CHAPTER 2

LITERATURE REVIEW

2.1. INTRODUCTION

Soil erosion by water is a major environmental issue, and thus, an important factor in managing land and water resources (Foster and Meyer, 1977; Morgan, 1995). Soil erosion decreases agricultural productivity, degrades water quality and ecological diversity in streams and wetlands, and reduces the capacity of conveyance structures (Pimental *et al.,* 1976; Foster and Meyer, 1977; Morgan, 1995). Although, soil erosion is a natural process, accelerated erosion, especially from agricultural lands, poses great environmental threat (National Academy of Science, 1974; Milliman and Meade, 1983). The impact of erosion could be grouped as on- and off-site effects (Morgan, 1995).

The on-site effects are particularly important on agricultural land where the redistribution of soil within a field, the loss of soil from the field, the breakdown of soil structure and decline in soil fertility result in a reduction of soil cultivatable depth, decline in soil fertility and available soil water. The on-site effects are necessarily borne by the farmer and are endured by the community in terms of higher food prices (Napier, 1987; Morgan, 1995). The off-site effects of erosion results from sedimentation downstream which reduces the capacity of the rivers and drainage ditches, enhances the risk of flooding, shortens the design life of reservoirs and blocks irrigation systems (Morgan, 1995). Furthermore, the sediment in its own right, is a pollutant, and as an adsorbing surface for polluting chemicals, can increase the level of N and P in water bodies and result in eutrophication. The off-site effects are mainly borne by local authorities, insurance companies and all landholders in the community. As a general rule, on-site effects are associated with soil loss, while off-site effects with sediment yield (Morgan, 1995).

The terms soil loss and sediment yield have distinct meanings in erosion technology. Soil loss refers to the soil material removed off from its original position, i.e., a particular slope or field (Mitchell and Bubenzer, 1980). It is defined as the detachment and movement of soil particles on a landscape profile, referred to as land facet, thus, distinguishing it from deposition and sediment transport in catchments (Nearing *et al.,* 1994). Sediment yield is the total sediment outflow from a catchment during any given time. In other words, it is the proportion of the soil loss that is not deposited before the catchment outflow or designated area of interest in the catchment (Mitchell and Bubenzer, 1980). Sediment yield is a net result of the complex processes of detachment and transport by raindrops and flowing water and deposition (Nearing *et al.,* 1994; Morgan, 1995). For small areas, e.g. land facets, runoff plots or field-sized areas, the soil loss corresponds with sediment yield. For bigger areas, sediment yield will be a proportion of soil loss. This is because temporal and permanent deposition takes place within the catchment.

2.2. PRINCIPLES OF WATER EROSION

Since erosion models are a simplification of erosion in reality, it is important to understand the principles underlying water erosion in order to evaluate erosion prediction tools.

2.2.1. Erosion processes

Soil erosion by water is a three-phase process, involving the detachment, transportation and deposition processes of sediment by the erosive agents (ASCE, 1975; Morgan, 1995). *Detachment* (phase 1) is the dislodging of soil particles from the soil mass, whereas, *transportation* (phase 2) is the entrainment and movement of the sediment from their original location (Foster and Meyer, 1977). The third phase (*deposition*) may occur when there's no longer sufficient energy to transport the particles (Morgan, 1995).

2.2.2. Types of erosion

2.2.2.1. Interrill

Detachment of soil particles from the soil matrix and / or entrainment of temporarily deposited sediment is primarily induced by rain-splash since the shallow flow depths are of negligible erosive power. As such, interrill flow (shallow overland flow) mainly accounts for sediment transportation from the interrill areas to the rill areas (Kinnel, 1988). Sometimes the interrill erosion is referred to as sheet erosion since flow is in the form of a thin layer.

2.2.2.2. Rill

Rill erosion occurs mainly when the shallow overland flow over the interrill areas break-up and develop into concentrated flow paths, i.e., small channels and micro-rills (Moss *et al.,* 1982). Rill flow detaches soil particles when its transport capacity is more than the sediment load and when the flow shear stress exceeds the soil resistance to detachment (Govers and Poesen, 1988).

2.2.2.3. Gully

In many respect gully erosion is like rill erosion, in that the processes are similar, except that gully erosion takes place at the wider scale. Furthermore, gullies are dynamically similar to small stream channel erosion except for their ephemeral flow and direct supply of sediment from the hillslope to the river stream. Sediment is produced in three ways in the gully; by scouring at the base of the scarp, supercritical flow at the heads of the depression, and by scouring action of running water on the banks of the gully channel, and also by flushing-out material in subsurface pipes by interflow. Tunnel erosion is normally associated with duplex soils where more runoff infiltrates the soil through small depressions, cracks and macropores, but on reaching clay-rich B-horizon moves along it as interflow. This is followed by piping, which occurs as clay disperses in areas of localised moisture regimes (Floyd, 1974).

2.2.2.4. Stream channel

Stream channel erosion closely resembles rill erosion and occurs due to the scour induced by stream flow or due to side slope instability. Stream channel erosion has two components (i) bank erosion and (ii) bedload erosion. Bank erosion occur when the boundaries of the channel are eroded, and bedload erosion result when the transported sediment at the base of the stream channel interact with the bed, thus causing it to erode.

2.2.3. Erosion factors

All the erosion factors can be grouped into the following major factors climate erosivity, soil erodibility, topography and vegetation characteristics.

2.2.3.1. Climate erosivity

Climate erosivity includes the erosivity of the rainfall and runoff. Rainfall erosivity is a function of two rainfall characteristics, intensity and its kinetic energy. Rainfall intensity is related to two types of rain events, the short-lived intense storm where the infiltration capacity of the soil is exceeded, and the prolonged storms of low intensity that saturate the soil (Morgan, 1995). Kinetic energy (KE) of the raindrops is the major factor initiating soil detachment and dispersal of soil particles that is a function of raindrop-size and its terminal velocity (Kowal and Kassam, 1976; Mihara, 1951; Morgan *et al.,* 1986; Rose, 1960). Most studies use IE_{30} to define the combined effect of 30 minute-intensity and kinetic energy of the rainfall (Laws and Parson, 1943; Mitchell and Bubenzer, 1980; Wischmeier and Smith, 1958,1978).

Runoff erosivity is a function of the shear stress of the flow exerted on the soil and the flow transport capacity (Neibling and Foster, 1977; Nearing *et al.,* 1994). The runoff erosivity is influenced by infiltrability of soil and the flow velocity (Morgan, 1995). Soil detachment and rill initiation by surface runoff (overland flow) is greatly related to flow depth or hydraulic radius and exceedence of the critical shear velocity, and depends on the inherent resistance of the soil (Meyer, 1965; Morgan, 1995).

2.2.3.2. Soil erodibility

The inherent resistance of the soil and surface roughness influences the flow velocity and requires attainment of velocity threshold value before erosion can commence. Soil's resistance, that is shear stress and strength, depends on soil aggregate stability, soil profile infiltration capacity, and in part on the topographic position, slope steepness and amount of disturbance (Kirkby and Morgan, 1980; Morgan, 1995). The important soil properties determining the soil susceptibility to erosion (erodibility) are chemical constituents, organic matter, texture and type of clay minerals present (Morgan, 1995).

2.2.3.3. Topography

Topography influences velocity and volume of the surface runoff. Erosion is expected to increase with slope length and slope steepness as a result of respective increases in velocity and volume of surface runoff. For erosion by rainsplash and surface runoff, soil loss increases curvilinearly as the slope steepens from gentle to moderate, reaching a maximum on slopes of approximately 8 - 10°. The curvilinear relationship does not apply for landslides, piping and gully erosion (Morgan, 1995). However, a single relationship cannot be used to describe soil loss. An increasing slope length limit erosion as the rate of detachment by shallow overland flow decreases downslope and becomes concentrated flow that may limit interrill-rainsplash processes (Gilley *et al.,* 1985; Abrahams, *et al.,* 1991) or as deposition occurs (Poesen, 1984).

2.2.3.4. Vegetation cover

The vegetation cover influences rainfall, runoff and slope stability and encompasses vegetation management and erosion control methods. The aboveground components of vegetation reduce the energy of raindrops and velocity of runoff so that less is directed to the soil. The effectiveness of vegetation cover in reducing erosion by raindrop impact depends upon the height and the continuity of the canopy, and the density of the ground cover (Morgan, 1995). Furthermore, vegetation cover dissipates the energy of the runoff by imparting roughness to the flow, thereby reducing its velocity, evacuate excess runoff (Troeh *et al.,* 1980), filter sediment from the runoff flow (Boubakari, 1992) and increasing infiltration (Melville, 1992). The level of roughness with different forms of vegetation depends upon the morphology and density of the vegetation, as well as height in relation to the flow depths. The belowground components of vegetation contribute to the mechanical strength of the soil against mass movements, thus enhancing slope stability (Morgan, 1995).

2.2.4. Impact of human activities

Human activities affect soil erosion by changing the soil erodibility. Tillage practices increase the rate of organic matter decomposition that leads to reduced soil organic matter content. This cause increased erodibility and reduced quality of vegetation cover. Human practices also affect soil erosion when vegetation cover is changed through various cropping and grazing practices. Conservation techniques that can reduce erosion include vegetation and soil management practices and erosion control practices by shortening slope length to reduce runoff rate and velocity (Morgan, 1995).

2.2.4.1. Vegetation management

Agronomic measures incorporate crop and vegetation management, and the use of the protective effect of various plant covers to reduce erosion, whilst maximising crop production (Morgan, 1995). Various forms of plant covers differ in their ability to protect the soil. The various agronomic practices include - crop rotation, cover crops, strip and multiple cropping, re-vegetation, agro-forestry, and mulching (Morgan, 1995; Lundgren and Nair, 1985).

2.2.4.2. Soil management

The types, frequency and timing of tillage operations influences the structure, surface roughness, cloddiness, compaction and microtopography of the soil, and thus susceptibility of the soil to water erosion. A sound soil management strategy involves the addition of organic matter and / or soil conditioners to improve soil aggregate stability, cohesion and infiltration capacity (Matthee and Van Schalkwyk, 2001). Tillage practices which leave the soil surface bare and exposed to direct heat, leads to erosion due to low or poor protective vegetation cover and increased rate of organic matter decomposition (Morgan, 1995).

2.2.4.3. Erosion support / control practices

Erosion support practices involve mechanical or physical field practices that are used to control the movement of water over the soil surface in order to control the transport phase of erosion. The various mechanical methods used to reduce erosion include: contouring, contour bunds, terraces, waterways, geo-textiles and stabilisation structures (Morin *et al.,* 1984; Vogel, 1991; Hurni, 1984; Morgan, 1995).

2.3. WATER EROSION MODELS

A model is of necessity a simplification of reality and is a scientific tool or technique designed to predict soil erosion under a wide range of specified conditions (Morgan, 1995). The erosion predictions are used to develop and evaluate crop management systems and can also be used to develop land-use strategies, to provide relative soil loss / sediment yield indices and to guide government policy and strategies on soil and water conservation (Smith, 1999). Consequently, erosion models are important evaluation tools in PM&E to make comparative measures (impact assessment) of physical characteristics, such as changes in soil loss under various land uses and management systems (Woodhill and Robins, 1998).

Soil erosion prediction techniques have evolved over many years since the recognition of erosion as a serious agricultural problem in the late 1920s (Figure 2.1). Early estimates were primarily qualitative in nature, and as more research was conducted, soil loss models were developed (Mitchell and Bubenzer, 1980). Soil loss models are strictly based on quantifying the on-site impact of erosion (Kirkby and Morgan, 1980), though many of these models are now being developed to incorporate sediment yield predictions (Morgan, 1995). However, efforts with the main aim of predicting sediment yield have been made since the late 1970s, as the non-point source pollution became more and more evident (Kirkby and Morgan, 1980; Elliot *et al.,* 1989). Due to greater prediction accuracy required, hence limitation posed by the soil loss models, physically based models were developed that incorporate spatial distribution of runoff and sediment over the land surface and during a single storm event. Most predictive models are empirical, but erosion models are increasingly being developed to reflect the erosion mechanics and become more spatially and temporally distributed (Figure 2.1).

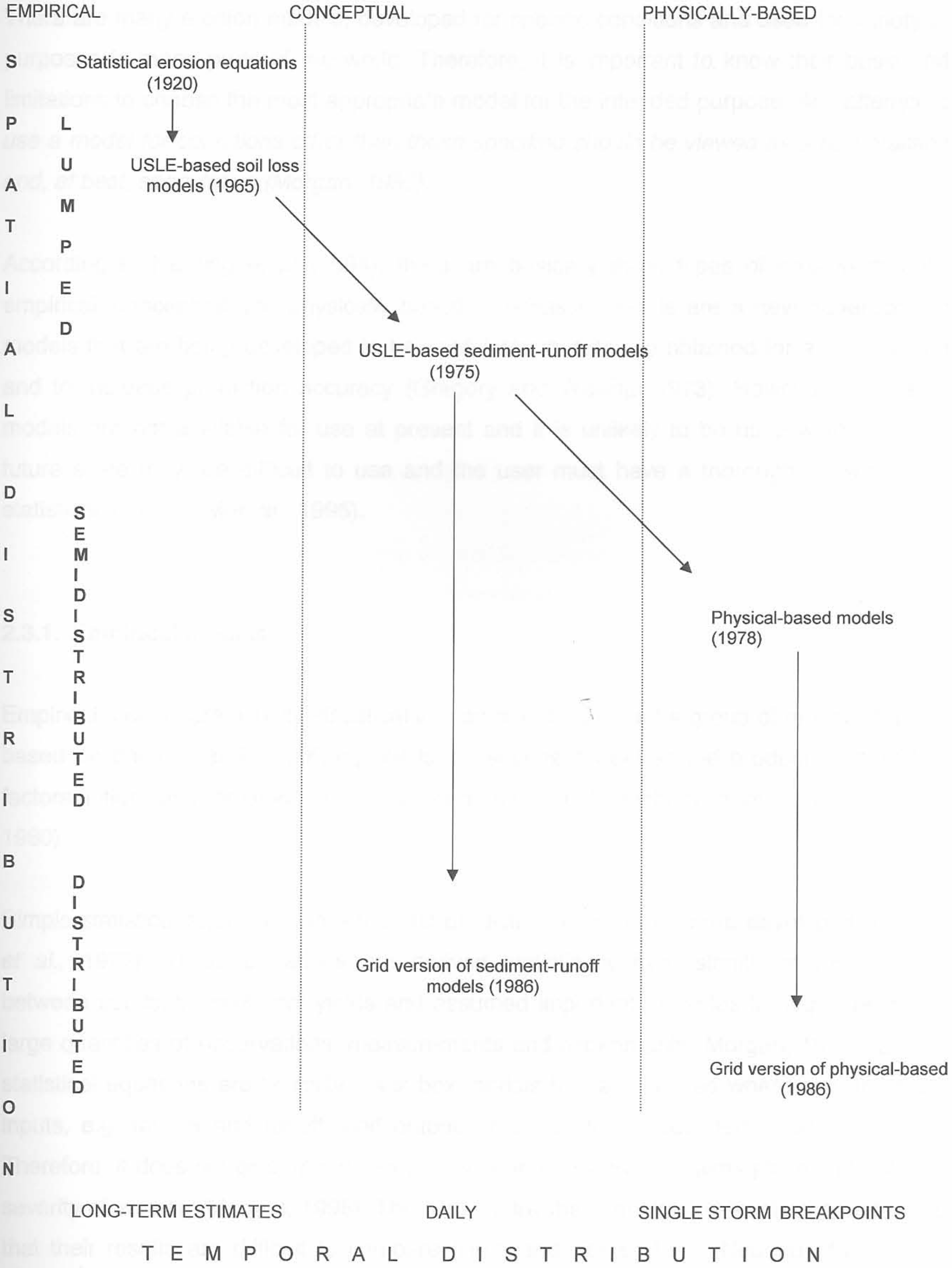

Figure 2.1. History of erosion models development (van Zyl and Lorentz, 2001)

There are many erosion models, developed for specific conditions and used for variety of purposes in many parts of the world. Therefore, it is important to know their basis and limitations to choose the most appropriate model for the intended purpose. *Any attempt to use a model for conditions other than those specified should be viewed as a bad practice and, at best, speculative (Morgan, 1995).*

According to Nearing *et al* (1994), there are basically three types of erosion models; empirical, conceptual and physically based. Stochastic models are a new generation of models that are being developed to be useful where data are obtained for a short period and to increase prediction accuracy (Gregory and Walling, 1973). However, stochastic models are not available for use at present and it is unlikely to be used widely in near future since they are difficult to use and the user must have a thorough knowledge of statistical analysis (Morgan, 1995).

2.3.1. Empirical models

Empirical models are usually statistical in nature and represent a group of models that are based on defining and identifying soil loss / sediment yield as the product of individual factors influencing erosion, hence ignoring the actual mechanism of erosion (Kirkby, 1980).

Simple statistical equations were the first predictive models that were developed (Onstad *et al.,* 1977). These equations are derived statistically from significant relationships between soil loss / sediment yields and assumed important variables through the use of large quantities of observations, measurements and experiments (Morgan, 1995). Simple statistical equations are typically black box models that are applied where only the main inputs, e.g. rainfall and runoff, and outputs, e.g. soil loss / sediment yield are known. Therefore, it does not give an indication of why and how erosion takes place, but only the severity of erosion (Morgan, 1995). The biggest drawback of simple statistical equations is that their results are difficult to compare (Herb and Yorke, 1976; Nearing *et al.,* 1994; Morgan, 1995).

With a large amount of experimental plot data (more than 10 000 plot-years), the Universal Soil Loss Equation (USLE) (Wischmeier and Smith, 1965) was developed to overcome the difficulty of comparing results of the various simple statistical equations. Thereafter, many models that followed were developed on the basis of USLE (Smith, 1999). Equation 2.1 describes the basis of the USLE model.

$$E = R.K.LS.C.P. \qquad \text{--------} \; (2.1)$$

where, E = Annual long term soil loss per hectare
R = Erosivity factor
K = Erodibility factor
LS = Slope length and steepness factor
C = Vegetative cover factor
P = Soil erosion control practices

The USLE is an empirical model since it considers erosion as a product of a series of the factors that does not allow any sort of interaction between the factors (Kirkby, 1980). Revision of USLE was initiated in 1987 that resulted in a computerised model RUSLE (Revised Universal Soil Loss Equation) (Renard *et al.,* 1991). Whereas the basic USLE has been retained, the algorithms used to calculate the individual factors have been improved to take account of seasonal variability and interaction between the factors (Renard, Laflen, Foster and McCool, 1994).

According to Smith (1999), the most widely applied soil loss models in South Africa are the USLE, its improved version the RUSLE and the Soil Loss Estimator Model for Southern Africa (SLEMSA) (Elwell, 1978).

Soil loss models are applied to differentiate between areas of high and low erosion potential and target efforts for conservation purposes. The weakness of soil loss models is that they:

- Do not account to why and how erosion occurs;
- Cannot be extrapolated beyond the data range with confidence, either to more extreme events, e.g. flooding, or to other geographical locations; and
- Can only be used to predict gross erosion (soil loss) over a long-term period from the rill and interrill areas, and fails to predict sediment yield and gully erosion (Smith, 1999).

2.3.2. Conceptual models

In order for empirical soil loss models such as RUSLE to determine the sediment yield, the gross erosion (soil loss) is multiplied by a sediment delivery ratio (sediment yield / soil loss). Numerous sediment delivery ratio concepts have been developed, but because of the empirical nature of these methods, relations should be applied with extreme caution. A prediction technology known as the sediment-runoff models (conceptual models) was developed to overcome the problems of determining a delivery ratio by replacing the rainfall erosivity factor of the USLE with a hydrological factor (storm runoff / peak storm runoff). These conceptual models are based on spatially lumped forms of water and sediment continuity equations by primarily using the concept of the unit hydrograph (Lane *et al.,* 1988, Nearing *et al.,* 1994, Lane and Nichols, 1996). Equation 2.2 describes the fundamental equation on which sediment-runoff models relies to predict sediment yield:

$$SY = a(Q/q_p)^b \, KLSCP \qquad \text{--------} \ (2.2)$$

where,
SY = sediment yield from an individual storm in metric tons
Q = storm runoff volume in m^3
q_p = peak runoff rate in m^3/sec
K = USLE soil-erodibility factor
LS = USLE slope length and slope gradient factor
C = USLE crop management factor
P = USLE erosion-control practice factor
a, b = model parameters (constants)

These models describe what happens in small catchments or part of a large catchment that have a constant variation of soil type and slope factors and only allows variation in runoff volumes, peak runoff rate and vegetation growth over time (Williams and Berndt, 1977; Morgan, 1995). Although sediment-runoff models have the same weaknesses of soil loss models, they have an advantage in that it can predict sediment yield on a daily basis (Williams, 1975). Most of the sediment-runoff models are spatially lumped models, but have lately become grid-based (Figure 2.1).

2.3.3. Physically based models

Subsequent advances in hydrology and erosion have provided the means of developing physically based erosion models, e.g. WEPP, based on infiltration theory, hydrodynamics of overland flow, interrill, rill and ephemeral gully sediment detachment, transport and deposition processes (Lane *et al.,* 1992). Physically based models provide several advantages over empirical and conceptual models, including most notably:

- Capabilities for estimating spatial and temporal distributions of net soil loss and sediment yield;
- More reliable extrapolation to ungauged areas;
- The ability to predict off-site delivery of sediment better; and
- Calculate in addition to erosion from interrill and rill areas, erosion from concentrated flows in ephemeral gullies, deposition in backwater and impoundments, deposition in concave slopes, and the enrichment of the fines caused by deposition (Nearing *et al.,* 1989).

Physically based models are intended to represent the essential mechanisms controlling erosion. The power of physically based models is that they represent a synthesis of the individual components that affect erosion, including the complex interactions between various factors and their spatial and temporal variabilities. However, erosion occurring in large gullies (dongas) and perennial streams is not considered, nor is stream bank erosion or erosion from wave action (Lane *et al.,* 1992; Nearing *et al.,* 1994).

2.4. EROSION ASSESSMENT METHODS

Erosion assessment methods are either aimed to quantify or classify current and potential erosion damage in the field. Assessment of current erosion involves direct measurements or estimation of soil loss from fields or parts thereof (Hudson, 1971). The methods ranges from the use of desktop aerial photography, satellite imagery, statistical methods and field inspection of current erosion (Wessels *et al.,* 2001). The biggest drawbacks of these methods are that they:

- Cannot account for erosion mechanics; and
- Could not have the same accuracy as the prediction models (Herweg, 1996).

2.4.1. Statistical approaches

Statistical approach methods consist of erosion ranking systems whereby individual forms of erosion are relatively weighted and ranked according to scores and compared with each other within a catchment or in different catchments. This method assumes a linear correlation of parameters and ignores the erosion factors (Taylor, 1998).

2.4.2. Field inspection methods

The field inspection methods involve direct and rapid assessment of erosion from the field. Examples of such methods are the Assessment of Erosion from Rangeland (Tongway, 1994) and the Assessment of Current Erosion Damage (Herweg, 1996). The assessment and monitoring procedures of field inspection methods vary from that of only soil and vegetation conditions to those that also include critical location of damage by erosion and the times of high erosion hazard on the field.

2.4.3. Desktop photographic and mapping assessment methods

This category of assessment methods includes erosion classification and erosion mapping systems. The erosion classification systems are a rapid but qualitative desktop-based aerial photographic assessment method of erosion to identify and classify erosion features and damage and are primarily used as decision-making tool (SARCCUS, 1981). The erosion mapping system is based on the use of 1: 10 000 scale orthophoto maps to map-out spatial patterns of erosion features, i.e., the type, location and intensity, and contributing erosion factors (Thwaites, 1986). This method provides a basis of qualitative (index and comparative) and quantitative (intensity and density) erosion assessment or prediction of changes in the nature and severity of erosion and environmental status at national and provincial or even regional scale (Wessels *et al.,* 2001). State of the art software programs such as OrthoBase are specifically designed to deal with orthophoto maps. Thus, erosion mapping systems can be used either manually as desktop or in *GIS* format.

2.4.4. Spatial modelling of erosion susceptibility and erosion hazard

Various qualitative assessment techniques, e.g. factor-based models (e.g. USLE/RUSLE and SLEMSA), remote sensing and geographical information systems, can be successfully applied to model the erosion susceptibility or erosion hazard of an area in a spatial context. A combination of remote sensing images and factor-based models were used to develop erosion susceptibility and hazard maps at a provincial level for the Mpumalanga and Gauteng (Wessels *et al.,* 2001; Wessels, 2001) Provinces of South Africa. These maps could be used to identify areas with a high erosion hazard in view of sustainable land use in the whole country.