Alternate stable states and threshold effects in semi-arid grazing systems

M. Rietkerk and J. van de Koppel


Models that explain the discontinuous behaviour of semi-arid grazing systems usually emphasize herbivore feeding characteristics or plant competition as possible mechanisms. Field studies indicate, however, that plant-soil relations could be more important. We show by means of a graphical model that the interactions between water infiltration or nutrient retention and plant density potentially give rise to the existence of alternate stable vegetation states and threshold effects in semi-arid grazing systems, even without the effect of a non-linear herbivore functional response or plant competition. These interactions may trigger a positive feedback between reduced plant density and reduced resource availability, and lead to a collapse of the system. The model results are in line with well-documented observations of spatial and temporal patterns such as two-phase mosaics and stably degraded grasslands.

Theoretical and empirical evidence for the existence of alternate vegetation states and threshold effects in grazing systems were presented two decades ago (Noy-Meir 1975). Alternate states may occur if at low plant density plant losses due to herbivory exceed plant growth, while at high density the reverse is true [see May (1977) for a review]. The shapes of the herbivore consumption functions (functional responses) and plant growth functions play a central role in Noy-Meir’s models, but the models do not address the underlying mechanisms of low plant growth rates at low plant density.

A vast body of literature exists on the relationship between high levels of herbivory and soil degradation (Elwell and Stocking 1974, Kelly and Walker 1976, Breman and De Wit 1983, Sinclair and Fryxell 1985, Le Houérou 1989, Prins 1989) which might contribute to a mechanistic explanation. A reduction of plant density often leads to a higher amount of rainfall running off instead of infiltrating into the soil and, because of that and wind erosion, to higher nutrient loss. The influence of vegetation on the infiltration rate of water into the soil and the rate of nutrient loss is more pronounced at the lower end of the plant density continuum than at the higher end (Walker et al. 1981, Stocking 1994). Plant growth will be affected by these processes, especially in semi-arid regions where growth is mainly governed by nutrient and water availability (Walker et al. 1981, Penning de Vries and Djitéye 1982, Breman and De Wit 1983). Hence, at low plant density, plant growth could be reduced below grazing induced plant losses, thereby adversely affecting the stability of semi-arid grazing systems.

In this paper we consider the main plant-soil relations occurring in semi-arid grazing systems and we investigate how they may influence the dynamics of these ecosystems by analysing a simple graphical model.
The model provides a new and conceptually effective explanation of the properties of these discontinuously stable grazing systems.

**Model grazing system**

**Assumptions**

A characteristic feature of semi-arid grasslands is that either water or nutrients limit plant growth (Penning de Vries and Djitéye 1982, Breman and De Wit 1983). In a water-limited system, plant growth is assumed to be determined only by water availability. In this case, we assume a constant rainfall and a constant proportion of soil water lost from the system per unit of time, through evaporation and percolation. In a nutrient-limited system, when plant growth is determined solely by nutrient availability, we assume a constant amount of nutrient released from the geochemical cycle per unit of time. For the sake of simplicity, a constant fraction of plant losses due to mortality (not grazing-induced) is assumed to be shunted directly into the soil nutrient pool and is subsequently available for plant growth (cf. DeAngelis 1992). The level of herbivory is kept constant, which is a good approximation to management practices in grazing systems (Noy-Meir 1975). A linear functional response of the herbivore to changes in forage availability is assumed, although we acknowledge that a saturating, sigmoid or hump-shaped response may be more realistic (Crawley 1983, Fryxell 1991, Van de Koppel et al. 1996). However, the effects of the plant-soil interactions on plant density are central in our investigation and not the shape of the functional response. The model we derive does not intend to capture the entire system but examines the consequences of dominant plant-soil interactions for the dynamics of the system. The model systems are mathematically defined in the Appendix.

**Zero-isoclines of plants and resources**

A simple way to analyse the dynamics of these water- and nutrient-limitation models is by plotting the zero-isoclines of the plant and its resources in a phase plane (e.g. Edelstein-Keshet 1988) (Fig. 1). The plant isocline is the line joining combinations of plant density and soil water or nutrients along which plant density does not change. We assume that plant production requires a certain minimum amount of soil water in case of water-limitation, or nutrients in case of nutrient-limitation, independent of plant density (see Armstrong and McGeehee 1980, Tilman 1982 for further details). Hence, the plant isocline is a straight, vertical line. At higher levels of soil water or nutrients, plant production is positive, while at lower levels plant production is negative. If the level of herbivory increases, the critical demand for soil water or nutrients increases and the plant isocline shifts to the right. Plant growth has to compensate for the increased level of herbivory in order to sustain production.

The soil water isocline is the line joining those combinations of plant density and soil water where the amount of soil water does not change. In the absence of vegetation the uptake rate of water by plants is zero and the soil eventually reaches an equilibrium value for the amount of soil water. Increasing plant density leads to a decreasing amount of soil water at equilibrium because the rate of total water uptake increases. When there is a constant infiltration rate and no runoff, the isocline has a negative slope (see Appendix).

---

![Resource availability](image)

**Fig. 1.** (A) The zero-isoclines for plant density ($dP/dt = 0$) and resources ($dR/dt = 0$) (soil water or nutrients) illustrated in the phase plane. The zero-isoclines for soil water and nutrients have the same shape and are plotted together. The vectors indicate the direction of change. The dotted line illustrates the path which the system follows during time, given the starting point as indicated. $R^*$ is the minimal amount of resources necessary for plant production. $R_s$ is the equilibrium amount of resources in the absence of plants. The dark circle is a stable equilibrium at $(R^*, P^*)$ and the light circle is an unstable equilibrium at $(R_s, 0)$. Note that $R^* > R_s$. (B) At a high level of herbivory, the system is overgrazed and the zero-equilibrium becomes stable because $R^* > R_s$. The system always shifts to the zero-equilibrium, indicated as a dark circle.
The nutrient isocline is the line joining those combinations of plant density and soil nutrients at which the latter does not change. In absence of vegetation the nutrient uptake rate by the plants is zero and the amount of soil nutrients eventually reaches an equilibrium. With increasing plant density the total rate of nutrient uptake increases. Thus, if a constant proportion of soil nutrients is lost from the system, this isocline has a negative slope too. Note that the shapes of the soil nutrient and soil water isocline are identical (see Appendix). Therefore, we combine the two isoclines; this results in a new isocline which we term “resource isocline” (Fig. 1).

Stability without feedbacks

When the plant and resource isoclines are drawn together in one phase plane as in Fig. 1, this results in a graphical representation of a grazed ecosystem with one or two equilibria, depending on the level of herbivory. Let \( P \) be plant density, \( R \) the amount of resources, \( R^* \) the minimal amount necessary for plant production and \( R_s \) the equilibrium amount of resources in the absence of plants. At low levels of herbivory, when \( R^* < R_s \), two equilibria exist, one unstable boundary equilibrium with no vegetation at \( (R^*, 0) \), and one stable internal equilibrium at \( (R^*, P^*) \) (Fig. 1A and Appendix). The equilibrium at \( P = 0 \) is unstable; any addition of seed (or any propagule) immediately leads to a further increase in plant density. Thus, for all initial conditions with a positive plant density, the system will move to the internal equilibrium.

An increase of herbivory shifts the plant isocline to the right, which leads to a decrease in equilibrium plant density. At high levels of herbivory, when \( R^* > R_s \), there is only one equilibrium, at \( (R^*, 0) \) (Fig. 1B). Consumption of plants by herbivores exceeds plant production, or, in other words, resource levels are insufficient for the plant to compensate for herbivore consumption. This causes the boundary equilibrium to be stable (see Appendix); the system is overgrazed.

Plant soil relations

Shape of the relationships

The mechanisms underlying the effect of vegetation on the capacity of the soil to absorb water and retain nutrients can be described by two simple relationships. Vegetation improves the structural and water-holding properties of the soil by forming root channels, by preventing crust formation through the interception of raindrops, and by stimulating biological activity in the soil, resulting in higher infiltration rates (Glover et al. 1962, Kelly and Walker 1976, Van Wijngaarden 1985, Scholte 1989, Kiepe 1995). A particular shape of the relationship between plant density and infiltration rate is proposed by Walker et al. (1981) (Fig. 2A). Empirical evidence for the shape of this relationship can be found in Van Wijngaarden (1985). The infiltration rate is defined in relative terms as the amount of water entering the soil expressed as a proportion of the amount which enters when plant density is at its maximum. (B) The shape of the feedback relation between the specific nutrient loss rate and plant density based on the results of experiments by Elwell and Stocking (1974, 1976) and Lang (1979).
plant density leads to a lower nutrient loss. Based on the results of experiments (Elwell and Stocking 1974, 1976, Lang 1979) and assuming a linear relationship between soil loss and nutrient loss, the form of the relationship between plant density and specific nutrient loss rate can be expressed graphically (Stocking 1994) (Fig. 2B). Where there is no vegetation, the specific nutrient loss rate is maximal. Initially, the specific nutrient loss rate rapidly decreases as density increases. The effect of increasing plant density on the specific nutrient loss rate diminishes at high biomass values. The interactive processes between plants and soil become increasingly effective at coping with erosive processes, hence the specific nutrient loss rate approaches zero with increasing plant density.

**Effect of plant-soil relations**

If a feedback relation between the rate of infiltration and plant density is incorporated in the water-limitation model, the shape of the soil water isocline in the phase plane alters. At low plant density, any increase in density results in a relatively large increase in the infiltration rate, possibly even exceeding the increase in the rate of water uptake by plants. In this case, the isocline has a positive slope (Fig. 3). Above a certain plant density, however, the effect of increasing plant density on water infiltration declines, and as a result the increase of water uptake by the plants will exceed the increase of water infiltration. Consequently, a hump appears in the soil water isocline (see Appendix). If the feedback relationship between plant density and nutrient loss is incorporated in the nutrient-limitation model, the shape of the nutrient isocline in the phase plane changes likewise. At low plant density, any increase in density results in a relatively large reduction of the specific nutrient loss rate. This results in a nutrient isocline that has a positive slope. Above a certain plant density, however, the decrease in nutrient loss plus the increase in nutrient release from plant mortality becomes lower than the increase in nutrient uptake by the plants. This produces the hump in the nutrient isocline (see Appendix). Again, the shape of the two isoclines are identical, so we can continue the analysis using the general resource isocline (Fig. 3).

**Stability with feedbacks**

When the plant and the humped resource isocline are plotted together in one phase plane as in Fig. 3, a graphical representation of a grazing system is obtained, with one, two or three equilibria, depending on the level of herbivory. At low levels of herbivory (if $R^* < R_1$) the system has two equilibria, comparable with a lightly grazed system in which feedbacks are absent (Fig. 3A, Fig. 1A).

If herbivory is increased, however, a certain threshold ($T_1$) will be crossed at $R^* = R_1$. There are now three equilibria: a stable one without vegetation ($R_1$, 0), a stable one at high plant density ($R^*$, $P^*$), and an unstable one at intermediate plant density ($R^*$, $P^*$) (Fig. 3B) (see Appendix). The two domains with differ-

---

**Fig. 3.** (A) The humped soil water and nutrient isocline drawn together as one resource (R) isocline in a phase plane. Note that $R^* < R_1$. The light circle indicates an unstable equilibrium and the dark circle a stable one. (B) If the level of herbivory crosses a certain threshold $T_1$, where $R^* = R_1$, there are three equilibria: a stable one at $P = 0$ ($R_1$, 0) and one at a high plant density ($R^*$, $P^*$), and an unstable one at a low plant density ($R^*$, $P^*$). (C) At a level of herbivory higher than the threshold $T_2$, the system is overgrazed and always shifts to the boundary equilibrium at $P = 0$, independent of initial conditions.
ent attracting equilibria that occur in the phase plane are separated by a separatrix (dotted line). For initial values of plant density and resource levels under the separatrix, the system shifts to the equilibrium at \( P = 0 \). For initial values above the separatrix, the system shifts to the equilibrium at high plant density. Under these conditions an environmental fluctuation or disturbance (e.g. fire) may carry the plant density or available resource levels below a breakpoint value (indicated by the separatrix), whereupon the system collapses. Note that both the differences in initial conditions and equilibrium resource levels may be very small and undetectable.

At levels of herbivory higher than the threshold \( T_2 \), the system is overgrazed (Fig. 3C). There is only one stable equilibrium, the no-vegetation equilibrium \( (R_i, 0) \). The consumption rate of the herbivores is so high that the system always shifts to that zero-equilibrium, independent of initial conditions. This situation is comparable with an overgrazed situation without feedbacks (Fig. 1B). However, when the level of herbivory is lowered in the situation where feedbacks do not operate, it leads to a continuous increase of plant density at equilibrium. Lowering herbivory to a level between the thresholds \( T_1 \) and \( T_2 \) in the situation where feedbacks operate has no effect at all. The system can only be restored if plant density or resource levels are increased dramatically beyond certain breakpoint values. Hence, the system is stably degraded (cf. Prins 1989).

**Model robustness**

In order to determine the consequences of the assumptions of a linear herbivore functional response and plant growth being limited by resource availability only, we investigated the effect of a non-linear functional response and density dependent plant mortality on the model results. Further, relaxing the assumption that the detritus and available nutrients are one nutrient pool, we investigated the effect of separating the available nutrients from the detritus on the model results by an analysis of the isocline surfaces of a three compartment model. All these models exhibit qualitatively similar behaviour, i.e. the interactions between water infiltration or nutrient retention and plant density may trigger a positive feedback between reduced plant density and reduced resource availability, and lead to a collapse of the system. Therefore, we think that the results derived in this paper are quite robust.

**Discussion**

It is recognized that herbivory affects plant growth by altering the supply of resources for the surviving plants (Owen and Wiegert 1976, McNaughton 1979). Several studies stress the importance of nutrients released from dung (Ruess and McNaughton 1984, Hik and Jefferies 1990) or increased mineralization (Holland and Detling 1990, Holland et al. 1991, Seagle et al. 1992), leading to an increase of plant growth. In this paper, we analysed two relations that cause the opposite, that is, resource levels to drop as a result of herbivory, leading to a reduction of plant growth. These feedback relations potentially give rise to the existence of alternate stable vegetation states and threshold effects in grazed ecosystems even without the effect of a non-linear herbivore functional response (Noy-Meir 1975) or plant competition (Walker et al. 1981).

The results of the model are consistent with well-documented observations of spatial and temporal patterns in grazed ecosystems. Two-phase mosaics (densely vegetated patches regularly alternating with almost bare areas) occur on smooth slopes and even flat landscapes of (semi-)arid lands (Boaler and Hodge 1964, Wickens and Collier 1971, McNaughton 1983, Belsky 1986, Montana 1992, Ludwig and Tongway 1995). Many different composite factors are involved in building and maintaining the two-phase mosaics, e.g., compaction of the soil by large herbivores, fluvial runoff-runon, aeolian saltation-deposition and below-ground biological activities. However, the existence of these patterns can probably be best explained in general terms by the positive feedback of water infiltration (cf. Belsky 1986) or nutrient retention (cf. Ludwig and Tongway 1995) and plant density. McNaughton (1983) and Belsky (1986) further found that a two-phase mosaic in the Serengeti disappeared if large herbivores were excluded. This means that, although the differences in water-infiltration rates between the two discrete vegetational phases was in this case the most important factor causing their stable existence, grazing ultimately triggered the discernible changes. Additionally, examples from grazing systems in Africa, North America and Australia, where overgrazing and soil degradation led relatively rapidly to a low (or bare) vegetation state, showed that improvement could not be attained on a practical time scale by simply lowering the level of herbivory (Sinclair and Fryxell 1985, Friedel 1991, Laycock 1991, Daily 1995, Rietkerk et al. 1996). The degraded grasslands can only be improved if plant density or resource levels are increased dramatically beyond certain breakpoint values by means of human intervention.

The patterns described above do not occur solely in semi-arid grazing systems. Similar vegetation mosaics are reported for arctic coastal plant communities in Canada (Jefferies 1988a, b). While summer grazing by geese, when plant growth is vigorous, results in a grazing lawn of graminoid species, spring grubbing for roots and rhizomes often leads to the destruction of existing plant communities and the creation of “bar-
rens”. Lacking vegetation cover, the soil is eroded and becomes hypersaline, resulting in the removal of organic matter and decreasing mineralization of nitrogen (Srivastava and Jeffries 1996, Wilson and Jefferies in press). The resulting edaphic environment is unsuitable for the colonization and establishment of the original plant communities, at least over considerable time spans.

Walker et al. (1981) modelled the competition between a woody vegetation and grasses for available soil water in a water-limited system. According to them, this competition, combined with the effects of the herbivore functional response and the positive interaction between grass biomass and infiltration rate, may lead to the development of two alternate stable states in semi-arid savannas: one with much woody vegetation, and one with a relatively large biomass of grass and rather little woody vegetation. We focus on a simpler case by omitting many complicating ecological factors that may occur in real semi-arid systems. Our point could therefore even be further generalized. Any positive plant-soil feedback that is strong enough to create a hump in the resource isocline may cause alternate stable states. Consequently, any factor inducing shifts in the position of the plant isocline (e.g. grazing) or resource isocline (e.g. environmental fluctuations) may trigger threshold effects. As plant-soil interactions serve as one of the most influential positive feedback loops in semi-arid grazing systems (Graetz 1991), we believe that our minimal modelling approach provides insights of general importance.

Further understanding of the behaviour of a system where feedbacks operate between plants and resources will help to establish whether the system may collapse under certain conditions, and whether there are warning signs of an imminent collapse. This is important for the management of pastoral or protected savanna areas.

Acknowledgements – We thank Rob de Boer and Frank van den Bosch for many valuable suggestions concerning the formulation and analysis of the model and Marten Scheffer for suggesting writing this paper. We are grateful to Jelte van Andel, Jan Bakker, Roy Behnke, John Fryxell, Jan van Grooten, Jef Huisman, Andrew Illius, Bob Jeffries, Han Olff, Herbert Prins, Marten Scheffer, Mark Stafford Smith, Leo Strooijnder and Franjo Weissing for useful comments and interesting discussions.

References


Appendix

Analytical justification of the graphical results

We analysed the properties of both models discussed above using the following condensed model:

\[
\frac{dR}{dt} = F(R, P) = R_{in}(P) - c(P)R - r(P)R
\]

\[
\frac{dP}{dt} = G(R, P) = g(R)P
\]

where \(R_{in}(P)\) describes the resource input as a function of plant density \(P\), and \(r(P)\) describes the resource losses as a function of \(P\). Both net specific resource uptake by plants \(c(P)\) and specific plant growth \(g(R)\) are monotonically increasing functions of \(R\). Specific plant growth \(g(R)\) equals zero at \(R = R^*\).

In case of the water-limitation model, \(r(P)\) is constant, while \(R_{in}(P)\) is a monotonically increasing function of \(P\) (Fig. 2A). \(R_{in}(P)\) levels off to the value of \(R_{in,max}\) at high values of \(P\). \(c(R)\) is assumed to be zero at zero resource availability \(R\), since water does not recycle. In case of the nutrient-limitation model, \(R_{in}(P)\) is constant, while \(r(P)\) is a monotonically decreasing function of \(P\) (Fig. 2B). \(r(P)\) levels off to the value of \(r_{min}\) at high values of \(P\). \(c(R)\) is assumed to be negative at low values of \(R\), due to recycling of plant nutrients, but is positive at resource levels higher than \(R = R_w\).

General shape of the resource isocline

The resource isocline intersects with the \(P = 0\) axis at the point \(R = R_{in}(0)/r(0)\). The slope of the isocline at \((R_s, 0)\) can be obtained by implicit differentiation of the isocline equation \((dR/dP = 0)\) with respect to \(P\):

\[
\frac{dR}{dP} = \frac{\frac{\partial F}{\partial R}}{\frac{\partial F}{\partial P}} = -\frac{\frac{\partial R_{in}(P)}{\partial P} - c(P)}{\frac{\partial c(P)}{\partial P}} - \frac{\frac{\partial c(R)}{\partial R} P - r(P)}{dR}
\]

The sign of \(dR/dP\) is determined by the numerator, since the denominator is always negative. So, the criterion for an isocline with a positive slope reads:

\[
\frac{dR_{in}(P)}{dP} - \frac{dr(P)}{dP} R > c(R)
\]

(A2)

In ecological terms, the isocline has a positive slope whenever, with increasing \(P\), the increase of net resource input into the system is higher than the increase in consumption. However, the left-hand side of (A2) decreases and approaches zero with increasing \(P\), since \(d^2R_{in}(P)/dP^2 < 0\) and \(dr(P)/dP = 0\) in case of the water-limited system, or \(d^2r(P)/dP^2 > 0\) and \(dR_{in}(P)/dP = 0\) in case of the nutrient-limited system. As \(dc(R)/dP = 0\), it follows that \(dR/dP\) will drop and become negative, resulting in an isocline that has a negative slope at high values of \(P\). Thus, for both models we expect an isocline that has a positive slope with respect to \(P\) at \((R_s, 0)\) whenever criterion (A2) is met, while the slope will decrease with increasing \(P\) and become negative at high values of \(P\).

If \(P\) tends to infinity, the isocline equation \((dR/dP = 0)\) reduces to:

\[
\frac{dR}{dt} = R_{in,max} - c(R)P - r_{min}R = 0
\]

which simplifies to:

\[
P = \frac{R_{in,max} - r_{min}R}{c(R)}
\]

Now \(P\) only approaches infinity if \(c(R)\) approaches zero, since the numerator is always positive at the
resource isocline. In case of the water-limitation model \( c(R) \) equals zero at \( R = 0 \), so the resource isocline approaches \( R = 0 \) as \( P \) approaches infinity. In case of the nutrient-limitation model, \( c(R) \) equals zero at \( R = R_u \), so the resource isocline approaches \( R = R_u \) as \( P \) approaches infinity.

**Stability analysis**
In order to establish the local stability of an equilibrium \((R^*, P^*)\), we investigate the Jacobian matrix of system A1:

\[
J = \begin{pmatrix}
    A_{11} & A_{12} \\
    A_{21} & A_{22}
\end{pmatrix}
\]

\[
= \begin{pmatrix}
    -\frac{dc(R^*)}{dR} R^* - r(P^*) & \frac{dR_m(P^*)}{dP} - c(R^*) - \frac{dr(P^*)}{dP} R^* \\
    \frac{dg(R^*)}{dR} R^* & g(R^*)
\end{pmatrix}
\]

We consider three equilibria: the boundary equilibrium \((R_s, 0)\), and the two internal equilibria \((R^*, P^*_1)\) and \((R^*, P^*_2)\). These equilibria are locally stable if (e.g. Edelstein-Keshet 1988)

1. \( \text{trace}(J) = A_{11} + A_{22} < 0 \),
2. \( \det(J) = A_{11}A_{22} - A_{12}A_{21} > 0 \).

**Boundary equilibrium**
At the boundary equilibrium \((R_s, 0)\) the stability criteria simplify to \( \text{trace}(J) = -r(0) + g(R_s) < 0 \) and \( \det(J) = -r(0)g(R_s) > 0 \). The determinant is positive as long as \( g(R_s) \) is negative, which is a more restrictive criterion for stability than the trace criterion. Hence, the boundary equilibrium is stable whenever \( R_s < R^* \), i.e. whenever the resource isocline intersects the \( P = 0 \) axis to the left of the plant isocline.

**Internal equilibria**
At the resource-plant equilibria \((R^*, P^*_1)\) and \((R^*, P^*_2)\) the specific growth rate of the plant is (by definition) zero, i.e. \( g(R^*) = 0 \). Hence, the stability criteria simplify to \( \text{trace}(J) = -(dc(R^*)/dR) R^* - r(P^*) < 0 \) and \( \det(J) = -(dg(R^*)/dR) R^* [(dR_m(P^*)/dP) - c(R^*) - (dr(P^*)/dP) R^*] > 0 \). Since the trace criterion is always satisfied, and \( (dg(R)/dR) R^* \) is always positive, it follows that any internal equilibrium is stable whenever \((dR_m(P)/dP) - c(R^*) - (dr(P)/dP) R^* < 0 \). This criterion is met whenever the resource isocline has a negative slope (see above). The slope of the resource isocline is negative at \((R^*, P^*_1)\), while it is positive at \((R^*, P^*_2)\). As a result, the equilibrium at \((R^*, P^*_1)\) is stable, while the equilibrium at \((R^*, P^*_2)\) is an unstable saddle point (see, e.g., Edelstein-Keshet 1988).