8 cm
coating

Deleted: 4

1135 **Figure 3: (a) Orthophoto map of part of the study area draped on a 3-D digital elevation model. Location in Fig. 1b. (b-h) Photographs of features of geomorphic interest taken at study area. The location of sample NZ13A is shown in f.**

4.3 Luminescence ages

Four sets of OSL ages are presented in Table 3. The $pIRIR_{290}$ ages are higher than the ages obtained by applying $pIRIR_{225}$ protocol. The cause of this difference is not yet fully understood, although it can partially be attributed to the results of the dose recovery test and the poor bleachability of the $pIRIR_{290}$ signals compared to $pIRIR_{225}$ signals (Buylaert et al., 2011). Considering that no anomalous behaviour of the investigated signals was observed (see Supplementary Materials), we are unable to explain the overestimation of the K-feldspar ages compared to the polymineral fine grain ages in the case of NZ13A, especially since the opposite behaviour is observed in the case of sample NZ14A. However, considering a 95% confidence level, ages obtained using different methods broadly overlap, the only exception being the $pIRIR_{225}$ ages obtained on K-feldspars on sample NZ14A, which we regard as an outlier.

Table 3: Summary of the $pIRIR_{225}$ and $pIRIR_{290}$ ages obtained on polymineral fine grains (4-11 µm) and coarse K-feldspars (63-90 µm). The ages were determined considering 15% water content. Uncertainties are given at 1σ, 68% confidence level. Further details are available in the Supplementary Materials.

	Age (ka)-$pIRIR_{225}$		Age (ka)-$pIRIR_{290}$	
Sample code	**Polymineral fine grains**	**K-feldspars (63-90 µm)**	**Polymineral fine grains**	**K-feldspars (63-90 µm)**
NZ13A	16.0±1.4	20.1±1.5	20.9±2.0	26.2±2.1
NZ14A	4.6±0.4	1.9±0.1	6.0±0.7	3.1±0.3

4.4 Morphological changes

4.4.1 Short term morphological changes

By comparing the orthophotos and digital elevation models generated from the UAV data acquired during the various site visits between May and October 2017, we document the formation of 3 new gullies (up to 30 m long, Figs. 4e-f) and the enlargement of 30 gullies (primarily by elongation, and occasionally by widening and branching) (Figs. 4a-d). The new gullies formed at locations where there was a small landslide scar in the middle of the cliff. There was no change in form in 3 of the gullies. Figure 5 shows the total area eroded between surveys (which amounts to 3273 m^2), the daily precipitation and the associated changes in water table height. Only three surveys recorded gully erosion. Two of these surveys happened soon after rainfall events of >40 mm in one day (Fig. 5). The most important of these covers the period between the 15th and 23rd July 2017, when 95% of the material was removed and the 3 new gullies were formed (Figs. 4-5). During this period, a total of 153 mm of rain fell (up to 120 mm on the 21st July 2017 alone, which was the most intense rainfall event since 1936), resulting in a 1.5 m rise in the water table. A third survey happened six days after the 21st July 2017 storm, with 22 mm of rain falling in one day. The material eroded from the gullies was deposited at the base of the cliffs as gravel cones, which were remodelled by debris flows during ensuing precipitation events and subsequently disappear from the orthophotos.

Formatted: Subscript
Formatted: Subscript
Deleted: box canyons
Deleted: 5e
Deleted: box canyons
Deleted: 5a
Deleted: box canyons
Deleted: box canyons
Deleted: 6
Deleted: box canyon
Deleted: 6
Deleted: . The
Deleted: box canyons
Deleted: 5
Deleted: 6
Deleted: canyons
Deleted: eroded by wave action

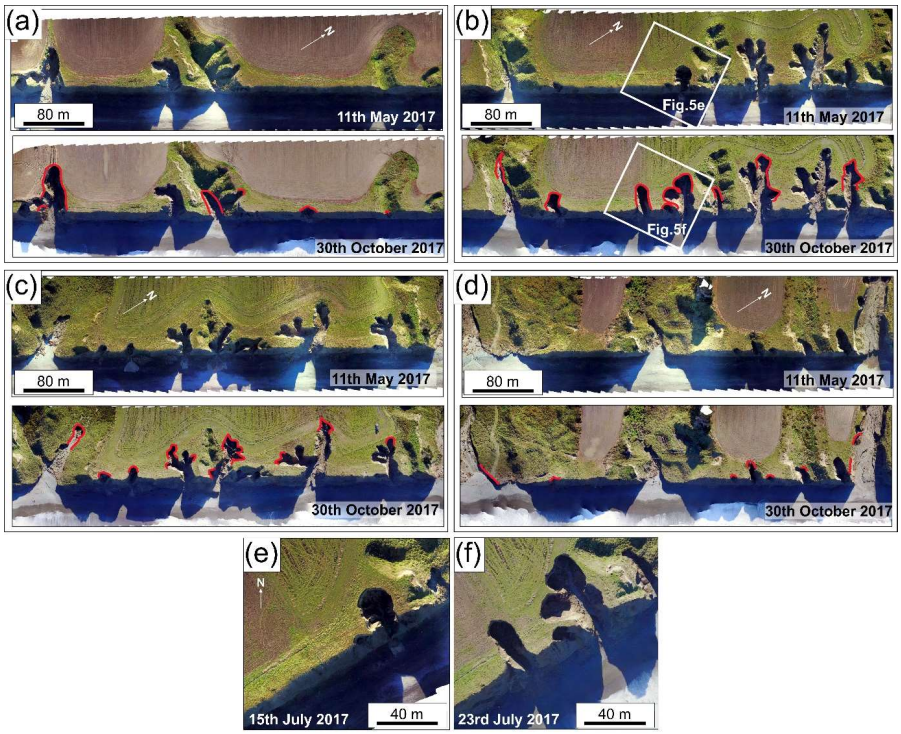


Figure 4: (a-d) Orthophotographs of the study area at the start and end of the UAV surveys, ordered from south-west to north-east. Red lines mark eroded areas. Location in Fig. 1b. Orthophographs from a part of the study area on the (e) 15th July 2017 and the (f) 23rd July 2017. Location in b.

Deleted: 5

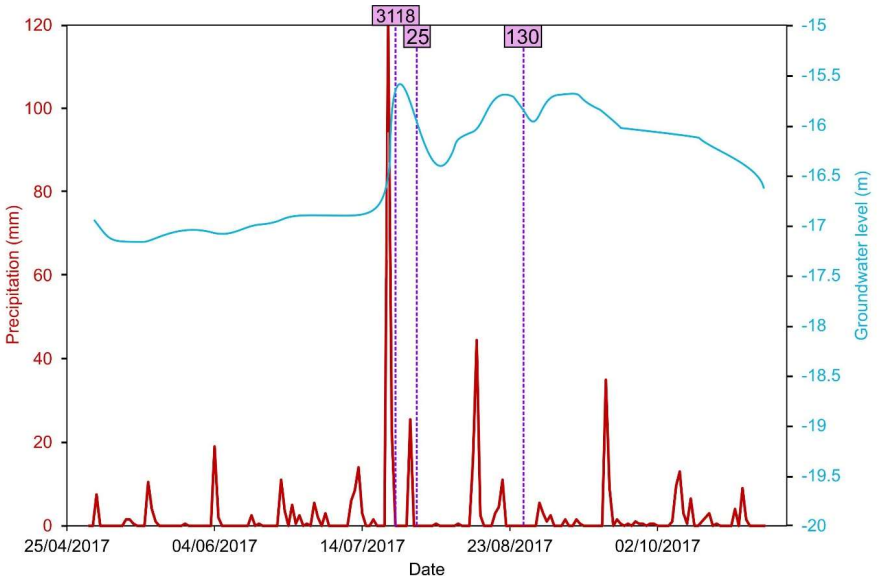


Figure 5: Daily precipitation (for Ashburton Council) and groundwater level records (from a well located 10 km north-east of the study area) for the period 1st May to 31st October 2017 (source: Environment Canterbury). The pink lines mark the surveys when gully erosion was observed (the value in the pink box corresponds to the eroded area in m²).

Deleted: 6

Deleted: box canyon

4.4.2 Long terms morphological changes

For the period 2004-2015 we have used satellite imagery to map the formation of 6 new gullies and the elongation of 22 gullies. 18 of these erosion episodes are recorded in the image taken on 26th August 2013 (Fig. 6a). This follows a major rainfall event between the 16th and 23rd June 2013, when 171 mm of rain fell in 7 days (with up to 51 mm falling in one day) (Fig. 6b). The other erosion episodes include the 5 gullies eroded by the 28th March 2009, after a storm of 46 mm per day on the 31st July 2008, and the 5 gullies eroded by the 19th October 2015, after a storm of 43 mm per day on the 19th June 2015.

Deleted: box canyons

Deleted: box canyons

Deleted: 7

Deleted: 7b

Deleted: box canyons

Deleted: box canyons

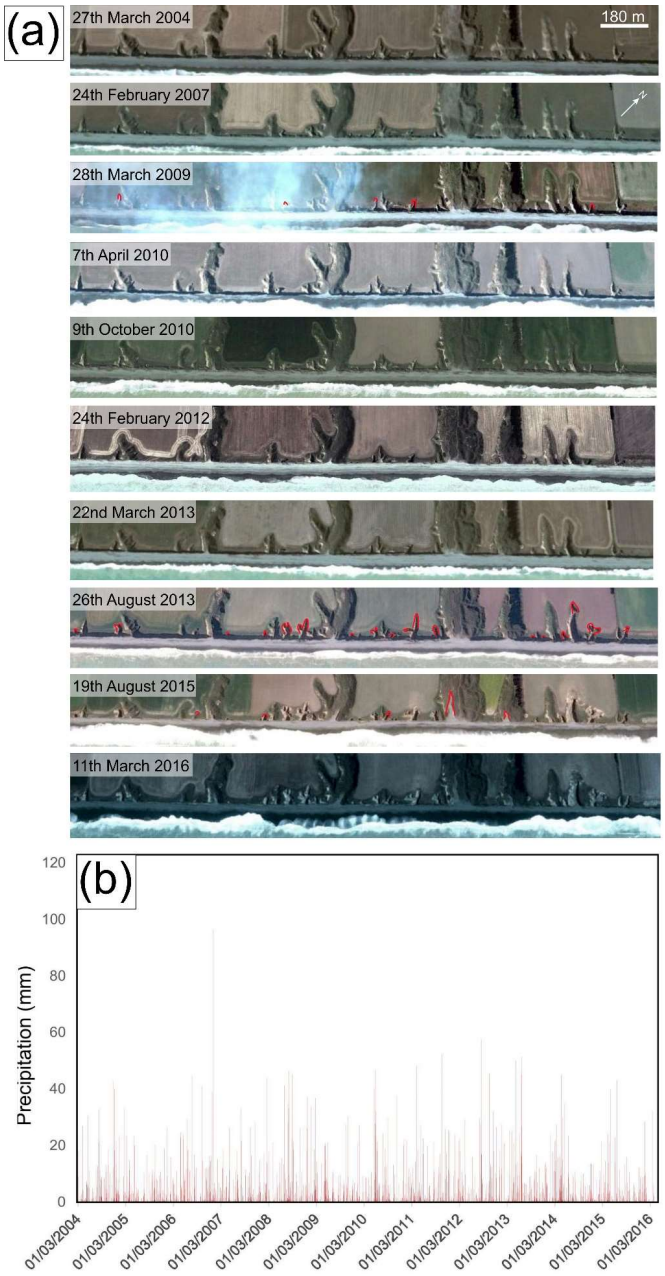


Figure 6: (a) Satellite imagery of the study area between the 27th March 2004 and 11th March 2016 (source: Google, Maxar Technologies). Eroded areas are marked by red lines. (b) Daily precipitation record for this period for Ashburton Council (source: Environment Canterbury).

Deleted: 7

4.5 Geophysical data

The location of the G-TEM transects is shown in Fig. 1b. An attempt was made to invert the G-TEM slingram-mode responses with 30-m TX-RX offset using 1-D Occam inversion. A representative inversion result is shown in Fig. 7a. The resistivity model is presented in the right panel, whereas the corresponding model-response with the actual data points is shown on the left. The best calculated smooth depth profile clearly does not fit well with the measured signal and there is excessive structure in the ~10-20 m depth range, including the very low resistivity layer (~10^{-4} Ωm) at depths in excess of ~12-15 m. The resistivity values between 40 and 100 m depth are lower than sea water resistivity (0.3 Ωm), which is not reasonable. The inability to fit a 1-D model to the slingram responses suggests that the geoelectrical sub-surface structure is strongly heterogeneous within the footprint of the G-TEM transmitter. As a result, we cannot trust 1-D inversions of the slingram-mode data in such a 3-D geological environment. We did not try to use the 1-D inversion software to further analyse and interpret the G-TEM data. However, even though the individual slingram-mode responses cannot be fit reliably by a 1-D model, we can still analyse lateral changes in the observed response curves along the slingram profiles to reveal information about subsurface heterogeneity; this is elaborated below.

Deleted: 8a

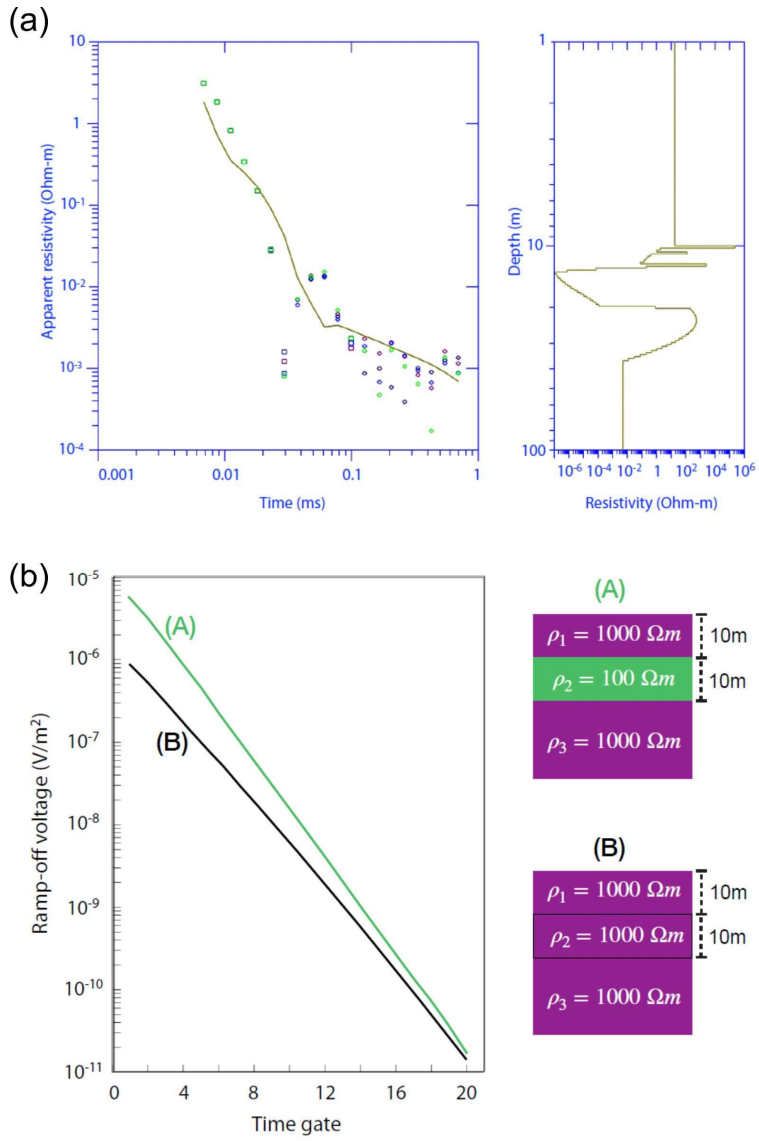


Figure 7: (a) 1-D inversion result for data at a station located 6 m from start of Profile May15-1 (location in Fig. 1b). (b) G-TEM slingram responses for model (A) containing a conductive zone at a depth of 10-20 m, and for model (B) without the conductive zone.

Deleted: 8

Instead of performing 1-D inversions, we present time-gate plots for all three transects. A time-gate plot is defined as a graph of the observed G-TEM voltage response, evaluated at a particular time-gate, as a function of position along a profile. Time-gate plots are a useful alternative to explore the lateral variability of the G-TEM

response along a profile in the event that the sounding curves at individual stations cannot be fit with 1-D models. It is presumed that variability in a time-gate plot is correlated with lateral heterogeneity in the subsurface geoelectrical structure, since a 1-D Earth structure would yield no spatial variability in a time-gate plot. Specifically, the amplitude of the G-TEM slingram response (in units of 10^{-10} V/m^2) at the first time-gate is plotted as a function of station number along a profile. Figure 7b displays a 1-D model (A) that contains a conductive layer of 100 Ωm between 10-20 m depths in a homogeneous 1000 Ωm background. The 1-D model (A) in Fig. 7b is motivated by the inversion results of deep-penetrating 40 × 40 m TX loop TDEM soundings carried out on top of the cliffs several tens of metres inland (Weymer et al., in review), which revealed such a conductive zone at these depths. Unlike the slingram profiles, the deeper-penetrating, larger-loop sounding curves are readily fit by a 1-D model. This model generates a G-TEM slingram response that has a substantially larger ramp-off voltage amplitude at all time gates than does the model (B) without the conductive layer, as shown in Fig. 7b. Thus we regard an enhancement of response at the first time gate as indicative of a conductive zone at depth beneath the slingram station. The spatial analysis of time-gate plots is not a conventional approach in time-domain electromagnetics, but it is somewhat analogous to the spatial analysis of apparent resistivity profiles in frequency-domain electromagnetics using terrain conductivity meters (e.g. Weymer et al., 2016). This is based on the idea that the G-TEM response at a fixed time-gate carries information similar to that of a terrain conductivity meter response at a fixed frequency.

The first-time-gate profile of transect May15-1 is located upslope of small but recently eroded gullies (Fig. 8a). In this figure, the 'first-time-gate profile' is a plot as a function of distance along the transect of the G-TEM ramp-off voltage at time gate number 1 (as shown, for example, in Fig. 7b), at the first sampled point of the transient response immediately after the TX current has been switched off. Near the middle of this transect there is a distinctive peak that is much higher than the background. The peak is ~20-30 m wide and it appears in a similar fashion on each of the gates 1 through 7 (not shown here), although it cannot be clearly observed after gate 7. Transect May 15-2 is located upslope of recently eroded gullies in the south-west and relatively less active gullies in the north-east of the investigated area (Fig. 8b). Lateral variations are evident along the 192 m length of the profile. The high amplitude response at the start of the profile (going from the south-west to north-east) is followed by a drop in amplitude near the midpoint of the profile, following which there is continuous fluctuation at a lower amplitude until the end of the profile. The time-gate plots for gates 2 to 7 remain similar in shape to that of the time-gate-1 plot and hence are not shown. After time-gate 7, the time-gate plots start to lose coherence due to the low signal-to-noise ratio of the decaying RX voltage at late times after TX ramp-off. G-TEM slingram profile May17-2 was acquired upslope of the tributary of a large gully covered by mature vegetation (Fig. 8c). As shown in Sect. 4.4, the size and location of this gully have been persistent over recent years, in contrast to the neighbouring, smaller gullies that are under active development. Transect May 17-2 shows a lower amplitude response in comparison to the previous two transects (Fig. 8c).

Deleted: 8b

Deleted: {Weymer, in review #2620}

Deleted: This model generates a G-TEM slingram response that has a substantially larger amplitude at all time gates than does the model (B) without the conductive layer.

Deleted: box canyons

Deleted: 9a

Deleted:

Deleted: box canyons

Deleted: canyons

Deleted: 9b

Deleted: The fluctuating signals remain similar in shape for time-gates 2 to 6 (not shown). G-TEM slingram profile May17-2 was acquired upslope of the tributary of a large box canyon covered by mature vegetation (Fig. 9c). As shown in Sect. 4.4, the size and location of this box canyon have been persistent over recent years, in contrast to the neighbouring, smaller box canyons that are under active development. Transect May 17-2 shows a lower amplitude response in comparison to the previous two transects (Fig. 9c). Based on all three profiles, a general observation that can be made is that the first-time-gate amplitude of the slingram response is higher upslope of the more recently active box canyons.¶

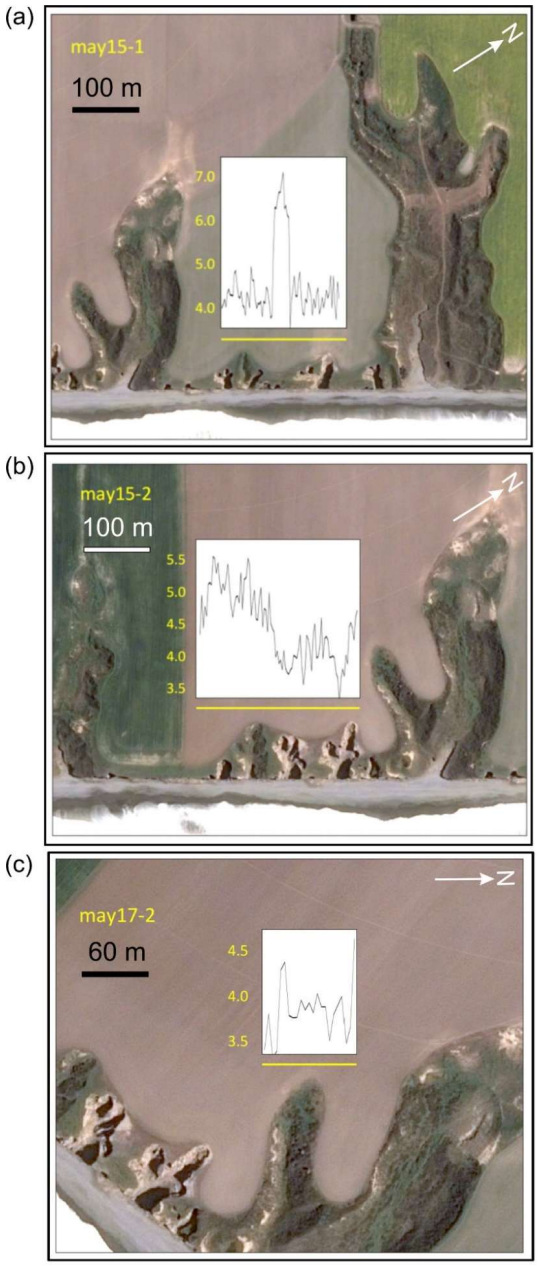


Figure 8: First-time-gate profiles of G-TEM slingram transects (a) May15-1, (b) May 15-2, and (c) May17-2 (units in yellow are 10^{-10} V m^{-2}). Source of background imagery: Google, Maxar Technologies. A yellow line marks a slingram transect, the length of which can be determined from the scale bar.

Deleted: 9

Based on all three profiles, a general observation that can be made is that the first-time-gate amplitude of the slingram response is higher upslope of the more recently active gullies.

4.6 Slope stability modelling

4.6.1 Slope with sand lens

The factory of safety of the slope prior to any rainfall event was 2.514. During the first scenario ($(I\text{-}D)_3$), the factor of safety decreased to 1.371 due to undermining by tunnelling associated to high pore pressures within the sand lens, and then to 0.614 as a result of a decrease in the shear strength of the lower slope material due to an increase in pore pressure (Figs. 9a, 10a). A rainfall intensity of 40 mm per day is required to bring the factor of safety below 1 (Fig. 10c). In the case of the second scenario ($(I\text{-}D)_{14}$), changes in pore-water pressure did not result in either tunnelling or slope failure. This only resulted in a decrease in the effective stress and in the factor of safety (1.216) (Figs. 9b, 10b).

Formatted: Right: 0.09 cm

Deleted: M

Deleted: stability modelling

Deleted: <#>*Slope with sand lens*¶
¶

Deleted: <#>10a

Deleted: <#>11a

Deleted: <#>11c

Deleted: <#>10

Deleted: <#>11b

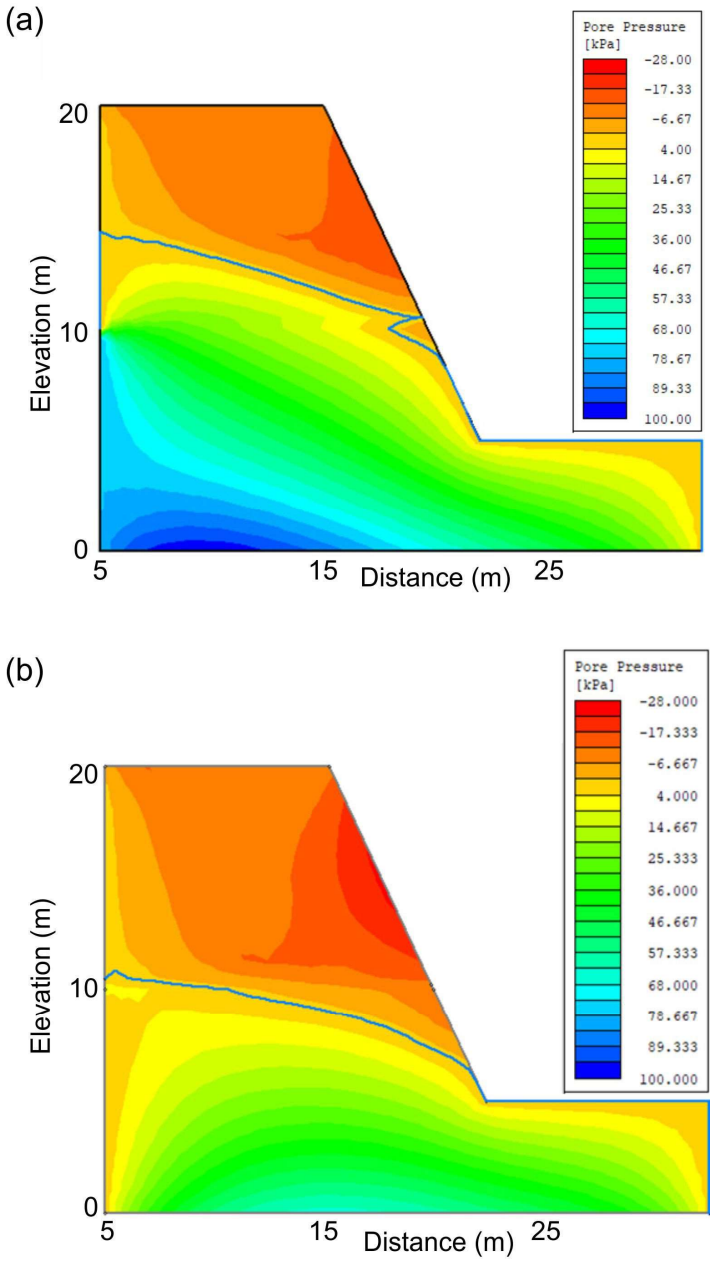


Figure 9: Model results for sandy gravel slope with sand lens. Estimated pore water pressure after (a) 3 days for first scenario ($(I\text{-}D)_3$) and (b) 14 days for second scenario ($(I\text{-}D)_{14}$). Blue line denotes wetting front.

Deleted: 10

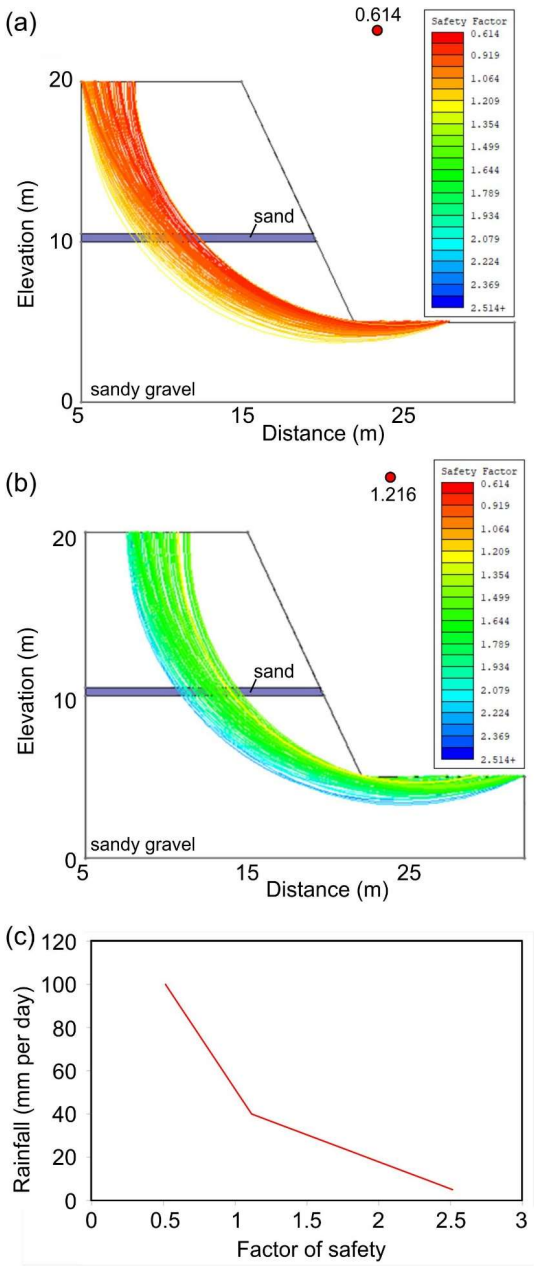


Figure 10: Model results for sandy gravel slope with sand lens. (a) Estimated factor of safety after 3 days for first scenario ($(I\text{-}D)_3$). (b) Estimated factor of safety after 14 days for second scenario ($(I\text{-}D)_{14}$). (c) Plot of rainfall intensity vs. factor of safety for the for first scenario ($(I\text{-}D)_3$ for the slope with sand lens.

Deleted: 11

4.6.2 Slope with gravel lens

The factor of safety of the slope prior to any rainfall event is 1.793. For the first scenario ($(I\text{-}D)_3$), the factor of safety decreased to 1.166, and neither tunnelling nor slope failure occurred (Figs. 11a-b). In the case of the second scenario ($(I\text{-}D)_{14}$), the outcome is the same, with the factor of safety decreasing to just 1.588 (Figs. 11c-d).

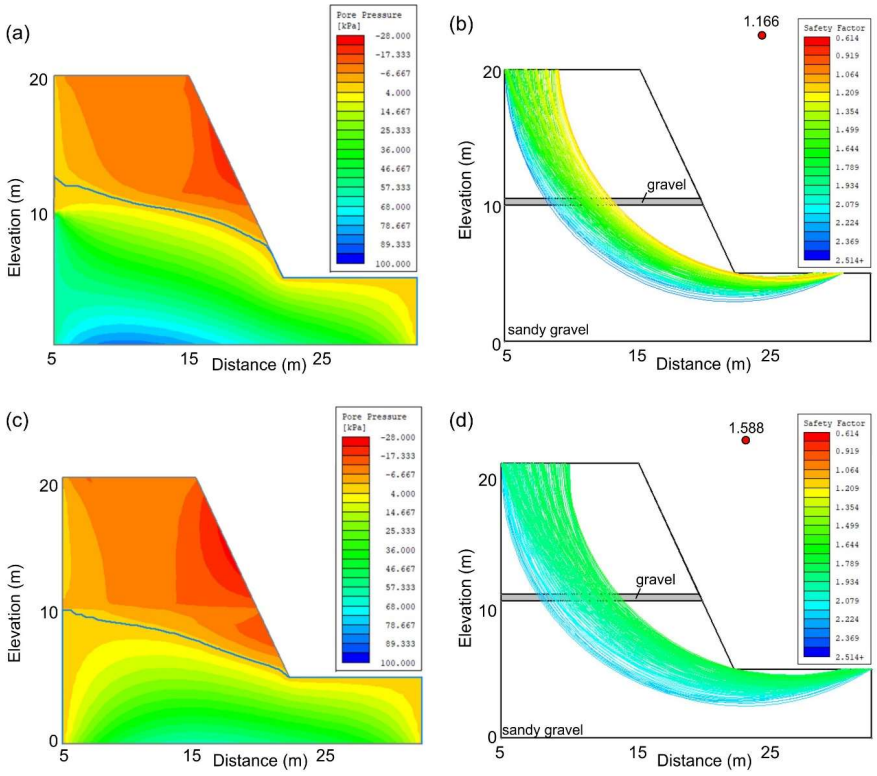


Figure 11: Model results for sandy gravel slope with gravel lens. (a) Estimated pore water pressure after 3 days for first scenario ($(I\text{-}D)_3$). (b) Estimated factor of safety after 3 days for first scenario ($(I\text{-}D)_3$). (c) Estimated pore water pressure after 14 days for second scenario ($(I\text{-}D)_{14}$). (d) Estimated factor of safety after 14 days for second scenario ($(I\text{-}D)_{14}$). Blue line in a and c denotes wetting front.

Deleted: <#>*Slope with gravel lens*¶
¶

Deleted: <#>12a

Deleted: <#>12c

Deleted: 12

Deleted: ¶

Deleted: 4.6.2 . LEM¶
¶
The model results for the two erosional modes and two rainfall scenarios are displayed in Figs. 13-16 and discussed in more detail below.¶
¶
The results of the stream power law model with a high intensity rainfall scenario are shown in Fig. 13. During the first hour and as a result of the low recharge value (1×10^{-9} m s^{-1}), the water table drops slowly and no erosion occurs. All parameters, except for shear stress, decrease rapidly. An intense storm (recharge of 2.8×10^{-5} m s^{-1}) starts after 1 hr; groundwater seeps out at the cliff base, where the shear stress is highest, forming a gully between the cliff base and the beach. In the ensuing 2 hr, the recharge rate stays the same but the water table height in the sandy area decreases due to the high permeability; as a result the shear stress decreases and erosion stops. At the third hour, the storm stops and there is no recharge; the water table drops slowly and the shear stress and groundwater seepage decrease. The results of the stream power law model with a low intensity rainfall scenario are similar to those of the high intensity rainfall scenario (Fig. 14). ¶
¶
¶
¶
... [5]

5 Discussion

Gullies are characteristic landforms along the Canterbury coast (Fig. 1a). They are an important driver of coastal geomorphic change as well as loss of agricultural land. In the following sections we integrate field observations with the modelling results to infer how coastal gullies are formed by groundwater erosion, the role that lithology and permeability play in gully initiation and evolution, and the temporal scale of gully formation.

5.1 Coastal gully formation by groundwater-related processes

The Canterbury gullies initiate and evolve via two types of groundwater-related processes. The first process is seepage erosion of sand, which leads to the formation of alcoves and tunnels. This inference is based on the exclusive occurrence of tunnels in sandy layers at the study site (Figs. 3c-e). Seepage erosion lowers the overall factor of safety of the slope, as demonstrated by slope stability model results for the slope with sand lens scenario, and is a precursor to the second process, which is slope failure (Fig. 3c). Site observations (Fig. 3h), the UAV data (Fig. 4) and satellite imagery (Fig. 5) show that gullies primarily evolve by retrogressive slope failure, which results in the elongation of the gullies and, to a lesser extent, in widening and branching along the gully walls. According to the slope stability model in Fig. 9, up to 4.4 m^3 of water are estimated to have seeped out of the cliff face to erode 1650 m^3 of material, which contrasts with the inference by Howard (1988) that 100-1000 times more water than volume of eroded sediment must be discharged in order to create a sapping valley. We infer that wave erosion is responsible for the removal of the failed material at the gully mouths and the base of the cliff. Isotropic scaling of length with width (Fig. 2c) suggests that gully planform shape is generally geometrically similar at consecutive stages of evolution.

5.2 Influence of geological/permeability heterogeneity on gully formation

Two factors control the location of gullies. The first factor is the occurrence of sand lenses across a sandy gravel cliff face. This geological framework is conducive to alcove formation, tunnelling and slope failure (Figs. 3c-e). The higher permeability of the sand and clean gravel lenses, in comparison to the surrounding sandy gravel (Table 2), facilitates faster water transfer to the cliff face; this is also corroborated by the weathering in the sandy layers and Fe and Mn deposits in the clean gravel layers (Fig. 3g). Alcoves and tunnels only form in the sand lenses, however, because the latter develop higher pore pressures, and sand is easier to entrain and remove in comparison to clean gravel in view of its lower shear strength (Table 2). Slope failure only occurs in sandy gravel slopes with sand lenses (Figs. 9-11). The higher pore pressure developed in the sand lenses is transferred to the sandy gravel slopes, resulting in a larger decrease in the shear strength and higher water table in comparison to the sandy gravel slope with gravel lens.

The second factor is a hydraulically-conductive zone upslope of the gully. This inference is supported by the following observations: (i) Braided river channel infills, which tend to comprise highly permeable, coarse grained materials (Moreton et al., 2002), lead into the gullies' heads (Figs. 2a-b); (ii) Clustered distribution of gullies between the two braided rivers with the highest flow rates (Rakaia and Rangitata Rivers (Environment Canterbury,

Deleted: Box canyons

Formatted: Font: (Default) Times New Roman, 10 pt, Font color: Black

Deleted: The morphologic attributes of the box canyon ... [6]

Formatted ... [7]

Formatted ... [8]

Formatted ... [9]

Deleted: ¶

Deleted: box canyons

Deleted: , which

Deleted:

Deleted: 4

Deleted: This

Deleted:

Deleted: Box canyons

Deleted: (Figs. 5, 7)

Deleted: box canyons

Deleted: canyon

Formatted ... [10]

Deleted: The shear stress of the water seeping out of ... [11]

Formatted ... [12]

Formatted ... [13]

Formatted ... [14]

Formatted ... [15]

Formatted ... [16]

Deleted: W

Deleted: removes

Deleted: canyon

Deleted: 3c

Deleted: canyon

Deleted: <#>¶ ... [17]

Formatted ... [18]

Formatted ... [19]

Deleted: The location of the box canyons is controlled by

Deleted: t

Deleted: , as confirmed by the LEM

Deleted: (Figs. 10-12)

Deleted: box canyon

Deleted: box canyons

Deleted: 3a

Deleted: box canyons

2019)) (Fig. 1a); (iii) Geophysical observations (Fig. 8). With regards to the G-TEM slingram time-gate plots (Fig. 8), we interpret the higher-amplitude responses on the time-gate-1 plots that are preferentially located upslope of recently active gullies as zones of relatively high electrical conductivity in the subsurface at depths of ~10 m. These zones are suggestive of buried groundwater conduits made up of gravel and/or sandy units (e.g. Weymer et al., in review), or tunnels formed by sub-surface groundwater flow in sand units. Further analysis of the G-TEM data, including 2-D modelling and inversion, is required to ascertain the sub-surface hydraulic geometry responsible for the along-profile amplitude variations. This is elaborated further in the Supplementary Materials. The above observations confirm the importance of spatial variations in hydrogeological properties as a factor controlling the location of a gully. This had initially been suggested by Dunne (1990) and has been documented for gullies in bedrock environments (Laity and Malin, 1985; Newell, 1970). Development of gullies downslope of permeable conduits may also explain why most of the erosion entails elongation of existing gullies, rather than formation of new ones (Figs. 4, 6). It also agrees with the results of experimental modelling by Berhanu et al. (2012), which suggest that channels grow preferentially at their tip when the groundwater flow is driven by an upstream flow. If seaward-directed groundwater conduits are responsible for the location of gullies, the G-TEM results predict that, along the Canterbury coast, we should generally observe active gully development downslope of peaks in slingram time-gate plots. If this is the case, G-TEM could be used to identify locations of incipient and even future gully development.

5.3 Temporal scale of gully formation

Morphological changes derived from time-series of UAV data (Figs. 4-5) and satellite imagery (Fig. 6), as well as the observations of suspended fences across gullies (Fig. 3b), suggest that gully formation is rapid (daily timescales) and recent (<3 years ago). It is an episodic process that occurs after a threshold is exceeded. This threshold entails a rainfall intensity of >40 mm/day, which occurs once every 227 days, on average. These values are derived from the comparison of daily precipitation with the area of gullies eroded (Figs. 5, 6), and the plot of factor of safety with rainfall intensity from the slope stability model for the first scenario ((I-D); for the slope with sand lens (Fig. 10). The erosion rate documented in our study area is up to 30 m per day (Figs. 4e-f), which is the highest rate documented for gullies formed by groundwater so far.

The majority of the gullies in our study area have shown evidence of erosion in the past 11 years (Figs. 4, 6). The OSL results (Table 3), however, suggest that the two largest gullies have largely been inactive during at least the last 2 ka; recent erosion is only documented in small gullies located in the central section of their mouths (Figs. 4, 6). This contrasts with the inference by Schumm and Phillips (1986) that they were formed by spillage of water from swamps behind the cliffs in the 19th century. We therefore propose that the short gullies (<200 m in length) are recently active features, whereas the largest gullies are relict features that formed as a result of higher groundwater flow, and possibly surface erosion, in the past. The age of sample NZ13A suggests that this may have occurred during the Last Glacial Maximum. Such difference in age, and possibly formation process, between gullies of different length may explain the different cross-sectional shape and higher scatter in the plot of length vs. width for the longer gullies (Fig. 2c).

Deleted: 9

Deleted: We interpret the higher amplitude response in G-TEM profiles located upslope of recently active box canyons as buried groundwater conduits

Deleted: box canyon

Deleted: box canyons

Deleted: box canyons

Deleted: canyons

Deleted: 5

Deleted: 7

Deleted: ¶

Deleted: Box canyon

Deleted: (Figs. 6, 7b, 11, 15)

Formatted: Font: Not Bold

Deleted: (Fig. 7b). According to our LEM, up to 100 m^3 of water are estimated to have seeped out of the cliff face to erode 2540 m^3 of material (Fig. 15), which contrasts with the inference by Howard (1988) that 100-1000 times more water than volume of eroded sediment must be discharged in order to create a sapping valley.

Deleted: 5e

Formatted: Highlight

Formatted: Highlight

Deleted: box canyons

Deleted: box canyons

Deleted: data

Deleted: box canyons

Deleted:

Deleted: (

Deleted: ,

Deleted: 5

Deleted: -7

Deleted: Instead w

Deleted: propose

Deleted: that the

Deleted: box canyons

Deleted: <#>¶

Formatted: Font: (Default) Times New Roman, 10 pt, Font color: Black

6 Conclusions

Gully erosion is a prevalent process shaping the Canterbury coast of the South Island of New Zealand. In this study we have integrated field observations, OSL dating, multi-temporal UAV and satellite data, time-domain electromagnetic surveying, and slope stability modelling to constrain the controlling factors and temporal scales of gully formation. Our results indicate that gully development in sandy gravel cliffs is a groundwater-related, episodic process that occurs when rain falls at intensities of >40 mm per day. At the study area, such rainfall events occur at a mean frequency of once every 227 days. Gullies have been developing, primarily by elongation, in the last 11 years, with the latest episode dating to 3 years ago. Gullies form within days and erosion rates can reach values of up to 30 m per day. Gullies longer than 200 m, on the other hand, appear to be relict features that formed by higher groundwater flow and surface erosion >2 ka ago. The key processes responsible for gully development are the formation of alcoves and tunnels in sandy lenses by groundwater seepage erosion, followed by retrogressive slope failure. The latter is a result of undermining and a decrease in shear strength due to excess pore pressure development in the lower part of the slope. The location of the gullies is controlled by the occurrence of hydraulically-conductive zones, which comprise relict braided river channels and possibly tunnels, and sand lenses exposed across the sandy gravel cliff. We also show that gully planform shape is generally geometrically similar at consecutive stages of evolution. The outcomes of our study can improve reconstruction and prediction of an overlooked geohazard along the Canterbury coastline.

Deleted: Box canyon

Deleted: formation

Deleted: and landscape evolution

Deleted: box canyon

Deleted: box canyon

Deleted: to

Deleted: Box canyons

Deleted: Canyons

Deleted: The two largest box canyons in the study area, however, have not been actively elongating or widening in the last 2 ka.

Deleted: canyon

Deleted: box canyons

7 Code and data availability

We used Drone Deploy (https://www.dronedeploy.com/), IXG-TEM (http://www.interpex.com/) and Slide2 (https://www.rocscience.com/software/slide2) in this paper. All data from this study appear in the tables, figures, main text and supplementary materials.

Deleted: Box canyon formation is best represented by a linear diffusive model and geometrical scaling.¶
¶

Formatted: Highlight

Deleted: ,

Deleted: and Landlab (https://landlab.github.io/)

Deleted: Only the latter is freely available.

8 Author contribution

A.M. designed the study and drafted the manuscript, which was revised by all co-authors. A.M., R.M, P.P., M.E. B.A.W. and P.W. participated in the fieldwork. R.M. and R.P.T. interpreted the UAV data and satellite imagery. N.S., and D.C. carried out the slope stability modelling. P.P. and M.E. processed the geophysical data. A.A. and A.T.G. were in charge of the optically stimulated luminescence dating.

Deleted: and

Deleted: R.C.G.

Deleted: and landscape evolution

9 Competing interests

The authors declare that they have no conflict of interest.

10 Acknowledgments

We are grateful to Robbie Bennett, Clark Fenton and Daniele Spatola for their assistance during fieldwork, and to Environment Canterbury for the provision of data.

Deleted: ,

Deleted: and Philippe Wernette

11 Financial support

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreements No 677898 (MARCAN) and No 678106 (INTERTRAP)).

References

Abotalib, A. Z., Sultan, M., and Elkadiri, R.: Groundwater processes in Saharan Africa: Implications for landscape evolution in arid environments, Earth-Science Reviews, 156, 108-136, 2016.

Abrams, D. M., Lobkovsky, A. E., Petroff, A. P., Straub, K. M., McElroy, B., Mohrig, D., Kudrolli, A., and Rothman, D. H.: Growth laws for channel networks incised by groundwater flow, Nature Geoscience, 2, 193-196, 2009.

Aqualinc Research Limited: Canterbury groundwater model 2. Christchurch (NZ), Aqualinc Research Limited, 2007.

Bal, A. A.: Valley fills and coastal cliff s buried beneath an alluvial plain: Evidence from variation of permeabilities in gravel aquifers, Canterbury Plains, New Zealand, Journal of Hydrology (New Zealand), 35, 1-27, 1996.

Berger, G. W., Tonkin, P. J., and Pillans, B.: Thermo-luminescence ages of post-glacial loess, Rakaia River, South Island, New Zealand, Quaternary International, 35/36, 177-182, 1996.

Berhanu, M., Petroff, A. P., Devauchelle, O., Kudrolli, A., and Rothman, D. H.: Shape and dynamics of seepage erosion in a horiztonal granular bed, Physical Review E, 86, 041304, 2012.

Browne, G. H., and Naish, T. R.: Facies development and sequence architecture of a late Quaternary fluvial-marine transition, Canterbury Plains and shelf, New Zealand: implications for forced regressive deposits, Sedimentary Geology, 158, 57-86, 2003.

Buylaert, J.-P., Murray, A. S., and Thomsen, K. J.: Testing the potential of an elevated temperature IRSL signal from K-feldspar, Radio Measurements, 44, 560-565, 2009.

Buylaert, J. P., Thiel, C., Murray, A. S., D., V., Yi, S., and Lu, H.: IRSL and post-IR IRSL residual doses recorded in modern dust samples from the Chinese Loess Plateau, Geochronometria, 38, 432-440, 2011.

Chu-Agor, M. L., Fox, G. A., Cancienne, R. M., and Wilson, G. V.: Seepage caused tension failures and erosion undercutting of hillslopes, Journal of Hydrology, 359, 247-259, 2008.

Coelho Netto, A. L., Fernandes, N. F., and Edegard de Deus, C.: Gullying in the southeastern Brazilian Plateau, Bananal, SP, Porto Alegre Symposium, Porto Alegre, 1988, 35-42,

Collins, B. D., and Sitar, N.: Geotechnical properties of weakly and moderately cemented sands in steep slopes, Journal of Geotechnical and Geoenvironmental Engineering, 135, 1359-1366, 2009.

Collins, B. D., and Sitar, N.: Stability of steep slopes in cemented sands, Journal of Geotechnical and Geoenvironmental Engineering, 137, 43-51, 2011.

Constable, S. C., Parker, R. L., and Constable, C. G.: Occam's inversion: Apractical algorithm for generating smooth models from EM sounding data, Geophysics, 52, 289-300, 1987.

Dann, R., Close, M., Flinto, M., Hector, R., Barlow, H., Thomas, S., and Francis, G.: Characterization and estimation of hydraulic properties in an alluvial gravel vadose zone, Vadose Zone Journal, 8, 651-663, 2009.

Davey, G.: Definition of the Canterbury Plains Aquifers, Environmental Canterbury, 2006.

Devauchelle, O., Petroff, A. P., Seybold, H. F., and Rothman, D. H.: Ramification of stream networks, Proceedings of the National Academy of Sciences of the United States of America, 109, 20832-20836, 2012.

Domenico, P. A., and Schwartz, F. W.: Physical and Chemical Hydrogeology, John Wiley, Chichester, 528 pp., 1997.

Dunne, T.: Hydrology, mechanics, and geomorphic implications of erosion by subsurface flow, in: Groundwater Geomorphology: The Role of Subsurface Water in Earth-Surface Processes and Landforms, edited by: Higgins, C. G., and Coates, D. R., Special Paper Geological Society of America, 1-25, 1990.

Environment Canterbury: River flow data, www.ecan.govt.nz/data/riverflow, 2019.

Fitterman, D. V.: Tools and techniques: Active-source electromagnetic methods, in: Resources in the Near-Surface Earth, Treatise on Geophysics, edited by: Slater, L., Elsevier, 295-333, 2015.

Fox, G. A., Wilson, G. V., Periketi, R. K., and Cullum, R. F.: Sediment transport model for seepage erosion of streambank erosion, Journal of Hydrologic Engineering, 11, 603-611, 2006.

Fredlund, D. G., and Krahn, J.: Comparison of slope stability methods of analysis, Canadian Geotechnical Journal, 14, 429-439.

Fredlund, D. G., Krahn, J., and Pufahl, D.: The relationship between limit equilibrium slope stability methods, 10th International Conference on Soil Mechanics and Foundation Engineering, Stockholm, Sweden, 1981,

Geonics: G-TEM Operating Manual, Geonics Ltd., Mississauga, Canada, 2016.

Gibb, J. G.: Rates of coastal erosion and accretion in New Zealand, New Zealand Journal of Marine and Freshwater Research, 12, 429-456, 1978.

Harrison, K. P., and Grimm, R. E.: Groundwater-controlled valley networks and the decline of surface runoff on early Mars, Journal of Geophysical Research, 110, E12S16, 2005.

Higgins, C. G.: Drainage systems developed by sapping on Earth and Mars, Geology, 10, 147-152, 1982.

Howard, A. D.: Groundwater sapping on Earth and Mars, in: Sapping Features of the Colorado Plateau, edited by: Howard, A. D., Kochel, R. C., and Holt, H. R., NASA, 1-4, 1988.

Howard, A. D., and McLane, C. F.: Erosion of cohesionless sediment by groundwater seepage, Water Resources Research, 24, 1659-1674, 1988.

Howard, A. D.: Case study: Model studies of ground-water sapping, in: Geological Society of America Special Paper, edited by: Higgins, C. G., and Coates, D. R., 257-264, 1990.

Howard, A. D.: Simulation modeling and statistical classification of escarpment planforms, Geomorphology, 12, 187-214, 1995.

Interpex: IXG-TEM Instruction Manual, Interpex Ltd., Golden, USA, 2012.

Kirk, R. M.: River-beach interaction on mixed sand and gravel coasts: A geomorphic model for water resource planning, Applied Geography, 11, 267-287, 1991.

Kline, S. W., Adams, P. N., and Limber, P. W.: The unsteady nature of sea cliff retreat due to mechanical abrasion, failure and comminution feedbacks, Geomorphology, 219, 53-67, 2014.

Kochel, R. C., Howard, A. D., and McLane, C. F.: Channel networks developed by groundwater sapping in fine-grained sediments: Analogs to some Martian valleys, in: Models in Geomorphology, edited by: Woldenberg, M., Allen and Unwin, St Leonards, Australia, 313-341, 1985.

Kochel, R. C., and Piper, J. F.: Morphology of large valleys on Hawaii - Evidence for groundwater sapping and comparisons with Martian valleys, Journal of Geophysical Research, 91, E175-E192, 1986.

Laity, J. E., and Malin, M. C.: Sapping processes and the development of theater-headed valley networks in the Colorado Plateau, Geological Society of America Bulletin, 96, 203-217, 1985.

Lamb, M. P., Howard, A. D., Johnson, J., Whipple, K. X., Dietrich, W. E., and Perron, J. T.: Can springs cut canyons into rock?, Journal of Geophysical Research, 111, E07002, 2006.

Laporte-Fauret, Q., Marieu, V., Castelle, B., Michalet, R., Bujan, S., and Rosebery, D.: Low-Cost UAV for High-Resolution and Large-Scale Coastal Dune Change Monitoring Using Photogrammetry, Journal of Marine Science and Engineering, 7, 63, 2019.

Lapotre, M. G. A., and Lamb, M. P.: Substrate control on valley formation by groundwater on Earth and Mars, Geology, 46, 531-534, 2018.

Leckie, D. A.: Modern environments of the Canterbury Plains and adjacent offshore areas, New Zealand — an analog for ancient conglomeratic depositional systems in nonmarine and coastal zone settings, Bullettin of Canadian Petroleum Geology, 51, 389-425, 2003.

Leyland, J., and Darby, S. E.: An empirical–conceptual gully evolution model for channelled sea cliffs, Geomorphology, 102, 419-434, 2008.

Leyland, J., and Darby, S. E.: Effects of Holocene climate and sea-level changes on coastal gully evolution: insights from numerical modelling, Earth Surface Processes and Landforms, 34, 1878-1893, 2009.

Limber, P. W., and Barnard, P. L.: Coastal knickpoints and the competition between fluvial and wave-driven erosion on rocky coastlines, Geomorphology, 306, 1-12, 2018.

Lobkovsky, A. E., Jensen, B., Kudrolli, A., and Rothman, D. H.: Threshold phenomena in erosion driven by subsurface flow, Journal of Geophysical Research, 109, F04010, 2004.

Mackey, B. H., Scheingross, J. S., Lamb, M. P., and Farley, K. A.: Knickpoint formation, rapid propagation, and landscape response following coastal cliff retreat at the last interglacial sea-level highstand: Kaua'i, Hawai'i, Geological Society of America Bulletin, 126, 925-942, 2014.

Malin, M. C., and Carr, M. H.: Groundwater formation of Martian valleys, Nature, 397, 589-591, 1999.

Micallef, A., Person, M., Haroon, A., Weymer, B. A., Jegen, M., Schwalenberg, K., Faghih, Z., Duan, S., Cohen, D., Mountjoy, J. J., Woelz, S., Gable, C. W., Averes, T., and Tiwari, A. K.: 3D characterisation and quantification of an offshore freshened groundwater system in the Canterbury Bight, Nature Communications, 11, 1372, 2020.

Moreton, D. J., Ashworth, P. J., and Best, J. L.: The physical scale modelling of braided alluvial architecture and estimation of subsurface permeability, Basin Research, 14, 265-285, 2002.

Nabighian, M. N., and Macnae, J. C.: 6. Time Domain Electromagnetic Prospecting Methods, in: Electromagnetic Methods in Applied Geophysics: Volume 2, Application, Parts A and B, 427-520, 1991.

Nash, D. J., Shaw, P. A., and Thomas, D. S. G.: Duricrust development and valley evolution: Process-landforms links in the Kalahari, Earth Surface Processes and Landforms, 19, 299-317, 1994.

Deleted: , https://doi.org/10.1038/s41467-020-14770-7

Nash, D. J.: Groundwater sapping and valley development in the Hackness Hills, North Yorkshire, England, Earth Surface Processes and Landforms, 21, 781-795, 1996.

Newell, M.: Canyonlands - modern history, Naturalist, 21, 40-47, 1970.

Onda, Y.: Seepage erosion and its implications to the formation of amphitheatre valley heads: A case study at Obara, Japan, Earth Surface Processes and Landforms, 19, 624-640, 1994.

Pelletier, J. D., and Baker, V. R.: The role of weathering in the formation of bedrock valleys on Earth and Mars: A numerical modeling investigation, Journal of Geophysical Research, 116, E11007, 2011.

Petroff, A. P., Devauchelle, O., Abrams, D. M., Lobkovsky, A. E., Kudrolli, A., and Rothman, D. H.: Geometry of valley growth, Journal of Fluid Mechanics, 673, 245-254, 2011.

Pillans, B.: Drainage initiation by subsurface flow in South Taranaki, New Zealand, Geology, 13, 262-265, 1985.

Pondthai, P., Everett, M. E., Micallef, A., Weymer, B. A., Faghih, Z., Haroon, A., and Jegen, M.: 3D characterization of a coastal freshwater aquifer in SE Malta (Mediterranean Sea) by time-domain electromagnetics, Water, 12, 1566, 2020.

Preusser, F., Chithambo, M. L., Götte, T., Martini, M., Ramseyer, K., Sendezera, E. J., Susino, G. J., and Wintle, A. G.: Quartz as a natural luminescence dosimeter, Earth Science Reviews, 97, 184-214, 2009.

Salese, F., Pondrelli, M., Neeseman, A., Schmidt, G., and Ori, G. G.: Geological evidence of planet-wide groundwater system on Mars, Journal of Geophysical Research, 124, 374-395, 2019.

Schorghofer, N., Jensen, B., Kudrolli, A., and Rothman, D. H.: Spontaneous channelization in permeable ground: Theory, experiment, and observation, Journal of Fluid Mechanics, 503, 357-374, 2004.

Schumm, S. A., and Phillips, L.: Composite channels of the Canterbury Plain, New Zealand: A Martian analog?, Geology, 14, 326-329, 1986.

Schumm, S. A., Boyd, K. F., Wolff, C. G., and Spitz, W. J.: A groundwater sapping landscape in the Florida panhandle, Geomorphology, 12, 281-297, 1995.

Scott, G. L.: Near-surface hydraulic stratigraphy of the Canterbury Plains between Ashburton and Rakaia rivers, New Zealand, Journal of Hydrology (New Zealand), 19, 68-74, 1980.

Sunderlin, D., Trop, J. M., Idleman, D., Brannick, A., White, J. G., and Grande, L.: Paleoenvironment and paleoecology of a Late Paleocene high-latitude terrestrial succession, Arkose Ridge Formation at Box Canyon, southern Talkeetna Mountains, Alaska, Palaeogeography, Palaeoclimatology, Palaeoecology, 401, 57-80, 2014.

Thiel, C., Buylaert, J.-P., Murray, A., Terhorst, B., Hofer, I., Tsukamoto, S., and Frechen, M.: Luminescence dating of the Stratzig loess profile (Austria) – Testing the potential of an elevated temperature post-IR IRSL protocol, Quaternary International, 234, 23-31, 2011.

Uchupi, E., and Oldale, R. N.: Spring sapping origin of the enigmatic relict valleys of Cape Cod and Martha's Vineyard and Nantucket Islands, Massachussets, Geomorphology, 9, 83-95, 1994.

Weymer, B. A., Everett, M. E., Houser, C., Wernette, P., and Barrineau, P.: Differentiating tidal and seasonal effects on barrier island hydrogeology: Testing the utility of portable multi-frequency EMI profilers, Geophysics, 81, E347-361, 2016.

Weymer, B. A., Wernette, P., Everett, M. E., Pondthai, P., Jegen, M., and Micallef, A.: Multilayered high-permeability conduits connecting onshore and offshore coastal groundwater aquifers, Frontiers in Marine Science – Coastal Ocean Processes, in review.

Whitaker, S.: Flow in Porous-Media I: a theoretical derivation of Darcy's-Law, Transport in Porous Media, 1, 3-25, 1986.

Wilson, G. V., Periketi, R., Fox, G. A., Dabney, S., Shields, D., and Cullum, R. F.: Seepage erosion properties contributing to streambank failure, Earth Surface Processes and Landforms, 32, 447-459, 2007.

Ye, F.-Y., Barriot, J.-P., and Carretier, S.: Initiation and recession of the fluvial knickpoints of the Island of Tahiti (French Polynesia), Geomorphology, 186, 162-173, 2013.

Yi, R., Cohen, Y., Devauchelle, O., Gibbins, G., Seybold, H. F., and Rothman, D. H.: Symmetric rearrangement of groundwater-fed streams, Proceedings of the Royal Society of London, Series A, 473, 20170539., 2017.

Page 22: [1] Formatted	Microsoft account	28/07/2020 16:14:00

Font: (Default) Times New Roman, 10 pt, Font color: Auto

Page 22: [2] Formatted	Microsoft account	28/07/2020 16:14:00

Font: (Default) Times New Roman, 10 pt, Font color: Auto

Page 22: [3] Deleted	Microsoft account	28/07/2020 16:22:00

using the IXG-TEM software from Interpex Limited

Page 26: [4] Deleted	Microsoft account	29/07/2020 12:31:00

Landscape evolution modelling

We also built a landscape evolution model (LEM) using the Python modelling environment Landlab (Barnhardt et al., 2020; Hobley et al., 2017). The model allowed us to simulate two main processes: groundwater flow and associated erosion (Fig. 2).

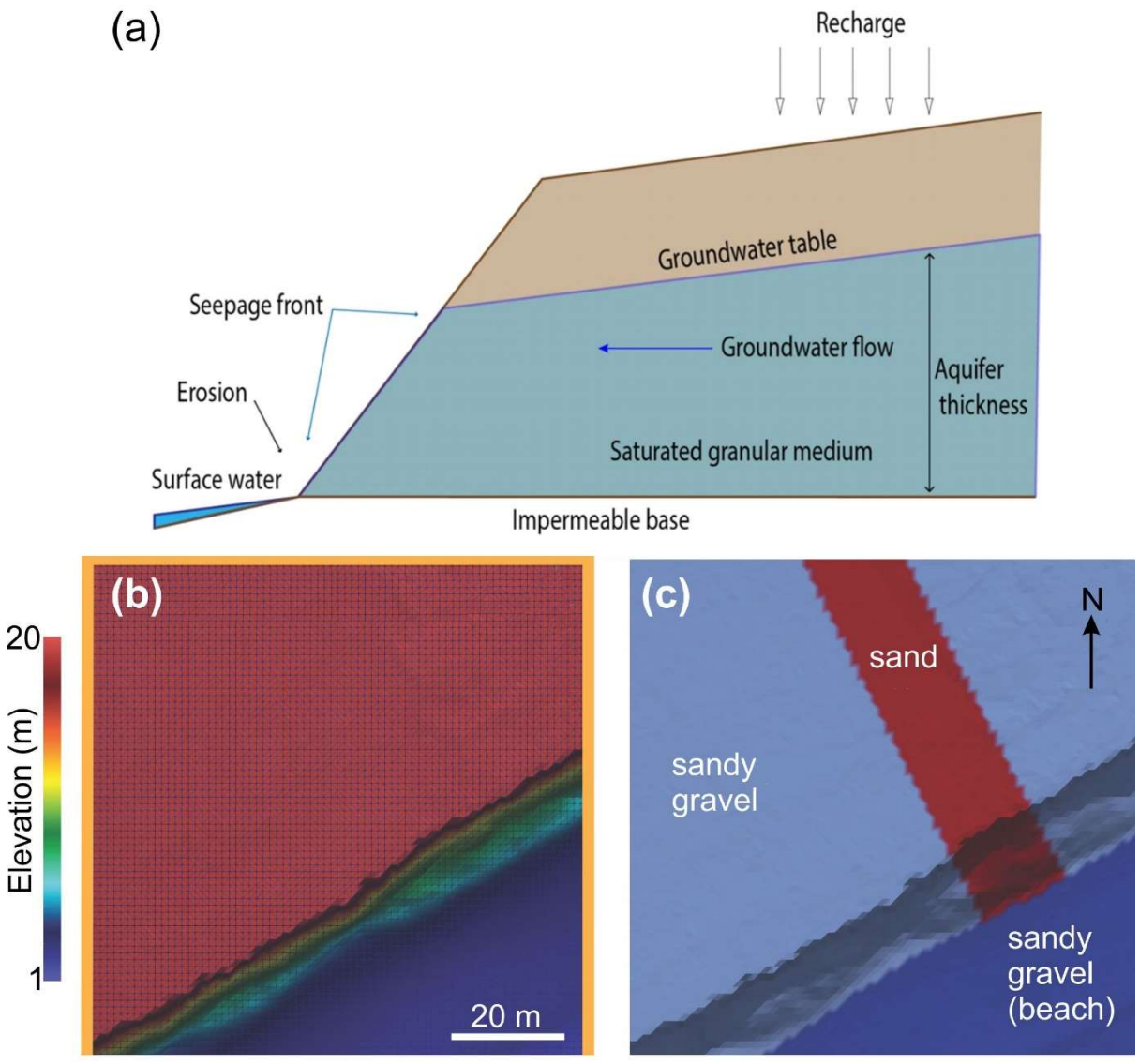


Figure 2: (a) Conceptual model of the LEM. (b) Initial topography used in the LEM. Orange line denotes closed boundary. (c) Spatial distribution of sediment type in the LEM.

The model uses the Dupuit-Forchheimer approximation to simulate groundwater flow, which was developed for Landlab by Litwin et al. (2020). This element of the model solves the Boussinesq equation for horizontal flow direction in an unconfined aquifer over an impermeable aquifer base. For a detailed description of the theory and its implementation in Landlab, the reader is referred to the documentation in https://landlab.readthedocs.io/en/master/reference/components/dupuit_theory.html. We computed the groundwater flow and the groundwater seepage; the latter occurs when the water table intersects the topographic surface and then becomes surface runoff. The model considers that all surface water is derived from groundwater seepage.

We used the model to test two erosional modes. The first mode is the stream power law model, which simulates surface erosion (Barnhart et al., 2019; Braun and Willett, 2013):

$$\frac{\partial \eta}{\partial t} = -KQ^m S^n \tag{7}$$

where η is the topographic elevation [L], t is time [T], K is an erosion coefficient [L^{1-3m} T^{m-1}], Q is the surface water discharge [L^3T^{-1}], S is the slope (dimensionless) and m and n are exponents (dimensionless). We assumed $m = 0.5$ and $n = 1$, due to a lack of data and a robust methodology to determine these two parameters (Harel et al., 2016).

The second mode is a linear diffusion to simulate erosion by gravitational sediment movement (Barnhart et al., 2019; Culling, 1963):

$$\frac{\partial \eta}{\partial t} = -D\nabla^2 \eta \tag{8}$$

where D is the diffusion coefficient [L^3T^{-1}].

For the sake of simplicity and to easily compare the results, K (Eq. (7)) is computed as D (Eq. (8)), which we consider to be proportional to surface water shear stress and seepage flux:

$$D = Q + M\left(\frac{\tau_b - \tau_t}{\tau_t}\right) \tag{9}$$

where τ_b is the shear stress [MLT^{-2}], τ_t is the threshold shear stress [MLT^{-2}] (assumed to have a value of 0.03 N m^{-2}), and M is an empirical parameter [L^3T^{-1}] that we assume to have a value of 0.1 m^3 s^{-1}. The values for τ_t and M are empirical and were estimated by trial and error.

Additionally, as a theoretical exercise, we tested the evolution of the surface using the stream power law approach with surface flow and without groundwater seepage. Equation (7) was changed by including surface water discharge (Q) from an upstream drainage area:

$$\frac{\partial \eta}{\partial t} = -KA^m S^n \quad (10)$$

The initial topography was extracted from a digital elevation model of part of the study area that has not been affected by canyon erosion (Fig. 2b), which allowed us to incorporate a realistic approximation of the cliff morphology. To simulate the groundwater flow, we arbitrarily defined the base of the aquifer at 0.1 m above sea level, to minimise vertical groundwater movement, and applied the Dupuit-Forchheimer approximation. The initial water table was placed at 5.9 m above sea level (in accordance with well data). Since Landlab is a 2-D modelling environment, it was not possible to include sand lenses. For this reason, we simplified the geology by considering two types of sediment: sandy gravels and with a NW-SE strip of sand. Their distribution and properties are shown in Fig. 2c and Table 2, respectively. In Eq. (10), we assumed that the erosion coefficient (K) of the sand unit is the double that of sandy gravel.

The aquifer recharge was modelled using two different precipitation scenarios. The 'high intensity' scenario is based on a first hour of low recharge (1×10^{-9} m s^{-1}) and a 2 hour long high recharge storm (2.8×10^{-5} m s^{-1}). The latter value was selected to replicate the rise in the water table reported after the 21st July 2017 storm (see Sect. 4.4.1). The 'low intensity' scenario consists of a first hour of low recharge (1×10^{-9} m s^{-1}), followed by a 5 hour long low recharge storm (1.12×10^{-5} m s^{-1}). The total modelling time was 24 hours.

Page 41: [5] Deleted **Microsoft account** **16/07/2020 12:40:00**

4.6.2 LEM

The model results for the two erosional modes and two rainfall scenarios are displayed in Figs. 13-16 and discussed in more detail below.

The results of the stream power law model with a high intensity rainfall scenario are shown in Fig. 13. During the first hour and as a result of the low recharge value (1×10^{-9} m s^{-1}), the water table drops slowly and no erosion occurs. All parameters, except for shear stress, decrease rapidly. An intense storm (recharge of 2.8×10^{-5} m s^{-1}) starts after 1 hr; groundwater seeps out at the cliff base, where the shear stress is highest, forming a gully between the cliff base and the beach. In the ensuing 2 hr, the recharge rate stays the same but the water table height in the sandy area decreases due to the high permeability; as a result the shear stress decreases and erosion stops. At the third hour, the storm stops and there is no recharge; the water table drops slowly and the shear stress and groundwater seepage decrease. The results of the stream power law model with a low intensity rainfall scenario are similar to those of the high intensity rainfall scenario (Fig. 14).

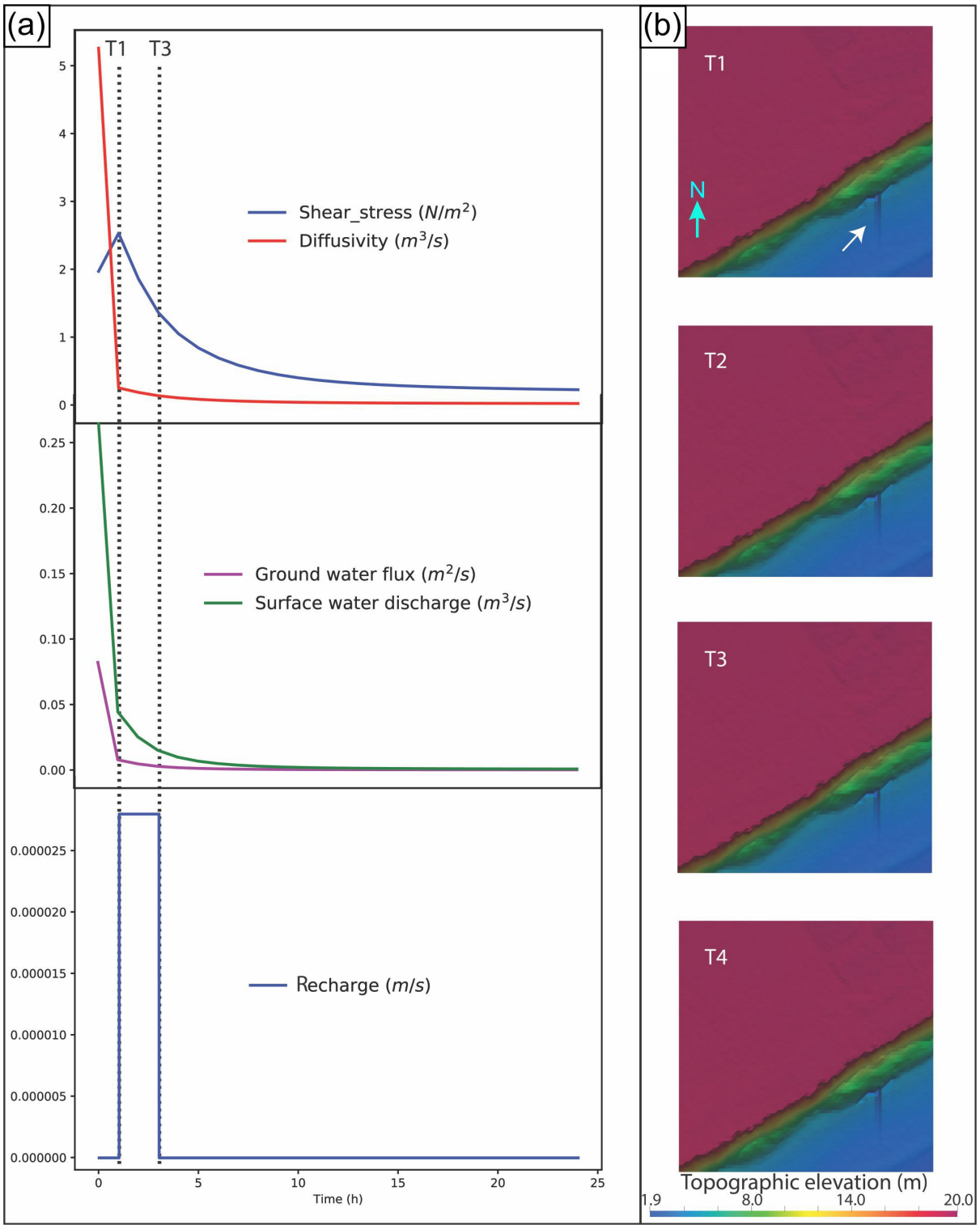


Figure 13: Evolution of the main simulated parameters for the stream power law with high intensity rainfall scenario. Section A shows the evolution curves of the maximum value of the main parameters. Section B shows the topographic evolution during the first 4 hours. White arrow indicates the gully.

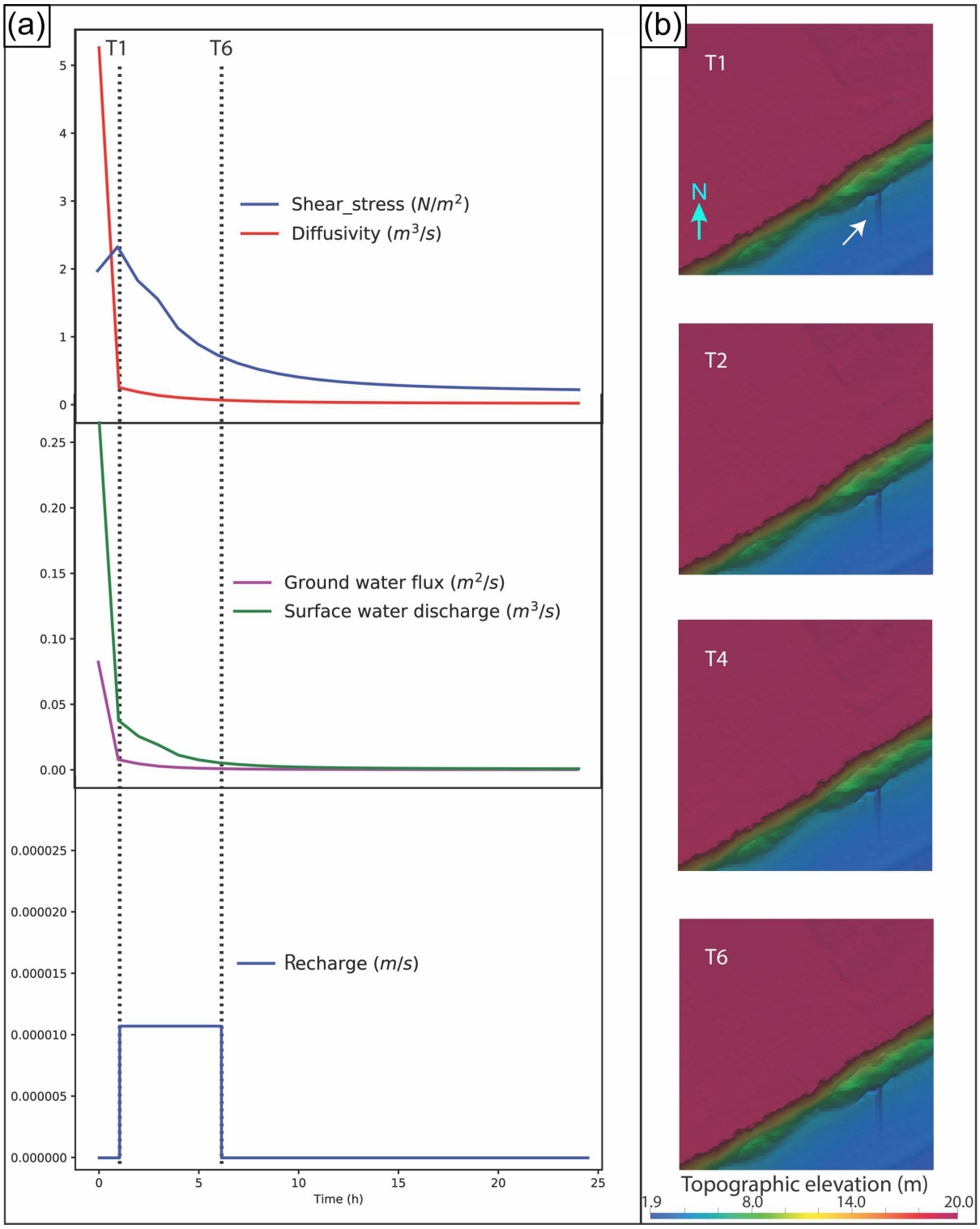


Figure 14: Evolution of the main simulated parameters for the stream power law with low intensity rainfall scenario. Section A shows the evolution curves of the maximum value of the main parameters. Section B shows the topographic evolution during the first 4 hours. White arrow indicates the gully.

The results of the linear diffusion model with high intensity rainfall scenario are shown in Fig. 15. During the first hour of the simulation, the recharge is low (1×10^{-9} m s^{-1}); the water table height, shear stress, diffusivity coefficient and groundwater flux decrease very slowly, whereas the surface water discharge increases. After the first hour, an intense storm increases the recharge to 2.8×10^{-5} m s^{-1}. The water table height rises quickly (from 5.9 to 6.5 m), as do all the parameters, reaching their maximum values. A small box canyon develops at the base of the cliff where the shear stress is highest. The intense storm continues for the next 2 hours, and the water table height rises quickly to its maximum value of 7.1 m. The main parameters decrease, and the box canyon enlarges to 16.5 m × 11.3 m, eroding into the cliff. At the end of the storm there is no recharge, the water table drops slowly, the shear stress and the groundwater seepage decrease rapidly, and erosion stops. During the rest of the simulation there is no recharge; the water table continues to drop, the shear stress and the groundwater seepage decrease, and no erosion occurs.

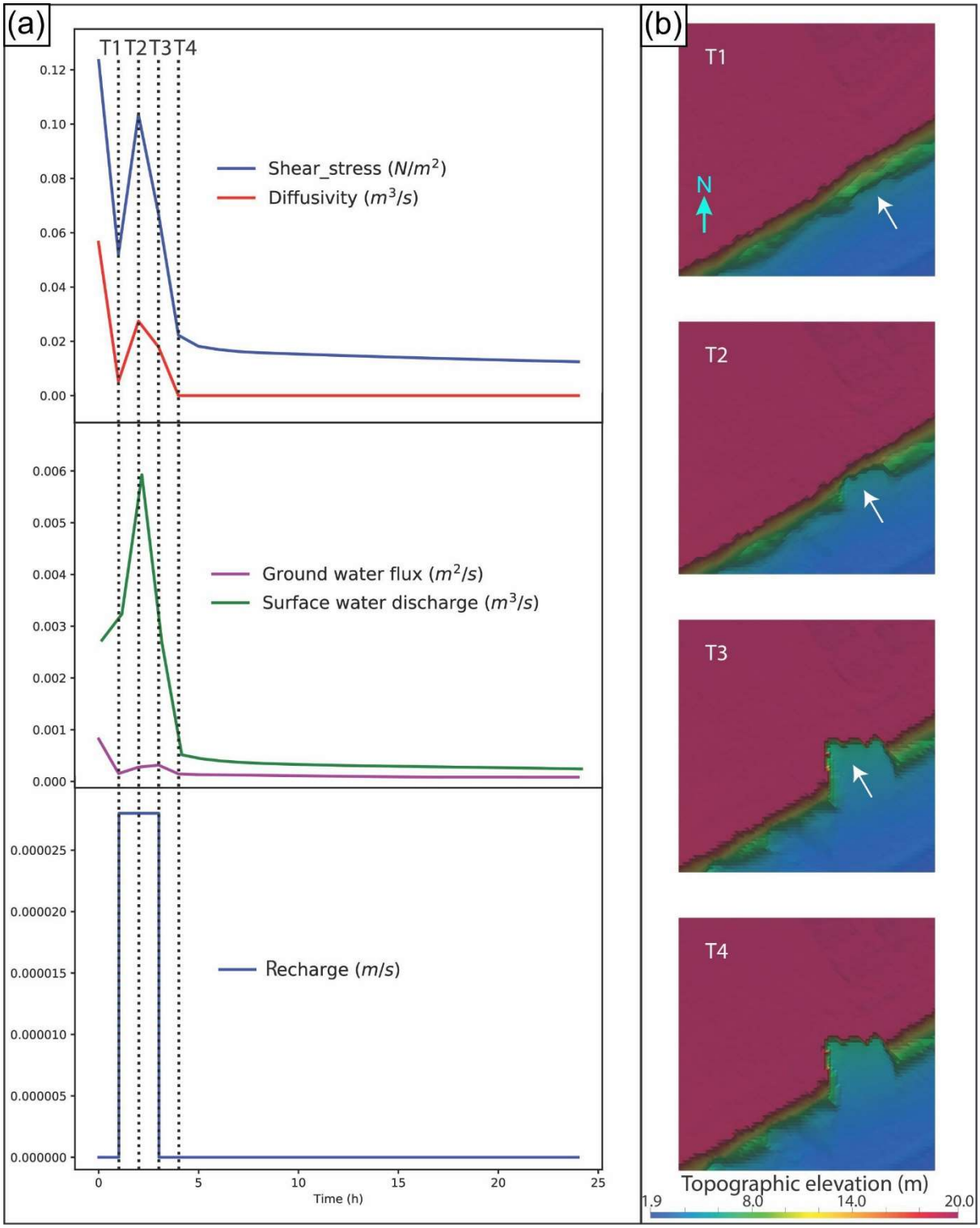


Figure 15: Evolution of the main simulated parameters for the linear diffusion with high intensity rainfall scenario. Section A shows the plots of the maximum value of the main parameters. Section B shows the topographic evolution during the first 4 hours. White arrow denotes the direction of erosion.

The linear diffusion with low intensity rainfall model (Fig. 16) starts with a low recharge (1×10^{-9} m s^{-1}) during the first hour. During this period, the water table drops slowly, the main parameters (shear stress, diffusion coefficient groundwater flux and surface water discharge) decrease rapidly and there is no erosion. After 1 hr, a low intensity storm starts (recharge of 1.12×10^{-5} m s^{-1}). The water table height increases slowly; shear stress, groundwater flux

and surface water discharge increase, whereas diffusion continues to decrease gradually. The erosion process forms a scarp at the base of the cliff. The storm continues for 4 hr. The water table rises slowly, but the high permeability in the sand unit does not allow the water table to increase in this area. All parameters continue to increase, but their values are not high enough to erode the cliff further. After 6 hr, the storm stops; there is no recharge, the water table drops slowly, and the values of the main parameters decrease.

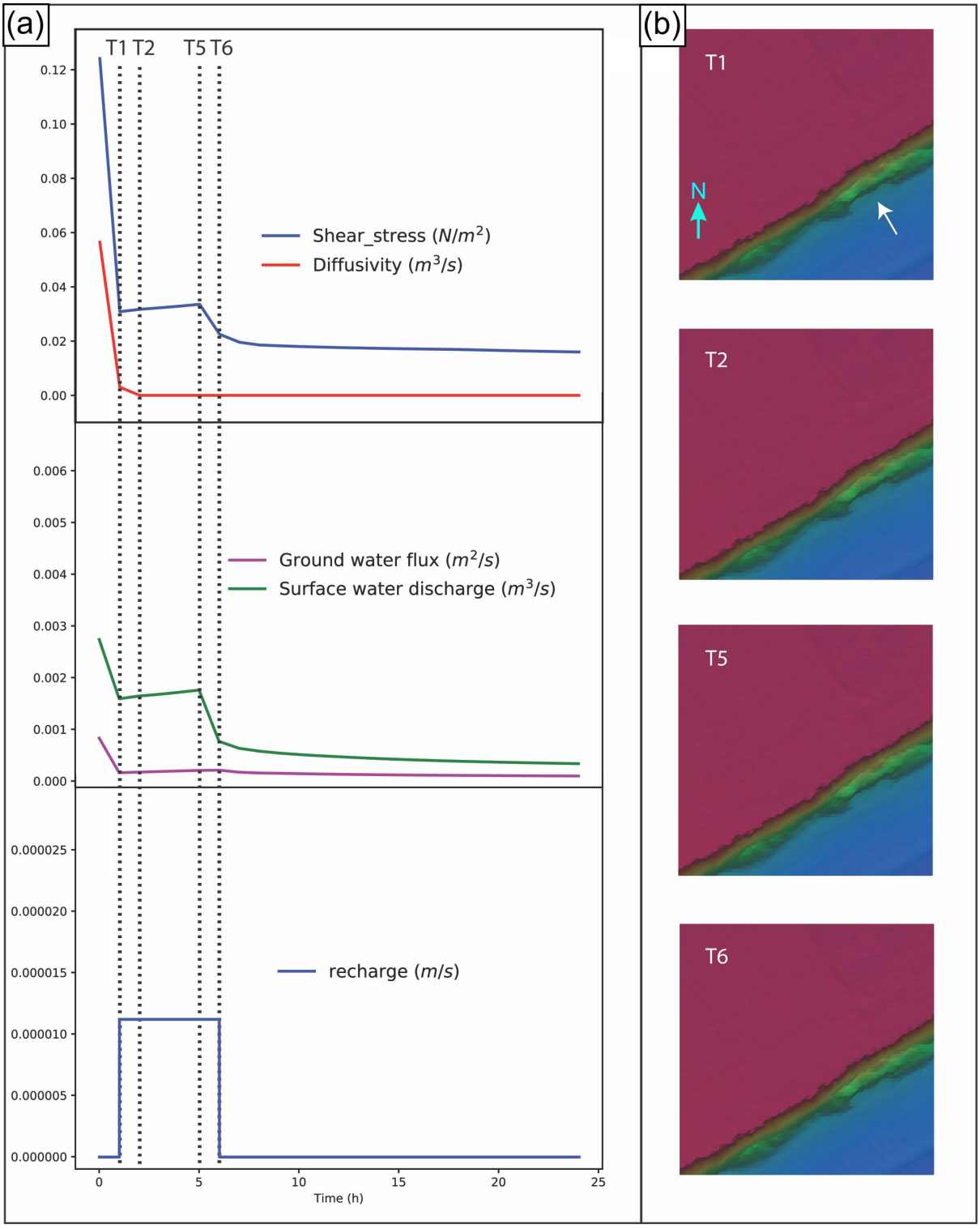


Figure 16: Evolution of the main simulated parameters for the linear diffusion with low intensity rainfall scenario. Section A shows the evolution curves of the maximum value of the main parameters. Section B shows the topographic evolution during the first 4 hours. White arrow marks the scarp.

Finally, the results of the stream power law model without groundwater seepage show a set of closely-spaced gullies, growing from the base of the cliff upwards. The largest of these are incised into the sand unit (Fig. 17).

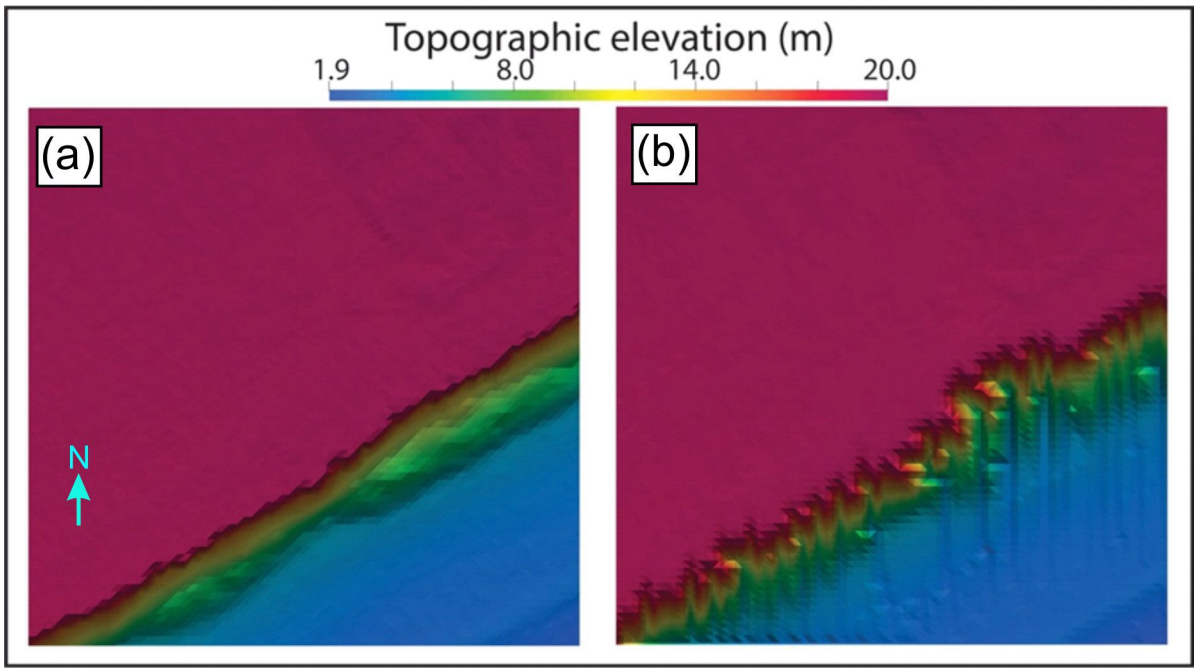


Figure 17: Topographic evolution using the stream power law without groundwater seepage. (a) Initial topographic conditions. (b) Topographic conditions after 24 hours.

Page 42: [6] Deleted	Microsoft account	30/07/2020 14:41:00

The morphologic attributes of the box canyons are similar to those associated with groundwater seepage elsewhere (Table 1).

Page 42: [7] Formatted	Microsoft account	29/07/2020 14:13:00

Font: (Default) Times New Roman, 10 pt

Page 42: [8] Formatted	Microsoft account	29/07/2020 14:13:00

Font: (Default) Times New Roman, 10 pt, Font color: Black

Page 42: [9] Formatted	Microsoft account	29/07/2020 14:13:00

Font: (Default) Times New Roman, 10 pt, Font color: Black

Page 42: [10] Formatted	Microsoft account	07/08/2020 16:25:00

Highlight

Page 42: [11] Deleted	Microsoft account	30/07/2020 17:01:00

The shear stress of the water seeping out of the canyon head and flowing along the base of the canyon is high enough to entrain the failed sandy material (shear stress of 7.5×10^{-5} in Fig. 15a; Petit et al. (2015))(Julien, 1998), which explains why the floors of box canyons are primarily covered by gravel.

Page 42: [12] Formatted	Microsoft account	07/08/2020 16:25:00

Highlight

Page 42: [13] Formatted	Microsoft account	07/08/2020 16:25:00

Highlight

Page 42: [14] Formatted	Microsoft account	07/08/2020 16:25:00

Highlight

Page 42: [15] Formatted	Microsoft account	07/08/2020 16:25:00

Highlight

Page 42: [16] Formatted	Microsoft account	07/08/2020 16:24:00

Not Highlight

Page 42: [17] Deleted	Microsoft account	30/07/2020 15:55:00

~~The field observations of box canyon morphology and development are most similar to the LEM results for the high intensity rainfall scenario of the linear diffusive erosion model (Fig. 16). There are two reasons for this: first, as observed in Culling (1963) and Barnhardt et al. (2019), the retrogressive slope failure is best simulated by a diffusive linear model; secondly, our linear diffusive erosion model links the diffusion co-efficient with the shear stress generated by the groundwater. The evolution of the different models and scenarios appears to be influenced by the initial topography - with the cliff preventing the upslope development of erosive features in the stream power law mode, regardless of the recharge values – and the water table height, which depends on permeability and recharge (Figs. 13-16). In the stream power law mode, erosion is linked to surface water, which depends exclusively on groundwater seepage. Since the water table can only intersect the beach or the cliff surface, surface erosion of the plain above the cliff top is not possible (Figs. 13-14). When the stream power law model does not depend on groundwater seepage, the surface erosion does affect the cliff face, but it results in a series of closely spaced and narrow gullies (Fig. 17). The conceptual model, the program configuration and the initial and boundary conditions make the LEM outcomes only applicable to box canyon formation at this site.~~

Page 42: [18] Formatted	Microsoft account	30/07/2020 14:41:00

Font: (Default) Times New Roman, 10 pt, Font color: Black

Page 42: [19] Formatted	Microsoft account	16/07/2020 12:40:00

Strikethrough