Stochastic modelling of phytoremediation

BY STEFANO MANZONI1,2,* ANNALISA MOLINI1,2 AND AMILCARE PORPORATO1,2

1 Department of Civil and Environmental Engineering, Box 90287, and 2 Nicholas School of the Environment, Duke University, Box 90328, Durham, NC 27708, USA

Leaching of heavy metals and other contaminants from soils poses a significant environmental threat as it affects the quality of downstream water bodies. Quantifying these losses is particularly important when employing phytoremediation approaches to reduce soil contamination, as contaminant escaping the system through leaching cannot be taken up by vegetation. Despite its undoubted importance, the role of such hydrologic forcing has seldom been fully considered in models describing the long-term contaminant mass balance during phytoremediation. The partitioning of contaminants between leaching and vegetation uptake is controlled by a number of biophysical processes as well as rainfall variability. Here, we develop a novel stochastic framework that provides analytical expressions to quantify the partitioning of contaminants between leaching and plant uptake and the probability of phytoremediation duration as a function of rainfall statistics and soil and vegetation characteristics. Simple expressions for the mean phytoremediation duration and effectiveness (defined as the fraction of contaminant that is recovered in plant biomass) are derived. The proposed framework can be employed to estimate under which conditions phytoremediation is more efficient, as well as to design phytoremediation projects that maximize contaminant recovery and minimize the duration of the remediation process.

Keywords: phytoremediation; stochastic rainfall; contaminant fate; leaching; coupled soil–plant model

1. Introduction

Phytoremediation has emerged as potentially effective and relatively inexpensive alternative to remove contaminants from soils, sediments and shallow groundwater (Salt et al. 1998; Dietz & Schnoor 2001; Pilon-Smits 2005; Gerhardt et al. 2009). Removal of heavy metals and organic compounds occurs thanks to plants being able to concentrate contaminants in their tissues and, in some cases, metabolize them to less toxic compounds (Baker & Brooks 1989; Burken & Schnoor 1997; Lasat 2002; Pilon-Smits 2005). Either through harvest of the above-ground plant biomass (phytoextraction), or thanks to phytodegradation, the amount of contaminant remaining on site is reduced. The effectiveness of the removal is largely dependent upon the bioavailability of the contaminant.

*Author for correspondence (stefano.manzoni@duke.edu).
and the partitioning between uptake by plants and other losses from the system. Uptake by vegetation depends on plant growing conditions, which are negatively affected by droughts and strong climatic variability (Larcher 2003; Rodriguez-Iturbe & Porporato 2004; Gerhardt et al. 2009). Leaching losses caused by intense rainfall or over-irrigation may be particularly important with soluble contaminants or when chelating agents are added to the soil to improve contaminant solubility and bioavailability (Lasat 2002; Schmidt 2003; Chen et al. 2004; Koopmans et al. 2007). These leaching events may contaminate aquifers and surface water bodies, thus representing a potential environmental hazard. Understanding and quantifying the partition of contaminant between plant biomass and leaching are thus necessary for proper design and management of phytoremediation efforts.

Several mathematical models have been developed to describe contaminant uptake as a function of chemical compounds involved, plant species used and environmental conditions at the remediation site. They are typically based on mass balance equations that describe contaminant transport from the soil to the roots and its subsequent accumulation in the plant. Such bioaccumulation models typically focus on transport, accumulation and chemical reaction within the plant (Paterson et al. 1994; Trapp & Matthies 1995; Burken & Schnoor 1997; Undeman et al. 2009). Recently, the focus has shifted towards vertically explicit uptake models to couple soil moisture and root dynamics and account for contaminant advection, dispersion, adsorption–desorption and reactions in the unsaturated zone (Brennan & Shelley 1999; Vogeler et al. 2001; Ouyang 2002; Mathur 2004; Verma et al. 2006). The modelled processes are inherently unsteady, as they depend upon time-varying environmental conditions (chiefly rainfall, and consequently soil moisture and uptake rate). While the role of rainfall variability has been recognized (Undeman et al. 2009), these models are driven by measured rainfall series and are thus site and rainfall-realization dependent. As a consequence, they lack the ability to address the wide range of possible hydro-climatic regimes and are not amenable to probabilistic risk analysis.

In contrast, stochastic models that treat rainfall depths and inter-arrival times as random variables are better suited to study the long-term fate of contaminants in soils, because they are able to determine probabilistically the frequency of deep percolation events as well as the rate of transpiration, which in turn control contaminant partitioning into leaching to groundwater and vegetation uptake. Such stochastic models have been developed in terrestrial ecology and hydrology to study vegetation and soil biogeochemical responses to rainfall variability (Rodriguez-Iturbe & Porporato 2004; Katul et al. 2007). While the dynamics of soil nutrients and soil salinization have been addressed under this framework (Botter et al. 2008; Manzoni & Porporato 2009; Suweis et al. 2010), phytoremediation models with stochastically varying rainfall have not.

Here, we develop a minimalist stochastic model that describes the mass balance of contaminant in the unsaturated zone, subject to phytoremediation and leaching losses triggered by intense rainfall events. This minimalist model yields an analytical, time-dependent probability density function (PDF) for the contaminant mass during the remediation process. Remediation durations and contaminant fate are also obtained analytically as a function of mean rain event frequency and intensity, as well as plant and contaminant characteristics.
Figure 1. (a) Modelling scheme including the soil water (left) and contaminant mass balances (right). Boxes represent compartments; arrows indicate mass fluxes for water (solid lines) and contaminant (dashed lines). Right panels illustrate (b) daily soil moisture, $s$ (solid black curve), driven by rainfall, $R$ (grey bars) and (c) long-term contaminant dynamics (normalized to the initial condition, solid black curve) driven by mean uptake and stochastic leaching events (equation (2.9); grey bars). The dot-dashed and grey solid curves are the contaminant trajectories without leaching and obtained from the fully coupled model of equations (2.1) and (2.2), respectively. The model is parametrized for a mesic climate ($\lambda = 0.2 \text{ d}^{-1}$, $\alpha = 0.015 \text{ m}$), and for trichloroethylene contamination in a sandy loam soil low in organic matter, with $K_d = 10^{-6} \text{ m}^3 \text{ g}^{-1}$ and $\kappa = 0.75$ (Office of Emergency and Remedial Response 1996; Dietz & Schnoor 2001).

### 2. Theory

In this section, we develop a coupled mass balance model of soil moisture ($s$) and soil contaminant concentration ($x$) subject to removal by plants and hydrologic processes (i.e. leaching), in the absence of further inputs to the system. Modelling scheme and symbol definitions are illustrated in figure 1, together with an example of modelled hydrologic and contaminant time series. Symbols are explained in table 1. The soil moisture balance at the daily time scale can be written as (Porporato et al. 2004)

$$nZ_r \frac{ds}{dt} = R(t) - ET(s) - LQ(s, t), \quad (2.1)$$

where $s$ is the relative volumetric soil water, $n$ is the soil porosity, $Z_r$ the mean rooting depth, $R$ the rainfall input, $LQ(s, t)$ the sum of surface runoff and deep infiltration and $ET(s)$ the evapotranspiration flux. For simplicity, in the following we assume that deep infiltration is the predominant water loss mechanism beyond evapotranspiration, as is often the case except for humid climates and very shallow soils (Laio et al. 2001).
Table 1. Definitions and symbol explanations. Subscripts 0 and c indicate initial and target concentrations, respectively; $\langle \cdot \rangle$ represents ensemble averaging and overbar long-term averaging.

<table>
<thead>
<tr>
<th>symbol</th>
<th>definition</th>
<th>units</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>soil porosity</td>
<td>—</td>
</tr>
<tr>
<td>$p_X(x; u)$</td>
<td>probability density function of variable $x$ conditional to $u$</td>
<td>$[x]^{-1}$</td>
</tr>
<tr>
<td>$s$</td>
<td>relative volumetric soil moisture</td>
<td>—</td>
</tr>
<tr>
<td>$s_w$</td>
<td>soil moisture at wilting point</td>
<td>—</td>
</tr>
<tr>
<td>$s_1$</td>
<td>soil moisture threshold for deep percolation</td>
<td>—</td>
</tr>
<tr>
<td>$w_0$</td>
<td>total soil water storage, $w_0 = nZr(s_1 - s_w)$</td>
<td>m</td>
</tr>
<tr>
<td>$x$, $x_c$</td>
<td>total contaminant concentration</td>
<td>g m$^{-3}$</td>
</tr>
<tr>
<td>$x_a$</td>
<td>contaminant concentration in solution</td>
<td>g m$^{-3}$</td>
</tr>
<tr>
<td>$x_a$</td>
<td>contaminant concentration in adsorbed phase</td>
<td>g g$^{-1}$</td>
</tr>
<tr>
<td>$x_{\text{max}}(t)$</td>
<td>trajectory of $x$ in absence of leaching</td>
<td>g m$^{-3}$</td>
</tr>
<tr>
<td>$y$, $y_c$</td>
<td>transformed contaminant concentration</td>
<td>m</td>
</tr>
<tr>
<td>$y_{\text{max}}(t)$</td>
<td>trajectory of $y$ in absence of leaching, $y_{\text{max}} = y_0 - \beta t$</td>
<td>m</td>
</tr>
<tr>
<td>$\text{ET}(s)$</td>
<td>evapotranspiration, $\text{ET}(s) = \text{ET}_{\text{max}}(s - s_w)/(s_1 - s_w)$</td>
<td>m d$^{-1}$</td>
</tr>
<tr>
<td>$\text{ET}_{\text{max}}$</td>
<td>maximum evapotranspiration</td>
<td>m d$^{-1}$</td>
</tr>
<tr>
<td>$F_{L,w}(t)$</td>
<td>Poisson process describing deep percolation events</td>
<td>m d$^{-1}$</td>
</tr>
<tr>
<td>$\text{LE}(x_s, t)$</td>
<td>leaching losses</td>
<td>g m$^{-2}$ d$^{-1}$</td>
</tr>
<tr>
<td>$K_d$</td>
<td>slope of the linear adsorption isotherm, $x_a = K_d x_s$</td>
<td>m$^3$ g$^{-1}$</td>
</tr>
<tr>
<td>$LQ(s, t)$</td>
<td>deep percolation</td>
<td>m d$^{-1}$</td>
</tr>
<tr>
<td>$\text{UP}(x_s, s)$</td>
<td>contaminant uptake by plants</td>
<td>g m$^{-2}$ d$^{-1}$</td>
</tr>
<tr>
<td>$R(t)$</td>
<td>rainfall</td>
<td>m d$^{-1}$</td>
</tr>
<tr>
<td>$T$</td>
<td>duration of the remediation project</td>
<td>d</td>
</tr>
<tr>
<td>$\bar{T}_{s_1}$</td>
<td>mean duration of excursions below the threshold $s_1$</td>
<td>d</td>
</tr>
<tr>
<td>$Z_r$</td>
<td>mean rooting depth</td>
<td>m</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>mean rainfall depth</td>
<td>m</td>
</tr>
<tr>
<td>$\alpha_L$</td>
<td>mean percolation event depth, $\alpha_L = \alpha$</td>
<td>m</td>
</tr>
<tr>
<td>$\beta$</td>
<td>drift term in equation (2.11), $\beta = \kappa \text{ET}(ns_1 + \rho_b K_d)/(n \bar{s} + \rho_b K_d)$</td>
<td>m d$^{-1}$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>normalized soil storage capacity, $\gamma = w_0/\alpha$</td>
<td>—</td>
</tr>
<tr>
<td>$\eta$</td>
<td>normalized maximum evapotranspiration, $\eta = \text{ET}_{\text{max}}/w_0$</td>
<td>—</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>transpiration stream concentration factor (equation (2.5))</td>
<td>—</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>mean rainfall frequency</td>
<td>d$^{-1}$</td>
</tr>
<tr>
<td>$\lambda_L$</td>
<td>mean percolation event frequency, $\lambda_L = 1/\bar{T}_{s_1}$</td>
<td>d$^{-1}$</td>
</tr>
<tr>
<td>$\rho_b$</td>
<td>soil bulk density</td>
<td>g m$^{-3}$</td>
</tr>
<tr>
<td>$\chi$</td>
<td>fraction of contaminant captured by plants</td>
<td>—</td>
</tr>
</tbody>
</table>

The mass balance equation for $x$ (mass of contaminant per unit soil volume, including adsorbed and dissolved fractions) can be written as (figure 1)

$$Z_r \frac{dx}{dt} = -\text{UP}(x_s, s) - \text{LE}(x_s, t),$$

where $x_s$ indicates the contaminant concentration in the soil solution (related to the total contaminant through an equilibrium isotherm, $x_a = x_a(x, s)$), which is affected by the hydrologic transport processes. Removal of contaminant occurs continuously through uptake by plants (UP) and leaching events (LE) triggered...
by deep infiltration. Plant uptake is controlled by plant growth and hence transpiration ($ET(s)$ in equation (2.1)), and both UP and LE depend on the concentration of contaminant in the soil solution.

The study of the long-term dynamics of equations (2.1) and (2.2) can be simplified thanks to the fact that soil moisture fluctuations are much faster than contaminant dynamics. Thus, instead of considering the coupled equations (2.1) and (2.2), we focus on the contaminant mass balance only, assuming that soil moisture is near steady state. We also assume that seasonal variability in climatic and plant characteristics is smaller than the variability at the rainfall event scale (extensions to account for seasonality are discussed in §3c).

To describe the specifics of the models, we start from the stochastic description of the soil moisture balance (§2a), which is instrumental to develop the plant uptake and leaching models that mathematically define the terms UP and LE in equation (2.2) (§§2b,c, respectively). After linking adsorbed and dissolved contaminant concentrations through an adsorption isotherm (§2d), the stochastic mass balance of equation (2.2) is solved under transient conditions (§2e) and the PDF of the phytoremediation duration is computed (§2f). Finally, the fraction of contaminant fixed by plants (a measure of the phytoremediation efficiency) is obtained (§2g).

(a) Stochastic soil moisture balance

Following Porporato et al. (2004), rainfall ($R$ in equation (2.1)) is modelled as a marked Poisson process with mean event frequency $\lambda$ and mean event depth $\alpha$. In response to this rainfall forcing, soil moisture fluctuates in time as illustrated in the example of figure 1b. In this simplified framework, it is assumed that above the threshold $s_1$ rainfall in excess to the storage $nZ_r(s_1 - s)$ is instantaneously lost by runoff and deep percolation. Evapotranspiration is described as a linear function of $s$ between the plant wilting point $s_w$ and $s_1$, i.e.

$$ET(s) = \frac{ET_{\text{max}}(s - s_w)}{s_1 - s_w},$$

where $ET_{\text{max}}$ is the maximum evapotranspiration rate. The total soil water storage capacity for this formulation is thus $w_0 = nZ_r(s_1 - s_w)$.

With these assumptions, the long-term mean evapotranspiration rate can be expressed as a function of rainfall statistics and soil and plant properties as (Porporato et al. 2004)

$$\bar{ET} = \frac{w_0}{\gamma} \left[ \frac{\lambda - \eta \gamma^{\lambda/\gamma} e^{-\gamma}}{\Gamma(\lambda/\eta) - \Gamma(\lambda/\eta, \gamma)} \right],$$

(2.3)

where $\eta = ET_{\text{max}}/w_0$, $\gamma = w_0/\alpha$, and $\Gamma(\cdot)$ and $\Gamma(\cdot, \cdot)$ are the gamma function and the incomplete gamma function (Abramowitz & Stegun 1972).

Percolation or leaching events are instantaneous and form a (non-Poissonian) renewal point process (Rodriguez-Iturbe & Porporato 2004; Tamea et al. submitted) with mean frequency $\lambda_L = \bar{T}^{-1}_{s_1}$. The mean inter-arrival time between percolation events can be computed as the mean duration of crossing of the
threshold $s_1$ (Rodriguez-Iturbe & Porporato 2004),

$$\bar{T}_{s_1} = \frac{1}{\lambda} I_1 \left[ 1; 1 + \frac{\lambda}{\eta}; \gamma \right], \quad (2.4)$$

where $I_1[\cdot ; ; \cdot]$ is the Kummer confluent hypergeometric function of the first kind (Weisstein 2011a). Moreover, thanks to the memoryless property of the exponential distribution, the PDF of percolation depths can be shown to be identical to the PDF of rainfall depths (i.e. exponential), with mean percolation depth $\alpha_0 = \alpha$ (Daly & Porporato 2010). These results are exact within the assumption of the minimalist soil moisture model of Porporato et al. (2004).

(b) Plant uptake

Uptake of organic compounds occurs via diffusion and transport in the transpiration flow, while inorganic contaminants are taken up actively, thanks to plant-induced gradients that drive nutrients and contaminants towards the roots (Pilon-Smits 2005). Active uptake mechanisms are triggered by plant nutrient requirements and are thus correlated to plant condition, with rapidly growing plants necessitating more nutrients. Transpiration is strongly correlated to photosynthesis and productivity across spatial (leaf to canopy) and temporal scales (hours to years; e.g. Larcher 2003), and can be thus used as a proxy for plant condition. As a result, the uptake of both organic and inorganic compounds may be approximated as passive processes simply proportional to the mean transpiration rate and the contaminant concentration in the soil solution at average soil moisture,

$$\text{UP}(x_s, \bar{s}) = \kappa \bar{ET} x_s, \quad (2.5)$$

where $\bar{ET}$ is computed from equation (2.3) and $\kappa$ is a non-dimensional coefficient often referred to as the transpiration stream concentration factor (TSCF) (Briggs et al. 1982; Dietz & Schnoor 2001). This approximation is supported experimentally (e.g. Walker 1971) and thus commonly employed in mathematical models for its simplicity and effectiveness in linking water availability, plant growth and contaminant accumulation in plant tissues (Trapp & Matthies 1995; Burken & Schnoor 1997; Mathur 2004). Uptake of heavy metals can be described by Michaelis–Menten kinetics (Verma et al. 2006), indicating that $\kappa$ decreases with increasing solution concentration. However, at low $x_s$, Michaelis–Menten kinetics can be approximated by a linear relationship, i.e. under these conditions and in the absence of phytotoxicity $\kappa$ is nearly constant (Vogeler et al. 2001; Mathur 2004).

This model is suitable to describe contaminant partitioning between leaching and biomass, but does not distinguish between allocation to belowground and aboveground compartments. However, in phytoextraction only the shoots are generally harvested, so that the actual contaminant removal is lower than that calculated from plant uptake. Moreover, a fraction of $\bar{ET}$ is encompassed by soil evaporation, which does not contribute to plant contaminant extraction. To account for these effects, we can interpret $\kappa$ as an effective contaminant bioaccumulation-harvesting efficiency, the numerical value of which may be lower than the measured TSCF.
(c) Leaching

Leaching events depend on water percolation intensity as well as on the contaminant concentration in the soil solution, $x_s$, which is linked to $x$ and $s$ by an equilibrium isotherm, that is

$$LE(x_s, t) = x_s(x, s)LQ(s, t). \tag{2.6}$$

Because of the form of the soil moisture model (§2a), percolation happens instantaneously at $s_1$. The fact that the percolation process takes place at $s_1$ allows us to use this value in the isotherm $x_s(x, s)$, so that $LE = x_s(x, s_1)LQ(s, t)$. Moreover, the instantaneous form of percolation raises the problem of interpretation of the stochastic forcing in the previous equation. In fact, although the leaching events are considered instantaneous, in reality they last for the finite if short period of time in which the soil is saturated. This corresponds to taking the white-noise limit of a coloured noise process (Van Kampen 1992; Suweis et al. in press), which in turn requires to interpret equation (2.2) in the Stratonovich sense (Van Den Broeck 1983).

Finally, in order to achieve analytical tractability, inter-arrival times may be assumed to be approximately Poissonian, with an exponential distribution of mean $\lambda_L$ (Botter et al. 2007; Suweis et al. 2010). As a consequence, leaching events can be modelled as a multiplicative (Stratonovich) process forced by marked Poisson events with parameters $\lambda_L$ and $\alpha_L$ denoted as $F_{L,w}(t)$,

$$LE(x_s, t) = x_s(x, s_1) \circ F_{L,w}(t), \tag{2.7}$$

where $\circ$ indicates that the equation is interpreted according to Stratonovich.

(d) Adsorption–desorption kinetics

Using equations (2.5) and (2.7), the mass balance (equation (2.2)) for contaminant concentration $x$ can be re-written as

$$Z_r \frac{dx}{dt} = -\kappa ET x_s - x_s \circ F_{L,w}(t). \tag{2.8}$$

To proceed, we need to relate the total contaminant concentration $x$ (expressed per unit soil volume) and concentration in solution $x_s$ (expressed per unit soil water volume), by specifying an adsorption isotherm $x_a = x_a(x, s)$, which partitions the contaminant in adsorbed (i.e. $x_a$, contaminant mass per unit soil mass) and dissolved fractions. A linear adsorption isotherm, $x_a = K_d x_s$, where $K_d$ is the equilibrium partitioning coefficient (Van Rees et al. 1990; Verma et al. 2006), has been used to describe adsorption kinetics of both heavy metals (Verma et al. 2006) and organic compounds (Trapp & Matthies 1995; Burken & Schnoor 1997).

In the more general case of nonlinear isotherm, the nonlinearities in $x_a = x_a(x, s)$ hinder the analytical tractability of the problem and will thus be neglected for simplicity. Because the total contaminant concentration $x = nx_s + \rho_b x_a = (ns + \rho_b K_d) x_s$ (where $\rho_b$ is the soil bulk density), we obtain $x_s = x/(ns + \rho_b K_d)$, indicating that dilution and adsorption significantly decrease the concentration.
in the soil solution at high $s$ and $K_d$, respectively. As a result, the contaminant mass balance is

$$Z_r \frac{dx}{dt} = -\frac{\kappa ET}{n \bar{s} + \rho_b K_d} x - \frac{1}{n s_1 + \rho_b K_d} x \circ F_{L,w}(t).$$  \hspace{1cm} (2.9)

An example of a simulated time series of $x$ is shown in figure 1c (black curve), along with a simulation where daily fluctuations in soil moisture are explicitly considered instead of the mean value (grey curve). While this slightly increases the variance of $x$, it does not appreciably affect the long-term dynamics, partly owing to the convergence of $x$ to zero and partly because of the saturating effect of the plant uptake in equation (2.9) at high $s$. Hence, our assumption that contaminant dynamics are controlled by the mean soil moisture between leaching events appears to be supported.

(e) Transient solution of the stochastic contaminant balance

Because of the Stratonovich interpretation of equation (2.9), it is possible to perform the change of variable following the usual rules of calculus (Van Kampen 1992)

$$y = Z_r (n s_1 + \rho_b K_d) \ln(x),$$  \hspace{1cm} (2.10)

which leads to a stochastic balance equation with additive noise,

$$\frac{dy}{dt} = -\beta - F_{L,w}(t),$$  \hspace{1cm} (2.11)

with initial condition $y_0$ and where $\beta = \kappa ET (n s_1 + \rho_b K_d)/(n \bar{s} + \rho_b K_d)$. A similar transformation can be applied when nonlinear adsorption isotherms are used (i.e. $y = \int dx/x(x)$).

In the absence of leaching, equation (2.11) simply gives $y_{\max}(t) = y_0 - \beta t$, which, using equation (2.10), leads to the upper bound for the trajectories, $x_{\max}(t) = x_0 \exp(-\kappa ET Z_r^{-1} (n \bar{s} + \rho_b K_d)^{-1} t)$. In general (when both uptake and leaching occur), equation (2.11) regulates how the contaminant is partitioned between the deterministic uptake and the random leaching events. By averaging equation (2.11) on both sides, and recalling that $\langle F_{L,w}(t) \rangle = a_L l_L$, the mean trajectory can be calculated as $\langle y(t) \rangle = y_0 - (\beta + a_L \lambda_L) t$, where the (average) partitioning in uptake ($\beta$) and leaching ($a_L \lambda_L$) appears clearly. The transient solution of equation (2.11) in terms of the PDF of the variable $y$ (i.e. $p_Y(y; t)$) can be obtained from the master equation of the process, as described in appendix A. The analytical solution in terms of the contaminant concentration $x$ is then found from equation (A5) using the transformation of variable in equation (2.10) as

$$p_X(x; t) = \delta(\xi) e^{-\lambda_L t} + \Theta(\xi) \frac{Z_r (n s_1 + \rho_b K_d)}{x} \sqrt{Z_r \frac{\lambda_L t}{\alpha_L \xi}} I_1 \left[ 2 \sqrt{\frac{\lambda_L \xi t}{\alpha_L}} \right] e^{-(\xi/\alpha_L) - \lambda_L t},$$  \hspace{1cm} (2.12)

where $\xi = Z_r (n s_1 + \rho_b K_d) \ln(x_0/x) - \beta t$, $I_1[\cdot]$ is the modified Bessel function of the first kind (Abramowitz & Stegun 1972), $\delta(\cdot)$ is the Dirac delta function and $\Theta(\cdot)$ the Heaviside function. The first term in the equation represents the probability of remaining on the trajectory $x(t) = x_{\max}(t)$, which is equal to the exceedance probability of a leaching inter-arrival time of duration $t$, i.e. $e^{-\lambda_L t}$. 

(f) Probability density of phytoremediation duration

The duration of the remediation process \( T \) is the time needed to reach an acceptable level of contaminant, say \( x_c \), starting from the initial condition \( x_0 \). Because of the randomness of rainfall inputs and leaching events, such duration is also a random variable, equivalent to the first passage time of the threshold \( x_c \) with initial condition \( x_0 \). The PDF of first passage times is derived in appendix B as

\[
p_T(T; x_c, x_0) = \delta(\zeta)e^{-\lambda_L T} + \Theta(\zeta)\lambda_L e^{-(\zeta/\alpha_L) - \lambda_L T} \times \left\{ 0F_1 \left[ 1; \frac{\lambda_L T \zeta}{\alpha_L} \right] + \frac{\beta T}{\alpha_L} 0F_1 \left[ 2; \frac{\lambda_L T \zeta}{\alpha_L} \right] \right\},
\]

where \( \zeta = Z_r(n s_1 + \rho_b K_d) \ln(x_0/x_c) - \beta T \), \( 0F_1[\cdot] \) is the confluent hypergeometric limit function (Weisstein 2011b), \( \beta \) is a function of \( \overline{ET} \) and \( \bar{s} \) (equation (2.11)) and \( \overline{ET} \) is given by equation (2.3).

The mean first passage time \( \langle T \rangle \) can be obtained from the moment generating function of \( p_T(T; x_c, x_0) \) (appendix B). In most practical applications, \( \langle T \rangle \) can be approximated as

\[
\langle T \rangle = \frac{Z_r(n s_1 + \rho_b K_d) \ln(x_0/x_c)}{\overline{R} + \overline{ET}(\kappa n s_1 + \rho_b K_d)/(n \bar{s} + \rho_b K_d) - 1},
\]

where we substituted \( \alpha_L \lambda_L = \overline{R} - \overline{ET} \) and the definition of \( \beta \) to express \( \langle T \rangle \) as a function of rainfall, contaminant and plant characteristics, and the ratio of initial to target contaminant concentration. Equation (2.14) can be further simplified when contaminant release to the soil solution is slow (i.e. \( \rho_b K_d \gg 1 \)) or under humid conditions (i.e. \( \bar{s} \approx s_1 \)), so that \( (n s_1 + \rho_b K_d)/(n \bar{s} + \rho_b K_d) \approx 1 \) and thus \( \beta \approx \kappa \overline{ET} \) and \( n s_1 + \rho_b K_d \approx \rho_b K_d \). As a result, a first-order approximation for the mean remediation duration becomes

\[
\langle T \rangle = \frac{Z_r \rho_b K_d \ln(x_0/x_c)}{\overline{R} + \overline{ET}(\kappa - 1)}. \quad (2.15)
\]

This expression predicts a logarithmic dependence of \( \langle T \rangle \) with the ratio of initial to target contaminant concentration and a linear effect of \( K_d \). In contrast, \( \langle T \rangle \) is inversely proportional to the mean percolation rate \( \overline{R} - \overline{ET} \) and plant uptake \( \kappa \overline{ET} \).

(g) Contaminant fate

Based on the stochastic mass balance described above, we can quantify how much contaminant is on average captured by plants and how much is lost through leaching. This can be done by directly averaging equation (2.9),

\[
Z_r \frac{d\langle x \rangle}{dt} = -\frac{\kappa \overline{ET}}{n \bar{s} + \rho_b K_d} \langle x \rangle - \frac{1}{n s_1 + \rho_b K_d} \langle x F_{L,w}(t) \rangle,
\]

where the first term represents the average plant uptake and the second the average leaching. This equation in principle allows us to obtain the mean trajectory \( \langle x(t) \rangle \) and in turn the other mean balance terms. However, the term \( \langle x(t) F_{L,w}(t) \rangle \) is not known in general, because in the Stratonovich
interpretation the stochastic variable $x$ is not statistically independent of $F_{L,w}(t)$ (see §§2d,e) so that their correlation structure is not known a priori. Nevertheless, owing to the particular nature of the multiplicative noise in equation (2.16), it can be shown that $\langle x(t)F_{L,w}(t) \rangle \approx \langle x(t) \rangle \langle F_{L,w}(t) \rangle$ (Hänggi 1978; Ridolfi et al. 2011), as also confirmed by numerical simulations of equation (2.9). Recalling that $\langle F_{L,w}(t) \rangle = a_{L}l$, we thus obtain

$$Z_{r} \frac{d\langle x \rangle}{dt} = -\left( \frac{\kappa ET}{n\bar{s} + r_{b}K_{d}} + \frac{\alpha_{L}\lambda_{L}}{nS_{1} + r_{b}K_{d}} \right) \langle x \rangle, \quad (2.17)$$

which is solved for the initial condition $\langle x(0) \rangle = x_{0}$ as

$$\langle x(t) \rangle = x_{0} \exp \left[ -\left( \frac{\kappa ET}{n\bar{s} + r_{b}K_{d}} + \frac{\alpha_{L}\lambda_{L}}{nS_{1} + r_{b}K_{d}} \right) \frac{t}{Z_{r}} \right]. \quad (2.18)$$

This time-dependent mean trajectory accounts for both bioaccumulation (represented by the first term in the exponential function) and leaching (the second term). Based on equations (2.5) and (2.7), the time-dependent mean plant-uptake $\langle UP(t) \rangle$ and leaching $\langle LE(t) \rangle$ can be calculated and used to quantify the total masses of contaminant that on average are either bio-accumulated or lost through leaching. The ratio of total bio-accumulated contaminant over the total mass loss (i.e. $Z_{r}[x_{0} - \langle x(t) \rangle]$) may be employed as a measure of phytoremediation efficiency,

$$\langle \chi \rangle = \frac{\int_{0}^{t}\langle UP(u) \rangle du}{\int_{0}^{t}[\langle UP(u) \rangle + \langle LE(t) \rangle] du} = \left[ 1 + \frac{\alpha_{L}\lambda_{L}(n\bar{s} + r_{b}K_{d})}{\kappa ET(nS_{1} + r_{b}K_{d})} \right]^{-1}, \quad (2.19)$$

which does not depend on time because the contaminant fluxes scale similarly with the concentrations. As expected, $\langle \chi \rangle$ increases with uptake efficiency $\kappa ET$ and decreases with percolation rate $\alpha_{L}\lambda_{L}$ (both affected by rainfall statistics), while it is only mildly dependent on the partition coefficient $K_{d}$ (which instead controls remediation duration, see equation (2.15)). For contaminants with high $K_{d}$, we can assume as before that $(nS_{1} + r_{b}K_{d})/(n\bar{s} + r_{b}K_{d}) \approx 1$, obtaining a first approximation for $\langle \chi \rangle$ that only depends on the ratio of mean rainfall to mean evapotranspiration and the TSCF $\kappa$,

$$\langle \chi \rangle = \left( 1 + \frac{\bar{R}/ET - 1}{\kappa} \right)^{-1}. \quad (2.20)$$

Thus, in the limiting case of $\kappa \approx 1$ equation (2.20) yields $\langle \chi \rangle \approx ET/\bar{R}$.

### 3. Results and discussion

(a) Temporal evolution of soil contaminants

Figure 2a illustrates how trajectories of contaminant concentration evolve in time from the initial condition because of random leaching events and continuous plant uptake. The PDF of contaminant concentration (i.e. $p_{X}(x,t)$, equation (2.12)) evolves accordingly from a highly asymmetric distribution
Figure 2. (a) Sample trajectories of contaminant concentration $x(t)$ (grey thin curves), mean trajectory $\langle x(t) \rangle$ (solid black curve), upper deterministic trajectory $x_{\text{max}}(t)$ (dashed curve), and examples of PDF $p_X(x; t)$ (equation (2.12)) at different times (the probability of remaining on $x_{\text{max}}$ at time $t$ is illustrated by black bars, not drawn to scale). (b) Evolution of the PDF of remediation duration, $p_T(T; x_c, x_0)$ (equation (2.13)) for different values of $x_c$ and under two rainfall scenarios (black curves and dark grey PDFs, $\lambda = 0.3 \text{ d}^{-1}$, $\alpha = 0.015 \text{ m}$; grey curves and light grey PDFs, $\lambda = 0.1 \text{ d}^{-1}$, $\alpha = 0.015 \text{ m}$). Solid and dashed curves, respectively, represent mean and upper trajectories for the two scenarios; parameter values as in figure 2 (a part from rainfall statistics in b).

with most of the probability mass on the deterministic trajectory $x_{\text{max}}(t)$ to a more symmetric shape with a mode close to the mean trajectory. As time progresses, the trajectories converge towards zero and the variance of $p_X(x; t)$ decreases. Figure 2b shows the PDFs of remediation duration ($p_T(T; x_c, x_0)$, equation (2.13)) for different target levels of contaminant. For target levels close to the initial condition ($x_c \approx x_0$), the probability of remaining on the trajectory $x_{\text{max}}(t)$ (i.e. no leaching event yet) is highest, resulting in a significant atom of probability on $x_{\text{max}}(t)$. For lower $x_c$, $p_T(T; x_c, x_0)$ becomes more symmetric and the variance increases.

Different climatic, ecological and contaminant features may affect the shape of $p_X(x; t)$ (not shown) and $p_T(T; x_c, x_0)$ (figure 2b). High contaminant solubility (low $K_d$) and rainfall ($\alpha$ and $\lambda$) tend to have the strongest effects, as they increase both uptake rates and leaching losses. Figure 2b illustrates the consequences of rainfall frequency alone, assumed to vary from $\lambda = 0.1 \text{ d}^{-1}$ (characteristic of a rather dry climate, light grey curves and PDFs) to $\lambda = 0.3 \text{ d}^{-1}$ (mesic conditions, black curves and dark grey PDFs). With respect to dry conditions, high rainfall accelerates the decrease of contaminant levels by both improving uptake (lowered $x_{\text{max}}(t)$) and increasing leaching (higher distance between $x_{\text{max}}(t)$ and $\langle x(t) \rangle$). Thus, the phytoremediation potential (e.g. the lowest pollutant concentration that can be reached for a given treatment duration) tends to decrease from mesic to dry conditions because of the inherently slower soil moisture dynamics in dry climates.

(b) Mean phytoremediation duration and efficiency

Rainfall regime affects the duration of the phytoremediation through its controls on soil moisture and in turn on the frequency and intensity of the leaching events (figure 2b). For a given threshold $x_c$, remediation is faster in humid
climates than in drier ones (respectively, grey and black curves in figure 2b), because of increased plant uptake (plant transpiration and growth increase with mean soil moisture) and leaching losses. As a consequence, the mean remediation duration is shorter, and the variance of $T$ is smaller. This rapid remediation is advantageous when the goal is to reduce the soil contaminant concentration to a required threshold in the shortest amount of time, but it may lead to significant contamination of groundwater and rivers downstream because of enhanced leaching. To quantify this tradeoff, in Figure 3 we compare the mean remediation duration $\langle T \rangle$ (equation (2.14)) to the mean remediation efficiency $\langle \chi \rangle$ (equation (2.19)).

Drier climates (with either low mean rain depth $\alpha$ or frequency $\lambda$, i.e. below the white dashed curve) on average need longer durations to achieve the same concentration goal, but (as expected) under these drier conditions bioaccumulation is the main contaminant export pathway (high $\langle \chi \rangle$). For example, durations of the order of few years with relatively high mean efficiencies ($\langle \chi \rangle > 0.6$) can be achieved for trichloroethylene in climates with mean growing season rainfall in the range 300–500 mm (central area in figure 3). Plants not only provide a sink for the contaminant, but also decrease soil moisture and hence lower or delay leaching losses (Chen et al. 2004). This important indirect effect of plants on contaminant fate is captured by our model through the coupling of plant transpiration and soil moisture in the soil water balance equation (equation (2.1)), as higher transpiration reduces in the long-term percolation and runoff. In humid climates (above the dashed curve in figure 3), the predictions of the model may be biased because we neglected runoff and assumed a transpiration function that is linear with soil moisture. On the one hand, increased runoff in these conditions may reduce the relative importance of leaching, resulting in higher remediation effectiveness than estimated in figure 3b. On the other hand, the lack of the often-observed plateau in transpiration at high soil moisture
Figure 4. (a) Mean duration, $\langle T \rangle$ (equation (2.14)), and (b) mean recovered contaminant fraction, $\langle \chi \rangle$ (equation (2.19)), as a function of TSCF, $\kappa$ and contaminant adsorption coefficient, $K_d$ (rainfall statistics as in figure 2; required reduction, $x_c/x_0 = 0.1$). The white dashed curves illustrate the relationship between $\kappa$ and $K_d$ for increasing values of octanol–water partition coefficient (left to right), for a soil with 1% organic carbon (the curves would shift towards higher $K_d$ as organic carbon content increases; see also (Office of Emergency and Remedial Response 1996; Dietz & Schnoor 2001)).

(Laio et al. 2001) might cause us to underestimate plant uptake, thus potentially decreasing remediation efficiency. The two effects to some degree may cancel out, but additional tests with more complete hydrologic models would be required to precisely quantify them.

Plant and contaminant properties, respectively, through the TSCF $\kappa$ and the partition coefficient $K_d$, also affect remediation duration and efficiency (figure 4). High values of $\kappa$ decrease $\langle T \rangle$ and improve $\langle \chi \rangle$, while high partitioning to the soil solution (low $K_d$) implies (similarly to humid climates) fast remediation and high leaching losses. As a consequence, $\langle \chi \rangle$ decreases as $K_d$ is decreased, for instance, by supplying chelating agents to the soil (Schmidt 2003). Because percolation events increase in number and intensity in humid climates, leaching increases more than linearly with increasing mean rainfall. Consistently with this pattern, experimental evidence showed doubled leaching losses when water added in a column study was increased by only 50 per cent (Grcman et al. 2001). For these reasons, selection of fast-growing, hyper-accumulator or tolerant species remains the most profitable way to improve phytoremediation, while use of chelating agents to increase solubility has major shortcomings (Salt et al. 1998; Dietz & Schnoor 2001; Lasat 2002; Schmidt 2003).

(c) Applications and limitations

The proposed modelling framework considers a continuous sequence of growing seasons and thus neglects the effects of seasonal climate on plant growth and on the statistics of rainfall events. This approximation limits only in part the applicability of the approach. The model can be applied to sites with nearly continuous vegetation cover throughout the year and where rainfall patterns do
stochastic modelling of phytoremediation

not show marked seasonality, i.e. in warm temperate and mesic tropical climates. However, it can also be applied to seasonal climates where the dormant seasons are characterized by limited leaching and plant uptake, as in semi-arid ecosystems where the dry season coincides with low plant activity, as well as limited leaching (Rodriguez-Iturbe & Porporato 2004). In cold temperate climates, plant activity is low during the winter, but sub-freezing temperatures inhibit percolation, thus reducing leaching as well. This allows us to apply the model in these strongly seasonal climates, under the assumption that climatic and soil moisture transients during spring and autumn have negligible consequences on the annual contaminant balances.

Temperate climates with high winter rainfall (e.g. Mediterranean climate) may be characterized by strong leaching events while plant uptake is limited. Under these conditions, the proposed model can be still applied if contaminant mobility is reduced during winter, as in sites contaminated by relatively insoluble compounds that are treated with chelating agents only during the growing season (e.g. Schmidt 2003), or by organic contaminants that are preferentially adsorbed to organic matter and roots (uptake occurs as root density increases during the growing season). Alternatively, markedly seasonal climates with a humid dormant season and in presence of relatively soluble contaminants could be modelled through numerical simulations of the coupled equations (2.1) and (2.8), with seasonally varying evapotranspiration and, in some cases, rainfall statistics. As a first approximation, however, the mean contaminant fate can be calculated from the weighted average of growing season conditions (large ET and relatively low leaching) and dormant season conditions (reduced ET and thus increased losses). This approach admittedly neglects transient effects between seasons, but would at least capture the general seasonal pattern of contaminant fate and how it relates to rainfall statistics and plant features. Future work will include explicitly the role of seasonality in the coupled dynamics of soil water and contaminant concentrations.

4. Conclusions

Our stochastic model of phytoremediation of contaminated soils accounts for the effects of rainfall variability on plant activity, which in turn controls bioaccumulation, and on leaching, which is a major, undesired contaminant loss pathway. Analytical solutions for the PDF of contaminant concentration and remediation duration for any given target contaminant level are obtained, allowing a rapid assessment of the dynamics of the phytoremediation project based on key measurable quantities (contaminant partition coefficient, TSCF and rainfall statistics). The approach is suitable for risk assessment, as it provides a probabilistic representation of both contaminant concentrations through time and remediation duration. The model can be employed to quantify the mean phytoremediation effectiveness (i.e. fraction of recovered contaminant) under different site management practices (e.g. choice of plant species, soil type, use of chelating agents), thus permitting an objective evaluation of the tradeoffs between contaminant recovery or degradation in vegetation, and the duration of the remediation.

This research was partially supported by the United States Department of Energy (DOE) through the Office of Biological and Environmental Research (BER) Terrestrial Carbon Processes (TCP) programme (NICCR grant: DE-FC02-06ER64156) and the National Science Foundation (CBET-1033467). We also thank Francesco Laio and Edoardo Daly for useful discussions.

APPENDIX A

In this appendix, we provide the derivation of the PDF for the transformed variable $y$ (equation (2.10)), $p_Y(y; t)$, for which the noise is additive (equation (2.11)). For simplicity of notation, we transform the problem into a process $z = y_0 - y$. This process grows monotonically according to a compound Poisson process with constant drift $\beta$ and jumps of mean frequency $\lambda_L$ and intensity $\alpha_L$. We start by writing the master equation for this process as (Cox & Miller 2001)

$$
\frac{\partial p_Z(z; t)}{\partial t} = -\beta \frac{\partial p_Z(z; t)}{\partial z} - \lambda_L p_Z(z; t) + \lambda_L \int_0^z [p_Z(u; t)e^{-\lambda_L(t-u)/\alpha_L} du, \quad (A 1)
$$

with initial condition $p_Z(z; 0) = \delta(z)$.

Laplace transformation in both state variable ($z \rightarrow k$) and time ($t \rightarrow s$) leads to

$$
^*p_Z^*(k; s) = \left[s + \beta k + \frac{\lambda_L}{k + \alpha_L^{-1}}\right]^{-1}, \quad (A 2)
$$

where we adopt the notation of Cox (1967), i.e. $^*p_Z(k; t)$, $^*p_Z^*(x; s)$ and $^*p_Z^*(k; s)$ indicate the Laplace transforms of $p_Z(z; t)$ in space, time and both space and time, respectively. Equation (A 2) can be anti-transformed to get the moment generating function of $p_Z(z; t)$,

$$
^*p_Z^*(k; t) = e^{-k[(\beta + \lambda_L)/(k + \alpha_L^{-1})]}, \quad (A 3)
$$

from which the central moments of $z(t)$ are obtained,

$$
\mu_r = (-1)^r \left. \frac{d^r p_Z(k; t)}{dk} \right|_{k=0}. \quad (A 4)
$$

For $r = 1$ (mean of $z$), we find $\langle z(t) \rangle = (\beta + \alpha_L \lambda_L)t$, in agreement with the expression obtained from averaging of the Langevin equation of the process (equation (2.11)), $\langle y(t) \rangle = y_0 - (\beta + \alpha_L \lambda_L)t$. Further anti-transformation of equation (A 3) yields the PDF of the process $z$, $p_Z(z; t)$, from which (after the change of variables $z = y_0 - y$), we obtain the mixed distribution,

$$
p_Y(y; t) = \delta(y_0 - y - \beta t)e^{-\lambda_L t} + \Theta(y_0 - y - \beta t)\sqrt{\frac{\lambda_L t}{\alpha_L(y_0 - y - \beta t)}}e^{-(y_0 - y - \beta t)/\alpha_L - \lambda_L t}, \quad (A 5)
$$

which is easily transformed in the PDF of the contaminant $p_X(x; t)$ (equation (2.12)) by means of a derived distribution using equation (2.10), $p_X(x; t) = p_Y(y; t) |dy/dx|$. Note that by imposing $\beta = 0$ the results for the simpler cumulative Poisson process (Cox 1967; Cox & Miller 2001) are recovered.

**APPENDIX B**

This appendix presents the derivation of the first passage time distribution, or the time necessary to reach a given threshold of contaminant (equation (2.13)). The process described by equation (2.11) leads to monotonically decreasing trajectories (i.e. the contaminant can only be lost). Thanks to this property, the probability to pass the threshold $y_c$ equals the probability that the first passage time has already been reached (Cox 1967; Daly & Porporato 2006). This provides the equation necessary to compute the first passage time distribution, $p_T(T; y_c, y_0)$. For clarity, we again analyse the process $z = y_0 - y$ (see appendix A) and finally change the variable back to the original $x$. For this process, the above condition reads

$$\int_T^{\infty} p_T(u; z_c) \, du = \int_0^z p_Z(v; T) \, dv. \quad (B\, 1)$$

Upon Laplace transformation, a general expression linking first passage time distribution and transient PDF of the process can be found (Cox 1967),

$$*p^*_T(s; k) = \frac{1}{k} - s *p^*_Z(k; s). \quad (B\, 2)$$

Using equation (A2) and anti-transforming equation (B2) yields the moment generating function of $p_T(T; y_c, y_0)$ (not shown for conciseness), from which the mean first passage time $\langle T \rangle$ is obtained. A second anti-transformation of equation (B2) and the change of variable $z = y_0 - y$ yields

$$p_T(T; y_c, y_0) = \delta(y_0 - y_c - \beta T) e^{-\lambda_L T} + \Theta(y_0 - y_c - \beta T) \lambda_L e^{-(y_0 - y_c - \beta T)/\alpha_L - \lambda_L T}$$

$$\times \left\{ 0 F_1 \left[ 1; \frac{\lambda_L T(y_0 - y_c - \beta T)}{\alpha_L} \right] + \frac{\beta T}{\alpha_L} 0 F_1 \left[ 2; \frac{\lambda_L T(y_0 - y_c - \beta T)}{\alpha_L} \right] \right\}, \quad (B\, 3)$$

which can be finally expressed as a function of contaminant concentration $x$ by means of equation (2.10) (equation (2.13) in the main text).

**References**


Stochastic modelling of phytoremediation


