The Concept of Transport Capacity in Geomorphology

John Wainwright,
Durham University, Department of Geography, Science Laboratories, South Road,
Durham, DH1 3LE, UK (john.wainwright@durham.ac.uk)

Anthony J Parsons
Department of Geography, University of Sheffield, Winter Street, Sheffield S10 2TN,
UK (a.j.parsons@shef.ac.uk)

James R Cooper
Department of Geography and Planning, School of Environmental Sciences,
University of Liverpool, Liverpool L69 7ZT, UK (james.cooper@liverpool.ac.uk)

Peng Gao,
Department of Geography, SUNY Syracuse, 144 Eggers Hall, Syracuse, New York
13244, USA (pegao@maxwell.syr.edu)

John A Gillies,
Division of Atmospheric Sciences, Desert Research Institute, 2215 Raggio Parkway,
Reno, NV 89512, USA (Jack.Gillies@dri.edu)

Luca Mao,
Department of Ecosystems and Environment, Pontificia Universidad Católica de
Chile, Avenida Vicuña Mackenna 4860, Santiago de Chile, Chile (lmao@uc.cl)

Julian D Orford
School of Geography, Archaeology and Palaeoecology, Queen’s University Belfast,
Belfast, BT7 1NN, Northern Ireland, UK (j.orford@qub.ac.uk)

Peter G Knight
School of Physical and Geographical Sciences
William Smith Building, Keele University, ST5 5BG. UK (p.g.knight@keele.ac.uk)
Abstract

The concept of sediment-transport capacity has been engrained in geomorphological literature for >50 years, although the earliest explicit definition is Gilbert’s work in fluvial geomorphology in the 1870s with implicit versions found in 18\textsuperscript{th} - 19\textsuperscript{th} century engineering. We review the development of transport capacity in the fluvial, aeolian, coastal, hillslope, débris flow and glacial process domains. Despite cross-fertilization between different process domains, there seem to have been independent inventions of the idea in aeolian and hillslope geomorphology. We demonstrate how different definitions have been used, which makes it both a difficult concept to test, and one that may lead to poor communications between those working in different domains.

The original relation between the power of a flow and its ability to transport sediment can be challenged for three reasons. First, as sediment becomes entrained in a flow, the nature of the flow changes and so it is unreasonable to link the capacity of the water or wind only to the ability of the fluid to move sediment. Secondly, the range of processes involved in most movements means that simple relations are unlikely to hold, not least because the movement of sediment often changes the substrate, which in turn affects the flow conditions. Thirdly, the inherently stochastic nature of sediment transport means that any capacity relationships do not scale either in time or in space. Consequently, new theories of sediment transport are needed to improve understanding and prediction, and guide measurement and management of all geomorphic systems.

Keywords: sediment transport; geomorphology; turbulence; complex systems; models; management
**Introduction**

The notion of sediment-transport capacity has been engrained in geomorphological and related geophysical literature for over 50 years. However, the definition has not always been consistent across the discipline. One reason for this inconsistency has been the tendency in recent decades for increasing specialization in studies relating to particular process domains. Together with the proliferation of literature, specialization means that only rarely are evaluations made of linkages between concepts in different process domains. This trend towards increasingly focussed specialization, however, contrasts with developments in complexity theory, Earth-Systems Science and integrated management of the environment that all emphasize the benefits of integrated approaches. In this review, we will first examine how the notion of transport capacity has been developed and applied in the various branches of geomorphology in order to evaluate commonalities and divergences in approach, in relation to conceptual developments and empirical support. In doing so, we assess the effectiveness of the concept by using advances across the different process domains to inform and interpret each other and thereby provide an overall critique of the concept. We take a critical realist approach (Richards, 1990; Sayer, 2000) to assess whether the use of the concept, both within a single process domain and between different process domains, enables the explanation of observations both in the field and laboratory. In other words, if the definitions of transport capacity that have been developed are real – mechanistic – characterizations of how sediment is transported in geomorphic systems, they should be able to explain observed phenomena. If not, there needs to be a reëvaluation of whether it is the concept (or concepts) that are problematic, or the experimental approaches and data that are flawed.

GK Gilbert was central to the development of the concept of transport capacity in geomorphology. In his *Report on the Geology of the Henry Mountains* [1877] he noted:

Transportation is favored by increasing water supply as greatly as by increasing declivity. When the volume of a storm increases, it becomes at the same time more rapid, and its transporting capacity gains by the increment to velocity as by the increment to volume. Hence the increase in power of transportation is more than proportional to the increase in volume.

[Gilbert 1877: 98]
Based on a steady-state conceptualization of river form, and an early formulation of a concept of stream power, he goes on to suggest:

Let us suppose that a stream endowed with a constant volume of water, is at some point continuously supplied with as great a load as it is capable of carrying. For so great a distance as its velocity remains the same, it will neither corrade (downward) nor deposit, but will leave the grade of its bed unchanged. But if in its progress it reaches a place where a less declivity of bed gives a diminished velocity, its capacity for transportation will become less than the load and part of the load will be deposited. Or if in its progress it reaches a place where a greater declivity of bed gives an increased velocity, the capacity for transportation will become greater than the load and there will be corrasion of the bed. In this way a stream which has a supply of débris equal to its capacity, tends to build up the gentler slopes of its bed and cut away the steeper. It tends to establish a single, uniform grade.

[Gilbert 1877: 106]

Implicit in this definition is the idea of supply limitation, that a channel will convey as much sediment as its capacity allows, unless there is insufficient material to be transported. More widely influential was Gilbert’s seminal USGS Professional Paper *The Transportation of Débris by Running Water*, published in 1914. It is this work that seems to provide the first formal definition of transport capacity in the English literature:

The maximum load a stream can carry is its capacity. ... Capacity is a function of various conditions, such as slope and discharge, .... When a fully loaded stream undergoes some change of condition affecting its capacity, it becomes thereby overloaded or underloaded. If overloaded, it drops part of its load, making a deposit. If underloaded, it takes on more load, thereby eroding its bed.

[Gilbert 1914: 35]

He goes on to say that capacity is a function of discharge, slope, débris size and channel-form (channel depth / width) ratio. While the concept of transport capacity is asserted without support from further reference in the 1877 report, his statement that “the maximum particles which streams are able to move are proportioned to the sixth powers of their velocities”
[Gilbert, 1877: 104] shows that he was at least aware of the work of Leslie [1823] or more likely Hopkins [1844] on entrainment by this time. In 1914, however, Gilbert suggests that the idea can be traced at least as far back as 1786 in the work of Dubuat, who demonstrated experimentally that the sediment discharge in a water flow was proportional to a square of excess velocity (flow velocity minus the flow velocity required for entrainment). These results were certainly still in use almost a century later, for example in the study of Lechalas [1871]. Both of these authors were working in a tradition of applied hydraulics, interested in the navigation of rivers and the building of canals.

Although Gilbert [1877, 1914] provides the first uses of the term in the geomorphological/geophysical literature, it was largely ignored or misunderstood by this community until championed by Strahler’s advocacy of a process-based geomorphology in his papers of the early 1950s [1950, 1952] (see, for example, Holmes’ review of 1915 in which he suggests the wide variability in observed rates mean that these sorts of observation are unlikely to be useful in dating geological deposits and thus unlikely to be of broader use. But, by contrast that “experiments of these kinds, while of considerable theoretical interest, serve only as a check to undue speculation” [1915: 134] and what is needed is more field study). However, the engineering aspects of his work meant that Gilbert influenced the hydraulic engineering literature more directly, for example being cited in the work of Hans Einstein [1942: 567], originally working in the Zürich school, but then later for the US Department of Agriculture (e.g., Meyer-Peter et al. [1934]; Einstein [1941/2; 1950] – but it is interesting that the first two of these papers do not mention the term capacity; it is only once Einstein was more fully embedded in the US system that he started using the term in print). From the 1950s, the term transport capacity is firmly established in fluvial geomorphology.

Apparently separately, in the aeolian literature Bagnold started to develop a concept of “sand-saturated wind” [Bagnold, 1936]. This concept arose from his experimental work to determine how much sand a given wind (shear) velocity could entrain. From this point on, the concept of capacity was established in the aeolian literature. Furthermore, because of Bagnold’s wider interests in both coastal geomorphology [Bagnold, 1946] and what he considered his “hobby” [Bagnold, 1990: 161] – fluvial geomorphology [Bagnold, 1966; 1973] – as well as granular flows [Bagnold, 1956], the terminology gained wider purchase. Bagnold’s work seems to be an example of an independent, parallel development of a concept, in that he
did not seem to be fully aware of Gilbert’s legacy until his own direct forays into hydraulic engineering in the 1950s (despite an earlier, practical start in this area: [Bagnold, 1990: 9]).

In the hillslope domain, a further apparently independent concept of transport capacity is introduced by Ellison [1947] in his work on soil erosion by water, although it was not taken further until the highly influential paper of Meyer and Wischmeier [1969]. The latter clearly uses concepts from the fluvial literature but cites neither Gilbert nor Einstein nor Bagnold as a basis for a broader concept of transport capacity in soil-erosion studies. The first explicit discussion of a transport capacity in relation to débris flows does not come until the 1990s, but also seems influenced by the fluvial literature. However, an example where the concept has been used much less, is in glacial geomorphology. In the rare case where it has been used, there seems to be an influence from the fluvial domain [Alley et al., 1997], and there needs to be an evaluation of the extent to which the concept is applicable to the complex glacial system.

Today, the concept of transport capacity influences most branches of geomorphology. It is widely presented in introductory texts, both in general terms:

the rate of transport is limited by the transport capacity of the process, which is defined at the maximum amount of material the process can carry

[Holden 2008: 302];

and in relation to specific processes, for example in the fluvial literature:

the transport capacity of the stream...can be viewed as being directly a function of flow discharge and slope

[Robert 2003: 11]

[the] rate of bedload transport is almost entirely a function of the transporting capacity of the flow

[Robert, 2003: 81]

capacity refers to the volume of material that can be removed for any given flow condition

[Robert, 2003: 146],
Most bedload formulae aim to determine the rate of bedload transport as a function of the transport capacity of the flow. This is done by calculating the excess flow capacity above a critical threshold, at which transport is initiated [Charlton, 2008: 110].

Stream power determines the capacity of a given flow to transport sediment. This is the maximum volume of sediment that can be transported past a given point per unit time [Charlton, 2008: 93].

Sediment transport rate (capacity) ... The sediment transport rate is the amount (weight, mass, or volume) of sediment that can be moved past a given width of flow in a given time [Bridge, 2002: 60].

and in relation to the hillslope domain:

Because soil creep moves the regolith material...transport is always at the transporting capacity [Holden, 2008: 309].

However, even within these examples, it is clear that there are differences in the terminology used. Furthermore, the application of the term in the research literature does not seem to be consistent within or among different process domains. For example, some authors use it to denote the maximum possible rate of sediment transport by a particular flow (e.g., Abrahams et al. [2001]; Abrahams and Gao [2006]; Foster and Meyer [1972]), while others use it to denote the potential for transport once grain resistance has been overcome (e.g., Davies [1988a]; Eaton et al. [2006]; Ferguson [2003]) and yet others as the transport rate under an equilibrium condition (e.g., Bagnold [1936]; Celik and Rodi [1991]; Gilbert [1877]; Gomez and Church [1989]; Zhang et al. [2009]). This flexibility in terminology is problematic. Not only does is make communication difficult both within the discipline, but also especially in interdisciplinary studies [Bracken and Oughton, 2006]. Furthermore, from an epistemic perspective, if transport capacity can refer to multiple real-world processes, then it becomes all
but impossible to test the idea, and it appears that only ancillary hypotheses are being evaluated rather than the core concept itself.

Wainwright et al. [2008] also criticized the use of the term in hillslope studies from both theoretical and practical considerations, and in particular the difficulties it poses when moving from one process domain to another. In the context of this review, they specifically note that erosion that shows a transition from concentrated overland-flow erosion and débris-flow erosion [as observed, for example, by Oostwoud Wijdenes and Ergenzinger 1998], could never be predicted by a transport-capacity-based model, as it would not allow sediment fluxes to become high enough to allow the transition to occur. They also demonstrated that both theoretical developments and the limited testing that had been carried out were limited to steady-state conditions, and were therefore unlikely to be representative of dynamic conditions.

In the following sections we provide an overview of how the concept has developed and been defined in the different process domains of fluvial, aeolian, coastal, hillslope, débris flow and glacial geomorphology areas of the discipline of geomorphology. We evaluate the conceptual development especially in terms of an undercurrent of changes in approach to measurement and monitoring: not least the shift from deterministic to stochastic underpinnings of process understanding; the dynamics in space and time, and thus the move away from simple, steady-state assumptions [Wainwright et al., 2008]; and the need to look at fluid-sediment systems in an integrated way in line with broader developments in complexity theory. By adopting a realist perspective, we argue that advances in all of these areas suggest that existing approaches are limited both conceptually and practically. We conclude by suggesting how the discipline might move forward for providing unified concepts of sediment-transport modelling.

Sediment-transport capacity in fluvial sediment transport

Gilbert’s concept of transport capacity in rivers was defined based on the assumption that sediment-laden flow is in equilibrium, so that the amount of sediment carried into a stream cross section or reach is equivalent to that transported out of it. Consequently, neither deposition nor erosion occurs in this cross section or reach and hence the associated channel morphology remains stable [Gilbert, 1877; 1914]. The linking of the ideas of capacity and equilibrium has been used to interpret the shapes of stream profiles (e.g., Mackin [1948]; but see discussions by Leopold [1980] and Bracken and Wainwright [2006]) and implied in régime
theory, which is the concept that river channels adjust towards a dynamic equilibrium when it is just able to convey the water and sediment supplied from upstream (e.g., Lacey [1930]; Knighton [1998]; Ferguson [2008]). The original idea of transport capacity was further developed by Einstein [1950] who considered capacity occurred when the rate of sediment supply was equal to the transport rate, and thus when the channel profile was in equilibrium. Most subsequent uses in the fluvial literature are based on one or both of Gilbert or Einstein.

However, the prevalent usage of sediment-transport capacity has been complicated for two main reasons. First, sediment in a river has been considered to be transported in different modes. Coarse particles travel through intermittent contact with the bed [Graf, 1971]. Thus, their movement is supported by the bed and the associated load is called bedload. Finer particles often move in suspension because their settling velocities are more easily balanced by turbulence. The associated load is referred to as suspended load [Vanoni, 1975]. Secondly, sediment transported in natural rivers is heterogeneous in size. Thus, both modes of transport can occur for a single flow condition. A further consequence of mixed grain sizes is that sediment-transport rates are controlled not only by river hydraulics, but also by the interaction between different sized particles within transported sediment and bed materials [Wilcock, 1988 - as already recognized in Gilbert, 1877], and the rate of upstream sediment supply. Sediment-transport capacity has thus been evaluated to date in terms of sediment – either homogeneous or heterogeneous – transported by bedload and in suspension.

Transport capacity for bedload

Gilbert explicitly indicated in his definition of transport capacity that his flume experiments were related to bedload transport only [Gilbert, 1914: 35]. The concept of bedload-transport capacity has been used as a benchmark against which the ability of a stream to transport sediment can be measured and compared. Furthermore, it has been widely used to determine degradation and aggradation rates on river-channel beds and to understand if bedload transport is primarily constrained by sediment supply (i.e., supply-limited rivers) or river-flow hydraulics (i.e., transport-limited rivers) [Gilbert, 1877; Andrew, 1979; Hicks and Gomez, 2003; Jackson and Beschta, 1982; Leopold et al., 1964; Lisle, 2007; Mackin, 1948; Reid and Dunne, 1996; Sear, 1996]. Quantitatively, the bedload-transport capacity has been regarded as the bedload-transport rate directly obtained from the flow in equilibrium, regardless of whether the flow is supply- or transport-limited [Ferguson et al., 1989; Gomez and Church, 1989; Wilcock et al., 2001] or more commonly inferred by comparing the measured transport rates
with those predicted by bedload-transport equations [Lisle and Church, 2002; Marti and Bezzola, 2006; Mueller and Pitlick, 2005; Reid et al., 1996; Warburton and Davies, 1998]. Determination of the bedload-transport capacity in these applications is fundamentally supported by the common assumption that bedload-transport capacity is the transport rate of a stream in equilibrium [Gomez and Church, 1989] that can be predicted by one of the established bedload-transport equations [Graf, 1971; Hicks and Gomez, 2003; Mueller et al., 2005]. However, numerous studies have shown that no single bedload-transport equation is applicable to all natural gravel-bed rivers [Almedeij and Diplas, 2003; Barry et al., 2004; Bathurst et al., 1987; Einstein, 1941/2; Gomez and Church, 1989; Martin, 2003; Reid et al., 1996] (Figure 1). This observation suggests that for given hydraulic and channel conditions, different equations may give rise to different bedload-transport rates, which is logically confusing as the ‘maximum load’ specified in the concept of transport capacity should be a unique value. However, even in Gilbert’s first statement on the concept, he suggested that “the capacity of a stream for transportation is greater for fine débris than for coarse” [Gilbert, 1877: 104], implying that the capacity is a function of particle size. Therefore, the concept of bedload-transport capacity needs to be revisited. Inasmuch as the behaviour of bedload transport is ostensibly different between beds with homogeneous and mixed sizes, the concept will be re-examined and re-quantified for the two types of grain-size distribution, respectively.

**Bedload-transport capacity for grains with homogeneous sizes**

Bedload with homogeneous grains mainly occurs in flume experiments. In a flow under the equilibrium condition and over a movable bed consisting of grains of identical size, bedload-transport capacity has been considered equal to the bedload-transport rate of the flow, because it has been assumed that in such flows only one transport rate is available for a given bed shear stress and grain size. Consequently, this transport rate is the ‘maximum’ rate that a given flow and sediment calibre can attain. This concept of transport capacity is consistent with Church’s theoretical maximum transport rate that occurs over an unstructured bed composed of the same material [Church, 2006: 336]. However, the hydraulics of these flows is often complicated by the emergence of bedforms such as ripples, dunes and pebble clusters wherein the shear stress responsible for the transport of bedload should be partitioned from the total shear stress [Engelund and Hansen, 1967; Fredsoe, 1993; Gao, 2012]. Although various correction methods for bed morphology have been developed [Baldwin et al., 2002; Engelund and Hansen, 1967;
Stone and Murdoch, 1989], selection of an appropriate method is not obvious in practice. To avoid such uncertainty, many experiments have been performed in flumes with flat, fixed beds [Abbott and Francis, 1977; Francis, 1973; Hu and Hui, 1996; Niño and Garcia, 1998; Sekine and Kikkawa, 1992]. However, by doing so, either these experiments do not test the underlying concept of capacity, or they can only ever represent the behaviour of sediment transport in very restricted conditions that seldom if ever occur in rivers. In the experiments of Niño and Garcia [1996], the saltation trajectories of individual bedload grains were measured using a high-speed video system, such that many dynamic properties of saltating grains (e.g., the saltation height, length and mean grain velocity) could be measured. These measured variables, together with measured flow hydraulics and sediment sizes have been subsequently used to solve governing equations established in terms of mass and momentum balance for individual bedload grains. This approach has led to various relationships between controlling variables such as shear stress and grain size (although issues relating to the interdependence of each of these variables is discussed below), and three dependent variables: $\delta_b$, the thickness of the bedload layer [m]; $C_b$, the mean volumetric concentration of bedload in this layer [kg kg$^{-1}$]; and $U_b$, the mean grain velocity [m s$^{-1}$] [Hu and Hui, 1996; Krecic and Hanes, 1996; Lee et al., 2000; Lee and Hsu, 1994; Sekine and Kikkawa, 1992; van Rijn, 1984a; Wiberg and Smith, 1985]. By definition, the unit volumetric bedload-transport rate, $q_b$ [m$^3$ m$^{-1}$ s$^{-1}$], is then given by:

$$q_b = \delta_b C_b U_b$$

Combining the relationships describing the controls on $\delta_b$, $C_b$, and $U_b$ in terms of equation 1 leads to various equations for bedload-transport rates.

However, many of the cited experiments that purport to determine transport capacity were performed by changing the sediment-supply rate while keeping the flow rate unchanged, implying that a given bed shear stress and grain size may be related to several different bedload-transport rates. Therefore, bedload equations developed in this way are not capable of estimating a unique transport rate for homogeneous grains, and are thus incompatible with the idea of a unique transport rate representing a capacity as suggested by this experimental approach. Again, either the underlying concept of transport capacity or the experiments used to try to demonstrate it have been found lacking.
The plane bed condition can also be achieved in flows over mobile beds [Fernandez Luque and van Beek, 1976; Graf and Suszka, 1987; Niño and Garcia, 1994; Rickenmann, 1991]. Bedload-transport rates measured in these flows have again been assumed to be at capacity. In these flows, bedload may be transported in either the saltation or the sheetflow régime [Gao, 2008a]. The former involves flows with low and medium shear stress in which bedload moves by sliding, rolling and saltating along the bed, while the latter contains flows with high shear stress where bedload is a loosely defined moving granular layer. In each régime, many bedload equations have been developed to predict bedload-transport capacities. For example, Fernandez Luque and van Beek’s [1976] equation and Smart’s [1983] equation are for capacities in the saltation régime, whereas Wilson’s [1966] and Engelund’s [1981] equations are for capacities in the sheetflow régime. However, none of these can predict bedload in both régimes [Wilson, 1989]. Thus, if the transport rates predicted by these equations do conform to transport capacity they do so only over the narrow range of hydraulic and sediment conditions under which they were developed and extreme caution should be employed when applying them more generally.

In an attempt to overcome this limitation, Abrahams and Gao [2006] compiled a large amount of bedload data with homogeneous grains that span a wide range of hydraulic and sediment conditions with mobile, flat beds. Using these data, they developed a general bedload equation aimed to predict bedload-transport capacities in steady, uniform, and turbulent flows transporting bedload of homogeneous grains over flat, mobile beds, which is hereafter termed the ideal condition:

\[ \phi_b = \theta^{3.5}G^{3.4} \frac{u}{u_*} \]  \hspace{1cm} (2)

where \( \phi_b = \frac{q_b}{\sqrt{(\rho_s/\rho - 1)gD^2}} \) is the dimensionless bedload-transport rate (Einstein [1950])

\( u \) is the mean flow velocity [m s\(^{-1}\)]

\( u_* = (gh_S)^{0.5} \) is the shear velocity [m s\(^{-1}\)]

\( g \) is the acceleration due to gravity [m s\(^{-2}\)]

\( G = 1 - \theta/\theta \)

\( \theta = \rho_S/(\rho_s - \rho)D \) is the dimensionless bed shear stress (also known as the Shields parameter)

\( \rho_s \) and \( \rho \) are the density [kg m\(^{-3}\)] of bedload grains and water, respectively
$D$ is the median size [m] of bedload grains

$h$ is the mean flow depth [m]

$S$ is the energy slope [m m$^{-1}$], and

$\theta_c$ is the critical value of $\theta$ for the initial movement of sediment.

The value of $\theta_c$ for each data set was estimated by plotting the volumetric bedload-transport rate $q_b$ against $\theta$, fitting a trend line to these data, and extending it to the $\theta$ axis where $q_b = 0$ [Abrahams and Gao, 2006].

Gao [2012] introduced a new dimensionless bedload transport rate $B = i_b / \omega$, where $i_b = q_b g (\rho_s - \rho)$ is the submerged bedload transport rate [J s$^{-1}$ m$^{-2}$], $\omega = \tau u$ is the stream power per unit area (or unit stream power) [W m$^{-2}$], and $\tau = \rho g h S$ is mean bed shear stress. He further demonstrated that equation (2) may be transformed into a much simpler form in terms of $B$:

$$B = G^{3.4}$$

The predictive ability of equation (3) was evaluated using a set of bedload data not only independent of that used in Abrahams and Gao [2006], but also covering a wide range of hydraulic and sediment conditions. Combination of these two groups leads to an unusual new group of bedload data that represent almost the entire spectrum of the saltation and sheetflow régimes under the ideal condition. The good fit of equation 3 to these data (Figure 2) suggests that under the ideal condition only, bedload-transport capacity can be generally determined by a single equation. Inasmuch as $G$ is quantitatively equivalent to $(P_g)^{0.5}$, where $P_g$ is the relative frequency of grain-to-grain collisions during bedload transport at a given flow intensity [Gao, 2012], the general good performance of equation (3) may be attributed to the perspective of treating bedload transport as a granular-flow phenomenon [Frey and Church, 2011]. Although equation (3) only applies to flows under the ideal condition, it demonstrates the ability of $G$ to characterize some aspects of the dynamics of bedload transport [Gao, 2012]. However, because the ideal condition rarely exists in natural rivers, it is of limited direct, predictive value. Furthermore, it could only ever be an incomplete component of a definition of capacity in more dynamic and variable settings.
Bedload-transport capacity for grains with heterogeneous sizes

Bedload of heterogeneous grains is typically transported in gravel-bed rivers. A well-known phenomenon in gravel-bed rivers is bed armouring, in which the bed coarsens due to the preferential transport of fine bed material. This coarsening causes a reduction in transport rate and for fines to become less exposed to the flow (the hiding effect) [Andrews and Parker, 1987; Egiazaroff, 1965; Einstein and Chien, 1953; Gomez, 1983; Lisle and Madej, 1992; Montgomery et al., 2000; Sutherland, 1987]. Therefore, for most of the time, only a fraction of grain sizes present on the bed of rivers is in motion [Gomez, 1995; Lisle, 1995], and so transport rates decrease compared to the values that have been assumed to represent theoretical capacity values as discussed above (Figure 3). During flooding, the armour layer may be broken [Ashworth and Ferguson, 1989; Clayton and Pitlick, 2008; Lisle and Smith, 2003; Parker and Klingeman, 1982; Wilcock and DeTemple, 2005; Wilcock and Southard, 1989] and hence surface grains of all sizes are able to move as bedload at similar levels of bed shear stress, the condition similar to the strong form of equal mobility [Parker and Toro-Escobar, 2002].

Bedload transported under this condition is generally believed to be at capacity [Gomez, 2006; Laronne et al., 1994; Parker, 2006; Powell et al., 1999; Powell et al., 2001; Wilcock and Crowe, 2003], which has been taken to imply that bedload transport capacity in gravel-bed rivers occurs only when the armour layer is broken and the transport rate is high, such as in peak flows during floods [Lisle and Church, 2002; Wilcock and McArdell, 1997].

However, as stated above, bedload-transport capacity has also been regarded as the transport rate under the equilibrium condition, which may also happen during relatively low flow with an armour layer. These two concepts of bedload transport capacity in gravel-bed rivers are inconsistent with each other if it is assumed that the flow alone controls the theoretical transport capacity. Furthermore, using the bedload data that were collected from a natural gravel-bed river [Reid et al., 1995] – that have been widely accepted to represent transport-capacity flow [Laronne et al., 1994; Powell et al., 2001] – Gao [2011] demonstrated by comparing the measured bedload-transport rates with those predicted using equation 2 that the former is significantly lower than the latter, suggesting that even the relatively high transport rates in flows without the armour layer are lower than the theoretical transport capacities of bedload with the equivalent homogeneous grains. Thus, the concept of bedload-transport capacity in gravel-bed rivers (i.e., for heterogeneous grains) appears to have been applied differently from that for flows transporting bedload of grains with homogeneous sizes. Such
an inconsistency reflects the problems with interpretation of the concept using a critical realist perspective.

In an attempt to avoid the inconsistency, bedload-transport capacity for heterogeneous grains was defined by Gao [2011] as the maximum possible transport rate a gravel-bed river can have for a given value of $\theta$ calculated using the median size of bedload grains $D_{50}$ [m], although this approach may confuse issues further by providing yet another definition of what capacity might be. During high flows when the armour layer is broken, average grain-size distributions in bedload and the bed substrate are similar [Andrews and Parker, 1987; Parker, 2006; Parker et al., 1982; Parker and Toro-Escobar, 2002]. When the channel bed is covered by an armour layer, bedload-transport rate has been related to the ratio of the median size of bedload grains ($D_{50}$) to the median bed-surface grain size ($D_{s,50}$ [m]) (eq. 3-48 in Parker [2006: 187]). Unfortunately, the relationship is uncertain due to variable sediment supply rates, which adjust both $D_{50}$ and $D_{s,50}$ [Buffington and Montgomery, 1999a; Buffington and Montgomery, 1999b; Dietrich et al., 1989; Gomez et al., 2001; Lisle and Madej, 1992; Whiting and King, 2003]. For example, the same ratio of $D_{50}/D_{s,50}$ may be caused by a low transport rate with small $D_{50}$ or a high transport rate with a large $D_{50}$ (Wilcock and DeTemple [2005]: see Figure 3). By examining a considerable body of bedload data obtained from natural gravel-bed rivers, Gao [2011] suggested that flows transporting bedload of heterogeneous grains at capacity should satisfy the following criterion:

$$D_{50} \geq D_{s,50}$$

where $D_{s,50}$ is the median size of grains in the bed substrate [m]. This criterion is compatible with the criterion above that the average grain-size distributions in bedload and the bed substrate are similar, and is valid for flows both with and without the armour layer in gravel-bed rivers. Using bedload-transport data from flows of various hydraulic and sediment conditions, Gao [2011] then suggested that bedload-transport capacity for heterogeneous grains may be generally determined using:

$$B = 0.9G^6$$

Equation 5 was further tested by showing its superior performance to the two well-known bedload-transport equations of Meyer-Peter and Müller [1948] and of Bagnold [1966], and its
ability to identify correctly flows that had been assumed to be at capacity in six gravel-bed rivers with armour layers in Idaho [Gao, 2011]. In practice, $\theta_c$ may be either derived from transport measurements [Church and Hassan, 2002; Mao et al., 2008] or estimated using a variety of statistical-, hydraulic- or lichenometric-based approaches [Gob et al., 2010; Recking, 2009; Thompson and Croke, 2008]. With a known value of $\theta_c$, equation 5 represents a single curve in the plot of $B$ against $\theta$ (the solid curve in Figure 4), which Gao [2011] suggests may be used to determine whether a given flow is transporting bedload at capacity (i.e., whether the flow is transporting bedload at a maximum rate). Comparing data representing four different flows measured in Simon River near Shoup, Idaho [Boise Adjudication Team, 2014] against the estimated maximum rate suggested that two achieved this rate and two did not (Figure 4). Further examining the ratio of the fractional transport rate $q_{bs}$ to the surface grain-size fraction, $f_i$, with respect to the associated median sizes of surface grains $D_i$ showed (Figure 5) that flows with higher $\theta_c$ values tend to have higher fractional transport rates for coarser grains ($D_i > 10$ mm in Figure 5), and that the distributions of the ratios for ‘capacity’ flows have similar patterns to those for ‘below-capacity’ flows. Thus, flows that are considered at capacity may have similar bed-surface conditions to below-capacity flows. For the concept of transport capacity to hold one would expect the bed surface in a capacity flow to be finer in order to transport bedload at a higher rate. Therefore the evidence suggests that the concept does not hold.

Significance of bedload-transport capacities for homogeneous and mixed grains

Gao [2011] suggests that bedload-transport capacities defined by equations 2 and 5 quantify the maximum loads a flow can transport for a given group of homogeneous and heterogeneous sediments. The definition is consistent with Gilbert’s, but more restrictive:

*bed-load transport capacity for heterogeneous grains is herein defined as the maximum possible transport rate a gravel-bed river can have for a given value of $\theta$ calculated using the median size of bed-load grains $D_{50}$. According to this definition, though a gravel-bed river with an armor layer may have several different available transport rates for a given $\theta$ value, it only has one maximum possible transport rate (i.e., the transport capacity)*

[Gao 2011: 298, emphasis original].
Since bedload in natural rivers is often transported at lower rates than these assumed capacity rates, as discussed in the previous section, these two equations (2 and 5) cannot be used directly for predicting bedload-transport rates. However, Gao [2011] suggested that they may serve as envelopes for all possible bedload-transport rates occurring in natural rivers and thus potentially have both theoretical and practical significance. In theoretical applications, by using the median grain size of the bed substrate ($D_{\text{sub}50}$), the dominant discharge, the representative channel slope, and the representative value of $\theta_c$, the putative transport capacity calculated from equation 5 seeks to predict the average bedload-transport rate of a river in equilibrium. Therefore, Gao [2011] suggests that equation 5 can act as a simple extreme hypothesis [Eaton et al., 2004; Knighton, 1998] to close the rational régime [i.e., that required to prevent erosion of the bed for design flow conditions: see Lacey, 1930], albeit in very restricted conditions. In unsteady, non-uniform flow, equation 5 has been treated as a conceptual framework in comparison with the standard approach of Schumm to help interpret channel response to changing flow régimes qualitatively [Phillips, 2013], although no testing of the framework was carried out.

In practical applications, it has been suggested that equation 5 provides a quantitative benchmark for identifying the hydraulic conditions under which an armour layer begins to break [Gao, 2011]. It has long been observed that bedload-transport rate increases abruptly at certain values of water discharge ($Q$ [m$^3$ s$^{-1}$]) in the plot of sediment-transport rate against $Q$. This increase has been attributed to additional sediment supply from the bed due to the breakup of the armour surface [Emmett, 1976; Emmett and Wolman, 2001] or the transition from selective transport (phase I) to equal-mobility transport (phase II) [Jackson and Beschta, 1982; Lisle and Smith, 2003]. However, the existing methods have generally focussed on determining the threshold values of $Q$, which are site specific [Bathurst, 2005; Ryan et al., 2002; Stoica and Sandgren, 2006]. Equation 5 provides a general tool for identifying such hydraulic conditions. Gao [2011] argued that it can be assumed that bedload in flows without an armour layer is transported at capacity (a point to which we will return in the discussion section), so that the hydraulic condition under which the armour layer breaks $\theta_c$ must be a point on the curve in the plot of $B$ against $\theta$ for a known value of $\theta_c$. Furthermore, the breakup of an armour layer must result in accelerated increase of bedload-transport rates, which should be consistent with the inflection point along the curve representing equation 5. Mathematically, determination of the inflection point leads to [Gao, 2011]: 
\[ \theta = 3.5 \theta_c . \] (6)

When \( \theta < 3.5 \theta_c \), bedload-transport capacities occur with an armour layer; when \( \theta > 3.5 \theta_c \), bedload-transport capacities occur without an armour layer. Based on these results, the two-phase model for identifying the breakup of an armour layer may be modified as shown in Figure 4. However, the asymmetry in observed transport rates as flows cross this threshold in increasing compared to decreasing flows [Mao, 2012; Tunnicliffe et al., 2000] implies that there is again a problem with interpreting the flow as having capacity conditions at that a single flow discharge. Besides leading to equation 6, equation 5 has also been used to estimate maximum possible sediment load for a potential extreme flood event as elaborated in Gao [2011], although this approach remains untested.

Transport capacity for suspended load

Since particles moving in suspension tend to have smaller sizes than those moving as bedload, suspended load most commonly occurs in alluvial rivers with sand beds. The suspended-sediment-transport capacity of a given flow has been defined as “the maximum amount of sediment this flow can carry in suspension under equilibrium conditions for a particular sediment material” [Celik and Rodi, 1991: 192]. In such a capacity flow, sediment size in suspension is the same as that on the bed and the flow is in sedimentary equilibrium – that is, deposited particles can be easily replaced by eroded ones and “any further addition of sediments to the flow leads to a deposition of sediments on the channel bed without an increase of the suspended sediment concentration” [Cellino and Graf, 1999: 455].

The complication of determining suspended-sediment-transport capacity arises from the fact that the transport of suspended sediment is almost always associated with bedload transport. For a given flow, distinguishing suspended sediment from bedload sediment is problematic and different criteria have been used to determine the two modes of transport [Abrahams and Gao, 2006; Allen, 2001; Turowski et al., 2010; van Rijn, 1984b]. Thus, the two modes of transport may be better described as a continuum rather than as two distinct classes [Parsons et al., 2015]. Therefore, it is often convenient to estimate the sum of suspended and bedload, which is conventionally referred to as the total load [Graf, 1971; Hicks and Gomez, 2003; Julien, 1998].
It has been assumed that these equations developed theoretically [e.g., using Bagnold, 1966; Einstein, 1950; Molinas and Wu, 2001; Yang, 2005] predict suspended sediment (or/and total sediment-) transport rates at capacity [Graf, 1971; Hicks and Gomez, 2003]. However, this assumption is not necessarily true for two main reasons. First, theoretically based equations often consist of several parameters that need to be determined using measured data. Unfortunately, in many equations, these parameters have been estimated using data obtained from flume experiments with fixed beds and variable sediment-supply rates [Galeone, 1996; Kronvang et al., 2012; Kronvang and Bruhn, 1996], which suggests that suspended sediment is not transported at a theoretical capacity (Celik and Rodi [1991]; see also the discussion above). Testing the concept of capacity in this way is thus about testing the auxiliary hypotheses relating to parameterization, rather than the concept itself. For example, the vertical profile of sediment concentration has been quantified using theories of diffusion-convection and vertical exchange of particles [Graf, 1971; Vanoni, 1975; Yalin, 1972], which led to various equations that include the Rouse number, \( R_0 = \frac{v_1}{\beta \kappa u_*} \), where \( v_1 \) is the settling velocity of suspended sediment [m s\(^{-1}\)], \( \kappa \) is von Kármán’s constant [-], and \( \beta \) is the ratio of the turbulent mixing coefficient of sediment to the momentum exchange coefficient [-] [Julien, 1998]. The value of \( R_0 \) is primarily controlled by that of \( \beta \), which is normally determined using measured profiles of sediment concentration in either flume experiments or natural rivers [Julien, 1998; Simons and Şentürk, 1992; Cellino and Graf, 1999; Graf and Cellino, 2002; van Rijn, 1984b].

The Concept of Transport Capacity in Aeolian Sediment Transport

In the context of aeolian sediment transport, the concept of transport capacity is most often expressed as being synonymous with the condition of steady state. Steady state implies that the flux of particles (e.g., sand in saltation) for a given wind shear is limited to an equilibrium value representing the saturation of the fluid flow with mobile particles. The assumption of steady state has been the foundation for developing models of sand transport by wind. A great deal of effort has been expended to define the theoretical relationship between the wind shear stress (\( \tau_a, \text{N m}^{-2} \)) or wind shear velocity (\( u_{*a}, \text{m s}^{-1} \), where \( u_{*a} = \sqrt{\frac{\tau_a}{\rho_a}} \), and \( \rho_a \) is air density [kg m\(^{-3}\)]) and the saturated flux of sand over a flat sand bed, typically denoted as \( q_{sat} \), which has the dimensions mass per unit width per unit time (e.g., Bagnold [1936, 1941]; Kawamura }
Bagnold [1936] derived the transport rate of sand-saturated wind \( q_{\text{sat}} \) \( \left( \text{kg m}^{-1} \text{s}^{-1} \right) \), which has come to be considered conceptually equivalent to transport capacity, as:

\[
q_{\text{sat}} = C_a \sqrt{\frac{d \rho_u u^3_a}{D_a g}} \tag{7}
\]

where \( C_a \) is an empirical coefficient for sorting [apparently defined based on wind-tunnel experiments: \( C_a = 1.5 \) for uniform sand, 1.8 for naturally graded sand, 2.8 for poorly sorted sand, and 3.5 for a pebbly surface], \( d \) is grain diameter [mm], \( D_a \) is the diameter of a standard grain (i.e., the one he first did his experiments on = 0.25 mm). The majority of theoretical models build on the Bagnold [1941] expression:

\[
q_{\text{sat}} = \frac{\rho u^2_a}{g} \tag{8}
\]

although a few models have been developed that are based on mean wind speed \( u_a \) at a reference height \( z \) (e.g., O’Brien and Rindlaub [1936]; Dong et al. [2011]) as opposed to the shear velocity. In both cases it is necessary to include a threshold term [Bagnold 1956], which sets a lower limit for transport as a function of either wind shear or mean wind speed.

In order to test the theoretical models, many experiments have been carried out using wind tunnels and atmospheric boundary layer flows to evaluate the predictive capability of models (e.g., Chepil and Milne [1939]; Bagnold [1941]; Zingg [1953]; Williams [1964]; Sarre [1988]; White and Mounla [1991]; Greeley et al. [1996]; Butterfield [1999]; Namikas [2003]; Li et al. [2009]). The measurement of the sand flux in these experiments has been accomplished using traps of various designs and efficiencies as well as active sensors based on correlating flux with acoustic, piezoelectric, or optical signals, thus reflecting an evolution from time-averaged to point measurements. Total horizontal flux, \( q \), is then calculated by integrating the vertical profile of point measurements of flux measured at specific heights (e.g., Shao and Raupach [1992]) or using traps with a continuous slot-like opening that extends to the approximate height of the saltation layer, thus integrating flux as a function of height during the measurement phase (e.g., Gillies et al. [2006]; Dong et al. [2011]).
Evaluations of how well measured horizontal flux rates compare with model-predicted values of saturated flux consistently show pronounced discrepancies (e.g., Sherman et al. [1998]). The reasons for the discrepancies can be traced to difficulties in accurate measurement of the sand flux [Ellis et al., 2009, 2012] and wind speed as well as in defining model coefficients [Sherman and Li, 2012], including coefficients in the “law of the wall”, which provides the basis for determining wind shear velocity [Bauer et al., 1992; Sherman and Farrell, 2008]. Li et al. [2010] observed variability in the von Kármán parameter $\kappa$ during sand transport, which is treated as a constant in the log-law expression for boundary-layer flow. One consistent observation from field studies is that the transport is highly variable in time (e.g., Stout and Zobeck [1997]; Baas [2004]; Davidson-Arnott et al. [2009]) and over space (e.g., Bauer et al. [1996]; Gares et al. [1996]; Ellis et al. [2012]), which suggests transport capacity is not a time-independent property of the system (see discussion below).

The degree of variability in wind flow and sand transport is much reduced in wind tunnels compared to the atmosphere, but wind tunnels create their own set of constraints on the transport system. The saltation process is altered by the dimensions of the wind tunnel, which introduce Froude-number effects [White and Mounla, 1991]. Durán et al. [2011] identify that a key difference between wind tunnel and atmospheric boundary layer flows is the integral turbulent time scales, defined as the time difference beyond which velocities at a single place and two different times become uncorrelated. In wind tunnels the turbulent time scale coincides with the transport time scale ($\approx 1$ s) while in a natural boundary layer with sand transport it is $10^3$ times larger [Durán et al., 2011].

With the observation that transport is typically unsteady in the atmospheric boundary layer, research has been directed towards an examination of this behaviour in the field and in wind-tunnel experiments. Stout and Zobeck [1997] characterize the intermittency of saltation with a basic turbulence intensity parameter and a simple wind-strength index. Stout [2004] advances the use of intermittency to estimate the threshold wind speed based on measurements of saltation activity, mean wind speed and the standard deviation of wind speed. The threshold is calculated during periods of saltation and reflects the conditions present at the time of measurement (e.g., grain-size distribution, surface-moisture content, relative humidity). The effect of unsteady wind conditions on flux was investigated in a wind-tunnel experiment by Butterfield [1998], who observed that transport was enhanced during introduced gusting periodicities between 6 s and 20 s, with rates in excess of those observed for steady winds of the same mean speed.
Spies et al. [2000] present a model developed from the work of McEwan [1991] that allows wind-velocity fluctuations to be forced upon the flow. The numerical simulations of Spies et al. [2000] wherein they impose several different types of unsteady wind behaviour to simulate the effect of a succession of gusts indicated that transport rate cannot respond to wind fluctuations greater than a frequency of 0.5 Hz. Their model also showed that the response time of $q$ to increased changes in wind speed was on the order of 2 to 3 s, while following a decrease in speed the response time of $q$ was approximately 1 s longer, which they note is in agreement with experimental data [Butterfield, 1991; Hardisty, 1993]. Based on their modelling results Spies et al. [2000] call in to question the legitimacy of modelling the transport system using the approach broadly defined by equation 8, as $u^*$ is an average quantity of the turbulent boundary-layer and cannot account for the effects of the fluctuating components of wind. They do note however that statistical properties of the flow such as the standard deviation and mean value of the fluctuations in wind speed scale with $u^*$, and suggest that the influence of turbulence on the transport mechanism be pursued further.

Researchers examining the aeolian transport system have conducted experiments in the field and in wind tunnels looking for links between turbulence parameters and sediment transport responses. In this alternative approach, $q$ is hypothesized to be governed principally by turbulent fluctuations and semi-coherent flow structures (e.g., Baas and Sherman [2005]; Baas [2004]; Walker [2005]; Baas [2006]) and does not invoke the concept of transport capacity. Baas [2006] carried out field measurements of wind and sand-transport activity at high frequency (20 Hz) span-wise to the flow and found that there was a complex interaction between turbulence and sand transport on three spatio-temporal time scales: (1) an external range on the order of 60 s, which represents longer term transport conditions that scale with time-averaged wind characteristics (i.e., $u^*$); (2) the integral time scale and below, which represent different transport patterns that show dependence on wind speed (streamer families, nested streamers, and clouds with embedded streamers as identified by Baas and Sherman [2005]); and (3) the scale of individual streamers at times less than 1 s. According to Baas and Sherman [2005], these streamers are a visual representation of near surface individual eddies that have translated down through the internal boundary-layer (IBL), skim across the surface and entrain/transport sand as they move downwind. The length scales of the streamers at their measurement location (i.e., width = 0.2 m, span-wise distribution = 0.9 m$^{-1}$) appeared to be stable and independent of wind speed (or wind shear). In this framework for analyzing sediment transport, boundary-layer turbulence and the gust cycles will, to a large extent, control...
the transport. However, Baas and Sherman [2005] caution that these properties of the flow may differ greatly for different surfaces even if $u_*$ were similar. Although research has demonstrated that turbulence and sediment transport are closely linked, there remains considerable challenge in developing transport models that are not based on mean flow and transport rates.

Aside from its use to identify maximal transport rates of sand, the concept of saturated flux, and the development of the system to reach an assumed saturation, has also been used in aeolian research to examine the time and length scales that the system manifests as it arcs from threshold to the attainment of saturated flux, or the distance it takes to adjust to a new saturated state following a change in surface conditions (e.g., surface topography, roughness). Originally identified by Shao and Raupach [1992] as an overshoot effect, the saturation length scale has been linked to the scaling behaviour of dunes affecting their profiles and wavelengths [Sauermann et al., 2001; Andreotti et al., 2002a, b; 2010], however the hypothesized scaling dependency for the elementary size of dunes remains controversial [Parteli et al., 2007a, b; Andreotti and Claudin, 2007; Andreotti et al., 2010].

Andreotti et al. [2002a] present a first-order relaxation differential equation to describe the behaviour of these saturation scales:

$$T_{sat} \frac{\partial q}{\partial t} + L_{sat} \frac{\partial q}{\partial x} = q_{sat} - q_a \quad (9)$$

where $t$ is time [s], $x$ is horizontal distance [m], $T_{sat}$ and $L_{sat}$ are the saturation time [s] and length [m], respectively, and $q_a$ is instantaneous sand flux [kg m$^{-1}$ s$^{-1}$]. $T_{sat}$ defines the time scale at which a saltation cloud adjusts from one saturated state to another following an abrupt or sudden change in wind speed, which is quite fast (order of seconds). According to Durán et al. [2011], $L_{sat}$ defines the length scale over which the flux reaches saturation as wind encounters an erodible surface, a change in roughness, or the adjustment from maximum $u_{*a}$ to maximum $q_a$ on the stoss side of a developing dune (Figure 7). Andreotti et al. [2010] argue that if grain inertia is the dominant dynamical mechanism limiting saturation then $L_{sat}$ should be independent of $u_{*a}$ and scale as a function of the drag length ($L_{drag}$ [m]), defined as the length needed for a grain to reach its asymptotic speed and scales as:

$$L_{drag} = \frac{\rho_g}{\rho_f} d \quad (10)$$
where \( \rho_p \) is particle density [kg m\(^{-3}\)] and \( \rho_f \) is fluid density [kg m\(^{-3}\)]. Using field data on sand-dune wavelengths and wind-tunnel data, Andreotti et al. [2010] demonstrate that once rescaled, \( L_{sat} \) is around \( 2 \times L_{drag} \) within a 50\% dispersion (Figure 7).

Hersen et al. [2002] offer a description of saturation length to conceptualize it and its rôle in sediment transport and bedform development. Briefly, as the entrainment threshold is reached and grains dislodge from the surface they will accelerate towards matching the wind speed. During this acceleration phase the grain covers a distance depending upon its inertia in the transporting fluid, which scales as equation 10. As a grain descends and collides with the surface it pushes some grains and splashes up others, and the splashed grains are, in turn, accelerated by the fluid flow, and as this process repeats, the sand flux increases until (assuming steady winds) grains cannot leave the surface except as a result of grains being deposited, i.e., the flux is saturated. The flux saturation length is proportional to this inertia length [Sauermann et al., 2001] and will increase if the length of the saltating trajectories increases (scaling as defined by equation 11).

Elbelrhitii et al. [2005] observed that the length of proto-dunes in the Sahara increased linearly with mean grain diameter for the same wind régime, providing evidence for the control of \( L_{drag} \) on initial dune size. The theoretical prediction of the length at which dunes emerge from a flat sand bed requires linear stability analysis (e.g., Andreotti et al. [2002a]) that incorporates two components [Andreotti et al., 2010]: (1) calculating the turbulent velocity field around an obstacle of low amplitude [Fourière et al., 2011]; and (2) describing the sand transport around the saturated state, wherein \( L_{sat} \) is critically important. The outcome of this analysis gives the relationship between the emergent wavelength of the most unstable mode, \( L_{sat} \), and the other parameters (e.g., \( u_{*a} \), aerodynamic roughness in the presence of saltation, \( z_0 \)). Dúran et al. [2011] note that the prediction of the emerging wavelength is essentially governed by \( L_{sat} \) and not sensitive to the flux relationship between \( q_a \) and \( u_{*a} \).

The aeolian transport system extends in a continuum from particles creeping along the surface to those in reptation (movement along the bed without significant movement up into the airstream) and saltation finally through to dust-sized particles (<70 \( \mu \)m diameter [Pye, 1987]) carried in suspension. These particles are released from the surface by aerodynamic entrainment [e.g., Kjelgaard et al., 2004a, 2004b; Sweeney and Mason, 2013], through ballistic impacts of saltating particles with the surface ejecting particles [Shao, 2000, 2004] or created by breakdown of soil aggregates [Kok, 2011]. These modes of transport are considered...
analogous to the bedload and suspended load for fluvial transport. The capacity for the wind to transport the particles in suspension scales with the wind shear velocity \( u_* \) balanced against gravitational settling velocity of the particle, which is a function of size and density. Particles will remain in suspension as long as the fluctuating vertical velocity component \( w' \) [m s\(^{-1}\)] exceeds the settling velocity \( v_{\text{sa}} \) [m], i.e.,:

\[
v_{\text{sa}} < \sqrt{\frac{w'^2}{\bar{u}_{\text{a}}}}
\]

(11)

and \( w' \) scales with \( u_* \) [Gillette, 1977]. Due to limitations on the supply and delivery of dust-sized particles to the atmospheric boundary-layer by constraints associated with surface conditions controlling the release of particles [Gillies, 2013], the transport of dust-sized particles will always be below capacity. During extreme dust-storm events, measured concentrations of particles ≤10 µm aerodynamic diameter have exceeded 14,000 µg m\(^{-3}\) (averaged over 24 hours) at Mono Lake, CA [Ono et al., 2011], which undoubtedly means that hourly concentration values were potentially much higher. Orlovsky et al. [2005] report that during dust storms in Turkmenistan visibility has been reduced to zero, which would require concentration of the suspended material to be much greater than any reported measurements. Even for light wind conditions, concentrations of dust (<50 µm geometric diameter) measured during haze events has exceeded 13,000 µg m\(^{-3}\) [Gillies et al., 1996]. The capacity of boundary-layer winds to transport dust-sized particles is likely never approached for terrestrial conditions.

The assumption of the aeolian sediment-transport system trying to or attaining saturation (i.e., capacity) has provided an important contribution to the explanation of transport rates, patterns and bedform development (e.g., Dúran et al. [2011]). It is clear however that this framework cannot explain all aspects of the mechanism of sediment transport by wind. Recent advances in observing the relationships between turbulence characteristics in the flow and responses of the sediment to these forcings has provided insights into the mechanisms of the transport processes (e.g., Baas [2006]), without invoking transport-capacity concepts. The challenge, however, of the turbulence-controlled framework for sediment transport is in developing predictive models to quantify sediment transport using some parameterizations for the key turbulence properties that control the flux of the particles in transport. At this juncture in aeolian process geomorphology it would seem that both saturation and non-saturation
approaches still provide critical insights into understanding the sediment-transport system. It remains to be determined if these frameworks can be unified, or even whether that is a necessary step to take.

**Sediment-Transport Capacity in Coastal Geomorphology**

Although the term sediment-transport capacity has not usually been explicitly employed in the field of coastal geomorphology, except in work directly by or influenced by Bagnold, the concept is implicit in the context of work on the rate of longshore sediment transport, which has been primarily developed from the standpoint of practical applications in coastal engineering. The history of the use of the term in coastal geomorphology mirrors the debate in the fluvial literature on the relationship of actual to theoretical values of transport capacity, and raises issues of scales of observation/measurement in the applicability of the term.

Early work developed empirical relationships between driving wave power and longshore sand transport (see Horikawa [1978: Ch. 5.4] for a summary). The major problem in Horikawa’s analysis is the substantial variance over orders of magnitude in the parameters associating the relationship between measures of transportability and sand transported (see both Fig. 5.4.23 in Horikawa [1978], and Table 4.2 in Horikawa [1992]). Such variation has to be seen in terms of the pragmatic attempts by investigators in establishing actual sediment transported as well as estimating the transport power involved at both prototype (i.e., full) and model scale. Most estimates of the relationship come from empirical statements of volume changes in profiles (i.e., some function of gross beach changes) versus wave measurements of often widely changing conditions and directions. The former, in particular, dealt with gross sediment changes and did not always differentiate between bed and suspended sediment loads (but assumed that median sand size was representative across the dynamic range), and had to make some assumptions as to limited or nil-offshore loss and dominant (if not total) longshore transport on actual beaches. It is also noticeable that initial measurements tended to derive from “infinitely” straight beaches, whereas most beach systems are not straight and as such induce subtle and scaled feedbacks in the variation of effective wave power especially after breaking where the strategic effects of morphodynamic relationships are dominant (see the discussion of Short’s [1999a, b] work). Likewise observations of wave power based on limited wave observations (often visual) or even from early forms of offshore wave recording, provide very poor characterization of available work, leading to major variance between reality and
observation. In particular later differentiation between direct wave thrust power (at the breakpoint), direct longshore tidal current, and secondary-current generation in the presence of breaking waves all combine to cause poor characterization by use of pre-breaking wave parameters. Set against these difficulties of measurement, it is unsurprising that relationships between transport rate/capacity and wave power were difficult to obtain. Nonetheless, a body of literature did emerge (especially from the west coast of the USA) that supported a pragmatic positive relationship (at log scale) between wave forcing and sediment transported.

Typical of this pragmatic approach is that of Caldwell [1956], which is often cited as the initial exemplar work of relating transport volume to what has become known as wave power:

\[ Q_l = 210[P \sin \alpha \cos \alpha]^{0.8} \]  (12)

in which \( Q_l \) is the longshore volume transport of sand [originally measured in cubic yards day\(^{-1} \)], \( P \) is the incident wave power [millions of ft-lb/day/ft of shoreline], and \( \alpha \) is the angle the breaking wave makes with the beach \(^{[\circ]} \). Caldwell’s work was trying to link a longshore sediment-transport rate (\( Q_l \)) to a generative process power (\( P \)), though later investigators use \( P_l \) (defined as the longshore-directed component of \( P \)) where \( P_l \) is defined as:

\[ P_l = (E C_w n)_b \sin \alpha_b \cos \alpha_b \]  (13)

where \( (E C_w n)_b \) is the wave-energy flux or power per unit wave-crest length at the breaker point \( (b) \) [as a power term, the units of \((ECn)_b\) are W m\(^{-1}\), though earlier definitions were in dynes s\(^{-1}\) m\(^{-1}\)]. Equation 13 is based on linear wave theory where the velocity of the wave form \( (C_w) \), the energy of the wave; \( (E) \) is the wave energy (proportional to the square of wave height); \( (n) \) a dimensionless calibration of wave-form velocity (which increases from 0.5 in deep water to 1 in shallow water); and \( \alpha_b \) is the angle between the breaker crest line and the shoreline. The sine and cosine terms translate the full onshore wave power to a directional thrust of wave power obliquely against the shoreline, i.e., the longshore component of wave power (\( P_l \)). This type of wave-power derivation has become central as the driver of shoreline transport in subsequent analyses (see Komar [1999] for a general development of this approach) where a blend of observed versus theoretical statements of actual longshore transport have been developed to match against this type of statement of potential longshore power (i.e., transport...
capacity). A major difficulty has been specifying a highly variable wave power (fluctuating at seconds to hours) against sediment-transport rates that at the prototype scale tend to be net statements based on months to years.

Inman and Bagnold [1963] pointed out that equation (13) is both dimensionally incorrect and provides only an empirical relationship between observed wave power and sand transport, rather than what may be thought of as a more theoretical transport capacity: a similar confusion to that of the ‘potential-actual’ debate of the fluvial dynamicists. Inman and Bagnold preferred to drop the $Q_l$ measurement for what they term the “immersed-sediment weight transport rate” ($I_t$ [m³ l⁻¹]), which can be related to $Q_l$ through:

$$I_t = (\rho_s - \rho)g\alpha'Q_l$$  \hspace{1cm} (14)

in which the use of both $\alpha'$ (the correction for pore space taken as 0.6) and the $(\rho_s - \rho)$ term means that only the immersed density of solid sediment transported is considered. The use of $I_t$ – rather than $Q_l$ – is based upon the sediment-transport theories of Bagnold [1963, 1966]. These authors proceeded to develop a theoretical underpinning for $I_t$ to derive:

$$I_t = K \frac{EC_w n \cos \alpha}{u_0 \tan \phi} \bar{u}_l$$  \hspace{1cm} (15)

in which $K$ is the ratio of the rate of work done in transporting sediment in relation to the total wave power available [\cdot], $EC_w n$ is the rate of transport of energy of a wave [W m⁻¹] as defined in equation 13, $u_0$ is the mean frictional velocity relative to the bed in the surf zone [m s⁻¹], $\phi$ is the inter-granular friction angle [°] and $\bar{u}_l$ is the mean longshore current velocity [m s⁻¹]. Inman and Bagnold further argued that if the position of rip currents can be assumed to be random in time and space along the beach then a general expression for the rate of longshore sediment-transport capacity over a straight beach could be derived. They provide the rationale by which the immersed sediment weight transported by wave power can be related by:

$$I_t = K P_l = K (EC_w n) \sin \alpha_b \cos \alpha_b.$$  \hspace{1cm} (16)
This relationship is plotted against field measurements of sand-transport rates (Figure 8) and shows quite good agreement, yielding a value for $K$ of 0.7 [Komar, 1999]. $K$ henceforth became a significant concept in relating alongshore sediment transport to available wave power, underpinning both the original theoretical work by Inman and Bagnold [1963] and the empirical relations such as that compiled by Horikawa [1978] and Komar [1999]. Unless the nature and controls on $K$ can be understood, any attempt to move beyond empirical predictions of longshore-transport rate towards any process-based notion of longshore-transport capacity will be hamstrung.

It has been suggested that $K$ might vary with grain size [Tanner 1978], and hence provide comparability with (limitations of) transport-capacity formulae in aeolian and fluvial systems. However, there is little support for the notion that $K$ decreases with $D_{50}$ within the range of sand-sized sediments [Komar, 1980], though some (but questionable) support that such a relationship might exist if the size range is extended to include gravel and shingle [Komar, 1999]. However, data from gravel beaches are limited, but the increasing influence of particle shape with increasing size (within the gravel-cobble size range: Orford and Anthony [2013]) means that a singular $K$ value is unlikely. Other research has proposed that $K$ depends on beach slope and wave steepness as well as grain size according to:

$$K \propto S \left( \frac{H_b}{D_{50}} \right) \left( \frac{H_b}{L_\infty} \right)$$

(17)

in which $S$ is slope [m m$^{-1}$], $H_b/L_\infty$ is a dimensionless measure of breaking wave steepness [breaking wave height [m] over offshore wave length [m]] and $H_b/D_{50}$ is the dimensionless scaled wave-sediment size parameter [l/l]. Furthermore, given that there is a positive relationship between sediment size and beach slope, the relationship will confound any attempt to find a simple relationship with grain size. Studies by Saville [1950], Özhan [1982] and Kamphuis [1991] have produced conflicting results on the relationship with wave steepness, but the suggestion is that $K$ decreases with wave steepness (i.e., $K$ reduces as short-period storm waves give way to long-period swell waves). Equation 17 also implies that transport rate increases with beach slope. However, given the advances in understanding beach morphodynamics [Short, 1999a] the above relationship also indicates that $K$ is proportional to whether the beach system is reflective (steep and coarse sediment) or dissipative (lowest slopes and smallest
The implications for secondary current generation within a dissipative morphodynamic régime and the ensuing associated three dimensional beach morphology changes with this energy domain, underlies the difficulty of asserting a single $K$ value across beach sediment size range. It also underlies the appropriateness of considering subsets of longshore transport capacity related to position in the tidal range and the ensuing potential of more consistent longshore beach morphologies where transport vectors both gross and net may be at base, uni-directional. Post-wave breaking on a reflective coarse clastic beach (sub-gravel size) may be the simplest context for $K$ analysis. The likelihood of this context is low and only serves to indicate the difficulty of assessing the connection between actual and theoretical transport capacity.

The range and debate about the value of $K$ may also reflect the possibility that the characterization of the dynamics of power to transport is still not adequately specified both in terms of actual parameters used and/or resolution of measurement of both sides of the power-transport equation.

In recent years there have been attempts to resolve coastal sediment-transport rates with better determination of transported sediment volumes as well as clearer characterization and measurement of process variables. Kamphuis [1991] proposed a transport-rate formula, the form of which was revised by Mil-Homens et al. [2013] as:

$$Q_l = \frac{0.15 \rho_s}{(\rho_s - \rho)} T_p^{0.89} (\tan \beta)^{0.86} d_{50}^{-0.59} H_{S,br}^{2.75} \sin(2 \theta_{br})^{0.5}$$

where $Q_l$ is longshore sediment [dry mass in kg s$^{-1}$], $H_{S,br}$ is significant wave height at breaker line [m], $\theta_{br}$ is wave angle at breaker line [°], $d_{50}$ is the median particle size in the surf zone [m], $\beta$ is the beach slope [°], and $T_p$ = peak wave period (s). Note that the immersed weight has been changed to solid weight of sediment transported, while the explicit value of $K$ has been lost, given the better specification of calculated driving power based on peak wave period and significant ($H_{67\%}$) breaking wave height. However a dimensionless calibration element (0.15) is still required to constrain the predicted transport rates.

van Rijn [2014] provides the latest comprehensive analysis of existing transport equations and transport rate data for both coastal sand and gravel systems. In this detailed
examination of process he provides the following sediment-transport predictor equation for
shorelines, based on the same approach as Kamphuis and Mils-Homens, but which has the
ability to be applied across the sand to coarse gravel (0.1-100 mm) sediment range:

\[ Q_t = 0.00018 K_{swell} \rho_s g^{0.5} (\tan \beta)^{0.4} d_{50}^{-0.6} H_{s:br}^{3.1} \sin (2 \theta_{br}) \]  

(19)

As regular swell waves yield larger longshore transport rates than irregular wind waves of the
same height (by a factor of 1.5), van Rijn proposed to take this effect into account by a swell-
correction factor \( K_{swell} \). This use of \( K \) is only loosely related to the earlier debate of \( K \), as this
\( K \) value is certainly not a constant and is designed to amplify the wave power potential of longer
period waves, rather than directly calibrate the potential versus actual wave power of the model.
He identifies the value of \( K_{swell} \) based on the percentage of swell waves \( (p_{swell}) \) within the
incident wave spectrum, such that as: \( K_{swell}=1.05 \) for \( p_{swell}=10\% \); \( K_{swell}=1.1 \) for \( p_{swell}=20\% \)
and \( K_{swell}=1.5 \) for \( p_{swell}=100\% \). If swell is absent (or unknown), then \( K_{swell}=1 \). The power of
this predictor is in its highly reduced variance between predicted and observed transport across
the sand and gravel size range (Figure 11 in van Rijn [2014]) where about 80% of the 22 data
points are within a factor of two of the 1:1 line. This is the current ‘state of art’ for predicting
longshore sediment transport and is a substantial improvement on the variance of the Horikawa

All the approaches above are what can be termed computing a “total energy” capable of
bulk transport. They apply a wave energy parameter and angle of wave incidence to produce
an “alongshore component” of wave power. Alternatively there are a few studies that attempt
to establish a modified shear stress/stream power formulae based on the formulae that were
originally used for estimating transport rates in rivers. Morfett [1990] explains the basic physics
of this approach. Investigators have to change the flow field from a stream to that of a wave
(e.g., Bijker [1971], Swart and Fleming [1980] and Baillard [1984]). This approach is an attempt
to model the combined bed and suspended loads as a consequence of the distribution of
transport rates in the surf zone and relating sediment transport to the rate at which energy is
being dissipated per unit area, \( D \). In an approach that draws directly upon the fluvial literature
Morfett [1990] proposed an equation for longshore sediment transport based upon the notion
of virtual wave power, drawing upon McDowell’s [1989] use of virtual stream power \( P_{str} \)
\([W \text{ m}^{-2}] \) in which:
where $u^*$ is shear velocity [m s\(^{-1}\)] and the subscript “cr” refers to the shear velocity at the threshold of sediment motion. The use of a third power on the shear velocity has indications of the continuity of Bagnold’s approach in thinking of effective powers of shear stress from other sediment-transport régimes. Morfett [1990] then derived the equation:

$$I_t = K[P (\sin \alpha)^{0.75} D_\ast]$$

(21)

where

$$P = P^* g^{-0.833} (\rho_s - \rho)^{-0.5} \nu^{0.167}.$$  

(22)

where $I_t$ is the transport rate [N s\(^{-1}\)], $P^*$ is the wave power that is dissipated in breaking and in overcoming friction [W m\(^{-2}\)], and $\nu$ is kinematic viscosity [m\(^2\) s\(^{-1}\)]. $D_\ast$ is a parameterized particle size (dimensionless), which Morfett developed from an empirical formulation based on the median grain size ($D_{50}$), relative density of immersed sediment and water viscosity. He derived $D_\ast$ as:

$$D_\ast = D_{50} \left( g \left( \frac{\rho_s}{\rho} - 1 \right) \nu^2 \right)^{0.333}$$

(23)

Morfett reduced the overall equation to a power-law relationship (using log-transformed best-fit linear regression) (Figure 7 in Morfett [1990]) of the form:

$$I_t = K P^* D_\ast$$

(24)

where $P^*$ is $P (\sin \alpha)^{0.75}$. Once again, however, the equation depends on the constant $K$ for which only an empirical determination based upon the range of observed transport rates is available. Morfett identifies $K$ as $10.5 \times 10^3$ as a best-fit regression constant. Hence his final transport based on a dissipative wave energy model is:

$$I_t = 10.5 \times 10^3 P^* D_\ast$$

(25)
Equation 25 attempts to provide a physical attempt at developing a transport-capacity statement for shorelines, however its determination is dependent on the ability to develop a statement of energy dissipation at the shoreline which is somewhat intractable considering the easier measurement required for the bulk transport equations (e.g., equation 19). This difficulty may explain why the literature has tended to expand on more defined and precise bulk rate equations (e.g., van Rijn [2014]), rather than the energy-rate determinations as exemplified by Morfett.

**Sediment-Transport Capacity in Hillslope Studies**

The study of sediment transport on hillslopes comes mainly from attempts to develop process-based models to predict soil erosion. Early work was carried out by the US Department of Agriculture Soil-Conservation Service as a result of the Dust Bowl of the 1930s, resulting in the studies of Zingg [1940] and Musgrave [1947] that predicted sediment yield as a function of discharge, slope and soil characteristics. The first explicit use of the concept of transport capacity seems to have been by Ellison [1947] in the first of his papers redefining soil-erosion studies. No citation is made of previous usage of the term and its derivation is related to empirical observations made in one of Ellison’s own experiments. Meyer and Wischmeier [1969] noted that Ellison’s ideas had not been adopted, so little progress had been made in the intervening two decades in developing the empirical data that might underpin his approach. They consequently developed his conceptual model more fully (Figure 9) and made the first attempt to apply it mathematically. Using the relationship of Laursen [1958] from fluvial experiments, they estimated the transport capacity of flow as a function of flow velocity or discharge, and thus the Meyer and Wischmeier paper seems to be the first instance when the fluvial literature is used to inform hillslope-erosion models. Although the idea of a transport capacity of rain is introduced by Ellison [1947] and then followed up by Meyer and Wischmeier, it has not subsequently been used, probably based on the assumption that measured rainsplash is always at “capacity”. The distinction between interrill and rill erosion, drawn by Foster and Meyer [1972; Meyer et al., 1972, 1975], is crucial in underpinning the theory of transport capacity as applied to studies of erosion. As such, it underpins all but the most recent erosion models, although the distinction probably hides more of a continuum of behaviours [Wainwright et al., 2008].

Transport capacity of interrill flows has been little investigated. Moore and Burch [1986] assumed unit stream power as defined by Yang could be used to predict capacity in a
continuum between interrill and rill flows, although most of their evaluations were on datasets relating to rilled slopes. They also demonstrated [Moore and Burch, 1987] that this form is conceptually equivalent to the standard form derived by Julien and Simon [1985] from dimensional analysis:

$$q_s = \zeta S^\lambda q^\mu i^\xi$$  \hspace{1cm} (26)

where $q_s$ is unit sediment discharge (assumed at capacity) [kg m$^{-1}$ s$^{-1}$], $q$ is unit flow discharge [m$^2$ s$^{-1}$], $i$ is rainfall intensity [m s$^{-1}$] and $\zeta, \lambda, \mu$ and $\xi$ are empirical parameters,

when $\lambda = 1.3 \psi$ or $1.375 \psi$ and $\mu = 1 + 0.4 \psi$ or $1 + 0.25 \psi$, depending on conditions; $\psi$ is the exponent in Yang’s relationship between transport capacity and (excess) unit stream power. More recent empirical studies [Everaert, 1991] and overviews [Prosser and Rustomji, 2000] have also supported the used of equation (26), although most experimental approaches have assumed that measured sediment transport, $q_s$, in equation (26), is equivalent to transport capacity without otherwise demonstrating the equivalence (e.g., Everaert [1991]).

The development of a process-based understanding of erosion and sediment transport by rillflow that is embedded in most process-based models of soil erosion of the later 20$^{th}$ century has taken its lead from the literature on river flow. The concept of transport capacity of rillflow is, therefore, an inherent component of these models. However, to transfer, largely empirical, equations developed for flow on very shallow gradients that is deep in relation to the size of most sediment transported to flow that is shallow and on typically much steeper gradients has proved problematic. Foster and Meyer [1972] developed a theory of erosion actually based on the assumption that under steady-state conditions:

$$\frac{D_F}{D_c} + \frac{G_F}{T_c} = 1$$  \hspace{1cm} (27)

where $D_F$ is the net flow-detachment rate [kg m$^{-2}$ s$^{-1}$]
By defining the detachment capacity as a linear relationship of transport capacity ($D_c = k \cdot T_c$), they produce the relationship:

$$D_F = k \left( T_C - G_F \right)$$

(28)

It should be noted that this approach was arrived at by (a) a qualitative appreciation of laboratory and field observations; and (b) an analogy that if transport capacity is a function of shear stress raised to a power (a simplified Yalin equation: see discussion below), then so too should detachment rate, with $k$ in equation (28) simply the ratio of the coefficients in the detachment- and transport-capacity relationships. At no point did they demonstrate that this analogy was supported empirically or theoretically. As early as 1974, Bennett pointed out that this argument was one that needed testing more fully, something that Wainwright *et al.* [2008] noted was yet to happen over 30 years later. Although some limited attempts have been made to verify the relationship [Rice and Wilson, 1990; Cochrane and Flanagan, 1996; Merten *et al.*, 2001], more detailed experiments by Polyakov and Nearing [2003] and Schiettecatte *et al.* [2008] demonstrate that the linear relationship in equation (28) does not hold (Figure 10).

In the Water Erosion Prediction Project (WEPP) model [Nearing *et al.*, 1989] soil detachment in rills is modelled as:

$$D_F = D_c \left[ 1 - \frac{G_F}{T_c} \right]$$

(29)

which is a straightforward rearrangement of the Foster and Meyer equation (equation 27). Transport capacity is then determined from a simplified form of the Yalin equation, of the form:

$$T_c = k_t \tau_f^{-1.5}$$

(30)
where $\tau_f$ [Pa] is the hydraulic shear acting on the soil and $k_f$ [m$^{1/2}$ s$^2$ kg$^{1/2}$] is a transport coefficient.

Similarly, in the European Soil-Erosion Model (EUROSEM) [Quinton, 1997; Morgan et al., 1998]:

$$D_{FE} = k_f w v_1 (T_C - C)$$  \hspace{1cm} (31)

where $D_{FE}$ [m$^3$ s$^{-1}$ m$^{-1}$] is the net detachment rate, $k_f$ [dimensionless] is a flow-detachment-efficiency coefficient, $w$ [m] is flow width, $v_1$ [m s$^{-1}$] is particle settling velocity, $T_C$ [m$^3$ m$^{-3}$] is transport capacity, and $C$ [m$^3$ m$^{-3}$] is sediment concentration. Transport capacity is determined from the work of Govers [1990] who found empirically that, for particles between 50 and 150 µm, transport capacity could be expressed as:

$$T_C = c(\omega_Y - \omega_{Ycr})^\eta$$  \hspace{1cm} (32)

where $\omega_Y$ [m s$^{-1}$] is Yang’s unit stream power (defined as $u S$), $\omega_{Ycr}$ is critical Yang’s unit stream power (= 4.0 mm s$^{-1}$), and $c$ and $\eta$ are experimentally derived coefficients that depend on particle size. LISEM (The Limberg Soil-Erosion Model: de Roo et al. [1996]) uses the same derivation, based on the work of Govers [1990]. GUEST (Griffith University Erosion System Template) bases its transport-capacity equation on Bagnold stream power [Misra and Rose, 1990; Rose et al., 1998].

Both equations 29 and 31 share the assumption made by Foster and Meyer [1972] that detachment varies linearly with the difference between sediment in transport and transport capacity, though no published evidence supports this assumption [Wainwright et al., 2008: 816]. Equation 29, taken from Smith et al. [1995], adds a modification to the detachment rate to take account of the fact that previously detached sediment will return to the bed, thereby reducing the sediment load. However, this modification is based upon the settling velocity of the particles, and, therefore, assumes that all transported sediment is in suspension, which Wainwright et al. [2008] have demonstrated is unlikely to be the case for all but the finest sediments on most hillslopes under overland flow.

Equation 27 was proposed by Foster and Meyer [1972] and has empirical support in the work of Alonso et al. [1981]. However, later work by Moore and Burch [1986] and Govers [1992] shows the Yalin equation to perform relatively poorly in laboratory experiments of
rillflow, and Govers further argues that all excess-shear formulae are unsuitable for rillflow. The form of the equation (30) assumes that shear stress is well in excess of the critical shear stress to entrain particles. Inasmuch as the initiation of rills is defined by the point at which critical shear stress is exceeded, an equation for transport capacity that does not apply to this situation would seem inherently inappropriate. Govers [1992] further suggests that the apparent success with the Yalin formula achieved by Alonso et al. [1981] may lie in the restricted range of conditions tested: specifically that no gradients exceeded 0.07 m m⁻¹, which is scarcely representative of hillslope erosion.

Huang et al. [1999] carried out experiments where they could control the bed hydrology, and found that identical flows where the slope had net drainage had lower estimated transport capacities than those with net seepage. Sander et al. [2007] used the Hairsine-Rose model to reinterpret these results as an emergence of an (effective) set of transport capacities under different conditions in that the Hairsine-Rose model does not define a transport capacity explicitly (as Huang et al. [1999] had also suggested for their results). They interpreted the results as reflecting the depositional layer protecting the bed once deposition starts to occur.

Polyakov and Nearing [2003] demonstrated that the Hairsine-Rose model would produce different values of transport capacity which they interpreted as hysteresis based on whether the value was reached from an excess or a deficit of sediment transport. They also noted that “transport capacity implies uniqueness” [ibid.: 42], and thus suggested that the above-mentioned variable transport capacities for a similar flow and sediment condition suggests that these flows are not all at a putative capacity, the familiar situation discussed earlier in the section on ‘transport capacity for bedload’ in alluvial rivers. Although one solution is to use equation 3 or 5 to determine the unique transport capacity, it is important to recognize that neither of these approaches has been tested in hillslope settings. Some reasons why fluvial approaches might not immediately transfer to the hillslope domain are elaborated in the discussion section.

Polyakov and Nearing also suggested that this mechanism might be an explanation for their results, but proposed that they could also be due to changes in friction following deposition, protection of the bed by moving bedload and differences in the energy required to start entrainment compared to continuing transport, as discussed in relation to aeolian sediment transport (and indeed, see Ellison [1947] on hillslopes). An alternative explanation of these patterns is the apparent emergence of multiple putative transport capacities as discussed in the
section on bedload in rivers, and it is impossible to evaluate the appropriate explanation for
hillslopes with the data available.

Equation 32 is derived from experimental work by Govers [1985, 1990, 1992]. However, in reporting this work, Govers noted that no equation that derives from the fluvial literature performed well over the full range of conditions that he tested, and that significant gaps in the empirical base remained. Of note, is the fact that Govers found the exponent \( \eta \) in equation 26 to be positive, a result also observed for interrill flow by Everaert [1991] implying that if critical stream power is fixed, then transport capacity increases with particle size, which is not intuitively obvious (and contradicts Gilbert [1877] as well as the later fluvial literature discussed above). Again, although further work is required to evaluate whether equations 3 or 5 could be used to overcome the limitations identified by Govers, there are limitations in relation to relative depth of flow that suggests such work would be fruitless (see discussion section).

The significance of particle size for transport capacity is addressed poorly for rillflow. Julien [1987] suggested that estimates of transport capacity will be highly sensitive to particle size. The problem can be well expressed in a statement taken from Govers [1990: 45]: “Flow incision … will only occur when the transport capacity of the flow is sufficiently high to evacuate all the material that is transported into the flow path from the interrill areas”. Since detachment in interrill areas is highly size selective and largely controlled by rainfall characteristics, its relevance to flow incision must be highly specific to particular soil characteristics. Ferro [1998] reinterpreted the grain-size dependence in Govers’ and others’ data by means of the Shields parameter and went on to suggest that the value should also be a function of flow depth relative to bed particle size (see discussion on relative roughness below). Beuselinck et al. [1999, 2002] suggested that the grain-size dependency increased above a threshold value and was not clearly distinguishable at lower flows. More recently, Zhang et al. [2011] estimated that transport capacity was proportional to \( d_{50}^{-0.345} \) for a series of laboratory experiments on steep slopes (8.7-42.3%) but without the presence of rainfall.

Sediment-Transport Capacity in Débris Flows

The transport-capacity concept as developed for fluvial applications is not stated explicitly in most débris-flow studies. The term “transport capacity” is often used without clear reference to established definitions or seminal papers in other literatures, and is sometimes used to
identify the overall volume of sediment (magnitude) of débris flows. The first explicit mention
of “sediment-transport capacity” in the case of a débris flow is given by Rickenmann [1990, 1991]. He studied hyperconcentrated flows and débris flows using a recirculating system in a steep flume and explored the effect of an increasing fluid density and viscosity on the flow behaviour and the sediment-transport capacity. Even if hyperconcentrated flows are usually considered two-phase, non-Newtonian flows, using a conventional (Newtonian) approach he showed that an increased fine sediment concentration and flow density produced an increased coarse transport rate. However, at the highest suspended sediment concentrations, a decrease in the bedload-transport rates was observed and related to macroviscous effects. In a recent paper, Shu and Fei [2008: 973] explicitly state that the sediment-transport capacity for a hyperconcentrated flow is “the total amount of sediments that water flow can carry”, and developed a formula for sediment-transport capacity of hyperconcentrated flow, in part based on Bagnold’s [1966] sediment-transport efficiency model for suspended transport. Based on the equilibrium equation of turbulent kinetic energy for solid and liquid two-phase flow they obtained a structural formula of sediment-transport capacity, which includes the flow viscosity and flowing resistance coefficient for hyperconcentrated flows:

\[ S_{v*} = p \left[ \frac{f(\mu_r)}{\kappa^2} \left( \frac{f_m}{\phi} \right)^{1.5} \frac{\gamma_m}{\gamma_s - \gamma_m} \frac{U^3}{g R v_1} \right]^N \]  

where \( S_{v*} \) is the sediment concentration at capacity [m$^3$ m$^{-3}$]; 
\( p \) is an empirically determined coefficient [-]; 
\( \mu_r \) is fluid viscosity [Pa s]; 
\( f_m \) is a coefficient of resistance [-]; 
\( \gamma_m \) is the specific gravity of the flow [kg m$^{-1}$ s$^{-2}$]; 
\( \gamma_s \) is the specific gravity of the sediment [kg m$^{-1}$ s$^{-2}$]; 
\( U \) is vertical mean flow velocity [m s$^{-1}$]; 
\( R \) is hydraulic radius [m]; and 
\( N \) is an empirically defined coefficient [-].

Note that despite the theoretical derivation of this equation, it still contains two terms – \( p \) and \( N \) – and a function – \( f(\mu_r) \) – that need to be derived empirically, is heavily dependent on
analyses of standard fluvial flows and assumes that the specific gravity of the flow is unchanging as sediment transport increases.

It should be stressed that there are no widely accepted criteria that unequivocally differentiate between hyperconcentrated- and débris-flows (and by extension between hyperconcentrated and other water flows). Sediment concentration has been used to provide a rough distinction between the phenomena, and débris flows are usually considered to contain more than half their particles coarser than sand [Pierson and Costa, 1987]. However, débris flows are pulsating events, often anticipated by liquid surges and followed by mud surges, which lead to rapid variation of coarse sediment concentration, velocity and viscosity (e.g., Iverson [1997]; Marchi et al. [2002]). Generally, in natural streams débris flows tend to occur when slope is higher than 0.2 m m$^{-1}$ [Takahashi, 1991], but at these same slopes bedload and hyperconcentrated flows can occur as well. By analyzing débris-flow, hyperconcentrated-flow and bedload-transport events that occurred in Switzerland in 2005, Rickenmann and Koschni [2010] found a smooth, increasing trend between processes, if the transported sediment volumes normalized by the effective runoff volume were plotted versus the channel slope. This smooth transition between transport processes seems also supported by comparative application of bedload- and débris-flow-derived formulae applied to calculate sediment concentration as a function of channel slope [Rickenmann, 2012]. Prancevic et al. [2014] suggest that this continuum can be understood in terms of a continuum of behaviours as Shields stress and slope angle vary, although more empirical work is required to develop their initial, experimental results.

The focus on débris-flow volume is well justified because its quantification is of primary importance for hazard mapping. In field studies débris flow volume can be roughly estimated from post-event field survey or accurately quantified using LiDAR data [Scheidl et al., 2008]. Débris-flow volume is required for developing empirical relationships between planimetric deposition area and débris-flow volume and ultimately for relating geomorphological characteristics of fans to potential runout areas [Iverson et al., 1998; Berti and Simoni, 2007, Scheidl and Rickenmann, 2010]. Furthermore, the volume is needed for assessing precisely the débris-flow volumetric sediment concentration ($C_V$; volume of sediment divided by volume of water and sediment), which is required as input parameter in widely used débris-flow- and hyperconcentrated-flow-routing models (e.g., Hsu et al. [2010]). Because in practical applications débris-flow volume is generally the unknown value, the volumetric sediment concentration needs to be estimated as the ratio between the equilibrium concentration ($C_D$)
and the volume concentration of solid fraction on the bed \((C^*)\), for which the porosity of sediment in the bed can be used as a proxy. Expressed in terms of discharges, the débris-flow discharge at equilibrium conditions \((Q_{DF} \text{ [m}^3\text{ s}^{-1}])\) could be calculated as:

\[
Q_{DF} = Q \frac{C^*}{C^*-C_D}
\]

where \(C_D\) is the equilibrium concentration \([\text{kg kg}^{-1}]\) and \(C^*\) is the volume concentration of solid fraction on the bed \([\text{kg kg}^{-1}]\).

The equilibrium concentration \((C_D)\) depends on the approach used to describe the rheological nature of the débris flow. It has been defined by Takahashi [1978, 1980], who suggested that the nature of débris flows is dominated by the collision between the coarse particles rather than the interstitial fluid. Takahashi’s inertial-type approach is based on Bagnold’s [1954] dispersive régime and allows the velocity profile to be calculated, as well as the equilibrium grain concentration and the depth of a quasi-steady-state débris flow. Takahashi defines the equilibrium concentration as the concentration of a débris flow that produces neither erosion of the bed nor deposition onto the bed. Assuming a uniform distribution of sediments over the whole depth, the equilibrium concentration \((C_D)\) is ultimately a function of the bed slope, and can be calculated as:

\[
C_D = \frac{\rho_f S}{(\rho_s-\rho_f)(\tan \phi - S)}
\]

in which \(\phi\) is the collision angle of grains which, according to Bagnold [1966], can be approximated with the internal friction angle of the sediments.

Even if in the model of Takahashi [1978] the rôle of interstitial fluid can be neglected, later studies suggested that the macroviscous flow régime should be taken into consideration (e.g., Davies [1988b]) and that cohesion and variable concentration and pressure over the flow depth should also be considered [Chen, 1988]. Attempts have been made to investigate intergranular stresses using a Coulomb mixture model, which accounts explicitly for stresses and interactions of distinct solid and fluid constituents and eliminates the need to specify rheologies of complex, multiphase moisture [Iverson and Denlinger, 1987]. Furthermore, from a series of experiments where liquid and solids were recirculated over a mobile bed, Armanini et al. [2008] showed an unstable stratification of collisional and frictional régimes across the flow depth, with implications for grain concentration and thus transport capacity. As to the
calculation of equilibrium sediment concentrations, another alternative and promising theoretical approach based on the maximum entropy principle has been proposed by Lien and Tsai [2003]. Overall, many investigators have modelled débris flows specifying different rheological rules that govern flow behaviour, and each model determines the sediment concentration at equilibrium.

In general, débris-flow routing can be simulated solving a momentum-conservation equation for the mixture of solids and fluid, mass-conservation equations for the liquid phase and solid phase, and an equation for bed-elevation change. Irrespective of the rheological approach to define the débris flow, which has to be reflected in the momentum-conservation equation [Takahashi, 2009], the débris-flow volume increases with bed erosion or decreases with deposition as calculated by the mass-conservation equations. Takahashi et al. [1987] implemented formulae predicting erosion or deposition depending on the deficit or excess of sediment concentration in the débris flow with respect to the equilibrium volume as calculated by:

$$\frac{E}{v_D} = \alpha_D \frac{C_B - C}{C^* - C_B} \frac{h}{d} \quad (E \geq 0)$$

$$\frac{E}{v_D} = \beta_D \frac{(C_B - C)}{d} \frac{h}{d} \quad (E < 0)$$  \hspace{1cm} (36)

in which $E$ is the erosion (positive)/deposition (negative values) rate [m s$^{-1}$], $v_D$ is the average velocity of the débris flow [m s$^{-1}$] and $\alpha_D$ and $\beta_D$ are experimental coefficients. Although Takahashi et al. [1987] describe some of their approach as having been derived by analogy with the bedload literature, they do not cite any of it explicitly. In the same way as the derivation of equation 35 is from theory (as with the analogous equation 27 developed for soil erosion by Foster and Meyer [1972], but which is not directly cited by Takahashi et al. [1987]), there seems to have been little empirical testing of the model, especially its dependence on the definition of empirical coefficients.

Subsequent numerical models added to this approach, introducing for instance the effects of bed sediment size on bed-erosion rate [Egashira et al., 2001]. As shown by Papa et al. [2004], the vertical erosion rate is inversely proportional to the grain size of the channel bed. Also, the critical condition for the entrainment of a grain from the bed depends on the difference between
the total shear stress and the yield stress acting on the bed surface, and depends on sediment concentration of the débris flow. Despite these flume-based insights on drag forces needed to entrain sediments from the channel bed, sediment are also destabilized by macro-scale processes due to undrained loading, impact loading, liquefaction, stream-bank undercutting [Hungr et al., 2005], and very little is known about the shear strength of bed sediments, bed stratigraphy, and pore pressure due to saturation. To overcome these mechanistic limitations, the magnitude of a débris-flow is often predicted using empirical algorithms based on the geometry of the routing colluvial channel (e.g., Ikeya [1981]; Hungr et al. [1984]) or the size of landslides-prone areas upstream. An alternative approach, which also allows the association of a frequency to the magnitude of the expected event, is based on the analysis of rainfall thresholds able to trigger débris flows [Jakob and Weatherly, 2003].

It should be noted that all numerical and flume-based studies on débris flows, and most of field empirical observations assume unlimited sediment supply and thus equilibrium concentration (and thus a steady-state approximation to capacity analogous to that used in some of the fluvial literature) is always supposed to be reached. This assumption appears reasonable in risk-based studies because hazard mapping or engineering countermeasures are better based on worst-case scenarios. However, even if such powerful events are likely to convey as much sediment as they can, relatively few geomorphological studies have questioned the supply-unlimited assumption. Bovis and Jakob [1999] stated that there are weathering-limited basins that require a certain “recharge period” prior to each débris-flow event and exhibit a lower frequency of débris flow activity if compared to transport-limited basins, where the frequency and magnitude of débris-flow events are controlled primarily by hydroclimatic triggering (see also Carson and Kirkby [1972]). In many cases, the interval from the occurrence of the last débris flow can help estimating the magnitude of the next event [Jackob et al., 2005], if activations of new sediment sources at the basin scale due to extreme events are excluded (e.g., widespread landslides). The recharge of sediments occurs mainly along the débris-flow transit channel, where sediment storage and availability can be inferred by cycles of aggradation and degradation [Jackob et al., 2005 Fuller and Marden, 2010; Berger et al., 2011].

**Sediment-Transport Capacity of Glaciers**

The concept of transport capacity has not traditionally been prominent in glacial geomorphology. Indeed, the term has been so poorly adopted by the glaciological community that it has sometimes been used in quite different ways. For example Hagen [1987] uses
“transport capacity” to describe the potential flux of ice through a surging glacier system, focusing on the glacier as a transporter of ice, rather than as a transporter of sediment.

Nevertheless, Boulton [1975] arguing that the lack of a theoretical framework for sediment transport by glaciers was inhibiting progress in the field, set about providing such a framework for temperate glaciers. Arguing from Glen [1952] and Weertman [1957, 1964], he derived an expression for the “transporting power” of a glacier as:

\[
P_t = \left( \frac{V_p' V_i}{(V_i - V_p')} \right) \left[ \frac{(V_i)}{A \rho R^4 \mu} \right]^{0.5} N \mu'
\]

where \( V_p' \) is the mean velocity of the débris entrained by the glacier [m a\(^{-1}\)], 
\( N \) = effective pressure [N] 
\( V_i \) = basal ice velocity [m s\(^{-1}\)] 
\( R \) = bed roughness [-] 
\( \mu' \) = average coefficient of friction [-].

The coefficient A is given as:

\[
A = \left( \frac{KB}{L \rho_i} \right)^{1/2}
\]

where \( L \) = latent heat of fusion of ice [J g\(^{-1}\)] 
\( K \) = thermal conductivity of ice [W m\(^{-1}\) K\(^{-1}\)] 
\( \rho_i \) = ice density [kg m\(^{-3}\)] 
\( B \) = the constant in Glen’s flow law [-], 
\( C \) = a constant relating pressure and resultant lowering of the freezing point [-].

However, this kind of work has more often focussed purely on entrainment (or deposition) mechanisms and has rarely given explicit consideration to sediment transport capacity. Iverson [2000] calculated the rate of infiltration regelation into subglacial sediments. For débris-rich ice formed by regelation around obstacles to flow, Lliboutry [1993] calculated the likely thickness of a steady-state layer of regelation ice. Calculations of this type could be put to the task of estimating some aspect of transport capacity, but this is not a direction that
has been taken to any significant degree. Subsequent work on sediment-transport mechanisms by ice (e.g., Melanson et al. [2008]) is not founded on concepts of transport capacity or power.

One review that does explicitly consider sediment-transport capacities of the glacial system is provided by Alley et al. [1997]. Their focus was on how glaciers entrain and transport basal sediment, and they focus exclusively on the transport capacities of various components of the basal environment: subglacial streams, débris-rich basal ice and the mobile subglacial débris layer. Although it does provide an illustration of some of the issues facing studies of glacier transport capacity this approach excludes major areas of the whole-glacier sediment transport system such as the supraglacial and englacial débris transport, which can be very important in many glacier settings. It also raises the problem of defining what we mean by glacier sediment transport: is the deforming subglacial layer part of the glacier? Is a meltwater stream within the ice part of the glacier? What about rock glaciers?

In the analysis by Alley et al. [1997] the concept of an actual “capacity” is focussed mainly on the water sections of the system. In other parts of the basal system (basal ice and subglacial débris) the focus is on entrainment capacity or potential entrainment rates. Sediment transport by glacial water streams involves essentially the same issues as fluvial sediment transport, except that subglacial environments have special conditions including sections under pressure and highly variable discharge (Figure 11). Alley et al. define transport capacity as “volume of sediment transported per time by the system, for a given distribution of grain sizes if the supply of sediment is not limiting.” [1997: 1018]. They consider a range of previous studies of bedload transport including the assessment by Gomez and Church [1989] that Bagnold’s [1980] equation best predicted bedload-transport rate. Alley et al. [1997] also consider pipe flow, which is relevant for flow in subglacial tunnels. In glacier systems typically steep gradients with high headwater pressures and high but variable discharges, including high frequency discharge pulses, make glacial meltwater streams particularly effective transporting agents. Subglacial water flow and sediment processes are very different if surface meltwater can reach the bed, so predicting the capacity of the system would depend on knowing about the connectivity of surface, englacial and basal drainage networks. This connectivity is likely to be variable over time and will evolve on a seasonal basis. It will also vary between glaciers. Nevertheless, Alley et al. [1997] do review a substantial set of potentially useful approaches to identifying at least some notion of a transport capacity for the fluvial components of the glacial
system. However, because glaciers involve four-phase flow (water, ice, air, sediment), it is unlikely that any simple approach to estimating transport rate or capacity could ever be achieved.

For other components of the basal sediment system, however, Alley et al. [1997], and others before and since, have greater difficulty even approaching a theoretical prediction of sediment fluxes. For example, on the question of predicting sediment transport through the deforming subglacial layer Alley et al. [1997: 1021] suggest that “without an accurate flow law including controls on viscosity, one cannot calculate debris fluxes from first principles.” On predicting sediment transport by glaciotectonic mechanisms they write: “Despite the fact the glaciotectonic structures are common in glaciated regions… no general theory has been advanced to account for the amount of debris transported in front of or beneath glaciers in this way.” As for transport of débris within basal ice, there is still no clear agreement even on the processes by which débris can be entrained, let alone any quantitative estimation of the potential débris flux through the basal layer. Several studies (e.g., Hunter et al. [1996]; Knight et al. [2002]) have measured actual débris flux through the basal layer by combining measurements of débris content and ice velocity, but these studies have not extended to identifying theoretical limits. The aim of these studies has been to explore how much débris glaciers do carry, not how much they can carry.

One reason for ignoring the concept of capacity might be that glaciers are immensely complicated systems for which we still lack an understanding of many fundamental processes. There is still neither a clear flow law for glaciers [Alley, 1992], nor a comprehensive understanding of the mechanisms of glacial débris entrainment and deposition [Evenson and Clinch, 1987; Alley et al., 1997]. Both would be required for an analysis of glacier transport capacity, which would need to address many variables beyond a simple fluid mechanics of ice-débris interaction. Evenson and Clinch [1987: 112] suggested that “glacier margins are probably the single most complicated clastic sedimentation system and that there are [sic] a bewildering array of potential sediment transport paths within any glacier system”. Alley et al. [1997: 1030] argued that “The complexity of glacial systems almost entirely precludes succinct rules describing erosion and sedimentation”. Sediment transport in glaciers is complicated because there are so many different transport paths within the glacier system (supraglacial,
englacial, basal, subglacial, fluvioglacial), so many débris sources, so many mechanisms of
débris transport, such a huge range of particle sizes and so many processes by which débris can
be entrained and released, few of which relate to the fluid-flow properties of the ice. In many
of these areas there is still no agreement even about the fundamental operation of processes.
For example, the significance of subglacial glaciohydraulic supercooling as a mechanism for
entraining substantial amounts of sediment is still being evaluated, and the details of the
environmental factors that control the mechanism are still being explored [Alley et al., 1998;
Cook et al., 2009; Creyts et al., 2013]. Therefore it would not yet be straightforward to
incorporate this process, which may be extremely important, into any working model of
entrainment potential, let alone one of transport capacity. We do not yet have all the parts to
put together into a comprehensive model of glacier-sediment transport. Arguably transport
capacity is not prominent in glacial geomorphology because the science is not yet sufficiently
advanced to address the problem.

The questions that arise in glaciology and glacial geomorphology surrounding débris
transport do not centre around the capacity of the system. There are many other issues being
considered, and sediment plays an important role in glacier dynamics, but transport capacity is
not often close to the forefront of thinking within the discipline. Nevertheless, glacial studies
have considered factors similar to those considered by geomorphologists looking at fluvial and
aeolian transport capacity. Ice velocity, basal shear stress, effective normal pressure and the
susceptibility of basal material to the tractive force applied by the movement of the overlying
material are central to many aspects of glacier sediment dynamics and have been treated in
detail by glaciologists and glacial sedimentologists. In particular, studies of how subglacial
sediments may be mobilised by stress imparted from the ice above come closest to linking ice
flow with sediment transport in a way similar to fluvial or aeolian studies.

Discussion
Transport capacity as a concept originally developed in fluvial systems and has received the
greatest and most enduring attention in this process domain. Gilbert introduced the term in his
study that aimed to address “the common needs of physiographic geology and hydraulic
engineering” [Gilbert, 1914: 9], and was directly influenced by engineers from France and
Germany. The need to underpin engineering decisions runs across a range of aspects of the development of the concept in fluvial, aeolian, coastal and hillslope geomorphology, and is reflected by common practice in all of these areas (Table I), even if uncertainty in results (see Figure 1) more often than not leads to calibration and uncertainty of how the concept does actually justify any of the results obtained. The idea that there should be a single value of sediment transport that corresponds to a particular discharge also underpins the régime theory developed by engineers to design stable channels over a similar time period to the original work on transport capacity (e.g., Kennedy [1895]; Lacey [1930]). However, it is intriguing to note that there seems not to have been a direct link between these literatures, at least until much later when the work on régime theory started to inform broader concerns of river management and restoration [White et al., 1982; Hey and Thorne, 1986]. This parallel development is also reflected in the apparent parallel inventions in the different sub-fields of geomorphology, although authors were often quick to point out the linkages (e.g., Einstein [1941] and his discussants [1942] recognized the link between the work on fluvial bedload and the work of Bagnold [1936] in the aeolian realm, even if the recognition was not mutual until much later).

Although many of these developments have been related to positivist and deterministic approaches relating to prediction, there is also an ongoing recognition that there is a need to engage with aspects of variability. For example, Einstein [1941:561] noted the need to base these predictions on the “new theories of turbulence” of Shields [1936] and Rouse [1939]. However, this need to address ongoing developments in the field of turbulence has not always been heeded (see discussion below). It also fails to recognize the conceptual limitations that may mean that there is no single concept of transport capacity that is applicable across all areas of the discipline, nor indeed to individual process domains under all conditions. Consequently, there are issues both in terms of whether any predictions based on the concept have a real, mechanistic understanding, and also in terms of how concepts of sediment transport are communicated between different parts of the discipline, or to those working in related areas in interdisciplinary projects.

Although the development of these different process domains can be considered as developments within geomorphology, they were principally carried out by engineers (Einstein was a hydraulic engineer; Meyer and Wischmeier were agronomic engineers; Caldwell worked for the Beach Erosion Board (now morphed into the Coastal Engineering Research Centre run by the US Army Corps of Engineers); Gilbert, although having a geological background, saw his research as having engineering applications; and Bagnold was something of an anomaly,
working at the time as an independent researcher [Bagnold, 1990], although his results were quickly put to practical application). The shift to the implementation of these engineering-based approaches in geomorphology came with the development of quantitative geomorphology in the 1950s and broader notions of the need for quantification and prediction. One of the main proponents of this development, Strahler, was clearly aware of Gilbert’s definition of transport capacity and the work on which it was based:

Had Gilbert’s philosophy of physical geology prevailed among students of landforms the analysis of slopes would not have been so long delayed. … An inevitable result … has been a gradual reduction in the physical science background of geomorphology students and teachers; and a consequent general sterility in original geomorphic research. But while academic geomorphology has been approaching stagnation important developments in the understanding of slope erosion processes have been made by hydrologists, hydraulic engineers, and soil erosion specialists concentrating upon soil conservation and sedimentation engineering

Strahler [1950: 210]

Once the idea made it into the discipline, further developments of the concept within the fluvial and (to a lesser extent) aeolian contexts went on to inform approaches within the coastal domain, soil erosion, and in débris flows and glaciers to a much more limited extent. As the transfer from one area to another occurred, there seems to have been restricted questioning of the core concept, and in many settings (especially in fluvial and slope studies) it seems to have become “black boxed” [Latour, 1987] and immune from critical evaluation. For example, Wainwright et al. [2008] discuss how the soil-erosion literature has avoided a critical evaluation of not only transport capacity but other issues such as the mode of sediment transport on hillslopes.

That there seem to have been multiple, independent inventions of the concept is not unusual (e.g., as with evolution, calculus and the periodic table). Indeed, there is a close parallel in multiple inventions of the notion of carrying capacity by game and range managers, population biologists, ecologists and demographers [Sayre, 2008]. All have their origins in the broader concept of engineering capacity (and what came to be known as payload of ships), and relate to idealized, static conditions. Sayre [2008] pointed out that all developed to relate to the control of environmental systems, but have all become difficult to sustain when issues of
scale, variability and system dynamics were addressed. These aspects will be considered further in relation to transport capacity below, not least because of the parallel in the idea of control based on the engineering underpinnings of the term.

In fluvial geomorphology, transport capacity is still very central despite the critiques discussed here in relation to turbulent processes. The importance of the concept has largely remained because of two reasons. First, it is has proven useful in river restoration, channelization and management for estimating the channel geometry required to transport sediment in design flows (e.g., White et al. [1982]). If the capacity estimate gives an overestimate of the maximum transport rate, although the engineering of structures in the channel will be more expensive, the structures would be secure beyond the design criterion. Secondly, the transport equations that have evolved due to the concept utilize river flow and channel variables that are commonly monitored. However, what is easily measured is not necessarily what is required in terms of the parameters that actually control sediment transport resulting in what are effectively empirical equations not rooted in physical understanding. The usefulness of the approach is thus more about being able to justify practical applications with recourse to an extended heritage of literature, rather than with a clear demonstration of understanding of the process, as would be required by a realist approach.

In the case of aeolian geomorphology, the decline of the concept of transport capacity may, in part, be related to the typical condition in the field of supply limitation. It was recognized early on in field work that the sediment transport in locations where the source material was not just composed of sand was not “at capacity” or “saturated” because so often very little sediment was observed to be moving [Gillette et al., 1980; Gillette et al., 1997; Gillette and Chen, 2001]. As information about turbulence filtered into the area from fluvial research (e.g., Bennett and Best [1995, 1996]; Jackson [1976]), aspects of turbulence controls on observed transport meant that spatial and temporal variability suggested there was not a single capacity value that had a meaning (e.g., Baas and Sherman [2005]).

In coastal studies the concept of transport capacity has been implicitly subsumed into the notion of longshore transport rate, which is still an important element in understanding beach development. It is applied more at an annual to decadal time scale whereby the general field state of a beach system might be gauged in terms of sediment input and transport potential out. However the research emphasis has now switched to the self-regulation of beach form by tuning incoming incident waves and secondary currents post wave breaking to short term
movement of sediment into distinctive beach morphologies where both cross-beach and
longshore beach transport may be co-dominant or negligible. This emphasis gave rise to
morphodynamics [Short, 1999b] as the principal mode of beach analysis over the last two to
decades. Transport equations are still considered, but analysis is now more dependent on
the detailed spectral breakdown of processes correlated with detailed empirical sediment
transport from acoustic sampling (sand grades). This means that detail of sediment transport as
a result of power potential is often sidestepped. Green and Coco (2014) have argued that in
the coastal process domain (as well as argued above in the aeolian), the reason is commonly to
do with supply limitation.

In hillslope studies, the concept is only just being re-evaluated. Wainwright et al. [2008]
pointed out the limitations of extrapolation of techniques from the fluvial literature to hillslopes
that were untested because of the difficulties of measuring the rapidly changing conditions in
very shallow flows, and because a number of the fundamental assumptions are not met in those
flows. There has been some debate about these arguments (e.g., Smith et al. [2010];
Wainwright et al. [2010]), but the community seems largely happier to calibrate existing
models than to address the fundamental basis of why that calibration is necessary (as indeed is
the case in fluvial examples, e.g., Xia et al. [2013]). Use of the concept is also starting to fall
out of favour with coastal geomorphologists because of the amount of calibration required to
produce predictions at any one location. As noted above, apart from some fluvially inspired
dabbling, neither glacial or débris-flow geomorphologists have been much influenced by the
concept, not least because of issues of complexity and variability.

The foregoing review demonstrates that much of the literature rests on the assumption
that a specific, unchanging capacity to transport sediment exists. To what extent is such an
assumption reasonable? Where the concept has been developed (and redeveloped) across the
different areas of geomorphology, there has been a convergence to the idea that the power of
the transporting medium uniquely allows the prediction of the amount of sediment that will be
moved. Presumably, the idea of a single power term was developed to avoid complications
from other sources of flow variability, but as has been seen, this simplification has proved
problematic. If that conceptualization held, then it would provide a powerful tool in the
management of environmental systems, or in the prediction of landscape evolution.
Notwithstanding the lack of consensus on how that power should be estimated, there are a
number of conceptual issues that call into question whether such a simple definition exists, and
therefore if there are multiple definitions, whether the core concept remains testable as opposed to the evaluation of ancillary hypotheses that simply protect that core from critique. These conceptual issues are: (i) whether or not the power of a flow can be characterized independently of the sediment transported by that flow and whether the power represents adequately the ways in which turbulent structures in different process domains control sediment transport; (ii) the complicated nature of transport systems in any geomorphic domain; and (iii) whether or not there is a scale independence of measured rates of transport.

Perhaps most importantly, the nature of flow is not independent of the sediment transported by that flow. As the flow starts to take on more of a two-phase nature (or even three-phase when air is rapidly trapped in water-sediment mixes for example in flood waves in ephemeral channels, on beaches or in débris flows, or four-phase flows when ice is included in glacial systems), the density and viscosity change so that the underlying assumptions of the equations used to make predictions diverge from the conditions in the field. In the fluvial and hillslope domains, these changes ultimately lead to the production of hyperconcentrated and then ultimately débris flows, which demonstrate non-Newtonian behaviour. Although this point is also recognized by Hessel and Jetten [2007], they overlook the practical implication that it means that it is conceptually impossible to parameterize a model based on a single capacity value. In aeolian-dominated systems, it has long been recognized that the initial movement of sediment by the air on its own requires higher rates of wind shear than initial movement when the air contains sand grains that impact the surface during rebound in saltation (called the fluid and impact thresholds, respectively) [Bagnold, 1941: 92-95]. The Owen effect also describes how the saltating sediment affects the effective roughness length of the surface, thus changing the transport behaviour of the system [Owen, 1964]. It is usually argued that this process is insignificant in the fluvial domain because of the lower difference between the density of the sediment and the fluid [Bagnold, 1973: 484], and is of course irrelevant by the time the flow has become non-Newtonian. Long et al. [2014] have shown that in rainsplash there can be an important effect of ballistic impact mobilizing grains as initially splashed particles impact on the surface. Only as techniques develop to look at the dynamics of multi-phase flow in more detail will it be possible to evaluate to what extent these secondary effects of particle-initiated entrainment are more than noise.

A further issue is the complicatedness of transporting systems. As already noted, this complicatedness means that the concept of transport capacity has provided limited explanatory power in glacial geomorphology, or indeed on hillslopes. The variability in processes in other
domains may also be an ultimate reason for the limitation. For example, on hillslopes, the
variation between raindrop detachment and splash, flow detachment and flow transport
introduces a wide range of spatial heterogeneity in observed transport [Parsons et al., 2004;
2006] so that spatial heterogeneity of surface and subsurface properties will introduce
significant variability in transport. Even if it were possible to make a single prediction of
“capacity” it might be so difficult to account for the underlying stochastic nature of surface
properties to render the practical application useless (e.g., because the range of potential values
is extremely broad, as suggested by the envelope curves in Figure 4), or as discussed above
that it would produce a significant overestimate leading to increased engineering costs.
Furthermore, the nature of transport also varies from creeping to rolling to saltation to
suspension, which will affect both the ability of the flow to transfer energy to particles and the
effect of particle-particle collisions in affecting the energy of the flow. The thresholds between
these different forms of transport are not necessarily clear [Parsons et al., 2015], further calling
into question the predictability of the system. Relations derived in one domain of transport
cease to hold in others in part because of the variability in mechanisms and the difficulty of
evaluating which mechanisms are in operation at which points in time and space. In coastal
studies, the complexity occurs as a result of multidirectional flows whether offshore under
oscillating waves, or onshore where swash and backwash are at odds with each other.
Consequently, transport power is often used as an explanatory variable in beach transport (i.e.,
relating to the potential power of a breaking wave), but the transformation post-breaking on
the beach face becomes a more complex issue that means for most geomorphologists, it is
easier to talk about macro-scale or bulk changes rather than the reductionist statements about
instantaneous power. Indeed, it is often stated that it is the hiatuses in flow that cause the main
interpretation problem for modelling long term coastal response. Again, it is not necessarily
the most straightforwardly measured variables that are the most significant in explaining
patterns of sediment transport.

If based on the assumption that a specific, unchanging capacity to transport sediment,
there is a further fundamental problem with the concept of transport capacity, in that in order
for it to hold, a flow must exhibit the same transport capacity at different temporal and spatial
scales. In other words, transport capacity must be independent of the scale of measurement, or
there needs to be an explicit representation of spatial and temporal variability in the prediction,
above and beyond the flow-power terms. Evidence from studies of sediment transport on
hillslopes, within rivers, by wind and along coasts suggest otherwise. This evidence can broadly be considered in terms of the temporal and the spatial variability of transport.

Sediment transport is a stochastic process whereas the transport-capacity concept is essentially deterministic. Especially for low concentrations, sediment transport is a granular phenomenon [Cooper et al., 2012] so understanding it has to be at the grain scale (see also Furbish et al. [2012]). Because of stochasticity, transport capacity varies with temporal scale and can only reflect certain time-averaged (steady-state) conditions. The evidence for the dependency of transport rate on temporal scale is as follows.

First, fluctuations in sediment transport under quasi-steady flow conditions occur commonly in rivers, hillslopes and aeolian transport. These sediment pulses occur over a range of temporal scales: (i) turbulence time scales due to the advection and propagation of turbulent flow structures (e.g., Drake et al. [1988]; Radice et al. [2013] for rivers and Baas and Sherman [2005]; Durán et al. [2011], for aeolian transport); (ii) time scales of tens to hundreds of seconds due to wind gusting [Baas, 2004; Butterfield, 1998], variability in wave power, migration of bedforms (e.g., Cudden and Hoey [2003]; Isey and Ikeda [1987]; Whiting et al. [1988] for rivers and; Andreotti et al., [2002a, b; 2010]; Sauermann et al. [2001] for aeolian transport; Kosov et al. [1978 cited in Sidorchuk, 1999] for hillslopes) or changes in the size and structure of river bed material (e.g., Gomez [1983]; Pender et al. [2001]); and (iii) at hourly scales in rivers due to processes that scale with the width of the channel, such as bar migration [Gomez et al., 1989], bank erosion [Cudden and Hoey, 2003], scour-fill sequences and changes in sediment supply (e.g., Gilbert [1917]; Knighton [1989]). Similarly, coastal sediment transport is now seen as being dominated by the concepts of secondary generation of periodic longshore and onshore currents generated by variable amplification of the shoaling incident wave spectrum [Huntley et al., 1977; Short, 1999b; Aagard et al., 2013]. Infragravity wave energy – i.e. those of a lower frequency than gravity waves – as well as incident wave energy is variably experienced on a beach face, depending on beach face reflective-dissipative status (often now indexed by the surf-similarity parameter [Battjes 1974], which is a function of the overall incident wave steepness (wave height over wave length) relative to overall shoaling shoreface slope [Huntley et al., 1977; Wright et al., 1979; Hughes et al., 2014]. Consequently, transport equations of the form proposed by Caldwell are only ever broad time-averaged at best, as well as being poor indicators of overall transport variability.
Secondly, these fluctuations in sediment transport imply that estimates of transport capacity are dependent upon sampling duration. For example, if particle movement is observed over a longer time frame, one would expect a higher likelihood of measuring large transport distances because more particles are transported. A particle that has travelled further may well be deposited in a more stable position than those that have travelled a shorter distance. For example in rivers, Ferguson et al. [2002] have demonstrated an apparent deceleration of tagged movement of fluvial gravel through a succession of floods, which may relate to the structure of bed material. Wainwright and Thornes [1991] and Parsons et al. [1993; 2010] make similar observations for particles moving on hillslopes following multiple storm events, as do Kirkby and Statham [1975; Statham, 1976] for rockfalls. Given that sediment flux is a product of the entrainment rate and transport distance, it follows that the sampling duration affects estimates of sediment flux, and thus estimates of the transport capacity of the flow [Bunte and Abt, 2005; Furbish et al., 2012]. In addition, sediment pulses that occur as the movement of individual grains exposes others to the flow (e.g., Cudden and Hoey [2003]; Drake et al. [1988]) are more likely to occur over longer sampling durations.

Thirdly, estimates of transport rate are dependent upon the frequency at which transport is sampled. If one were to observe the movement of transported particles at a temporal frequency that was comparable to the frequency of turbulent flow structures, single entrainment events would be observed. Thus, entrainment rate and transport distance, and therefore their product, which gives sediment flux [Parsons et al., 2004], would be estimated directly. If particle motion were sampled infrequently (e.g., before and after a single flow, storm or wind event), it would not be possible to determine whether the measured movement is due to single or multiple transport events. Thus, the entrainment rate could not be measured, and the sediment flux would be estimated indirectly by the virtual velocity (ratio of distance travelled to sampling duration), as for example in equation 1. Thus the two estimates would likely give different values for the transport capacity of the flow.

Fourthly, sediment transport is often assumed to adapt to local conditions instantaneously, and accordingly sediment-transport rate is modelled as a capacity value based on some local flow condition. The fluvial literature has clearly demonstrated phase differences between flow and sediment-transport rate for both suspended and bedload transport (e.g., Alexandrov et al. [2003]; Kuhnle [1992]; Laronne and Reid [1993]; Lee et al. [2004]; Marcus [1989]; Reid et al. [1985]; Sutter et al. [2001]; Williams [1989]). The deviation between bedload-transport rate and capacity increases with time lag, and the time lag is longer for higher
transport rates [Cao et al., 2010]. As highlighted earlier, a similar situation also exists with wind-blown transport [Butterfield, 1991; Hardisty, 1993; Spies et al., 2000], in which the time lag is greater with an increase in wind speed than a decrease. Thus, suspended and bedload transport in fluvial and aeolian environments cannot adapt to local flow conditions faster than changes in flow due to turbulence [Cao et al., 2007; Cao et al., 2010; Stout and Zobeck, 1997] and hence the transport-capacity concept cannot provide a mechanistic understanding of sediment transport at turbulence time scales, nor at scales smaller than the time lag.

Fifthly, transport rates change with the duration of the flow event. As a surface erodes a grain is likely to come to rest in a more sheltered position than its original location so a higher shear stress will be required to remobilize the grain, increasing its subsequent resting duration and travel distance, and virtual velocity. In rivers, transport rates fall to negligible levels, even during steady flows, due to bed armouring (e.g., Pender et al. [2001]). Hairsine and Rose [1992a, 1992b] suggest that the probability of detachment on hillslopes is dependent upon whether the particle resides in the deposited layer or in the unshielded original soil layer. Furthermore in rivers, flow history also influences transport capacity. Both sub-threshold flows (e.g., Paphitis and Collins [2005]) and above-threshold flows [Hassan et al., 2006; Mao, 2012; Reid et al., 1985] have been shown to increase detachment thresholds for sands and gravels over a succession of flow events.

In terms of spatial dependency, transport rates are dependent upon the spatial distribution of fluid shear stress and critical shear stress. Thus transport capacity is dependent spatially. This dependency can be demonstrated in the following ways:

First, spatial variations in transport rates occur at a range of scales; from the grain-scale due to differences in the surface microtopography, such as packing and grain exposure (e.g., Darboux and Huang [2005]; Drake et al. [1988]; Jackson et al. [2006]; Radice et al. [2009]; Wainwright and Thornes [1991]), at the bedform-scale (e.g., Baas and Sherman [2005]; Brayshaw [1984]; Church et al. [1998]; Farres [1987]; Gares et al. [1996]; Laronne and Carson [1976]; Richards and Clifford [1991]; Torri [1987]) and at the plot- (e.g., Vandaele and Poesen [1995]; Favis-Mortlock et al. [2000]) or reach-scale (e.g., Hooke [1975]; Dietrich and Smith [1984]) due to changes in morphology (Bauer et al. [1996]; Gillies et al. [2006]; Gillies and Lancaster [2013]). Thus estimates of transport capacity vary according to the spatial scale over which the measurements of transport are made. This dependency on sampling area can also be
nicely illustrated with the following evidence. Measurements of travel distances of individual particles during runoff events on hillslopes and of tagged gravel in rivers show that transport distances are small and have a heavy-tailed distribution [Hassan et al., 1991; Wainwright and Thornes, 1991; Parsons et al., 1993; Hill et al., 2010; Lajeunesse et al., 2010]. Thus, only the smallest eroded particles, or a fraction of larger ones, are likely to be transported large distances, after even a very large runoff or flow event. Hence, estimates of sediment flux, and therefore transport capacity, will vary with sampling area. The spatial dependency between roughness scale and aeolian transport can also result in a given wind condition not producing the same transport rate at different spatial scales. For equivalent roughness (as defined by roughness density), the size of the roughness elements dramatically affects the transport. For the same shear velocity one can get quite different flux rates as the saltating particles interact with the roughness [Gillies et al., 2006; Lancaster and Gillies, 2013]. Thus, although the shear velocity may be the same the flux may be very different due to the size of the roughness, suggesting that in cases where sediment supply is not limiting, there would either have to be two capacity states for the same shear velocity, or that the concept of capacity is not a mechanistic description of the process.

Secondly, transport-capacity equations are not scalable from one river to another, nor from one hillslope to another. Transport capacity is scaled by, amongst other things, bed shear velocity (or bed shear stress) and grain size. Two flows can have the same bed shear stress but occur over rills or channels with differing slopes, and therefore have differing relative submergence (flow depth: bed roughness size). To illustrate this difference, consider the following example. Take two river channels with a slope of 0.005 (-) and 0.02 (-), identical bed shear velocity of 0.1 m s⁻¹, and composed of the same material ($D_{50} = 0.02$ m). Given that $u_* = \sqrt{gh_w S}$, the flow depths $h_w$ for the two river channels are 0.20 m and 0.05 m, respectively, so the relative submergence is 10.2 and 2.5. This difference in submergence may be important when extrapolating the transport capacity relationships from one river to another, or one from hillslope to another, for the following reason. Within the fluvial literature there is strong evidence that the mean fluid shear stress at which sediment is entrained is inversely correlated to relative submergence (e.g., Bathurst et al. [1983, 1987]; Bettess [1984]; Shvidchenko and Pender [2000]; Mueller et al. [2005]; Parker et al. [2011]) because the structure of the near-bed flow changes with submergence [Ashida and Bayasit, 1973; Graf, 1991; Lamb et al., 2008; Cooper, 2012]. Wainwright and Thornes [1991] found a similar pattern for coarse particles on hillslopes. One further problem in trying to use the concept for different slopes is the concept
only applies strictly to steady and uniform flows. Therefore the degree to which the flow can transport its capacity depends on the steadiness and uniformity of the flow. As slope increases the flow is more likely to be non-uniform, and locally unsteady. Thus, these differences in flow submergence, uniformity and steadiness on hillslopes or within channels with differing slopes means that flows with the same mean bed shear stress will not have the same transport “capacity”.

Thirdly, transport-capacity equations do not scale across different process domains. If the concept is physically robust this scaling problem should not arise. For example, the extrapolation of transport-capacity equations from the fluvial literature to processes occurring on a hillslope, though commonplace, poses potential problems. The flow relative submergence under which transport occurs is usually higher for a river so the transport relationships developed for rivers may not scale to overland-flow conditions. For a given bed shear velocity and grain size, the slope of a hillslope is likely to be greater than in a river, and the flow depth and relative submergence will therefore be lower. To illustrate this issue, consider the following example. Take a hillslope with a typical slope of 0.2 (-) and a river with a slope of 0.005 (-), an identical bed shear velocity of 0.1 m s^{-1}, and composed of the same material ($D_{50} = 0.0005$ m). The flow depths $h_w$ for the hillslope and river are 0.005 m and 0.2 m, respectively, so the relative submergence is 20 and 815. Therefore the difference in relative submergence between river and overland flows is likely to result in transport rates on hillslopes being underestimated by fluvial transport-capacity equations. Furthermore, the use of bed shear velocity as the scaling parameter from the river and to the hillslope relies on the assumption of steady and uniform flow. These conditions are unlikely to occur on the higher gradients commonly found on hillslopes. A similar or somewhat analogous situation happens in aeolian transport as affected by roughness scale. Based on results presented in Brown et al. [2008] and Raupach et al. [2006], for equivalent roughness densities ($\lambda = \text{total roughness element frontal area/area occupied by the elements}$), the same shear stress will be exerted on the surface among the roughness elements, regardless of their size and distribution. The saltation flux for similar roughness densities however can be quite different, which Gillies et al. [2006] and Lancaster and Gillies [2013] attribute to the interaction of the particles in transport with the roughness elements. Large elements reduce the flux beyond that which can be solely attributed to the effect caused by the partitioning of shear stress between the roughness elements and the surface. This observation suggests that there can be multiple “capacity” states for the same
surface shear stress in aeolian sediment transport, due to roughness effects related to their physical size.

If the issues highlighted in this discussion mean that it is difficult to retain a concept of transport capacity, what are the implications for understanding and predicting sediment transport? The experience from especially the fluvial, aeolian and coastal process domains suggest that a concept of capacity is not a prerequisite for predicting transport rate. If we focus on the grain scale, then employing the terminology of Furbish et al. [2012], the flux represented by an ensemble of sediment particles at a particular point in space and time can be defined as:

\[
\bar{q}_x(x,t) = \bar{u}\bar{\gamma} - \frac{1}{2} \frac{\partial}{\partial x} (\bar{k}\bar{\gamma})
\]

where \( \bar{q}_x \) is the mean unit flux relative to the direction of flow [L² T⁻¹], \( \bar{u} \) is the mean particle velocity [L T⁻¹], \( \bar{\gamma} \) is the mean “particle activity” or volume of particles in the fluid per unit area of the bed [L], \( x \) is distance in the direction of the flow [L] and \( \bar{k} \) is the mean diffusivity of particle movement [L² T⁻¹].

Each of these terms has an implicit scale of variability, and only at low transport rates can they be considered correlated [Furbish et al., 2012]. Particle velocity is controlled by the position of the particle in the vertical profile, which is a function of initial conditions of entrainment and its trajectory, both related to turbulence and to sediment particle size, spin (Magnus force) and interactions with other particles. The particle activity is a function of relative entrainment and deposition rates, and thus depends on local conditions in relation to the former, but conditions upflow for the latter [see Parsons et al., 2004]. Diffusivity is related to the autocorrelation of variability in the particle velocity [Furbish et al., 2012], and thus again on turbulence structures over distances upflow corresponding to different lengths as a function of different particle sizes. Figure 12 illustrates how these different scales might vary through a vertical flow profile. Transport capacity would only hold if, averaged over a suitable time scale (e.g., over timescales of turbulent variations or of bedform translation: Furbish et al., [2012]),
the sediment flow into the control volumes at different points in the fluid flow balanced the sediment flow out. However, because turbulence in environmental flows is anisotropic, i.e., the flow structures have a downstream directional preference and thus the mean flow velocity has a gradient (see Grant and Marusic [2011] for a useful review), changes in particle activity, velocity and diffusivity will relate to different space and time scales in the flow field and thus supply into – and the delivery out of – the control volume cannot be steady and thus transport capacity cannot be a meaningful description of the process. Transport capacity could only hold in conditions of homogeneous turbulence (i.e., turbulence has the same structure quantitatively in all parts of the volume so the velocity fluctuations are random and the mean fluctuation is zero) or isotropic turbulence (the statistical features have no directional preference and thus there is no mean flow velocity gradient, and the mean velocity is either zero or constant throughout the volume. As particle activity increases, the probability of particle-particle collisions will increase (e.g., Bagnold [1954]; Sommerfeld [2001]), the nature of the fluid will change, and thus too will the correlation lengths over which the terms in equation 38 need to be averaged.

The grain-based perspective suggests, then, that what is observed is a time-averaged sediment flux. In this setting, models that relate entrainment to underemployment of “capacity” or deposition to an exceedance of “capacity” (e.g., equations 28, 29, 31, 32 and 35) are physically unreasonable, because they confound the instantaneous processes of transport with a phenomenological representation of time-averaged conditions. Such a perspective also requires a more nuanced approach to the idea of supply limitation. At the very least, there needs to be a recognition of the temporal and spatial variability of sediment supply, which also relates to conditions upflow, in the same way as the variability in transport does. Improved models of sediment transport thus need to focus on direct or indirect characterizations of the terms in equation 38 to allow predictions – whether averaged or including stochastic fluctuations – to be more reliable and transferable because of their process basis. The general nature of equation 38 means that it is a valid basis for developing general, universal models for predicting sediment-transport rates, and thus overcome some of the limitations highlighted in this review as a result of working in specific process domains.

This perspective also has implications for what is measured and how those measurements are carried out. At a specific scale, different measurements will reflect different characteristics and states of the sediment-transport process within different components of the geomorphic system. Figure 13 demonstrates how different techniques have been used at different scales.
More long-standing techniques tend to emphasize averaging over longer time intervals, and we would argue have tended to support concepts relating to steady state, such as transport capacity, rather than the inherent variability of sediment-transport processes. Newer technologies are increasingly able to evaluate shorter temporal and spatial variability, and it will be will the application of these techniques that the more robust and transferable models of sediment transport will be developed.

**Conclusions**

Transport capacity is a concept that is used across a wide range of domains within geomorphology. Although initially defined in fluvial geomorphology there were subsequently a number of independent inventions of similar concepts that were subsequently reabsorbed into the discipline. This process happened at an accelerating pace from the 1950s following the drive to make the discipline more quantitative, but was also underpinned from perspectives needed to manage and control environmental systems. The above review suggests that the multiple different ideas and applications of the term add potential for confusion and mean that ways of testing ideas of capacity are unclear. Unless the use in different areas is clarified or common terminology is defined (e.g., Bracken and Oughton, [2006]), there is a risk that coherent testing of the idea will be impossible. If the fact that the complicatedness of environmental systems and that estimates of capacity do not transfer across scales means that the concept is fundamentally limited, there will be serious practical consequences of using it to make predictions, especially as environmental management increasingly moves towards integrated approaches at catchment or landscape scales (e.g., Integrated River Basin Management and related policies such as the EU Water-Framework Directive or aspects of the Clean Water Act in the USA). If the existing models need significant calibration and are thus poor representations of the process, the continued use of the model is more about a justification of truth given the need to expend effort to carry out that calibration. For practical use, much simpler models could be calibrated if all that is required is an empirically based prediction of a rate. As noted in the overview on carrying capacity in a range of disciplines by Sayre [2008: 131]: “If carrying capacity is conceived as static, it is theoretically elegant but empirically vacuous; but if it is conceived as variable, it is theoretically incoherent or at best question-begging.”
The recognition of the complicatedness of sediment-transport systems and the effects of spatial and temporal dynamics in them – both as a result of turbulence and of environmental heterogeneity – should mean that new approaches are needed that are not underpinned by an ideal transport capacity that is virtually impossible to produce outside of controlled, laboratory conditions. Just because an effect is isolatable in laboratory conditions does not imply that it is a useful way of approaching an understanding of the real world [Hacking, 1983]. There is a need to take on board the complicatedness of the environment and of process. In the latter case, there is a strong implication from the comparisons here that an approach that recognizes that different types of flow form continua would be a useful way forwards, recognizing that some of the institutional distinctions made in the discipline hinder the development of geomorphological understanding overall. If so, the implication is strongly that we need to move away from the idea of transport capacity, and certainly that a single capacity for any set of condition is practically impossible. As the advances discussed above suggest, the way forward requires fundamental characteristics of sediment transport to be reëvaluated using an integrated approach that combines both fundamental theory with empirical observations, and that the latter should be driven by the former.

Acknowledgements
Discussions that led to this paper have been supported by a University of Sheffield LR Moore Fund award to JAG and a Durham University COFUND project supporting PG. We would like to thank the Editor, Mike Kirkby and two anonymous reviewers who have helped in clarifying the ideas in this paper since our first draft. As a review paper, no new data have been used in its writing; all data presented are appropriately cited and included in the reference list.

References


249, 235–297.
2073
Bagnold RA 1963 ‘Beach and nearshore processes Part I. Mechanics of marine
2074
sedimentation’, in MN Hill (ed.) The Sea – Ideas and Observations on Progress in the
2075
2076
Bagnold RA 1966 An Approach to the Sediment Transport Problem from General Physics. US
2077
2078
Bagnold RA 1973 ‘The nature of saltation and of "bed-load" transport in water’, Proceedings
2079
of the Royal Society of London A 332, 473–504.
2080
Bagnold RA 1980 ‘An empirical correlation of bedload transport rates in flumes and natural
2081
2082
2083
2084
Baillard JA 1984 ‘A simplified model for long shore sediment transport’, ASCE Proceedings
2085
2086
Baldwin DS, AM Mitchell, JM Olley 2002 ‘Pollutant-sediment interactions: sorption,
2087
reactivity and transport of phosphorus’, in PM Haygarth, SC Jarvis (eds) Agriculture,
2088
2089
Barry JJ, JM Buffington, JG King 2004 ‘A general power equation for predicting bed load
2090
2091
2092
Engineering 133, 1192–1205.
2093
Bathurst JC, WH Graf, HH Cao 1983 ‘Initiation of sediment transport in steep channels with
2094
2095
213, AA Balkema, Brookfield.
2096
Bathurst JC, WH Graf, HH Cao 1987 ‘Bed load discharge equations for steep mountain rivers’,
2097
2098
477, John Wiley & Sons, Chichester.
2099
Battjes JA1974 ‘Surf similarity’, Proceedings 14th International Conference on Coastal
2100
Engineering, 466–480. doi:10.9753/icce.v14
2101
Bauer BO, RGD Davidson-Arnott, KF Nordstrom, J Ollerhead, NL Jackson 1996
2102
‘Indeterminacy in aeolian sediment transport across beaches’, Journal of Coastal Research
2103
12, 641–653.


Cao, Z, Z Li, G Pender, P Hu 2012 ‘Non-capacity or capacity model for fluvial sediment transport’, Water Management 165(WM4), 193–211.


Chepil WS, RA Milne 1939 ‘Comparative study of soil drifting in the field and in a wind tunnel’, _Scientific Agriculture_ 19, 249–257.


Cooper JR 2012 ‘Does flow variance affect bedload flux when the bed is dominated by grain roughness?’, _Geomorphology_ 141, 160–169.


Egiazaroff IV 1965 ‘Calculation of nonuniform sediment transportations’, *Journal of Hydraulics Division, American Society of Civil Engineers* 91, 225–247.


Einstein HA 1942 ‘Formulas for the transportation of bed load’, *American Society of Civil Engineers, Transactions* 107, 561–577 (as 1941 paper but with discussion by JW Johnson, AA Kalinske, OG Haywood Jr, S Shulits, JS McNown and HA Einstein, pp. 578–597.)


Einstein HA, N Chien 1953 *Transport of Sediment Mixtures with Large Ranges of Grain Sizes*, US Army Corps of Engineers, Missouri River Division, Missouri River Division Sediment Series No. 2, Omaha.


Emmett WW, MG Wolman 2001 ‘Effective discharge and gravel-bed rivers’, *Earth Surface Processes and Landforms* 26, 1369–1380.


Gomez B, M Church 1988 ‘A catalogue of equilibrium bedload transport data for coarse sand gravel-bed channels’, *Department of Geography, University of British Columbia*, 1–90.


Govers G 1985 ‘Selectivity and transport capacity of thin flows in relation to rill generation’, *Catena* 12, 35–49.


Hagen JO 1987 ‘Glacier surge at Usherbreen, Svalbard’, Polar Research 5 n.s., 239–252


Kawamura R 1951 ‘Study on sand movement by wind’, Reports of the Physical Science Research Institute of Tokyo University 5, 95–112.


Laursen EM 1958 ‘The total sediment load of streams’, *Journal of Hydraulics Division, American Society of Civil Engineers* **84, HY1**, 1–36.


Lisle TE, M Church 2002 ‘Sediment transport-storage relations for degrading, gravel bed channels’, *Water Resources Research* 38, 1219–1232.


Lliboutry L 1993 ‘Internal melting and ice accretion at the bottom of temperate glaciers’, *Journal of Glaciology* 39, 50–64.


Misra RK, CW Rose 1990 Annual for Use of Program GUEST. Division of Australian Environmental Studies Report, Griffith University, Brisbane, Qld.


Ono D, P Kiddoo, C Howard, G Davis, K Richmond 2011 ‘Application of a combined measurement and modeling method to quantify windblown dust emissions from the exposed playa at Mono Lake, California’, *Journal of the Air and Waste Management Association* 61, 1036–1054.


Tanner WF 1979 ‘k-type coefficients in the littoral drift system’, Marine Geology 33, M15–M19.


Vanoni VA 1975 Sedimentation Engineering. ASCE Manuals and Reports on Engineering Practice, No. 54, Reston VA.


Williams G 1964 ‘Some aspects of the aeolian saltation load’, *Sedimentology* 29, 409–417.


Wilson KC 1966 ‘Bedload transport at high shear stress’, *Journal of Hydraulic Division, American Society of Civil Engineers* 92, 49–59.


Zingg AW 1940 ‘Degree and length of land slope as it affects soil loss in run-off’, *Agricultural Engineering* 21, 59–64.

Table captions

Table I: A list of commonly used models that employ sediment-transport-capacity relations across different process domains.
Figure captions

Figure 1: Comparisons between observed and calculated bedload transport in Elbow River, Alberta, Canada data for different bedload formulae, illustrating each formula produces contrasting estimates. (HRS: a group of formulae developed by researchers associated with the United Kingdom Hydraulics Research) [from Gomez and Church, 1989].

Figure 2: Measured submerged bedload transport rate, $i_b$, versus predicted values using equation 3. The data were compiled from Johnson (1943), Smart and Jaeggi (1983), Gomez and Church (1988), Recking (2006), and those compiled by Gao (2003). These data include all of the available experiments to date that transport bed load of homogeneous grains under the ideal condition and cover the full range of both the saltation and sheetflow régimes.

Figure 3: Schematic illustration of grain-size changes in the bed-surface ($D_{s50}$) and transported ($D_{50}$) sediment. In (a), a feed flume, only the bed surface changes, while in a recirculating flume (b), the change is primarily in the transported sediment. In both cases (c), transport coarsens relative to the bed surface. Thus, the same ratio of $D_{50}/D_{s50}$ may be caused by (1) a low transport rate with small $D_{50}$ or (2) a high transport rate with big $D_{50}$ (after Wilcock and DeTemple, 2005).

Figure 4: The modified two-phase model. The two solid curves represent equation 5 with $\theta_c = 0.03$ and 0.06, respectively. The two dashed lines denote the boundary between the two régimes for the same two $\theta_c$ values. The areas between these curves and lines reflect the influence of the uncertainties in the determination of $\theta_c$ values. The dots are the bedload data reported in Hayes (1999) from a gravel-bed river significantly affected by a recent volcanic eruption (the data that have values of $B$ greater than 1 are not included). Régime I is the area below the horizontal zone that includes two parts, the narrow area bounded by the two solid curves and the one on the right representing bed load transported at and below capacities, respectively. In the below-capacity area, bedload transport rate is relatively low for a given flow meaning the transport efficiency is relatively low and the median size of bed load $D_{50}$ is small comparing to that of the bed surface, $D_{s50}$ and substrate, $D_{sub50}$. In the at-capacity area, the transport rate is relatively high for the same flow suggesting the relatively high transport efficiency and $D_{50}$ is between $D_{s50}$ and $D_{sub50}$. Régime II is the area above the dashed horizontal zone. It also has below-capacity and at-capacity areas. Flows in
the former have a bed with an armour layer, while in the latter do not. $D_{50}$ in the
former is relatively small, while in the latter is equivalent to both $D_{a50}$ and $D_{sub50}$.

**Figure 5:** Plot of surface-based fractional transport rates ($q_{bi}/f_i$) against grain size fractions, $D$.
Each curve represents a flow transporting bed load at capacity in one of four gravel-
bed rivers in Idaho, USA.

**Figure 6:** Schematic of the streamlines above a low amplitude undulation of a sand surface in
an aeolian setting. The maximum $u^*$ is located at a distance upwind from the crest
(maximum $\zeta$) proportional to the wavelength $\lambda$. The sand flux maximum $q_{sat}$ is
located at a distance $L_{sat}$ downwind, which separates the zones of erosion and
deposition (after Durán et al., 2011).

**Figure 7:** Saturation length $L_{sat}$, rescaled by the drag length ($L_{drag} = \frac{\rho_p}{\rho_f} d$), as a function
of the wind shear velocity $u_{sat}$, rescaled by the threshold $u_{sat}$. Direct measurements,
performed in a wind tunnel (◊) and in the field (△), are compared to those determined
from the initial dune wavelength (storms: (☆) and slipfaceless dunes (○)) (after
Andreotti et al., 2010).

**Figure 8:** Different equations used to predict sediment transport under longshore conditions,
showing the wide range of potential values for a specific wave energy (after Komar,
1999).

**Figure 9:** Conceptual model of soil erosion derived by Meyer and Wischmeier (1969) from
Ellison (1947) and other sources.

**Figure 10:** Comparison of experimental data with derived transport capacity and detachment
capacity of overland flow on bare soil without sediment input at the top of the slope,
but with flow addition at the rates specified (after Schiettecatte et al., 2008).

**Figure 11:** Schematic diagram of possible subglacial conditions, showing a plan view of the
bed of an active ice sheet or glacier flowing from the top to the bottom of the page.
The key sediment-transport mechanisms are illustrated. The thicknesses of the curves
on the left represent the ice flux (linear scale), the water flux (logarithmic scale), and
total flux of débris in the ice plus deforming or stream-transported sediment below
the ice (logarithmic scale) (after Alley et al., 1997).

**Figure 12:** Relationships as a function of flow depth ($z$) between mean flow velocity ($v$), mean
sediment velocity ($u$), mean particle activity ($\gamma$ – see equation 38) and mean sediment
flux ($q$) for: a. flow dominated by coarse particles moving near to the substrate; b.
flow dominated by fine particles moving throughout the flow column; and c. hyperconcentrated or débris flows where the velocities of the fluid and the sediment converge. In moving from condition a. to condition c., the upstream effects of flow variability cause both $u(z)$ and $\gamma(z)$ to become more variable because of anisotropy in the turbulent flow.

**Figure 13:** Spatial and temporal scales of measurement of sediment transport and their implicit or explicit uses. The area shaded in blue has been the one typically used to infer transport-capacity rates. Measurements at other scales and for other purposes have come to challenge the underlying concept of transport capacity.
Table I: A list of commonly used models that employ sediment-transport-capacity relations across different process domains.

<table>
<thead>
<tr>
<th>Model</th>
<th>Transport-capacity equations</th>
<th>Reference</th>
</tr>
</thead>
</table>
Figure 1: Comparisons between observed and calculated bedload transport in Elbow River, Alberta, Canada data for different bedload formulae, illustrating each formula produces contrasting estimates. (HRS: a group of formulae developed by researchers associated with the United Kingdom Hydraulics Research) [from Gomez and Church, 1989].
Figure 2: Measured submerged bedload transport rate, $i_b$, versus predicted values using equation 3. The data were compiled from Johnson (1943), Smart and Jaeggi (1983), Gomez and Church (1988), Recking (2006), and those compiled by Gao (2003). These data include all of the available experiments to date that transport bed load of homogeneous grains under the ideal condition and cover the full range of both the saltation and sheetflow régimes.
Figure 3: Schematic illustration of grain-size changes in the bed-surface ($D_{s50}$) and transported ($D_{50}$) sediment. In (a), a feed flume, only the bed surface changes, while in a recirculating flume (b), the change is primarily in the transported sediment. In both cases (c), transport coarsens relative to the bed surface. Thus, the same ratio of $D_{50}/D_{s50}$ may be caused by (1) a low transport rate with small $D_{50}$ or (2) a high transport rate with big $D_{50}$ (after Wilcock and DeTemple, 2005).
Figure 4: The modified two-phase model. The two solid curves represent equation 5 with $\theta_c = 0.03$ and 0.06, respectively. The two dashed lines denote the boundary between the two regimes for the same two $\theta_c$ values. The areas between these curves and lines reflect the influence of the uncertainties in the determination of $\theta_c$ values. The dots are the bedload data reported in Hayes (1999) from a gravel-bed river significantly affected by a recent volcanic eruption (the data that have values of $B$ greater than 1 are not included). Régime I is the area below the horizontal zone that includes two parts, the narrow area bounded by the two solid curves and the one on the right representing bed load transported at and below capacities, respectively. In the below-capacity area, bedload transport rate is relatively low for a given flow meaning the transport efficiency is relatively low and the median size of bed load $D_{50}$ is small comparing to that of the bed surface, $D_{s50}$ and substrate, $D_{sub50}$. In the at-capacity area, the transport rate is relatively high for the same flow suggesting the relatively high transport efficiency and $D_{50}$ is between $D_{s50}$ and $D_{sub50}$. Régime II is the area above the dashed horizontal zone. It also has below-capacity and at-capacity areas. Flows in the former have a bed with an armour layer, while in the latter do not. $D_{50}$ in the former is relatively small, while in the latter is equivalent to both $D_{s50}$ and $D_{sub50}$. 
Figure 5: Plot of surface-based fractional transport rates ($q_{bd}/f_i$) against grain size fractions, $D$. Each curve represents a flow transporting bed load at capacity in one of four gravel-bed rivers in Idaho, USA.
Figure 6: Schematic of the streamlines above a low amplitude undulation of a sand surface in an aeolian setting. The maximum $u^*$ is located at a distance upwind from the crest (maximum $\zeta$) proportional to the wavelength $\lambda$. The sand flux maximum $q_{sat}$ is located at a distance $L_{sat}$ downwind, which separates the zones of erosion and deposition (after Durán et al., 2011).
Figure 7: Saturation length $L_{sat}$, rescaled by the drag length ($L_{drag} = \frac{\rho_p}{\rho_f} d$), as a function of the wind shear velocity $u^*_{a}$, rescaled by the threshold $u^*_{at}$. Direct measurements, performed in a wind tunnel (◇) and in the field (△), are compared to those determined from the initial dune wavelength (storms: (☆) and slipfaceless dunes (○)) (after Andreotti et al., 2010).
Figure 8: Different equations used to predict sediment transport under longshore conditions, showing the wide range of potential values for a specific wave energy (after Komar, 1999).
Figure 9: Conceptual model of soil erosion derived by Meyer and Wischmeier (1969) from Ellison (1947) and other sources.
Figure 10: Comparison of experimental data with derived transport capacity and detachment capacity of overland flow on bare soil without sediment input at the top of the slope, but with flow addition at the rates specified (after Schiettecatte et al., 2008).
Figure 11: Schematic diagram of possible subglacial conditions, showing a plan view of the bed of an active ice sheet or glacier flowing from the top to the bottom of the page. The key sediment-transport mechanisms are illustrated. The thicknesses of the curves on the left represent the ice flux (linear scale), the water flux (logarithmic scale), and total flux of debris in the ice plus deforming or stream-transported sediment below the ice (logarithmic scale) (after Alley et al., 1997).
Figure 12: Relationships as a function of flow depth (z) between mean flow velocity (v), mean sediment velocity (u), mean particle activity (γ – see equation 38) and mean sediment flux (q) for: a. flow dominated by coarse particles moving near to the substrate; b. flow dominated by fine particles moving throughout the flow column; and c. hyperconcentrated or débris flows where the velocities of the fluid and the sediment converge. In moving from condition a. to condition c., the upstream effects of flow variability cause both u(z) and γ(z) to become more variable because of anisotropy in the turbulent flow.
Figure 13: Spatial and temporal scales of measurement of sediment transport and their implicit or explicit uses. The area shaded in blue has been the one typically used to infer transport-capacity rates. Measurements at other scales and for other purposes have come to challenge the underlying concept of transport capacity.