Optimal control solutions to sodic soil reclamation

Yair Mau\textsuperscript{a,}\textsuperscript{*}, Amilcare Porporato\textsuperscript{a,b}

\textsuperscript{a}Department of Civil and Environmental Engineering, Duke University, Durham, NC 27708, USA
\textsuperscript{b}Nicholas School of the Environment and Earth Sciences, Duke University, Durham, Box 90328, NC 27708, USA

\textbf{A R T I C L E  I N F O}

Article history:
Received 15 January 2016
Accepted 23 February 2016
Available online 11 March 2016

Keywords:
Sodicity
Reclamation
Salinity
Optimal control
Irrigation
Rehabilitation
Amelioration

\textbf{A B S T R A C T}

We study the reclamation process of a sodic soil by irrigation with water amended with calcium cations. In order to explore the entire range of time-dependent strategies, this task is framed as an optimal control problem, where the amendment rate is the control and the total rehabilitation time is the quantity to be minimized. We use a minimalist model of vertically averaged soil salinity and sodicity, in which the main feedback controlling the dynamics is the nonlinear coupling of soil water and exchange complex, given by the Gapon equation. We show that the optimal solution is a bang–bang control strategy, where the amendment rate is discontinuously switched along the process from a maximum value to zero. The solution enables a reduction in remediation time of about 50\%, compared with the continuous use of good-quality irrigation water. Because of its general structure, the bang–bang solution is also shown to work for the reclamation of other soil conditions, such as saline–sodic soils. The novelty in our modeling approach is the capability of searching the entire “strategy space” for optimal time-dependent protocols. The optimal solutions found for the minimalist model can be then fine-tuned by experiments and numerical simulations, applicable to realistic conditions that include spatial variability and heterogeneities.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

In arid and semi-arid areas, where good-quality water for irrigated agriculture is scarce, inadequate management of soil and water resources can lead to soil sodification [1]. Soil sodicity is characterized by a relative high concentration of sodium cations in the exchange complex or in the soil water, causing negative effects on the physical and chemical properties of the soil. Among the physical changes are the breakdown of macroaggregates (slaking), the release of individual clay platelets from aggregates (dispersion), and surface rupturing, which have a detrimental impact on hydraulic conductivity, infiltration rate, seedling emergence and water holding capacity [2–4]. Chemical effects of sodicity include specific ion deficiencies and toxicities [5].

Several amelioration methods exist for reducing the relative amount of sodium in the soil, including the use of chemical amendments (most commonly calcium-based) [6,7], application of high electrolyte water [8], and phytoremediation [9]. For a review on the amelioration of sodic soils see [10]. Sodic soil reclamation can be very costly and resource intensive (in terms of water, amendments, time, etc.), although arguably not as expensive as a “no action” policy [11]. Therefore, from an economic and environmental perspective, it is important to find strategies that optimize resource usage.

The main approaches used so far to tackle this challenge are experiments and computer simulations. On the one hand, field and laboratory experiments have been employed to compare different reclamation and management strategies for specific soil conditions, and determine the ones that “most effectively” reduce sodicity and improved soil structure [12–19]. On the other hand, computer simulations of unsaturated water flow and solute transport have the advantage of being able to examine the effectiveness of different reclamation methods for various soil conditions with little costs and in much faster times, compared to experiments [20–22].

However, both experiments and cumbersome computer simulations have limited probing power with regard to time-dependent reclamation strategies. Of the uncountable ways one can choose to reclaim sodic soils, these two approaches can usually only compare the outcomes of changing one or more parameters (called control parameters), while keeping them fixed in time throughout the experiment or simulation.

In order to be able to compare a continuum of scenarios and reclamation protocols, in this paper we make use of optimal control theory. This allows us to find optimal reclamation strategies in the entire “strategy space”, where the control parameter can be continuously changed in time. For this, we introduce a minimalist model for the dynamics of soil salinity and sodicity [23], that
is amenable to analysis and that is suitable for solving the optimal control problem, while retaining the essential elements of the physical problem. In this model, we assume that irrigation is the main water input (e.g., negligible precipitation during a Mediterranean dry season), and that the agricultural soil is relatively homogeneous, so that the variables may be averaged over the spatial dimensions. Our main goal here is to show that optimal control theory can be a very useful tool in the problem of degraded soil reclamation, and the specific model in question provides a convenient test case for our approach.

Optimal control theory has been used in many modern environmental challenges, such as the management of soil and water resources [24], soil erosion [25], pollution [26], forest carbon sequestration [27] and urban drainage systems [28]. In the case of agricultural systems, its use has been mainly focused on irrigation water allocation [29–31] and greenhouse management [32,33]. To the best of our knowledge, however, this is the first study on optimal control of sodic soil reclamation.

Within the broader topic of sodic soil reclamation, this work concerns the problem of finding a rehabilitation protocol that allows the reclamation of a sodic soil in the least possible time, by means of irrigation with calcium-based amendments. Thus, the amelioration process is viewed as an optimal control problem [34]: the amount of calcium added to irrigation water is the control, and the optimal strategy is the time-dependent addition of calcium that takes the system from a sodic condition to a desired “normal” soil target, while minimizing a “cost”, in our case, the total time. In the course of this paper we will show that the optimal strategy is able to cut by about half the reclamation time (also the amount of irrigation water used), compared to simply using good-quality water to flush sodium cations from the root zone. Although highly idealized, this theoretical estimate provided by control theory offers an upper bound that can serve as a benchmark in actual rehabilitation efforts in the field [10].

The paper is structured as follows. Section 2 describes the model for the dynamics of soil salinity and sodicity. In Section 3 the model is linearized and, by using the tools of optimal control theory, the optimal strategy for the control is calculated analytically. Appendix A, at the end of the paper, describes the more cumbersome arguments involved in the analysis of Section 3. Finally, in Section 4, we discuss the conclusions and future directions.

2. Dynamics of salinity and sodicity in the soil

We will briefly review the differential equations that govern the dynamics of soil salinity and sodicity. A detailed description of the model can be found in [23].

The dynamics of salt and sodium cations in the root zone is modulated by the dynamics of water in the soil. We assume here that irrigation is the dominant input of water. This assumption is reasonable when precipitation is negligible, as in a Mediterranean dry season or in case of a greenhouse, and that there is no upward flux of water from the water table due to capillarity. The irrigation rate, denoted by I, is constant in time, and its electrolyte concentration is C_{I}. For analytical tractability, we also consider the evapotranspiration rate T to be constant in time, a reasonable assumption in irrigated soils. Both I and T are given in mm/d or L/m²/d. We model here a flat and relatively homogeneous agricultural plot, therefore all the variables are vertically lumped over the rooting depth Z_r = 0.4 m. For moderate irrigation rates, no ponding will occur, and the steady-state percolation reads simply I – T. Writing the percolation function as L(s) = rK_s s^{c} [35], we can find the steady-state soil water volume per unit area w (L/m²) as

\[ w = n Z_r \left( \frac{I - T}{rK_s} \right)^{1/c}, \]

where \( n \) is the porosity, \( K_s \) is the saturated hydraulic conductivity, and the parameter \( c \) depends on the soil properties [36]. The parameter \( r \) modulates the saturated hydraulic conductivity as the salinity and sodicity of the soil change. We will leave the discussion on this dependence to Section 2.2, after the salinity and sodicity variables are introduced.

By writing the balance equation for salt, represented by the total amount of sodium and calcium cations in the root zone, one obtains an equation for the electrolyte concentration of soil water C [23],

\[ \frac{dC}{dt} = \frac{K_I}{w} \frac{I - T}{w} C, \]

where both C and \( C_I \) are in millimoles of charge per liter \(^1\), or mmol/L. The first term in the right-hand side of Eq. (2) is the salt input due to irrigation, and the second term is the salt output due to water percolation \( I - T \) to deeper soil layers. We consider the rehabilitation efforts to take place before the growing season, therefore the equation above does not include salt uptake by plants.

This linear equation has a characteristic time scale \( t_c = w/(I - T) \) in which the system converges to its steady state \( C^* = K_C/(I - T) \).

Cations in the soil water can replace readily exchangeable cations adsorbed to soil particles, in a process called cation exchange. This reaction has a time scale of minutes to hours, while salinity and sodicity processes take place at much longer time scales of weeks to months. For this reason, we assume the exchange reaction to be in a local thermodynamic equilibrium, which in turn warrants the use of the well known Gapon equation [23,37], linking the sodium cation in the soil water to that in the exchange complex.

The relative amount of sodium in the exchange complex is simply the ratio of sodium cation concentration to the total concentration of adsorbed cations (or the cation exchange capacity, \( C_{exc} \)). This quantity is called the equivalent fraction of sodium in the exchange complex, denoted here by E, or when expressed as a percentage, the exchangeable sodium percentage (ESP). The relative amount of sodium in the irrigation water is the equivalent fraction \( E_I \), while the equivalent fraction of sodium in the soil water is denoted here by \( E_c \).

Using the equivalent fractions of sodium defined above, the Gapon equation can be written as

\[ \frac{E}{1 - E} = K_E \sqrt{2C} \frac{E_I}{\sqrt{1 - E_I}}, \]

where \( K_E \) is the Gapon selectivity coefficient. The value of \( K_E \) is soil-specific, ranging from 0.0072 to 0.01740 (mmol/L)^{-1/2} [38,39]. Here we use \( K_E = 0.01475 \) (mmol/L)^{-1/2}, which represents a mean behavior of 59 soil samples of varied origin, as reported in 1954 by the United States Salinity Laboratory Staff [40]. In this report, Eq. (3) is presented as \( E = K_E \) SAR, where \( E = F(1 - E)^{-1} \) is the exchangeable sodium ratio, and \( SAR = [Na^+] / [Ca^{2+}] \) is the sodium adsorption ratio, and the brackets denote molar concentration in mmol/L. The sodicity of a solution is usually measured by the SAR, but in this paper we use the equivalent fraction of sodium for modeling convenience. However, the SAR can be easily related to the right-hand side of Eq. (3), by using the expression \( SAR = \sqrt{2CE_r(1 - E_r)^{-1/2}} \), where \( C_r \) and \( E_r \) are the electrolyte concentration and exchangeable sodium fraction of the solution z in question (either soil solution or irrigation water, in our case).

---

1 Note that the quantity mmol/L coincides with the SI units mol/m³.
Writing the balance equation for sodium cations, and using Eq. (3) to eliminate the variable $E_s$, one obtains [23]

$$\frac{dE}{dt} = \left\{ I C E - (I - T) C E (C, E) \right\} - \frac{dC}{dt} \left[ w E (C, E) + wC \frac{\partial E (C, E)}{\partial C} \right] \times \left[ C_{ec} M + wC \frac{\partial E (C, E)}{\partial E} \right]^{-1}.
$$

where

$$E (C, E) = 2 \left( 1 + \sqrt{1 + 8 K_s C (1 - E^{-1})^2} \right)^{-1}
$$

is achieved by solving Eq. (3), and $M$ denotes the dry soil mass. Table 1 shows the symbols for all the variable and parameters, together with their definitions, units, and default values for the parameters.

Eqs. (2) and (4) constitute the salinity–sodicity system. It can be shown [23] that they admit one linearly-stable fixed point

$$C^* = \frac{I C}{I - T}$$

$$E^* = \left( 1 + \sqrt{\frac{(I - T)(1 - E_s)}{2 I C} \frac{1}{K_s E_s}} \right)^{-1}.
$$

The flow in phase space ($C$, $E$) defined by Eqs. (2) and (4) is shown in Fig. 1 by gray arrows, together with the stable fixed point ($C^*$, $E^*$), denoted by the black circle. The nullclines, $dC/dt = 0$ and $dE/dt = 0$, are represented by the black solid curve and black dashed curve, respectively. The irrigation parameters $C_r$ and $E_r$ used in this graph put the steady state in the point ($17$, $0.20$) of the phase space, thus characterizing a sodic soil. The flow in phase space is predominantly in the horizontal direction, suggesting that in this case the dynamics of salinity $C$ is much faster than that of sodicity $E$.

### 2.1. Calcium flux as a control parameter

In the derivation of the equations above, we assumed that sodium (Na$^+$) and calcium (Ca$^{2+}$) are the only cations that participate in the exchange reaction. This assumption is valid for soils with negligible potassium levels, and because the Gapon equation treats calcium and magnesium indistinguishably with respect to the cation exchange [41], these divalent cations are effectively lumped together in our description of Ca$^{2+}$. Fig. 2a shows the exchange isotherm for Na$^+$ and Ca$^{2+}$, based on the Gapon Eq. (3), where it can be seen that calcium has a larger affinity for the exchange complex, compared to sodium cations. For a soil water concentration of 20 mmol/L, a relative concentration of about 95% of Na$^+$ in the soil water means that the relative concentration of Na$^+$ in the exchange complex is only 28% (denoted by the black pentagon).
Dissolving calcium cations into irrigation water aims at decreasing the relative concentration of Na⁺ in the soil water, and thus reducing sodicity levels in the exchange complex. As an illustration, the gray diamond in Fig. 2a represents the condition after a given amelioration process with calcium, where the relative concentration of Na⁺ in the soil water was reduced to about 47%, while the electrolyte concentration C was increased to 70 mmol/L, thus bringing the sodicity in the exchange complex to ESP = 10%. This concomitant increase in C and decrease in E is clearly mediated by a change in the irradiation water quality.

Denoting by u (mmol/L) the amount of calcium cations added to irrigation water, and calling \( C_I \) and \( E_I \) the respective electrolyte concentration and equivalent fraction of sodium of unamended irrigation water (i.e., for \( u = 0 \)), the dependence of the irrigation parameters \( C_I \) and \( E_I \) on \( u \) is

\[
C_I(u) = \frac{IC_I + u}{I} = \frac{C_I}{1 + u/C_I}, \tag{7a}
\]

\[
E_I(u) = \frac{IE_I}{IC_I + u} = \frac{E_I}{1 + u/C_I}. \tag{7b}
\]

Eq. (7a) shows that the total cation flux \( IC_I \) from irrigation water is simply the sum of the cation flux of the unamended irrigation water \( IC_I \) plus the calcium flux \( Lu \) added to it. Eq. (7b) follows from the fact that the equivalent fraction of Na⁺ in the irrigation water is the flux of sodium \( IC_I E_I \) (clearly not affected by the calcium flux \( u \)) divided by the total cation flux \( IC_I \). Both functions \( C_I(u) \) and \( E_I(u) \) are shown in Fig. 2b and c, respectively. The good-quality irrigation water used here has \( C_I = 15 \text{ mmol/L} \) and \( E_I = 0.40 \), which translates to SAR = 2.8.

It is important to note that \( u \) is the amount of calcium cations dissolved in the irrigation water and stay dissolved in the soil water. The choice of using different calcium-based amendments such as calcium chloride (CaCl₂), calcium carbonate (CaCO₃), and calcium sulfate (CaSO₄), should take into account their cost, relative effectiveness in soil rehabilitation, and application difficulty [10]. Most importantly for this study, the different amendments have widely varying solubilities in water, so we define \( u_{\text{max}} \) as the maximal amount of Ca that can be dissolved in the irrigation water, and stay dissolved once it is applied to the soil. This quantity will appear as an important control constraint in deriving an optimal solution for the rehabilitation problem (Section 3). Throughout the paper we will use \( u_{\text{max}} = 20 \text{ mmol/L} \), a reasonable value assuming a non-calcareous soil and choosing gypsum as the calcium-based amendment.

By substituting the expressions for \( C_I \) and \( E_I \) found in Eq. (7) into Eqs. (2) and (4), we obtain a system that depends linearly on the control parameter \( u \):

\[
\frac{dC}{dt} = \frac{IC_I + u}{w} - \frac{I - T}{w} C \tag{8a}
\]

\[
\frac{dE}{dt} = \left[ IC_I E_I - (I - T)C E_I(C, E) \right] - \frac{dC}{dt} \left[ wE_I(C, E) + wC \frac{\partial E_I(C, E)}{\partial E} \right] \times \frac{C_{\text{calc}} + wC \frac{\partial E_I(C, E)}{\partial E}}{\partial E}. \tag{8b}
\]

The control parameter \( u \) appears linearly in Eq. (8a) and in \( dc/dt \) in Eq. (8b). The first term \( IC_I E_I \) in Eq. (4) became \( IC_I E_I \) in Eq. (8b). This latter result is somewhat intuitive, since \( IC_I E_I \) is the flux of sodium cations due to the irrigation water, which in unaffected when a flux \( Lu \) of calcium cations is incorporated to it.

In order to simplify the notation, we define the dynamical variable vector \( x = (C, E)^T \), also called state vector, and thus Eq. (8) can be rewritten as

\[
\frac{dx}{dt} = \mathbf{h}(x, u), \tag{9}
\]

where function \( \mathbf{h} = (h_1, h_2)^T \) is composed of the right-hand sides of Eqs. (8a) and (8b), respectively.

2.2. Varying saturated hydraulic conductivity

Soil structure can be greatly influenced by salinity and sodicity conditions, and among the most important effects is the decrease in saturated hydraulic conductivity \( K_s \) for higher exchangeable sodium fraction \( E \) or lower electrolyte concentrations \( C \) [42,43]. In the equations introduced above, we assumed a steady-state water content \( w \), but it can vary as the system evolves in time, especially given the link between soil water conductance and the variables \( C \) and \( E \). This steady-state assumption is justified in two ways.

First, the equilibration time scale for soil water is much faster than the processes of salinization and sodification, so effectively it can be described merely as a function of \( C \) and \( E \) (in the dynamical-system parlance, \( s \) can be called an “enslaved” variable). The second reason pertains to the little change in \( s \) even when relative saturated hydraulic conductivity \( r \) changes significantly. Fig. 3a shows how \( r \) changes in the phase space \((C, E)\), for a sandy loam soil [44]. The triangle denotes the sodic soil initial condition \((C, E) = (10, 0.30)\) used throughout the paper, while the square indicates the target \((17.1, 0.08)\) for reclamation. While the relative saturated hydraulic conductivity of the sodic soil and the target state are very different \((r = 0.35 \text{ and } r = 0.72, \text{ respectively})\), their steady-state relative soil moisture do not differ much. Fig. 3b shows the percolation function \( L(s) \) for these two states. As long as the water input \( I - T \) (denoted by the dotted line) does not surpass the value \( rK_s \), the daily percolation will equal \( I - T \), and the soil moisture will change according to \( L(s) \). As shown in the graph, even a large increase in \( r \) (from 0.35 to 0.72), causes \( s \) to vary by only 5.5% (from 0.849 to 0.804).

3. Optimization

Our goal is to rehabilitate a sodic soil in minimal time, and at the end of the process have a non-saline, non-sodic soil, which we call here “normal” soil. Minimization of total time is chosen because reclamation costs can be recovered sooner if the soil is fully productive at an earlier date. A complete economic analysis would
be required to determine the best relative allocation of resources (time, applied water volume, amount of amendment, etc.). Because of the assumption that irrigation rate $I$ is constant, in this paper, minimal time is equivalent to minimal water usage.

For simplicity, we divide the phase space $(C, E)$ into four quadrants, each representing a soil type. They can be visualized in Fig. 4a, starting from the bottom left quadrant, in the clockwise direction: normal soil, $C < 40$ mmol/L, $E < 0.15$; sodic soil, $C < 40$ mmol/L, $E > 0.15$; saline-sodic soil, $C > 40$ mmol/L, $E > 0.15$; and saline soil, $C > 40$ mmol/L, $E < 0.15$.

3.1. No control

We start from a sodic soil state $x_1 = (10, 0.30)$. In conditions where the irrigation parameters are $C_l = C_l = 15$ mmol/L and $E_l = E_l = 0.40$, i.e., there is no control (no $Ca^{2+}$ is added to irrigation water, $u = 0$), the resulting steady state $x^*$ is in the normal soil quadrant (see Fig. 4a). The dashed red curve shows a simulated trajectory that starts from $x_1$ and converges to $x^*$. Fig. 4b shows the evolution of $C$ and $E$ for this trajectory as function of time, where the vertical axis on the left corresponds to $C$ (dark curve), and the vertical axis on the right corresponds to $E$ (light curve). Note that both axes were rescaled so that the initial and final position of the two curves coincide.

An analysis of the time scales associated with this process [23] shows that salinity $C$ converges to its steady state $C^*$ on a time scale about 2.9 times faster than that related to the convergence of sodicity $E$ to its steady state $E^*$, which is attested by Fig. 4b.

The target for sodicity reduction is set to $E = 0.08$. Following the dashed trajectory shown in Fig. 4, it takes approximately 41.6 days for the system to reach this target, marked with a square at point $x_2 = (17.1, 0.08)$. The $\simeq$ symbol indicates that only asymptotically the electrolyte concentration reaches the steady-state $C^* = 17\frac{1}{2}$ mmol/L.

As discussed in Section 2.2, every point in the phase space $(C, E)$ is associated with a relative saturated hydraulic conductivity $r$. Therefore, our ultimate goal of setting point $(17.1, 0.08)$ as the target state is not merely the reduction of sodicity from $E = 0.30$ to $E = 0.08$, but rather a doubling of the hydraulic conductivity (see Fig. 3), which will have positive effect on agriculture.

Introducing a time dependent control $u(t)$ results in different trajectories in the phase space, starting from $x_1$. Among the infinite control possibilities that take the system from $x_1$ to $x_2$, the optimal control $u^*(t)$ is the one that does so in the least possible time.

3.2. Linearization

Eq. (9) is nonlinear, and it is not possible to reach an analytical solution $x(t)$, even for a constant control $u$. One can approximate the system’s dynamics by linearizing Eq. (9) around its steady-state $x^*$, yielding an equation that is amenable to analysis. As will be seen, the solution of the linearized problem does not deviate much from the behavior of the nonlinear system [23]. In addition, it is very convenient to work with a linear system from an optimal control standpoint, for there is a vast literature on this topic [34,45–47], with clear steps to follow in order to achieve an analytical optimal solution.

Accordingly we perform some simplifications to Eq. (9) with regard to two sources of nonlinearities. The first simplification is the
linearization of Eq. (9) with regard to \( \mathbf{x} \), around the stable fixed point \( \mathbf{x}^* \), according to

\[
\frac{d\mathbf{x}}{dt} = J'(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) = g(\mathbf{x}; u)
\]

(10)

where \( J'(\mathbf{x}^*) \) is the \( 2 \times 2 \) Jacobian matrix \( \partial h/\partial \mathbf{x} \) calculated at \( \mathbf{x}^* \). Because \( h_1 \) is a linear equation in \( C \), we have that \( g_1(\mathbf{x}; u) = h_1(\mathbf{x}; u) \), and effectively the procedure above only linearizes \( h_2 \).

Before the linearization, the functions \( h_1 \) and \( h_2 \) had a linear dependence on the control \( u \), as discussed just after Eq. (8). However, the linearization of \( h_2 \) made \( g_2 \) nonlinear with respect to \( u \). To see this consider that the element of the Jacobian \( J_{22} = \partial h_2/\partial \mathbf{x} \mathbf{x}^* \) is extremely convoluted, and the simple linear relation between \( h_2 \) and \( u \) now becomes an equally convoluted relation between \( g_2 \) and \( u \). To avoid this, we carry out a second linearization around the point \( u = 0 \), as follows

\[
\frac{d\mathbf{x}}{dt} = g(\mathbf{x}; u = 0) + \frac{\partial g(\mathbf{x}; u)}{\partial u} \bigg|_{u=0} u
= \begin{pmatrix}
    f_1(\mathbf{x}, u) \\
    f_2(\mathbf{x}, u)
\end{pmatrix}
= f(\mathbf{x}, u)
\]

(11)

The system above is now linear both in the dynamical variables \( \mathbf{x} \) and in the control parameter \( u \). We can rewrite the right-hand side \( f(\mathbf{x}, u) \) of Eq. (11) as

\[
\frac{d\mathbf{x}}{dt} = f(\mathbf{x}, u) = A(\mathbf{x}) + B(\mathbf{x})u,
\]

(12)

where the functions \( A(\mathbf{x}) \) and \( B(\mathbf{x}) \) read

\[
A(\mathbf{x}) = \begin{pmatrix}
    A_0^0 + A_1^1 C \\
    A_1^0 + A_2^2 C + A_2^2 E
\end{pmatrix}
B(\mathbf{x}) = \begin{pmatrix}
    B_0^0 \\
    B_1^0 + B_2^2 + B_2^2 E
\end{pmatrix}.
\]

(13)

Using the parameter values shown in Table 1, the coefficients in (13) are

\[
A_0^0 = 4.11 \\
B_0^0 = 0.27 \\
A_1^0 = -0.24 \\
B_1^0 = 0 \\
A_0^0 = 2.97 \cdot 10^{-3} \\
B_0^0 = 7.76 \cdot 10^{-6} \\
A_0^0 = 2.27 \cdot 10^{-17} \\
B_0^0 = -1.93 \cdot 10^{-18} \\
A_0^0 = -6.95 \cdot 10^{-2} \\
B_0^0 = -3.88 \cdot 10^{-3}.
\]

(14)

Note that \( f_2(\mathbf{x}, u) \) depends on \( C \) through the terms \( A_2^2 C \) and \( B_2^2 C \) in \( A(\mathbf{x}) \) and \( B(\mathbf{x}) \), respectively, where \( B_2^2 \) and \( B_2^2 \) are several orders of magnitude smaller than the other coefficients. It is thus possible to neglect these coefficients, thus effectively decoupling the two linear equations with respect to the variables \( C \) and \( E \), while still leaving both equations linearly dependent on the control \( u \).

### 3.3. A bang–bang optimal control

The task of rehabilitating a sodic soil in minimal time is now translated to a suitable mathematical language: find the time-dependent control \( 0 \leq u(t) \leq u_{\text{max}} \) that takes the two-dimensional linear system of Eq. (12) from state \( \mathbf{x}_1 \) to \( \mathbf{x}_2 \), while minimizing the cost functional \( J = t_f \), that is, the total time.

It can be shown with optimal control theory that this problem has a bang–bang optimal control [45]. This means that the optimal control \( u^*(t) \) assumes only the extremal values in the range \( 0 \leq u \leq u_{\text{max}} \), with discontinuous switches between them. This behavior follows from two properties Eq. (12): (a) it is an input-affine system, i.e., the right-hand side \( f(\mathbf{x}, u) \) is linear with regard to the control \( u \), and (b) it is a normal system, i.e., there is a switching function \( \psi(t) \) that equals zero only at isolated instants of time [45].

Appendix A shows the derivation of the time-optimal control \( u^*(t) \) and its associated time-optimal solution \( \mathbf{x}^*(t) \). From \( t = 0 \) to a switching time \( t = t_f \), the optimal control is \( u^* = u_{\text{max}} \), resulting in the trajectory \( \mathbf{x}^*(t) \). As the name suggests, at the switching time \( t = t_f \) the control is “switched off” to \( u^* = 0 \), resulting in the second part of the trajectory, \( \mathbf{x}^*(t) \), that arrives at the target state \( \mathbf{x}_2 \) at \( t = t_f \).

Fig. 5a shows in solid black the time-optimal trajectory \( \mathbf{x}^*(t) \) in the phase space \( (C, E) \). The first part of the trajectory \( \mathbf{x}^*(t) \) takes the system from a sodic soil quadrant to a region near the ESP=15% threshold (and a much higher electrolyte concentration \( C \)), and the second part of the trajectory \( \mathbf{x}^*(t) \) then takes the system to the target \( \mathbf{x}_2 \) in the normal soil quadrant, in a total of 19.9 days. The red dashed curve indicates the no-control case, i.e., no calcium addition \( (u(t) = 0) \), where the trajectory starts at \( \mathbf{x}_1 \) and reaches the target after 41.6 days. The red circle indicates day 19.9 in the no-control trajectory.

The optimal control protocol reduces the amount of time required to decrease sodicity to \( E = 0.08 \) to about half of the length of time necessary for the scenario without control.

Fig. 5b and c respectively show the evolution of the variables \( C \) and \( E \) as function of time. For this initial condition, a rapid decrease in sodicity \( E \) can only be accomplished by an initial fast increase in salinity, which in the second part of the solution can then be rapidly remediated by the flushing of salt from the root zone using good-quality irrigation water.

Also plotted in Fig. 5a is the optimal control \( u^*(t) \) applied to the original nonlinear system Eq. (9). The dotted black curve shows that despite the linearizations carried out in Section 3.2, the trajectory of the nonlinear system follows quite closely that of the linear.
system, arriving at the end of the control process at a neighborhood of the target state $x_2$.

Fig. 6a shows two different (albeit close) target states. The optimal trajectory cannot reach the target $x_2$ described in Section 3.1, because it lies to the left of the nullcline $h_1(x,0) = 0$. We therefore set the point (17.6, 0.08) (to the right of the nullcline) as the target for the optimal control. The proximity between the targets of the no-control trajectory and of the optimal control trajectory justifies the comparison between their rehabilitation times above.

The circle in Fig. 5a indicates day 19.9 in the no-control trajectory, calling our attention to the very slow dynamics on the $E$ axis (vertical direction) near the $f_1(x,0) = 0$ nullcline (not shown to avoid clutter, consult instead Fig. 4a). The advantage of the optimal control is that, by taking a "zig-zag" path in the phase space, it is able to avoid this slow dynamics.

### 3.4. Generalization for other initial conditions

Because of the bang–bang solution obtained, our results can be easily generalized for other initial soil conditions in the $(C, E)$ plane. The key lies in constructing the switching curve associated with the target $x_2$. Fig. 6a shows the switching curve (thick black), dividing the phase space into two regions. For initial conditions above it, the optimal control is $u^* = u_{\text{max}}$ and the system evolves along the trajectories denoted by thin lines, until it meets the switching curve, when the control is then switched off to $u^* = 0$.

After the switch, the system will evolve along the switching curve until it reaches the target $x_2$. The target $x_2$ is unreachable to initial conditions that lie below the switching curve.

The switching curve is formed by two segments. The segment to the right of $x_2$ is the solution of $dx/dt = f(x, u_{\text{max}})$ that passes through the point $x_2$, while the segment on the left is the solution of $dx/dt = f(x, u_{\text{max}})$ that passes through the point $x_2$. We can now reinterpret $x^*(t)$ shown in Fig. 5a: $\ddot{x}(t)$ is simply the solution of $dx/dt = f(x, u_{\text{max}})$ that passes through the point $x_1$, while the second part $\dddot{x}(t)$ is just a segment of the switching curve.

The same reasoning is valid for other initial conditions, e.g., Fig. 6b shows the optimal trajectory that takes the system from a saline-sodic condition, at point $x_1 = (70, 0.35)$ (denoted by a star), to the target $x_2$. In solid black we have the trajectory of the optimal solution that starts at $x_1$, arriving at the target in 22.8 days, while the no-control trajectory, shown by the red dashed curve, arrives at the target in 48.2 days. Once more, the optimal control $u^*(t)$ applied to the nonlinear system of equations yields a trajectory (black dotted curve) that does not deviate significantly from the trajectory of the controlled linear system (black solid curve).

### 4. Conclusions

The tools of dynamical systems and optimal control theory were applied here to the environmental problem of reclamation of a sodic soil. Because of the simplicity of the salinity–sodicity model used in this paper, the optimal control strategy presented should be viewed as giving a first-order estimate for the duration of the calcium amendment application and for the duration of application of fresh water in real conditions, where the problem is complicated by spatial and temporal heterogeneities not included explicitly here. However, the very simplicity of our minimalist model is also the strength of this approach: it dispenses with cumbersome numerical calculations, allowing us to shed light on the response of salt dynamics in the soil to variations in irrigation water quality (i.e., the control protocol).

The modeling approach presented here can be of best use when combined with traditional approaches such as numerical simulations and experiments. These approaches can help fine-tune the optimal strategy in order to find the best parameter values for each specific case. We believe that this dialog between different approaches to the problem of sodic soil reclamation can be very fruitful, each approach contributing with its strengths. This paper is an early step to bridge the knowledge gap in the analytical modeling aspects of the problem.

In field conditions, the reduction in rehabilitation time should actually be even greater than that found in this paper. The dynamics of water in the root zone, and most importantly, the dependence of the hydraulic conductivity on salinity and sodicity levels, when taken into account, indicate that it is the ‘longer path’ in phase space that increases the ability of irrigation water to flush salt from the root zone, when compared to a ‘shorter path’ that is located in a low hydraulic conductivity region of phase space, i.e., the sodic quadrant. The role of the feedback between soil salinity–sodicity conditions and water dynamics within the context of soil rehabilitation will be our focus in future studies.

It is also interesting that the bang–bang solution found here is in agreement with existing amelioration techniques, such as the high-salt water dilution method [10]. This method consists in applying to the soil high electrolyte water with a high relative fraction $R$ of divalent cations (in the notation used here, $R = 1 - E_1$). The soil is then equilibrated with successive dilutions of this saline water, until salinity, sodicity and the soil physical properties reach acceptable conditions. The bang–bang protocol presented here, achieved through the explicit modeling of the dynamics of soil salinity and sodicity, can be a useful guide for soil rehabilitation, while reducing the depth of water required in the amendment process and minimizing the total amelioration time.

In summary, by analyzing the coupled dynamics of salinity and sodicity, and their associated time scales, it becomes possible to...
devise strategies to rehabilitate degraded sodic and saline-sodic soils in an optimal fashion. Secondly, doing so with a model amenable to analytical techniques is helpful to sharpen our intuitions for more complex problems.

Acknowledgments

YM acknowledges support from BARD, the United States-Israel Binational Agricultural Research and Development Fund, Vaadia- BARD Postdoctoral Fellowship Award No FI-517-14. AP acknowledges NSC Grants: CBET 1033467, EAR 1331846, EAR 1316258, FESD 1338694, as well as the US DOE through the Office of Biological and Environmental Research, Terrestrial Carbon Processes program (de-sc0006967), the Agriculture and Food Research Initiative from the USDA National Institute of Food and Agriculture (2011-67003-30222), USDA Agricultural Research Service through cooperative agreement 58-6408-3-027. The Mathematica code used to calculate the optimal trajectories is available in the supplementary material.

Appendix A. Derivation of time-optimal solution

Here we show a concise derivation of the solution for our sodic soil reclamation problem. For a detailed discussion on optimal control and on Pontryagin’s maximal principle the reader is referred to [34,45].

For this time-optimal fixed-endpoint control problem, the cost functional \( J(u) \) simply reads

\[
J(u) = \int_{t_0}^{t_f} L(u,t) dt = t_f, \tag{A.1}
\]

where the final time \( t_f \) is to be minimized, and thus the Lagrangian (or running cost) is \( L = 1 \).

The Hamiltonian is defined as

\[
H(x, p, u) = (p, f(x,u)) + p_0L, \tag{A.2}
\]

where the brackets \( \langle \cdot, \cdot \rangle \) denote the inner product; \( p = (p_1, p_2)^T \) is the adjoint vector (also called costate), and \( p_0 \leq 0 \) is the abnormal multiplier, which can be normalized to \( p_0 = -1 \). The canonical equations prescribe the evolution of the optimal state trajectory \( x^* \) and costate trajectory \( p^* \), according to

\[
\frac{dx^*_i}{dt} = \frac{\partial H(x^*, p^*, u^*)}{\partial p_i}, \tag{A.3a}
\]

\[
\frac{dp^*_i}{dt} = -\frac{\partial H(x^*, p^*, u^*)}{\partial x_i}, \tag{A.3b}
\]

where the superscript \( * \) indicates that the variables are calculated for the optimal control \( u^*(t) \). Solving Eq. (A.3) gives

\[
x^*_1(t) = x_1(t) e_1(t, \tau) + \left[ A^0_1 + B^0_1 u^*(t)(e_1 - 1) \right] / A^1_1 \tag{A.4a}
\]

\[
x^*_2(t) = x_2(t) e_2(t, \tau) + \left( A^0_2 + B^0_2 u^*(t) \right) e_2(1) / \left( A^2_2 + B^2_2 u^*(t) \right) \tag{A.4b}
\]

\[
p^*_1(t) = p_1(t) / e_1(t, \tau) \tag{A.4c}
\]

\[
p^*_2(t) = p_2(t) / e_2(t, \tau). \tag{A.4d}
\]

where

\[
e_1(t, \tau) = \exp \left[ A^1_1(t - \tau) \right] \tag{A.5a}
\]

\[
e_2(t, \tau) = \exp \left[ \left( A^2_2 + B^2_2 u^*(t) \right)(t - \tau) \right]. \tag{A.5b}
\]

and \( x(\tau), p(\tau) \) are the boundary conditions for the state \( x \) and costate \( p \) at time \( \tau \).

Because in our case \( u^*(t) \) is a bang–bang control, we can simply solve Eqs. (A.4a) and (A.4b), assuming the trajectory \( x^*(t) \) is formed by consecutive segments with alternating values of \( u^* = 0 \) and \( u^* = u_{max} \), and then “stitch” them together to ensure continuity in \( x \). However, we still do not know how many switches occur along the trajectory, the switching times, and the initial value of the control \( u^*(0) \). For that, we need to study the switching function \( \psi \), given by

\[
\psi(t) = \left\{ \begin{array}{ll}
\langle p^*(t), B(x^*(t)) \rangle = B^0_1 p^*_1(t) + \left( B^0_2 + B^2_2 u^*(t) \right) p^*_2(t) & (A.6)
\end{array} \right.
\]

The value of \( u^* \) will then vary according to

\[
u^*(t) = \begin{cases} 
0 & \text{if } \psi(t) < 0 \\
\text{undetermined} & \text{if } \psi(t) = 0 \\
\text{u}_{\max} & \text{if } \psi(t) > 0.
\end{cases} \tag{A.7}
\]

The condition above is a result of Pontryagin’s Maximum Principle, which says that the Hamiltonian \( (A.2) \) has a global maximum for \( (x^*, p^*, u^*) \) [34].

We start analyzing \( \psi(t) \) by substituting Eq. (A.4) into Eq. (A.6), so that \( \psi \) now depends explicitly on time \( t \), the control \( u^*(t) \), and on the boundary conditions \( x(t) \) and \( p(t) \):

\[
\psi = \psi(t, x^*(t), p^*(t), u^*(t)) \tag{A.8}
\]

Because in our approach we solve a fixed-endpoint control problem, i.e., the target soil condition is a well-defined point in the space, the values for \( x(0) \) and \( x(t_f) \) can be used as boundary conditions. However, we cannot know in principle the values that \( p(t) \) assumes, we only know that it is nonzero for all \( t \) [34]. We use the fact that for free-time problems, the Hamiltonian \( H = 0 \) along the optimal trajectory described by \( (x^*, p^*, u^*) \), so we can express \( p_1(t) \) as a function of \( p_2(t) \), and thus eliminate \( p_1(t) \) in expression (A.8), so we finally have \( \psi = \psi(t, p_2(t), t, u) \). A careful analysis of this expression yields that there is only one switching of control, and that the initial value \( u^*(0) = u_{\max} \).

With this information, one can find the optimal trajectory \( x^*(t) \) Eqs. (A.4a) and (A.4b) are evaluated for the two parts of the trajectory, namely, the first part \( \vec{x}^*(t) \) is evaluated for \( = 0, \ x(0) = x_1 = (10, 0.30) \) and \( u^*(t) = u_{\max} \); and the second part \( \hat{x}^*(t) \) is evaluated for \( = t_f, \ x(t_f) = x_2 = (17.6, 0.08) \) and \( u^*(t) = 0 \). These two pieces of the trajectory meet at the switching time \( t_s \), i.e., \( \vec{x}^*(t_s) = \hat{x}^*(t_s) \), so one can find now the values for both \( t_s \) and \( t_f \). For the parameters given in Eq. (14) one finds that \( t_s \approx 5.6 \) d and \( t_f \approx 19.9 \) d. Using as initial condition \( x_3 = (70, 0.35) \), we find that \( t_s \approx 5.6 \) d and \( t_f \approx 22.8 \) d.

The effect of different initial conditions can be explored in the Mathematica code found in the supplementary material.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.advwatres.2016.02.014

References


edu/handle/32264.