Rock damage and regolith transport by frost: an example of climate modulation of the geomorphology of the critical zone

Robert S. Anderson,1,2* Suzanne P. Anderson1,3 and Gregory E. Tucker2,4
1 Institute of Arctic and Alpine Research (INSTAAR), University of Colorado, Boulder, CO. USA
2 Department of Geological Sciences, University of Colorado, Boulder, CO. USA
3 Department of Geography, University of Colorado, Boulder, CO. USA
4 Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, CO. USA

Received 7 December 2011; Revised 11 September 2012; Accepted 13 September 2012

*Correspondence to: Robert S. Anderson Institute of Arctic and Alpine Research (INSTAAR), University of Colorado, Boulder, CO. USA. E-mail: andersrs@colorado.edu

ABSTRACT: In this article we craft process-specific algorithms that capture climate control of hillslope evolution in order to elucidate the legacy of past climate on present critical zone architecture and topography. Models of hillslope evolution traditionally comprise rock detachment into the mobile layer, mobile regolith transport, and a channel incision or aggradation boundary condition. We extend this system into the deep critical zone by considering a weathering damage zone below the mobile regolith in which rock strength is diminished; the degree of damage conditions the rate of mobile regolith production. We first discuss generic damage profiles in which appropriate length and damage scales govern profile shapes, and examine their dependence upon exhumation rate. We then introduce climate control through the example of rock damage by frost-generated crack growth. We augment existing frost cracking models by incorporating damage rate limitations for long transport distances for water to the freezing front. Finally we link the frost cracking damage model, a mobile regolith production rule in which rock entrainment is conditioned by the damage state of the rock, and a frost creep transport model, to examine the evolution of an interfluve under oscillating climate. Aspect-related differences in mean annual surface temperatures result in differences in bedrock damage rate and mobile regolith transport efficiency, which in turn lead to asymmetries in critical zone architecture and hillslope form (divide migration). In a quasi-steady state hillslope, the lowering rate is uniform, and the damage profile is better developed on north-facing slopes where the frost damage process is most intense. Because the residence times of mobile regolith and weathered bedrock in such landscapes are on the order of 10 to 100 ka, climate cycles over similar timescales result in modulation of transport and damage efficiencies. These lead to temporal variation in mobile regolith thickness, and to corresponding changes in sediment delivery to bounding streams. Copyright © 2012 John Wiley & Sons, Ltd.

KEYWORDS: critical zone; frost processes; hillslope evolution; weathering; mobile regolith production; climate change

Introduction

The geomorphologist views the critical zone as a three-dimensional (3D) structure formed by weathering and transport processes, which are themselves powered by tectonics, climate, biota, and gravity. Understanding the complex interactions that govern the evolution of this structure is among the ‘grand challenges’ in Earth surface processes (National Research Council, 2010). A slice through the critical zone (Figure 1) reveals three interfaces of significance: (1) the boundary between atmosphere and the ground surface, (2) the boundary between disaggregated and displaced material that comprises mobile regolith, and the weathered but in-place bedrock beneath it, and (3) the boundary between weathered and fresh bedrock. To understand the evolution of critical zone structure requires process models that describe transformations and mass transport. Together these processes produce the structure we see and control the relative movement of these three interfaces. At a minimum, models of hillslope evolution require rules for the rate of detachment of rock into the mobile regolith layer, for the rate of mobile regolith transport, and for channel incision or aggradation rates that serve as boundary conditions for the hillslopes. A more complete model would also describe the evolution of material properties (e.g. mineralogy, grain size and derived quantities such as porosity, hydraulic conductivity and thermal diffusivity) within the weathered bedrock and on the conveyor belt of the mobile regolith layer. Here we present an approach to understanding hillslope evolution and critical zone structure in which the rules are based upon specific processes, rather than on generic descriptors of the system such as slope angle and soil thickness.

A note on terminology

Because standard terminology varies somewhat between the geomorphology, geochemistry and soils communities, it is useful to start by defining several key terms. We use the word ‘bedrock’
to mean fresh, unweathered bedrock and often use the qualifier ‘fresh’ to further reduce ambiguity. This is the geochemists’ protolith. To the geomorphologist, weathered material can be subdivided into material that has been physically displaced by transport or mixing processes and material that has not been transported or disrupted by mixing processes. We refer to the mobile material, what often is called soil, as ‘mobile regolith’, honoring the recent use of that term (e.g. Lebedeva et al., 1999, 2011). Mobile regolith is ultimately delivered to the stream channel, whose elevation and transport capacity serve as boundary conditions for the hillslope. This figure is available in colour online at wileyonlinelibrary.com/journal/espl.

**Weathering as Damage**

Weathering describes everything from solute flux from watersheds (e.g. Gaillardet et al., 1999), to changes in soil mineralogy (White et al., 1996), to generation of fractures (Hall, 1999; Fletcher et al., 2006), to production of mobile regolith (e.g. Anderson and Humphrey, 1989). Here we address weathering as a suite of processes that damage rock. We focus specifically on rock damage that accrues through initiation and growth of fractures. In other words, we will focus on physical processes rather than chemical processes, although we acknowledge that physical damage can promote chemical processes (e.g. Fletcher and Brantley, 2010; Jin et al., 2011), while chemistry rather than physics leads the way in many settings. Our focus on rock damage through fracturing has parallels in the analysis of the effect of stress on rock, for instance in landsliding, rock cliffs, and faults (e.g. Amitrano and Helmstetter, 2006). Weathering processes that fracture rock include thermal stress (Hall, 1999), biotite hydration (e.g. Fletcher et al., 2006; Fletcher and Brantley, 2010), and frost cracking (Matsuoka and Murton, 2008). Growth of microcracks increases rock porosity and permeability (Sousa et al., 2005; Graham et al., 2010; Jin et al., 2011); greater fluid circulation leads to a cascade of processes that further alter and/ or damage the rock. Indeed, the full spectrum of cracks is important. Larger cracks can govern rock mass strength, and will strongly influence bulk permeability because they are likely to be more interconnected and have larger apertures than microcracks.

The damage that rock accrues on its way toward the surface can be thought of as a reduction in the rock’s ability to resist shear, and ultimately to resist disintegration. This resistance is likely related to the density of flaws in the rock parcel. In this view, the rate at which damage accumulates in a rock mass may be cast as the rate at which new crack surface area is generated per unit volume of rock \( \frac{dV}{dt} \) \( = \) \( L^2LT = 1LT \). (Note that we use \( = \) to mean ‘has dimensions of’ throughout this paper, with \( M \), \( L \) and \( T \) representing mass, length, and time, respectively.) Damage, \( D \), which is the time integral of damage rate, therefore has dimensions of \( 1/L \). Its inverse is a length scale, and may be thought of as the mean spacing between microcracks. We argue that the susceptibility of weathered bedrock into the mobile regolith ought to increase with increasing damage; this is equivalent to assuming that the smaller the pieces are, or the weaker the mass is, the more easily the weathered bedrock will become part of the mobile regolith. Note that the concept of damage as used here is similar in spirit to Kirkby’s (1977, 1985) ‘soil deficit,’ which represents the amount of parent rock remaining at a particular level within a soil profile. It is also consistent with the concept of weathering-related damage used by Hoke and Turcotte (2002) in an analysis of surface disintegration through mineral dissolution.

Damage therefore results in reduction of physical strength (ability to resist deformation under stress), which depends upon the numbers and strengths of the bonds within the material. One may then argue that (1) flaws matter, (2) the density of flaws affects rock porosity and permeability, as well as strength (Graham et al., 2010; Luque et al., 2011), (3) strength is modified by a rock’s tectonic history (Molnar et al., 2007; Koons et al., 2012), which can produce flaws, (4) strength (and damage) can be quantified with tensile or shear strength (e.g. Aydin and Basu, 2006), as measured by tools such as splitters and shear vanes, (5) the huge range in strength among geopolymers, and notably between fresh bedrock and saprolite, raises a practical challenge in quantifying strength and damage across this range, and (6) strength may be related to mineralogy, cement, bulk density (in the sense that it reflects grain proximity and bond strength), degree of chemical alteration, connectedness of the fracture network, and so on. We note that the properties of the
rock mass that matter to entrainment in mobile regolith will likely
differ from the strength that matters to the disintegration or failure of a cliff (e.g. Moore et al., 2009).

In the remainder of the paper we will describe a generic view of the damage zone, which we follow by working a specific example in which damage occurs by frost cracking of the rock. Armed with this model of damage, we link damage, entrainment, and transport processes to produce a model of hillslope evolution relevant to landscapes in which frost-related processes dominate.

Reactor on a Hillslope

We first describe the critical zone in its hillslope context, focusing on the concept of a reactor on a hillslope that is emerging as an appropriate analog. Consider the path taken by a parcel of rock as the landscape surface erodes down toward it (Figure 1). From the point of view of the rock, it moves toward the surface at a rate governed by the rate of exhumation (lowering) of the landscape. In effect, the rock advects through the weathering zone en route to the base of the mobile regolith. As the rock advects vertically through the weathering zone, it will begin to be altered by surface-related critical zone processes. These processes are physical, chemical, and biological, and the relative roles of these processes in altering the rock mass differ by climate, rock type, and the density of flaws in the rock that reflects a legacy of its tectonic past. The rock accumulates damage as it passes through the critical zone. If the dominant changes are chemical, the analogy with a ‘reactor’ is direct. If they are dominantly physical, perhaps the better term would be ‘mill’ or ‘grinder’. However, because the comminution of the material by any physical processes will in turn alter its hydrologic properties, with attendant feedbacks on the rates of chemical alteration, the term reactor perhaps remains appropriate. Upon emerging at the top of the weathered bedrock, the rock parcel is available to be released into the mobile regolith. The probability of release may depend on the degree of damage. For example, release into mobile regolith may require the parcel to be smaller than some critical size, or to possess a sufficient density of cracks to allow plant roots to pry it from the subjacent rock. In any case, the state of the rock parcel upon entrainment on the mobile regolith conveyor belt governs the initial mineralogical and grain size distribution of the mobile regolith. Subsequent motion of the parcel down the hillslope is governed by the processes that both translate and stir the mobile regolith. Here the chemical and biological reactor intensifies as hydrologic, thermal, and biological activity are all amplified in the near surface. During translation, other parcels of rock are added to the base of the mobile regolith, their own character upon release being dependent upon the particular damage profile through which they were advected within the base of the critical zone. We expect the mineralogy and grain size distribution of the mobile regolith to evolve as it translates downslope. The rate of motion of the mobile regolith conveyor on which these reactions take place is governed by the local slope, and by climatically modulated chemical, physical and biological processes.

The soil ultimately delivered to the adjacent stream therefore reflects the evolution of rock as it moved through the intact rock portion of the critical zone – the weathering zone - and further digestion during downhill transport on the mobile regolith conveyor. The timescale over which this evolutionary play takes place is thousands to hundreds of thousands of years (Yoo et al., 2011). However, the chemistry of the water delivered to the stream reflects the operation of present-day hydrologic, chemical and biological processes integrated over timescales of minutes to years. The rocks and minerals through which the water passes, and the reactions that these allow, will of course condition the chemistry, allowing examination of the rates and types of weathering reactions taking place.

The timescales over which the rock parcels transit this critical zone system are long. This requires that we acknowledge deep time in studies of the critical zone. One timescale may be derived from steady state landscape analysis, in which the horizontal length scale of the hillslope, $L$, and the effective diffusivity of the landscape, $\kappa$, combine to yield a hillslope timescale $T = L^2/\kappa$ (e.g. Koons, 1989). Given hillslope lengths of order hundreds of meters, and landscape diffusivities of order $10^{-2}$ m$^2$/yr (e.g. Fernandes and Dietrich, 1997), this suggests that the time to develop steady hillslope forms is on the order of a million years. Another timescale that is more relevant to the critical zone that forms the carapace of such hillslopes can be constructed from the residence time of mobile regolith on a hillslope. This is scaled by $T = H/w$, where $H$ is the thickness of mobile regolith, and $w$ the rate of mobile regolith production (rate of lowering of the mobile regolith–weathered bedrock interface). Given mobile regolith thickness of order 1 m, and mobile regolith production rates constrained by concentrations of cosmogenic radionuclides of order $10^{-2}$–$10^{-3}$ m/yr, we expect residence times of order 10–100 thousand years. If the thickness of the zone of chemical and/or physical weathering in the rock beneath the soil is several times the thickness of the soil, the residence time within the weathering zone will be commensurately longer. It is therefore clear from even this simple scaling that the critical zone evolves on timescales that are similar to those of the major swings in climate that characterize the Plio-Pleistocene glacial ages, with accompanying major changes in physical, chemical and biological processes and process rates. It is therefore inevitable that the architecture of the critical zone, and the mineralogical, grain size, and hydrologic properties of the materials presently residing in the critical zone, reflect a legacy of past climates, reaching back at least one glaciation. The present is therefore not a proper or complete key to the past. Any study of the critical zone must honor this legacy.

Damage Profile – A Generic Case

Consider the changes in a parcel of initially unweathered bedrock as it is brought toward the weathered bedrock–mobile regolith interface by exhumation (Figure 1). We expect that many processes responsible for altering the rock mass strength will decline in intensity with distance below the surface. For now, suppose for simplicity that damage rate, $D$, falls off exponentially with depth and is independent of the damage state of the rock mass:

$$D = D_0 e^{-z/D}$$

where $z_{damage}$ is a length scale characterizing the rate of decay of damage rate with depth. This would take the place of the length scale $L$ in our discussion of timescales earlier. In later sections of the paper, we acknowledge specific processes that exert depth profiles of damage rate that are not necessarily exponential.

The rock parcel moves through this damage zone at a rate governed by the steady landscape lowering rate, $\dot{z}$. The accumulated damage as the bedrock arrives at the base of the mobile regolith is therefore simply the integral of the damage rate, and is

$$D = \int_0^{z_{damage}} D_0 e^{-z_{damage}/\dot{z}} dt = \int_0^{z_{damage}} \frac{D_0 z_{damage}}{\dot{z}} + D_0$$

Note that the rock will arrive in the near-surface zone with some initial flaw density and hence initial state of damage, $D_0$, setting
the ‘intact rock strength’ of the fresh bedrock. In general, this will vary with rock type and with the tectonic history to which the rock has been subjected prior to interaction with the surface (Molnar et al., 2007; Clarke and Burbank, 2010, 2011), and therefore will vary spatially. The additional accumulated damage upon reaching the base of the mobile regolith (at \( t = 0 \)) depends directly on the damage process rate at the mobile regolith–weathered bedrock interface, \( D_{\text{crit}} \), and the characteristic length scale over which it declines with depth, \( z_{\text{damage}} \), and depends inversely on the lowering rate of the mobile regolith–weathered bedrock interface, \( \dot{e} \). This formalism is perfectly analogous to that used to constrain local erosion rates from cosmogenic radionuclide concentrations in rock at the base of the regolith (e.g. Small et al., 1999). Just as more rapid erosion yields lower beryllium-10 \(^{10}\text{Be}\) concentrations in the rock that emerges at the base of mobile regolith, the more rapidly the rock moves through the damage zone, the less damage it accumulates. The framework is also similar in spirit to the model of granite weathering by albite dissolution explored by Lebedeva et al. (2010).

Accumulated damage could reflect more than one factor. For example, if pyroxene grains in a gabbrro dissolve, both the bulk tensile strength and the bulk density would decrease. They might do so at different rates and with different length scales, \( z_{\text{damage}} \). In another case, root stresses may weaken intergranular bonds without producing much change in bulk density. Thus, one might expect variations in the damage rate profiles with landscape position.

We now acknowledge the potential role of a limiting damage, \( D_{\text{crit}} \), beyond which it becomes much more difficult to generate additional damage. For example, \( D_{\text{crit}} \) might represent a state of complete transformation of a reactive mineral species, such as albite to kaolinite (Lebedeva et al., 2010), or a case in which all or most intergranular bonds have been broken by physical stresses from root growth, ice lens formation, or similar processes. The rate of damage then becomes

\[
D = D_{\text{crit}} \left( 1 - e^{-z/\chi_{\text{damage}}} \right) \tag{3}
\]

Here we have assumed the simplest linear dependence on damage, although one could consider more complex dependence if the physics warrants it. [Note that a formulation analogous to this has been used to model the rate of damage to vegetation by erosion (Collins et al., 2004; Tucker et al., 2006).]

Integrating Equation 3 from great depths to the surface yields the state of damage in the rock at the base of mobile regolith

\[
D = D_{\text{crit}} \left( 1 - e^{-z/\chi_{\text{damage}}} \right) \tag{4}
\]

Exhumation rates high enough to assure that the ratio in the argument of the exponential is much less than unity, in other words, \( \dot{e} > \frac{\dot{z}_{\text{damage}}}{\chi_{\text{damage}}} \), result in the dependence captured in Equation 2 (because in this case \( D/D_{\text{crit}} < 1 \) in Equation 3). Lower exhumation rates [or thicker zone (\( z_{\text{damage}} \)] or greater maximum damage rate \( D_{\text{crit}} \) allow sufficient damage to accumulate that the limiting damage \( D_{\text{crit}} \) begins to play a role. Numerical integration of Equation 3 for a number of exhumation rates results in steady state damage profiles (Figure 2) that illustrate this dependence. Under high exhumation rates, the time spent in the damage zone is insufficient to achieve the limiting damage \( D_{\text{crit}} \). When exhumation rates are low, the limiting damage is achieved; further lowering of the exhumation rate results in profiles that differ only in the depth at which this limiting damage is attained.

Consider now the top and bottom of the damage profile. One may ask at what depth one would encounter ‘fresh’ rock. Operationally, one might assign a particular threshold damage below which the rock is considered intact, or fresh, and we would call that the base of the weathered bedrock. One might document this depth in rock cores. Less expensively, the depth to intact rock may be remotely sensed using shallow seismic methods (e.g. Clarke and Burbank, 2011) in which velocity gradients may be used to estimate fracture density profiles. If however profiles exist in the form of cores or pits, one may document the depth scale, \( z_{\text{damage}} \), for decay of some strength measure of damage, and define fresh rock as some multiple of \( z_{\text{damage}} \) at which depth the damage falls below some level.

At the top end of the damage profile, the weathered bedrock–mobile regolith interface, a potential coupling exists between the weathering of rock in the subsurface and the likelihood that a parcel of rock will be detached from its neighbors to become entrained in the mobile regolith. This might occur, for example, if the detachment were accomplished by the burrowing of rodents, the growth of tree roots, or growth of ice lenses, and these processes could only loft rocks below a given size. This critical size will depend upon the available entrainment mechanisms, and will therefore be site-specific and dependent upon the climate.

The damage profiles shown in Figure 2 are analogous in many ways to the one-dimensional (1D) profiles developed by geochemists to capture the essence of weathering. As recently summarized by Brantley and Lebedeva (2011), chemical depletion profiles display dependence on both the rate of lowering of the land surface (erosion rate) and the depth-dependence of the process involved in altering the rock and the minerals within it. The damage-limited case could also be analogous to the geochemist’s kinetically limited case, in which the shape of the weathering front is governed by the reaction constants in the mineral transformations.
Similarly, recent models of corestone development in weathering profiles (Fletcher and Brantley, 2010) acknowledge the roles of initial flaw densities, and the rate of propagation of reaction fronts into the blocks between these fractures as setting the sizes of the corestones as they encounter the earth’s surface. They differentiate between incompletely developed profiles, in which corestones remain in the profile, and completely developed profiles in which no remnants of the original blocks remain in the top of the profile.

While the approach outlined in this section is simple, it is impoverished in that it fails to address any particular process. Next we develop models based on a specific process, namely frost cracking, by which rock can be damaged during its passage through the weathered bedrock zone toward the surface. We argue that frost cracking is relevant to the evolution of many regions, in that while modern temperatures may not reach deeply into frozen conditions, peri-glacial conditions were widespread in the last glacial maximum (LGM) when mean annual temperatures were depressed by at least 6 °C and in places by 12 °C.

Damage by Frost Cracking

Theory of frost cracking

We now consider the case in which the damage imparted to the rock mass as it passes through the weathering zone is due solely to frost cracking (e.g. Murton et al., 2006). Ice lens growth in pre-existing flaws in rock can expand openings and extend fractures, leading to shattering of rock (Matsuoka, 1990). Studies of freezing soils (e.g. Taber, 1930; Williams, 1967) have shown that the transformation of water into ice in a porous medium proceeds somewhat differently from the phase transformation in bulk water. In porous material such as soil or rock, ice forms over a range of temperatures, so that unfrozen water and ice are both present at sub-zero temperatures (Dash et al., 2006). Ice growth begins in the largest pores, and continues in progressively smaller pores as temperatures decline. Unfrozen water migrates in thin films through partially frozen porous media to feed the growth of discrete ice lenses in the larger voids of soil or rock at sub-zero temperatures. Frost cracking is important at temperatures just below freezing, a temperature range called the ‘frost cracking window’; although it varies with rock properties and freezing rate, we take this range to be −3 to −8 °C (Walder and Hallet, 1985). Experiments confirm that cracking continues in rock within the frost-cracking window as long as liquid water is available to move to the freezing front (Walder and Hallet 1985; Walder and Hallet, 1986; Hallet and Hallet, 1991; see also Murton, 1996; Murton et al., 2006; Matsuoka and Murton, 2008). Finally, we follow Hales and Roering (2007) in supposing that the rate of frost cracking by growth of segregation ice is proportional to the temperature gradient (see also Worster and Wettlaufer, 1999).

We develop a model of the depth profile of frost cracking in several steps. We begin with a consideration of temperature variations in the subsurface, as the duration of sub-freezing temperatures and the rate of freezing are important parameters in frost cracking. We then present several models of increasing complexity that quantify the effect of thermal state on frost cracking.

Temperature variations in the critical zone

Measured temperature time series such as those illustrated in Figure 3 reveal several features that we must capture in any proper model of the system. Temperature varies on both annual and daily cycles, with comparable amplitudes. Snowfall dramatically damps the amplitude of surface temperature oscillation as long as snow cover exists. And the freezing and thawing of moisture in the ground releases and consumes latent heat that must be accounted for in any calculation that conserves heat.

In Figure 3, the shallow ground temperatures are depicted from the Gordon Gulch site of the Boulder Creek Critical Zone Observatory (CZO), Colorado, USA, from four depths. When snow covers the ground, the surface temperature approaches 0 °C, and is held there while the snow persists. Temperatures at all other depths more slowly approach 0 °C as this serves as the new boundary condition.

Inspired by such temperature records, and in the hope of generating a thermal model of the near subsurface that permits exploration of the dependence of surface and subsurface processes on temperature, we take the temperature of the surface to be sinusoidal at both daily and annual timescales. We prescribe the mean annual temperature, and amplitudes of both daily and annual variations around the mean to yield

\[
T_s = \bar{T} + \Delta T_a \sin \left( \frac{2\pi t}{365} \right) + \Delta T_d \sin \left( \frac{2\pi t}{P} \right) \quad (5)
\]

where \( \Delta T_a \) and \( \Delta T_d \) are the amplitudes of annual and daily temperature oscillations, respectively, and \( t \) is time in days.

The temperatures at all depths and times are then calculated by conserving heat, assuming that conduction dominates (no heat transport by moving water), and acknowledging the role of latent heat at the phase boundary. In one dimension, the rate of change of temperature \( T \) with time \( t \) reflects conservation of heat:

\[
\frac{\partial T}{\partial t} = - \frac{1}{\rho c z} \frac{\partial Q_h}{\partial z} \quad (6)
\]

where \( z \) is the depth into the ground, \( \rho \) is local material density, \( c \) is its heat capacity, and \( Q_h \) the vertical heat flux. Ignoring the role of motion of water as a heat transport mechanism, heat transport is governed by conduction through Fourier’s Law:

\[
Q_h = -k \frac{\partial T}{\partial z} \quad (7)
\]

where \( k \) is the local (vertical) conductivity of the substrate. We allow thermal properties to vary with water and ice content. We account for the role of phase change of water using a strategy employed by Ling and Zhang (2003), and mimicked by West and Plug (2008) (see also Matell et al., 2011). The thermal
conductivity and specific heat capacity are updated in each time step according to whether the material is thawed ($T > 0 \, ^\circ C$) or frozen. For temperatures within a ‘phase change envelope’ (defined here as the 1 °C envelope between −1 °C and 0 °C), the thermal constants are calculated differently (Ling and Zhang, 2003; see also West and Plug, 2008) to incorporate the effects of latent heat, as well as to acknowledge the fact that the freezing of water in soil and rock does not take place entirely at the freezing temperature. Thermal conductivity and volumetric specific heat capacity are calculated as

$$k = \begin{cases} k_i & T < T_i - DT \\ k_i + \frac{k_u - k_i}{DT} [T - (T_i - DT)] & T_i - DT < T < T_i \\ k_u & T > T_i \end{cases} \quad (8)$$

and

$$C = \begin{cases} C_i & T < T_i - DT \\ C_i + L_p b \left( W - W_i \right) & T_i - DT < T < T_i \\ C_u & T > T_i \end{cases} \quad (9)$$

where $T_i$ is the freezing temperature (in kelvin), $DT$ the width of the phase change envelope (in kelvin), $T_f$ the latent heat of freezing for water (in J/kg), $W$ the fractional total water content of the soil by mass, $W_i$ the unfrozen water content that remains at $T_i$ (here taken to be 5% of $W$), and the subscripts ‘f’ and ‘u’ refer to the frozen and unfrozen values of the thermal conductivity and specific heat capacity, both of which are calculated as inputs to the model based on soil properties and water/ice content (Lunardini, 1981). Thermal conductivities are calculated separately for frozen and unfrozen soils as geometrically weighted values

$$k = k_i^{\nu_i} k_u^{\nu_u} k_v^{\nu_v} \quad (10)$$

where $k_i$, $k_u$, and $k_v$ are the thermal conductivities of ice, water, and mineral or rock, and where $\nu_i$, $\nu_u$, and $\nu_v$ are the fractions of ice, water, and mineral in the substrate (Ling and Zhang, 2004). Specific heat capacity $c$ (in J/kg/K) is calculated as

$$c = c_i \nu_i + c_u \nu_u + c_v \nu_v \quad (11)$$

where $c_i$, $c_u$, and $c_v$ are the specific heat capacities of ice, water and mineral (Ling and Zhang, 2004). While $\nu_v$ does not vary in time, the fractions of ice and water evolve as the substrate freezes and thaws. Volumetric specific heat capacity $C$ (in J/m$^3$K) is converted into specific heat capacity $c$ by dividing by the density of the ice/water/soil combination.

As shown from the data depicted in Figure 3, the presence of snow can greatly influence the subsurface temperature histories (e.g. Sturm et al., 2001; Bartlett et al., 2004). We do not perform a full energy balance model to compute the temperatures within the snowpack. Instead, we employ the surface air temperature to melt the snow surface when air temperature rises above 0 °C, and redeposit the latent heat associated with that melt within the snowpack. If air temperature is below 0 °C, heat is transported purely by conduction within the snowpack with a thermal conductivity appropriate for snow. We emphasize that this is not a full snowpack thermal model, but wish through this exercise to illustrate the essence of the role of snow in altering subsurface temperature histories.

In the example calculation shown in Figure 4, the top 40 cm is taken to be mobile regolith with 40% porosity, whereas the underlying bedrock is taken to have 3% porosity. Both are assumed to be 90% saturated. While water content of rock is very difficult to measure directly, recent work in alpine rock masses has shown that water content in rock is quite high and relatively steady at depth (Sass, 2005; Graham et al., 2010; Langston et al., 2011). At shallow depths, the assumption of 90% saturation certainly overestimates the role of water in stalling the penetration of heat into the subsurface. We employ this value as a means of highlighting the effects here. To illustrate the role of snow cover, we have prescribed a 40 cm snowfall event that occurs at night on day 60.

That the snow persists for two months allows this case to serve as an illustration of the expected effects of snow cover. We assumed that snow cover of 40 cm fell at midnight, when the surface temperature was at its lowest. Given the high thickness of the snowpack, the amplitudes of the temperature variations at all depths are subsequently highly damped. When the daily oscillations of air temperatures begin to go above 0 °C, on day 80, melting of the snow begins to occur. Re-deposition of that heat in the snowpack allows it to warm, which is reflected in the rise of ground surface temperatures (Figure 4, red line). The snowpack becomes isothermal (at 0 °C; not shown) on day 90, after which the temperatures at all depths asymptotically approach 0 °C; the problem has now become an instantaneous warming problem. The snowpack then remains isothermal until the snowpack disappears on day 120. At each depth the temperature stays within the −1 °C to 0 °C phase change envelope, DT, until sufficient heat to melt all ice in the pores has been supplied. The stalling of subsurface temperatures during phase change, called the ‘zero curtain’ effect, reflects the latent heat required

![Figure 4](image-url)
to thaw the ice. The zero curtain persists for roughly 10, 20 and 30 days for depths of 20, 30 and 40 cm, respectively. The zero curtain occurs again when the ground freezes in the fall, such that near-surface temperatures drop well below freezing while the temperatures at depth are stalled by the release of latent heat in the freezing process. The duration of this effect is dependent upon the total water content of the soil and subjacent rock. Our example, using 90% saturation maximizes the duration of the zero curtain.

Another way to view temperature histories is shown in Figure 5, where we depict temperature profiles at many different times within an annual cycle. Note the asymmetry of the envelope within which temperature varies. In the absence of latent heat effects, this envelope would be a symmetrical exponential funnel with length scale $\sqrt{\kappa P/\pi}$ (e.g. Gold and Lachenbruch, 1973), where $\kappa$ is the thermal diffusivity ($=k/C_p$), and $P$ is the period of the temperature oscillation. For a typical diffusivity $k$ of $10^{-6} \text{ m}^2/\text{s}$, the length scale for daily fluctuations would be 0.17 m, while that for the annual cycle would be 3.2 m. In the case illustrated, with mean annual temperature of $-2^\circ \text{C}$ and daily and annual oscillations of 6 $^\circ \text{C}$, temperatures never rise above $0^\circ \text{C}$ at depths below about 2 m. The area above this depth, where temperatures do rise above $0^\circ \text{C}$, is known as the active layer. Rock spends some amount of time in the cracking window for all depths down to 5 m. But given that the easily available water in the larger pores is permanently frozen below the active layer, access to water is a problem for these greater depths. As water must come from the surface in permafrost situations, it becomes less and less likely that cracking occurs for greater depths. Inspection of specific temperature profiles suggests that the primary times at which water is readily available to drive ice growth and hence cracking are in the spring, when daily surface temperature rises above $0^\circ \text{C}$.

**Frost cracking models**

We present several models of frost damage (Figure 6) with increasing levels of complexity and realism. For the first exercise, meant to allow comparison with past models, we omit daily temperature oscillations, we employ a uniformly low porosity material (with no distinction between mobile regolith and bedrock), which minimizes the effects of latent heat discussed earlier, and we ignore effects of snowpack. These simplifications allow us to compare with the models of Anderson (1998) and of Hales and Roering (2007), and to focus on the first order controls on frost cracking.

In the simplest model, put forth by Anderson (1998), the time rock spends in the frost cracking window (middle panels of Figure 6) serves as a proxy for frost damage. Hales and Roering (2005, 2007) improved this model by acknowledging that this simple proxy does not capture the role of water as a limiting factor in frost lens growth in rock. In their model, two conditions must be met for frost cracking to occur: the rock is in the frost-cracking window, and a source of liquid water ($T>0^\circ \text{C}$) is available in a pathway along which temperatures monotonically increase (unfrozen water at the surface and downward decreasing temperatures; or water at depth and upward decreasing temperatures). Reasoning that cracking intensity is proportional to temperature gradient (as is growth rate of segregation ice; Worster and Wettlauer, 1999), Hales and Roering calculate crack intensity as the product of time in the frost cracking window and temperature gradient, which should yield a measure with units of temperature-time/length. The calculation is carried out for each time step, and summed for one year. Hales and Roering (2007) report crack intensity in units of $\text{C/cm}$, however, leaving the time increment implicit. We can reproduce their results by recalling that they employ a one day time step and noting a factor of 100 difference that arises from temperature gradients in $\text{C/m}$ rather than $\text{C/cm}$. We track cracking intensity in units of $\text{C-days/m}$, although on several of the plots we have normalized against the maximum value to allow comparison of the patterns between models that generate very different magnitudes of cracking intensity.

Comparison of these models is illustrated in Figure 6. We show two cases, one in which the mean annual temperature is well above $0^\circ \text{C}$, and one in which it is well below $0^\circ \text{C}$. On the left we show temperature profiles at many times of the year, as in Figure 4; in these cases, however, the profiles define a symmetric exponential funnel, because latent heat effects are ignored. In the middle panels we show the expected time spent in the cracking window [here taken to be $(-8^\circ \text{C} \text{ to } -3^\circ \text{C})$]. Following Anderson (1998), these patterns are well predicted with an analytic solution (dashed black line) as long as the time series of surface temperatures is sinusoidal, and the material properties governing thermal diffusivity are uniform with depth and constant in time. The numerical tracking of time spent in the frost-cracking window (solid red line) reproduces well the expected analytic solution in both cases.

In the right panels, we show in gray the temporal sum of thermal gradient [the proxy for cracking intensity proposed by Hales and Roering (2007)], here normalized against its maximum value. We show in the left panel with heavy gray lines the segments of the instantaneous temperature profiles with the appropriate combination of temperature gradient and presence of unfrozen water elsewhere in the profile such that frost cracking can occur. In Figure 6a, with $T_s = 3^\circ \text{C}$, the predicted pattern of annual cracking intensity is clearly governed by time in the cracking window. But given that the only times when the temperatures dip into the frost cracking window correspond with times when the surface temperatures are $<0^\circ \text{C}$, the (unfrozen) water available for driving cracking must always come from depths that are well below the site of potential frost cracking. In the sub-zero mean annual temperature case (Figure 6b), there are times in summer when water is available from the surface, and times in winter when it is available at depth (two sets of heavy gray lines in the left panel, one dipping to the left, one dipping to the right).
The resulting crack intensity profile is significantly different from the simpler but less defensible proxy of time in the cracking window (shown for comparison in the middle panel, Figure 6b). We now add a further element of realism to the Hales and Roering frost cracking intensity model in the water limited or ‘penalized’ case shown in cyan on the right panels of Figure 6. Here we acknowledge that the rate of growth of cracks should be limited not only by the presence of liquid water, as in the Hales and Roering algorithm, but also by the distance that water must be moved through cold materials to get to the site of potential frost cracking. This effect is analogous to the process by which a growing ice lens in soil is abandoned in favor of a new ice lens when the water supply to it becomes limited (Rempel, 2007). This ‘penalty function’, here represented by a simple exponential of the form $e^{-l/l^*}$, where $l^*$ is a specified length scale (we take for this example $l^* = 0.5$ m), significantly alters the predicted pattern of cracking intensity. In general, this modification moves the locus of predicted cracking damage to greater depth in the non-permafrost case (Figure 6a), and to a single maximum in the permafrost case (Figure 6b).

In the case of negative $T_s$ (in Figure 6b, $-3^\circ$C), deep sites have access to unfrozen water in the spring, as the surface temperature rises (left-dipping gray profiles on the left panel), while shallow sites have access to water in the fall, as surface temperature drops (right-dipping profiles). These give rise to the two peaks in the predicted Hales and Roering cracking rate profile. However, when the system is penalized for water transport distance, the resulting pattern displays a single peak. In the fall, water must be drawn from depths of 2 to 3 m to near-surface sites, while in the spring water must be drawn from the surface to sites of increasing depth. But early in the spring (uppermost left-dipping gray lines), sites at ~1 m depth begin to have access to

---

**Figure 6.** Comparison of models for frost cracking intensity. Profiles of temperature (left), time in the frost cracking window (middle), and frost cracking intensity (right) for (a) $T_s = +3^\circ$C, and (b) $T_s = -3^\circ$C. In both cases, the annual temperature amplitude ($\Delta T_a$) is $12^\circ$C, and both daily oscillations and latent heat effects are ignored. Left panels: Calculated temperature profiles at weekly intervals, with frost cracking window shaded gray. Parts of profiles that lie in the frost cracking window are red; of these, cases where some portion of the profile is above $0^\circ$C are shown with heavy gray lines. Middle panels: Analytical solution (dashed line), and numerical solution (solid line) for time spent in the frost cracking window. Right panels: Frost cracking intensity as calculated from the magnitude of the thermal gradient (as suggested by Hales and Roering, 2007) and water availability in the direction of heat flow (bold gray lines). This pattern is further modified when the rate of cracking is taken to be dependent on the distance to the unfrozen water (‘penalized’ cases, blue lines). This figure is available in colour online at wileyonlinelibrary.com/journal/espl
Additional elements of reality

We now relax the assumptions that allowed us to compare directly with past models of Anderson (2002) and Hales and Roering (2005, 2007), so that we may incorporate elements of reality we consider important in the hillslope problem. In the following models, we (1) include a daily temperature cycle; (2) use a non-uniform porosity structure (and hence k and c structure) in the subsurface to represent a typical critical zone architecture of high porosity mobile regolith overlying low porosity bedrock; (3) incorporate latent heat, which stalls the freezing front; (4) acknowledge and attempt to illustrate the role of snowcover.

As in the case depicted in Figure 4, we assume that 0.4-m thick mobile regolith, with 40% porosity, overlies rock with 3% porosity, and we assume that both bedrock and mobile regolith are 90% saturated. We perform a no-snow case, and a case in which the snow is 0.5 m deep and remains on the landscape for two months. The daily surface temperature amplitude is taken to be 4 °C.

In Figure 7, we show the resulting cracking intensity profiles when these elements of reality are included, for cases both with and without the penalty for distance that water travels to drive frost cracking. The profiles for \( \Delta T_s \) of +3 and −3 °C can be compared with the right panels of Figure 6 to visualize the effects of all elements of reality listed earlier. In general the patterns are all compressed toward the surface, reflecting the stalling of the freezing front by latent heat effects (as illustrated in Figure 4) and by the high porosity of the mobile regolith. These effects are maximized by our choice of high water content. In addition, the daily oscillations in temperature, which are felt in only the top few tens of cm of the surface, allow additional cracking in these shallow depths.

The predicted profiles differ not only in the shapes, but also in their integrals, and hence in the predicted total crack damage accomplished as a rock passes through the cracking rate profile. We plot this integral metric of frost damage for a wide range of mean annual temperatures (\( \bar{T}_s \)) in Figure 8 for models with and without a daily temperature cycle, and with and without the penalty for water travel distance. Consider first the models that omit daily temperature oscillations. Total frost damage increases monotonically as \( \bar{T}_s \) drops from +1 to −8°C (the lower edge of the frost cracking window) in the penalized case, while it displays two peaks in the unpenalized case. The principal reason for the reduction of predicted cracking as \( \bar{T}_s \) rises is that during times in which temperatures are in the frost cracking window, water is available only at depths that are far below all sites of potential cracking. The penalty on frost cracking for inefficient water transfer is very high, an impact well documented in field studies (Hall et al., 2002; Sass, 2005). This is reflected in the bias toward cracking at the base of the frost cracking window in Figure 6. Also note that these profiles of damage are normalized to emphasize the difference in the spatial pattern. The non-normalized values are tremendously different owing to the penalty; this accounts for the 35-fold difference in y-axis scales in Figure 8.

Adding in the daily temperature cycles produces an even greater increase in the accumulated damage sustained by weathered bedrock when it reaches the base of the mobile regolith as the mean annual temperature declines, in the most realistic case (including daily cycles and a transport-distance penalty) (Figure 8). Not much damage is sustained for mean annual temperatures greater than 0 °C, whereas significant damage has occurred for bedrock exhumed in conditions with

---

**Figure 7.** Profiles of cracking intensity for more realistic cases for unpenalized (a) and penalized (b) cases, for a range of \( \bar{T}_s \) (mean annual temperature) at intervals of 2 °C. In both cases, we include a daily cycle of temperatures, different porosities and hence thermal properties of the mobile regolith and bedrock, and latent heat associated with phase changes. We have ignored the role of snow, and assume that the bedrock and mobile regolith are 90% saturated at all times of the year. Profiles for mean annual temperatures of +3 °C (green-blue) and −3 °C (brown) are heavy and dashed. Note the large difference in the calculated magnitudes of cracking intensity between the two cases.
The mean annual ground surface temperature is presently about −6 °C. This is much higher than conditions during glacial times. More precisely, the Gordon Gulch surface temperatures would have dropped into a cooling climate regime (e.g. Brugger, 2010), meaning that the mean annual temperatures between −2 and −12 °C. The damage peaks at \( T_s = −12 °C \) because time spent in the frost cracking window is greatest at the outer edges of the cycle; when the mean is −12 °C, the temperatures in the summer turn around within the frost-cracking window, and therefore spend significant time in the window. The expected total damage again drops off with further decline in \( T_s \), as the times when the bedrock is both in the frost-cracking window and has access to unfrozen water declines. At \( T_s < 0 °C \), permafrost develops, and the hydrologic depth scale shortens to the active layer; the permafrost table becomes an aquitard. This will alter hydrologic flow paths to become essentially surface-parallel. Deep flow paths that dominate many alpine fracture-flow systems will be eliminated, with consequences for chemical weathering and for solute chemistry of the rivers. Decreased near-surface permeability on frozen ground may also lead to increased surface runoff, with potential impacts on sediment transport and landscape evolution (e.g. Bogaart et al., 2003).

In addition, the hillslope conveyor system into which weathered bedrock is entrained upon reaching the mobile regolith interface may be changed significantly in a cooler regime. One might imagine that the active layer may more efficiently pass mobile regolith downhill by solifluction and frost creep (e.g. Matsuoka and Murton, 2008). The sign and magnitude of the change will likely also reflect any change in vegetative cloak on the landscape. Loss of trees may for example release mobile regolith to be more efficiently transported due to frost processes. However, in some cases loss of trees may result in lower transport rates if the transport is governed by tree throw and root-related transport (e.g. Hughes et al., 2009). In the following section we attempt to link damage processes with transport processes in order to explore their co-dependence in a model setting.

**Numerical Hillslope Model**

We now explore the emergent behavior of an interfluvial in the face of climate-controlled variations in both the damage and
transport processes. In Gordon Gulch of the Boulder Creek CZO, opposing slopes that face dominantly north (N) and south (S) receive very different radiation and snow cover, and are cloaked with contrasting vegetation. We observe a broad asymmetry in weathering on these opposing slopes. The depth to fresh rock, as measured with shallow seismic refraction, is greater on N-facing slopes than on S-facing slopes (Befus et al., 2011). Thin mobile regolith of order 0.5 m depth is found throughout the catchment.

Our modeling strategy is to link the processes responsible for damaging bedrock, and for producing and transporting mobile regolith, with the meteorological characteristics that govern these processes. In the present model we simplify the process set to one in which damage in the subsurface, production of mobile regolith, and transport of mobile regolith are all driven by frost-related processes. The required meteorological forcing therefore reduces to the thermal state and moisture balance of the hillslope, which will importantly vary in both time (climate change) and space (radiative setting and hence slope aspect). We acknowledge that other processes we have discussed earlier (such as the roles of trees in all three of these hillslope processes) will also vary with climate and with aspect. It remains a target for future research to abstract the roles of trees, and the linkage between the vegetative landscape and the frost-related processes, so that they can be incorporated in critical zone evolution models. We describe first the climate forcing, followed by how this is incorporated in the hillslope processes rule set.

Climate

We impose an oscillation in climate with 40 ka period. As the LGM glacier lengths in the headwaters of the Boulder Creek watershed may best be modeled by imposing a 6°C drop in $\bar{T}_s$ from present conditions (Dühnforth and Anderson, 2011), we impose a 40 ka oscillation in mean annual surface temperature at all locations with an amplitude of $3^\circ C$ ($6^\circ C$ full range). Dependence on aspect is incorporated by aligning our 1D hillslope model to be N–S oriented, such that left is S and right is N. We have not incorporated an energy balance model to evaluate the surface temperatures as a function of position on the slope transect, although this is an excellent target for future studies. Such a model would need to incorporate the roles of trees of different species and density, and how they affect the snow cover, as well as the more straight-forward radiative problem. In the absence of a full energy balance model for our landscape, we simply assert a slope dependence of $\bar{T}_s$; that is reasonable. This is accomplished for northern hemisphere settings by allowing the $\bar{T}_s$ to depend linearly on hillslope gradient in the model, which is positive on S-facing slopes (left side of the simulations) and negative on N-facing slopes. The spatial and temporal distribution of mean annual temperature is therefore taken to be

$$\bar{T}_s(x, t) = \bar{T}_{s0} + \Delta T \sin(2\pi t/\tau) + \frac{\partial \bar{\theta}}{\partial x}$$  \hspace{1cm} (12)

where the second term on the right hand side represents the temporal variation, with $\Delta T$ the amplitude of the oscillation, and the third term represents spatial variation. The parameter $\tau$, which has units of temperature per unit slope gradient, controls the sensitivity of mean annual temperature (MAT) on the slope of the landscape. In our current model we employ $\tau = 2^\circ C$ per unit slope gradient, which results in slope-related differences of about $2^\circ C$.

The distribution of $\bar{T}_s$ across the landscape serves to modulate both the depth and intensity of the frost cracking in the weathered bedrock, and the number and depths of the freeze–thaw events in the mobile regolith. Hence it governs both the weathering and transport components of the hillslope evolution. We assume that sufficient moisture is available in the unfrozen parts of the subsurface throughout the year.

As in many hillslope evolution models (e.g. Dietrich et al., 1995; Anderson, 2002), the basic statement of mobile regolith conservation in one dimension is

$$\frac{\partial H}{\partial t} = w - \frac{\partial Q}{\partial x}$$  \hspace{1cm} (13)

where $H$ is thickness of mobile regolith, $w$ the rate of release of weathered bedrock into the mobile regolith layer, $Q$ the volumetric discharge of mobile regolith per contour length (=) $L^3/T$, and $x$ the distance from the hillcrest. In order to proceed, we require rules for the rate of conversion of weathered bedrock to mobile regolith, $w$, and for transport of mobile regolith downslope, $Q$. We add to this list the rule discussed earlier for the progressive damage of rock within the weathering zone, and propose one possible linkage between damage and mobile regolith release or production rate.

Mobile regolith transport

In keeping with our argument that frost-related processes dominate in this landscape, the transport rule for mobile regolith is based upon earlier models of mobile regolith motion by frost heave (Anderson, 2002). As illustrated in the work of Matsuoka and Moriwaki (1992), the volumetric downslope transport by an individual frost heave event may be taken to be

$$q = -\frac{1}{2} \frac{\partial^2 \bar{\theta}}{\partial x^2} \frac{\partial \bar{\theta}}{\partial x}$$  \hspace{1cm} (14)

where $\beta$ is the slope–normal strain induced by frost and $\bar{\theta}$ is the depth of frost penetration [L]. The dimensions of $q$ are therefore $L^2$, the cross-sectional area of material displaced per event. This may be extended to include the frequency of frost events and to embrace a distribution of frost penetration depths to become

$$Q = -\frac{1}{2} \int \beta \frac{\partial \bar{\theta}}{\partial x} \left[ 1 - e^{-H/\bar{\theta}^*} \frac{1 - H/\bar{\theta}^*}{1} \right] \frac{\partial \bar{\theta}}{\partial x}$$  \hspace{1cm} (15)

for the total volumetric discharge [=] volume/LT. Here we have integrated over an exponential probability density function of frost depths characterized by a depth scale $\bar{\theta}^*$ (see Anderson, 2002). The parameter $\tau$ represents the frequency of freeze–thaw (frost heave) events [s] number/year. The strain induced by frost, $\beta$, is in turn related to the water content of the soil and the frost-susceptibility of the soil. The complexity of this equation largely reflects the fact that it accounts for the role of mobile regolith thickness, $H$; when $H = 0$, $Q = 0$, and as $H >> \bar{\theta}^*$, $Q$ approaches the limit

$$Q = -\frac{1}{2} \int \beta \frac{\partial \bar{\theta}}{\partial x} \frac{\partial z}{\partial x}$$  \hspace{1cm} (16)

Climate therefore comes into the mobile regolith discharge rule through its governance of $\tau$, $\beta$, and $\bar{\theta}^*$. High values of each increase the mobile regolith discharge on a given slope; given the square dependence on frost depth, this exerts the strongest control. The equation is of the common form $Q = - k (dz/dx)$ that lies behind the diffusive nature of hillslopes (e.g. Culling, 1965). Here we may identify the transport efficiency, $\kappa$, as
This formulation captures the dependence of transport efficiency on both climate and mobile regolith thickness for this particular frost-driven surface transport process. In Figure 9 we show the results of thermal models in which we have tracked the number of freeze–thaw events in an annual cycle, $t$, and their depths, $z$, so that the total mobile regolith discharge may be assessed as a function of the mean annual temperature. Here rather than impose a probability density function for freeze depths, we simply tally those that occur in the model, and sum over all events, so that the efficiency becomes

$$\kappa = \frac{1}{2} e^{-h(z)} \left[ 1 - e^{-H(z)} \left( 1 + \frac{H(z)}{z} \right) \right]$$  

(17)

Here, $a$ indexes depth into the mobile regolith, and there are $N$ such depth intervals. A frost event is said to have occurred if the temperature at that depth passes both downward and upward through $0\,^\circ C$ (i.e. we count the zero-crossings and divide by two). In the frost-creep algorithm, therefore, no account is made for how much time is spent in these excursions into subfreezing temperatures; in this sense it differs from the algorithm used for frost damage of rock, in which time spent in the frost cracking window matters. This acknowledges the fact that frost heave associated with ice lens growth in soil can occur much more rapidly than frost cracking, which entails transport of water through much less permeable material. In the models depicted, the mobile regolith thickness is taken to be 0.4 m, and the amplitudes of the annual and daily oscillations are taken to be 8°C and 3°C, respectively, in accord with temperature records in Gordon Gulch. Two sets of models were run, one with no snow, and the other with an imposed snowfall of 0 cm. In the sprit of the present emphasis on thermal processes, we capture the fact that it should depend upon the state of the weathered bedrock at the mobile regolith–weathered bedrock interface (its susceptibility to entrainment), and the processes available to entrain it. In this sense the problem becomes similar to sediment transport mechanics, in that the theory builds on a quantification of resisting and driving stresses. As discussed earlier, the susceptibility to entrainment is quantified by its initial state (e.g. how cracked the original rock is, or how bedded it is) and the damage it has sustained en route to the surface.

We take the mobile regolith production rate to be the product of two terms, one capturing the dependence on distance from the surface to the mobile regolith interface (i.e. the soil depth), the other representing the damage-dependent susceptibility of bedrock at that interface. We propose that the susceptibility to entrainment depends upon the damage in a fuzzy thresholded way (Figure 10). We employ a hyperbolic tangent to switch smoothly from very low probability to very high probability of entrainment over a specified range:

$$P(w) = \frac{1}{2} \left[ 1 + \tanh\left( \frac{D - D_s}{\sigma_D} \right) \right]$$  

(19)

where $D_s$ is the damage scale at which the probability of entrainment is 50%, and $\sigma_D$ is a damage scale over which the entrainment probability changes significantly. The processes responsible for entrainment are taken to be dependent upon depth of mobile regolith, which can be conceived as a proxy for either root dislodgement or frost-lofting of rock. In the sprit of the present emphasis on thermal processes, we choose an exponential function to capture the essence of the decay of the number of frost events with depth. Hence, the mobile regolith production function takes the form:

### Mobile regolith entrainment

Here we craft a rule for mobile regolith production in which we capture the fact that it should depend upon the state of the weathered bedrock at the mobile regolith–weathered bedrock interface (its susceptibility to entrainment), and the processes available to entrain it. In this sense the problem becomes similar to sediment transport mechanics, in that the theory builds on a quantification of resisting and driving stresses. As discussed earlier, the susceptibility to entrainment is quantified by its initial state (e.g. how cracked the original rock is, or how bedded it is) and the damage it has sustained en route to the surface.

We take the mobile regolith production rate to be the product of two terms, one capturing the dependence on distance from the surface to the mobile regolith interface (i.e. the soil depth), the other representing the damage-dependent susceptibility of bedrock at that interface. We propose that the susceptibility to entrainment depends upon the damage in a fuzzy thresholded way (Figure 10). We employ a hyperbolic tangent to switch smoothly from very low probability to very high probability of entrainment over a specified range:

$$P(w) = \frac{1}{2} \left[ 1 + \tanh\left( \frac{D - D_s}{\sigma_D} \right) \right]$$  

(19)

where $D_s$ is the damage scale at which the probability of entrainment is 50%, and $\sigma_D$ is a damage scale over which the entrainment probability changes significantly.

The processes responsible for entrainment are taken to be dependent upon depth of mobile regolith, which can be conceived as a proxy for either root dislodgement or frost-lofting of rock. In the sprit of the present emphasis on thermal processes, we choose an exponential function to capture the essence of the decay of the number of frost events with depth. Hence, the mobile regolith production function takes the form:
Figure 10. Rule for susceptibility to entrainment of weathered bedrock into mobile regolith. The more damaged the parcel of bedrock, the more likely it is to be entrained in the overlying mobile regolith. The functional dependence is here captured using a damage scale \( D \) and a sensitivity scale \( \sigma_D \).

\[
\psi = (w_0e^{-N/H_0}) \frac{1}{2} \left[ 1 + \tanh \left( \frac{D - D_0}{\sigma_D} \right) \right]
\]  

(20)

Here the first factor represents the dependence of the efficiency of the entrainment process on mobile regolith thickness, and the second factor represents the efficiency of the weathered bedrock immediately below the mobile regolith to entrainment (Equation 18). Note that for damage \( D \gg D^* \), we have chosen this form such that it reduces to the oft-used exponential dependence of mobile regolith production rate on mobile regolith thickness (e.g. Heimsath et al., 1997). We could also entertain other forcing functions in which the dependence on mobile regolith thickness takes on a humped rather than exponential form (e.g. Riggin et al., 2011; see also discussion in Humphreys and Wilkinson, 2007, and an example from tree-related processes in Gabet and Mudd, 2010).

Initial and boundary conditions

We initiate all models with a triangular hillcrest with small roughness (0.01 m amplitude random perturbations on grid with 1 m spacing) and uniform mobile regolith thickness of 0.1 m. The bounding channels are lowered at a prescribed steady rate, \( e_{\text{fs}} \). The reported steady state results are not sensitive to the choice of initial conditions, although the transient route to steady state form is. The rock mass may be assigned any initial state of damage, \( D_0 \), reflecting the initial concentration of flaws in the rock that in turn reflect both the rock type (the density of bedding planes, for example) and the tectonic history that has produced additional flaws prior to interaction with surface processes (e.g. Molnar et al., 2007; Clarke and Burbank, 2011; Koons et al., 2012). In the present model we assume zero initial damage in the rock mass.

Results

In Figure 11 we show an example of the hillslope model after 1 Ma model time. The straight-sloped initial conditions have been effectively ‘forgotten’, and the hillslope form has evolved toward a convex shape with nearly uniform mobile regolith thickness. This is in accord with many earlier models (e.g. those reported in Anderson, 2002). We focus here on the more novel aspects of the model results.

Figure 11. Example result of hillslope evolution model. Simulated critical zone architecture after 1 Ma, with initial conditions of a triangular symmetrical interfluve centered at \( x = 0 \). Resulting morphology is asymmetric, showing divide migration toward the warmer S-facing slope. More efficient transport and damage on colder N-facing slope requires lower slopes to accomplish mobile regolith transport than on warmer less efficient S-facing slopes. Mobile regolith thickness (pink layer) is nearly uniform, although it varies in time through imposed 100 ka climate cycles (see Figure 12). Degree of damage in weathered rock depicted with color shading. Damage occurs to greater depths on the N-facing slopes, as depicted by the depth to weathered bedrock with 10% of the damage at the mobile regolith–weathered bedrock interface (bottom plot).

Lower mean annual temperatures on N-facing slopes (by a maximum of \( -2 \) °C at all times; Equation 12 with \( a = 2 \) °C/unit slope) result in deeper and more intense frost damage profiles on N-facing slopes (bottom plot in Figure 11). In addition, and more importantly to the form of the hillslope, the efficiency of mobile regolith transport is much greater on the N-facing slopes. This results in an asymmetry of the ridge profile, with steeper slopes facing south. A corollary to this asymmetry in efficiency is that in order to achieve a steady form, divide migration is required: the slopes must everywhere be steeper on the S-facing slope to compensate for the inefficiency of transport there. We discuss this asymmetry further later.

The imposed 100 ka oscillation in thermal forcing (3 °C amplitude; 6 °C range) results in oscillation of mobile regolith thickness (shown in Figure 12 for one slope position). We find that the amplitudes of change are greatest nearest the bounding streams.

Given that the rule for mobile regolith production entails a damage threshold, albeit a fuzzy one (Equation 18), the damage of the weathered bedrock just below the mobile regolith becomes essentially uniform at the damage threshold value. What differs between N- and S-facing slopes is the depth to which weathering damage has proceeded (Figure 12). This length scale is greater on the N-facing slopes.

Discussion

We have attempted to craft a model in which each of the components of the system is linked to the climate. Not surprisingly, we find that both the damage and transport processes are strongly modulated by mean annual temperature when frost-related processes dominate. We hope that this effort can serve to show the way forward when other geomorphic and climatic
of saprolite to mobile regolith, are thought to be accomplished by reactive transport of water and oxygen to sites of mineral dissolution within the shale (e.g., Jin et al., 2011). The density and efficiency of the flowpaths through which these reactants are transported are in turn governed both by tectonically generated fractures, and by fractures generated by near-surface processes, including the peri-glacially generated frost cracks on which we have focused in this article.

Damage-limited versus entrainment-limited cases

The pace of lowering of the mobile regolith–weathered bedrock interface may be controlled by either the degree of damage of the rock, or by the efficiency of the available entrainment processes. Consider the case in which the rock mass quality is so poor that even before reaching the near-surface it has sufficient density of flaws to be entrained by the processes available at the base of the mobile regolith. In this case, the rate of release and hence lowering of the interface will be governed entirely by the rates of processes active at the base of the mobile regolith [e.g. upfreezing of clasts (Anderson, 1988a, 1988b) or tree-root prying]. This we may call the entrainment-limited case. However, if the rock mass possesses so few flaws to begin with, or is so tough that it is difficult for new cracks to form as it approaches the surface, the rate of release of pieces of bedrock into the mobile regolith will be governed entirely by the rate at which new damage occurs. We may call this case the damage-limited case. In a steady landscape the rate of lowering of the mobile regolith–weathered bedrock interface must be uniform; in this damage-limited case, the degree of rock damage in the weathered bedrock at the interface must therefore be uniform, as damage controls the rate of lowering. In the entrainment-limited case, the bedrock damage at the interface may vary considerably, but everywhere exceeds that required to release it into mobile regolith. In this case, whatever attribute of the weathered rock is responsible for governing the entrainment process must be uniform. Most earlier models of hilltop evolution have appealed to mobile regolith thickness as the proxy for the process intensity; the likelihood of upfreezing (progressive upward motion of rocks in soil due to frost heave; e.g., Anderson, 1988a, 1988b), the density of roots, and the availability of water to drive chemical alteration of the substrate indeed all depend upon thickness of mobile regolith.

Divide migration

The relief, $R$, of a hilltop of steady form above its bounding stream is a function of the lowering rate of the landscape (both the hillslope and the stream), the mobile regolith transport efficiency (here $x$), and the length of the hillslope from the divide to the stream, $L$ (e.g., Anderson, 2002):

$$ R = \frac{\rho L}{\rho b} \frac{W}{2K} L^2 $$

(21)

On a steady asymmetric hilltop, with bounding streams at the same elevation, the relief calculated from both streams is identical

$$ R_1 = R_2 = \frac{\rho L}{\rho b} \frac{W}{2K} L_1^2 = \frac{\rho L}{\rho b} \frac{W}{2K} L_2^2 $$

(22)

so that we may deduce the required shift in the divide due to differences in the landscape diffusivity on one versus the other slope:
In the case illustrated in Figure 11, the mean diffusivities on the two slopes differ by 32% at the end of the simulation (maximum difference 71%), while the divide has migrated so that the lengths of the two slopes differ by ~22%. The calculated divide migration should result in a length ratio of $\sqrt{\frac{1.32}{1.15}}$; in other words, one slope is 15% longer than the other. The present position of the divide at least crudely follows the expected scaling. Reasons for departure from the expected scaling include the fact that landscape diffusivity is non-uniform on both slopes (following the spatial pattern of mean annual surface temperature), and varies through time as climate dictates.

The asymmetry of peri-glacial hillslopes has been studied by a number of workers, as summarized by French (2007; see section 13.4.2). The steepest slope is most commonly west to southwest in northern hemisphere sites. These will be the warmest slopes, as in our simulation. Two explanations have been advanced. One conforms with what we suggest, that the steeper slope corresponds to the one on which the mobile regolith transport process is less efficient. In our simulations, this reflects higher mean annual temperatures that result in fewer and less deep frost-heave events. A second explanation involves the lateral movement of the stream that bounds the slopes; in other words, it has to do with the boundary condition at the base of the slope. The steep slope then corresponds to that toward which the stream is shoved, presumably by rapid mobile regolith deposition from the more efficient slope. The stream is not allowed to migrate laterally in our model, and hence this mechanism cannot explain the asymmetry produced. While slope asymmetry is a prominent outcome in our simulation, we do not see marked slope asymmetry in the Gordon Gulch watershed.

Information transfer across the landscape

While the models here are most pertinent to the Gordon Gulch section of the Boulder Creek CZO, we have made use of information gleaned from other portions of the Front Range landscape. The glacial headwaters to Boulder Creek lie only 5 km to the west of Gordon Gulch. The 6 °C amplitude reduction in mean annual temperature required to explain ice extent in the Front Range alpine zone (Dühnforth and Anderson, 2011) likely describes the thermal cycle to which Gordon Gulch was subjected; these temperature swings take the landscape into and out of peri-glacial conditions, and strongly modulate the importance of frost-related mobile regolith transport and bedrock damage of the substrate. Outboard of the crystalline core of the Front Range lie scraps of alluviated surfaces that appear to have widened and aggraded during glacial times (Dühnforth et al., 2012). Formation and subsequent abandonment of these surfaces has been interpreted to reflect significant swings in sediment supplied to the streams from their non-glaciated headwaters; high sediment supply leads to widening and alluviation, and low sediment supply leads to abandonment of the surfaces. Our models of thermally modulated, frost-heave mobile regolith transport operate with the proper sign, with low discharge of mobile regolith during inter-glacial times, and many times higher discharge of mobile regolith during cold glacial times.

Future Work

Despite these successes, the present model has a long way to go before we can consider it a viable model of landscapes such as the one that we target in the Boulder Creek CZO. There is much room for improvement. We have ignored the following complexities and feedbacks:

- In crystalline rocks, the density of flaws in the rock mass will govern the permeability of the rock. As it is the access of water to the rock mass that in turn governs the chemical attack of the rock, which in turn weakens it to further cracking, we expect a strong feedback here. To the degree that permeability structure controls the efficiency and the pathways available for water in the landscape, this feedback will also exert control on the chemistry and the biological activity in the subsurface. Indeed, in most 1D geochemical models of weathering front form and advance, hydrologic feedbacks in the system are either ignored or simply parameterized. For example, the evolution of the surface area on which chemical reactions play out, and which strongly governs the advection velocity (and the role of lateral flow) of fluids within the profile, is rarely incorporated (see Brantley and Lebedeva, 2011). We have similarly ignored this potentially important feedback in that our model is not coupled to a model of the hydrology of the hillslope. As the problem of hillslope evolution is not 1D but (at least) two dimensional (2D), full treatment requires a 2D transient hydrologic model that is capable of addressing the evolving spatial distribution of crack-induced porosity and permeability.

- In order to simplify the problem and demonstrate how rich even this simple system can be, we have focused solely on the physical processes of frost cracking and frost heave. But there remains a parallel effort that could and should be taken to explore the roles of other damage processes, and other transport processes. This reflects a broader effort of the geomorphic community in the last decade toward development of geomorphic transport rules (e.g. Heimsath et al., 2001; Dietrich et al., 2003). We hope that the attention to the role of climate in modulating the processes we have tried to capture will inspire similar attention to the roles of climate change in governing the roles of trees in the landscape. For example, in the Boulder Creek CZO we know that trees differ across the current landscape in terms of both species and density, and the elevation of the sites suggests that the vegetative cloak on the landscape will have shifted dramatically from a potentially treeless tundra to the present montane forest as the climate swings from glacial to inter-glacial conditions.

- The frost-related damage that we have attempted to capture in our model provides only one of the next steps in moving toward models that address both the climatic and geologic reality of any particular site. We have broken the problem of cracking into two pieces: the initial condition associated with rock structure and tectonically-generated cracks; and additional frost-cracking that is always strongly related to distance from the instantaneous earth surface. We have ignored the influence of topography on the state of stress in the near-surface rocks, which may either enhance or dampen the rate of crack growth. This is addressed in another paper in this volume (Slim et al., in press) in which the superposition of topographic and farfield stresses in the Shale Hills CZO, Pennsylvania, is hypothesized to enhance crack growth in valley bottoms and reduce it on ridgetops.

- We have simplified the climate to a sinusoidal history of mean annual temperature. In fact, the climate has likely spent much more time in the middle range of temperatures than such an algorithm produces (for example, the δ18O record from deep sea suggest that the global ice volume is normally distributed about the mean; see Anderson et al., 2006).

- Our thermal model is also inadequate in its characterization of aspect-related dependence of mean annual temperature. We have employed a linear dependence of mean annual
temperature on S-directed slope angle (last term in Equation 12). In reality, we require closer attention to the radiation field that will govern the thermal differences between the thermal states of N- versus S-directed slopes. While the mean annual direct radiation is easily calculated on any arbitrary slope, the additional roles of non-uniform tree cover and of non-uniform snow cover make calculation of the anticipated ground surface temperature histories quite complex. We have resorted to a simple algorithm that can be constrained empirically in the present day, and that at least has the correct sign.

- We have greatly simplified the role of snowfall, and for that matter of water in the landscape. In the present model we have at least attempted to capture the fact that water is required to drive frost cracking, and have done so by enacting a penalty on the cracking rate that increases with the distance the water would have to be pulled. But there is no true assessment of the availability of water. The wetness in the mobile regolith and in the bedrock mass does not evolve with time. As we have used high values for both mobile regolith and weathered bedrock in the majority of our calculations (90% saturation), our calculations of frost cracking should therefore be considered overestimates, although the present algorithms are far more defensible than the simple tallying of time in the frost-cracking window used in some earlier models.

- There is great need for documentation of the damage profile in weathered bedrock. While we have suggested that the crack surface area per unit volume is an appropriate metric, this is a difficult quantity to measure, and has not often been documented in the real world. We seek proxies for the degree of damage in rock specimens, which may include the tensile strength of rock (e.g. Kelly et al., 2011). On the landscape scale, Clarke and Burbank (2011) demonstrate the utility of shallow seismic methods, from which the velocity structure can be inferred and the effective fracture density can be calculated. Documentation of cracking frequencies in boreholes (e.g. Slim et al., in press) is indeed helping to inform models of cracking in other critical zone observatories.

- Again for simplicity, and to isolate the role of bedrock damage evolution in this hillslope system, we have assumed that mobile regolith production proceeds in a continuous fashion. In reality, discrete blocks are released whose size is governed by the degree of damage (the inverse of which is a length scale, or distance between flaws). While this element of reality is captured in a model presented by Riggins et al. (2011), there is utility in merging these approaches.

- The initial rock mass is assumed in the present model to be uniform in all of its properties. In reality, rock mass quality, or ‘intact rock strength,’ will vary due to stratigraphic architecture (bedding planes) and in its susceptibility to cracking by both tectonic and surface processes (Molnar et al., 2007; Clarke and Burbank, 2011). This variability will in turn exert considerable control on real hillslopes by governing variations in the initial condition of a rock as it begins to feel the surface ($D_0$ in Equation 2) and the rate of damage that subsequently accures.

- 1D versus 2D. In the present model, mobile regolith must move in one direction only; there can be no component of lateral motion. As discussed in Anderson (2002), this will eliminate the potential for formation of bedrock outcrops on midspotes; these are effectively killed by mobile regolith being forced to move over bare outcrops. This is particularly the case if the rate of rock damage is minor on bare rock outcrops (e.g. the humped rule; see Anderson, 2002), or if there are large spatial variations in the initial state of damage of the rock. Preliminary 2D models of mobile regolith motion indeed demonstrate a more complex motion of mobile regolith as it interacts with bedrock knobs.

- Finally, the boundary conditions in the present model are held steady: the bounding streams are made to lower at a prescribed steady pace. Yet in many of our stream systems, there is ample evidence in the form of small terraces that the streams have experienced episodes of aggradation. Where dated, these imply aggradation during glacial times. This is a failing of 1D calculations: stream dynamics depend not only on the local delivery of sediment to the stream, but to the delivery at all points along the channel, and the ability of the channel to transport the sediment delivered to it. The latter will clearly vary with climate. In a very real sense the hillslopes are slaves to the history of the stream. But it is a two-way coupling. The streams respond to climate change by not only handling the water from rain and snowmelt, but by attempting to pass on the sediment delivered to them by the hillslopes.

Conclusions

We espouse the view of the critical zone as one in which rock is attacked by damage processes en route to the mobile regolith interface, and is subsequently released into the hillslope conveyor system. Progress in understanding this system occurs when we can construct self-consistent models that acknowledge specific processes that damage the bedrock of a specified character, release it into mobile regolith, and transport it under specific climatic conditions. The time scales involved in the exhumation and the transport down typical hillslopes suggest that the legacy of past climates will be strong, emphasizing the need to characterize the modulation of these processes, and in some cases the switching of one suite of processes with another, by climate.

We have worked a specific example in which frost-related processes both exert damage on the rock in the subsurface by frost-cracking, and transport it by frost creep once released into mobile regolith. Results suggest that persistent aspect-related differences in microclimate translate into differences in hillslope transport efficiency that can lead to asymmetries in critical zone architecture and hillslope form (divide migration). Modulation of climate between glacial and inter-glacial conditions can result in significant variation in the mobile regolith cloak on the landscape. A direct corollary to this is a history of mobile regolith delivery to the streams that fluctuates significantly, with peaks in delivery during glacial maxima. This can affect both aggradation events (fill terraces) within the mountain front, and lateral planation to produce broad strath surfaces on the adjacent plains. There remains much to do in crafting process-based rules for bedrock damage, weathered bedrock release into mobile regolith, and mobile regolith transport in different climates and different rock types. Advances are needed to better our treatment of frost-related processes, specifically the explicit linkage to hydrologic processes that govern the degree of saturation of the soils through annual cycles. In other settings, other suites of processes, some of them biological, will require a parallel treatment.

Acknowledgments—This research was funded by a grant from the National Science Foundation (Boulder Creek CZO: NSF-0724960). The authors thank Taylor Peron, Sue Brantley, and two anonymous reviewers for insightful comments that greatly clarified the manuscript.

References


