Supplementary material
Model predicts the impact of root system architecture on soil water infiltration

Andrew Mair[*], Lionel X Dupuy[†], Mariya Ptashnyk[‡]

1 Computation of a rotation matrix R for a particular segment of a root system

For a segment $\mathfrak{R}_i$ of the root system $\mathfrak{R}$, let $\mathbf{v}$ be the direction vector for $\mathfrak{R}_i$. A rotation matrix $R_i \in \mathbb{R}^{3\times 3}$ can be computed, which by left multiplication rotates the vector $\mathbf{e}_3 = (0,0,1)^\top$ into the direction of $\mathfrak{R}_i$. We compute R_i for a given $\mathfrak{R}_i$ using the formula of Rodrigues (1840) as follows:

1. Normalised direction vector of the segment: $\hat{\mathbf{v}} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$.
2. Normalised cross product of the segment: $\mathbf{y} = \frac{\mathbf{e}_3 \times \hat{\mathbf{v}}}{\|\mathbf{e}_3 \times \hat{\mathbf{v}}\|}$.
3. Angle between $\mathbf{e}_3$ and $\hat{\mathbf{v}}$ around the axis $\mathbf{y}$: $\gamma = \cos^{-1}(\mathbf{e}_3 \cdot \hat{\mathbf{v}}) \in [0, \pi]$.
4. Cross product matrix of $\mathbf{y}$:
$$Y = \begin{pmatrix} 0 & -y_3 & y_2 \\ y_3 & 0 & -y_1 \\ -y_2 & y_1 & 0 \end{pmatrix}.$$
5. Rotation matrix:
$$R_i = \begin{cases} I & \text{if } \hat{\mathbf{v}} = \mathbf{e}_3, \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} & \text{if } \hat{\mathbf{v}} = -\mathbf{e}_3, \\ I + \sin(\gamma) Y + (1 - \cos(\gamma)) Y^2 & \text{otherwise.} \end{cases} \tag{1.1}$$

[*]The Maxwell Institute for Mathematical Sciences, Edinburgh, Scotland, United Kingdom (adm13@hw.ac.uk)
[†]Department of Conservation of Natural Resources, Neiker & IKERBASQUE, Basque Foundation for Science, Bilbao, Spain (ldupuy@neiker.eus)
[‡]The Maxwell Institute for Mathematical Sciences, Edinburgh, Scotland, United Kingdom (m.ptashnyk@hw.ac.uk)

2 Numerical method for simulations of empirical and PF models

The PF model is formulated as

$$\partial_t\big(\eta(x)\theta(h)\big) - \boldsymbol{\nabla}\cdot\big(\eta(x)\big(\eta(x)I + H(c_a;x)\big)K_s\kappa(h)\boldsymbol{\nabla}(h+x_3)\big) = 0 \quad \text{in } \Omega\times(0,T]. \tag{2.1}$$

with boundary and initial conditions

$$\begin{aligned} \mathbf{j}\cdot\mathbf{n} &= q_{\text{in}} && x\in\Gamma_1,\ t\in(0,T],\\ \mathbf{j}\cdot\mathbf{n} &= 0 && x\in\Gamma_2,\ t\in(0,T],\\ h &= 0 && x\in\Gamma_3,\ t\in(0,T],\\ h(x,0) &= h_0(x) && x\in\Omega, \end{aligned} \tag{2.2}$$

where $\mathbf{j} = -\eta(x)\big(\eta(x)I + H(c_a;x)\big)K_s\kappa(h)\boldsymbol{\nabla}(h+x_3)$ and

$$\theta(h) = \theta_r + \frac{\theta_s-\theta_r}{(1+|\alpha h|^n)^m}, \tag{2.3}$$

$$\kappa(h) = \left(\frac{\theta(h)-\theta_r}{\theta_s-\theta_r}\right)^l\left[1-\left(1-\left(\frac{\theta(h)-\theta_r}{\theta_s-\theta_r}\right)^{\frac{1}{m}}\right)^m\right]^2. \tag{2.4}$$

The constant $T>0$ is the final time being considered and $\Omega\subset\mathbb{R}^3$ is the Lipschitz domain of soil. See the main text for definitions of all other functions and constants included in (2.1)-(2.4). To discretise in time, $N\in\mathbb{N}$ is set as the number of time steps and the step size is $\tau = T/N$. Then for $n\in 1,...,N$ with $t_n=\tau n$ and $h^n = h(\cdot,t_n)$, the time derivative is approximated using the backward Euler method

$$\partial_t\theta(h^n) \approx \frac{\theta(h^n)-\theta(h^{n-1})}{\tau}.$$

Applying the L-scheme, which exploits the Lipschitz continuity of θ, we write $\theta(h^n)$ in the following form

$$\theta(h^{n,j}) = \theta(h^{n,j-1}) + L(h^{n,j}-h^{n,j-1}) \qquad \text{for } j=1,\dots J,$$

where J is the total number of iterations, $h^{n,0}=h^{n-1}$, and $L\geq L_\theta = \sup_{h\leq 0}|\theta'(h)|$ (List and Radu, 2016). The time discretised and linearised variational form of (2.1), (2.2) then reads as follows: for $n\in\{1,...,N\}$ and $j\in\{1,...J\}$ find $h^{n,j}\in W^{1,2}_{\Gamma_3}(\Omega)$ such that

$$\begin{aligned} &\left\langle\eta(x)Lh^{n,j},\phi\right\rangle + \tau\left\langle\eta(x)[H(c_a;x)+\eta(x)]K_s\kappa_{\tilde{e}}(h^{n,j-1})\left(\boldsymbol{\nabla}h^{n,j}+\mathbf{e}_3\right),\boldsymbol{\nabla}\phi\right\rangle\\ &= \left\langle\eta(x)\theta\left(h^{n-1}\right),\phi\right\rangle - \left\langle\eta(x)\theta(h^{n,j-1}),\phi\right\rangle + \left\langle\eta(x)Lh^{n,j-1},\phi\right\rangle - \tau\left\langle\eta(x)\,g_{\text{in}},\phi\right\rangle_{\Gamma_1}, \end{aligned} \tag{2.5}$$

for all $\phi\in W^{1,2}_{\Gamma_3}(\Omega)$. Here we use the notation $\langle\phi_1,\phi_2\rangle = \int_\Omega\phi_1\phi_2\mathrm{d}x$ and $\langle\phi_1,\phi_2\rangle_{\Gamma_1} = \int_{\Gamma_1}\phi_1\phi_2\mathrm{d}\gamma_x$, along with $W^{1,2}_{\Gamma_3}(\Omega) = \left\{\phi\in L^2(\Omega)\,|\,\partial_{x_i}\phi\in L^2(\Omega),\ \text{for } i=1,2,3,\ \phi=0 \text{ on } \Gamma_3\right\}$, where $L^2(\Omega)$ is the space of all square integrable functions on Ω. The discretisation in space is achieved by projecting (2.5) onto a first-order continuous Galerkin finite element space V_h. The space V_h is defined by the decomposition of Ω into a finite family of disjoint simplexes $\{\mathcal{T}_i\}_{i=1}^{N_h}$ so that $V_h = \left\{\phi\in W^{1,2}_{\Gamma_3}(\Omega)|\,\phi_{|\mathcal{T}_i}\in\mathcal{P}(\mathcal{T}_i),\ \text{for } i=1,...,N_h\right\}$, where $\mathcal{P}(\mathcal{T}_i)$ is the space of first-order polynomials on $\mathcal{T}_i$, see e.g. (Thomée, 2006) for more details on continuous Galerkin finite element discretisations. This numerical scheme was implemented in Python 3 using the FEniCS library (Alnæs et al., 2015).

In (2.5) we use a regularisation κ_ε of (2.4) to satisfy conditions on θ and κ which ensure the convergence of the numerical method used here (List and Radu, 2016). These conditions are as follows:

(A1) The retention function $\theta(\cdot)$ is non-decreasing and Lipschitz continuous.

(A2) The function $\kappa(\cdot)$ is Lipschitz continuous and there exist $\kappa_{\min}$ and $\kappa_{\max}$ such that $0 < \kappa_{\min} \leq \kappa(h) \leq \kappa_{\max} < \infty,\ \forall h \in \mathbb{R}$.

For proof that $\theta(\cdot)$ from (2.3) satisfies (A1) see Section 3. The exact formulation of the regularisation $\kappa_\varepsilon(h)$ and verification of (A2) can be found in Section 4. In addition to these conditions, there exists $\eta_0 > 0$ such that $0 < \eta_0 \leq \eta(x) \leq 1$ for $x \in \Omega$. Furthermore, for any root system $\mathfrak{R}$ and axial facilitation constant $c_a \geq 1$, the flow-anisotropy matrix $H(c_a; x)$ in (2.1) is symmetric and uniformly positive definite, see Section 5 for the proof.

3 Lipschitz continuous nondecreasing retention function (A1)

From (2.3) it can be seen that

$$\left|\theta'(h)\right| = \frac{(\theta_s - \theta_r)\,\alpha mn\,|\alpha h|^{n-1}}{(1+|\alpha h|^n)^{m+1}} = \frac{(\theta_s - \theta_r)\,\alpha mn\,|\alpha h|^{n-1}}{(1+|\alpha h|^n)^{2-\frac{1}{n}}} \quad \forall h \in \mathbb{R}.$$

With the exact values for θ_r, θ_s, α and n stated in the parametrisation and calibration section of the main text, it can be seen that $0 < (\theta_s - \theta_r)\alpha mn < 1$ and $2 - \frac{1}{n} > 1$. It then follows that

$$\left|\theta'(h)\right| \leq \frac{|\alpha h|^{n-1}}{1+|\alpha h|^n} \leq 1 \quad \forall h \in \mathbb{R}.$$

In order to show that the retention function is nondecreasing, the domain of θ is first restricted to the physical situation upon which this paper focuses, namely an unsaturated soil. This translates as nonpositive pressure head $h \leq 0$ and it can then be seen from (2.3) that

$$\theta'(h) = \frac{(\theta_s - \theta_r)\,\alpha mn\,(-\alpha h)^{n-1}}{(1+(-\alpha h)^n)^{m+1}} \geq 0 \qquad \forall h \leq 0.$$

4 Condition (A2) for the regularised κ_ε

The function κ given in (2.4), is bounded above by 1 but it is not Lipschitz continuous in h nor is it bounded below by a value greater than 0. In order to rectify this, a regularisation of κ was constructed and denoted by κ_ε, where $0 < \varepsilon < (\theta_s - \theta_r)$. The first step in this regularisation is to define θ_ε as follows:

$$\theta_\varepsilon(h) = \theta_r + \frac{\theta_s - (\theta_r + \varepsilon)}{(1+|\alpha h|^n)^m} + \frac{\varepsilon}{2} \qquad \forall h \leq 0. \tag{4.1}$$

It can be seen that θ_ε is non-decreasing for $h \leq 0$ and retains the Lipschitz continuity of θ. However, unlike θ, θ_ε exhibits the following limiting behaviour in $h \leq 0$:

$$\theta_\varepsilon(h) \to \theta_s - \frac{\varepsilon}{2} \ \text{ as } \ h \to 0, \quad \theta_\varepsilon(h) \to \theta_r + \frac{\varepsilon}{2} \ \text{ as } \ |h| \to \infty.$$

Then κ_ε is defined as

$$\kappa_\varepsilon(h) = \Big(\frac{\theta_\varepsilon(h) - \theta_r}{\theta_s - \theta_r}\Big)^l \Big[1 - \Big(1 - \Big(\frac{\theta_\varepsilon(h) - \theta_r}{\theta_s - \theta_r}\Big)^{\frac{1}{m}}\Big)^m\Big]^2. \tag{4.2}$$

To show Lipschitz continuity of κ_ε it can first be seen that $|\kappa_\varepsilon'(h)| = |\omega_1(h) + \omega_2(h)|$ for $h \in (-\infty, 0]$, where:

$$\begin{aligned}
\omega_1(h) &= l\Big(\frac{\theta_\varepsilon(h)-\theta_r}{\theta_s-\theta_r}\Big)^{l-1}\Big[1-\Big(1-\Big(\frac{\theta_\varepsilon(h)-\theta_r}{\theta_s-\theta_r}\Big)^{\frac{1}{m}}\Big)^m\Big]^2\theta_\varepsilon'(h),\\
\omega_2(h) &= \omega_{2a}(h)\omega_{2b}(h)\omega_{2c}(h)\omega_{2d}(h)\theta_\varepsilon'(h)\,,\\
\omega_{2a}(h) &= 2\Big(\frac{\theta_\varepsilon(h)-\theta_r}{\theta_s-\theta_r}\Big)^{l}, \qquad \omega_{2b}(h) = 1-\Big(1-\Big(\frac{\theta_\varepsilon(h)-\theta_r}{\theta_s-\theta_r}\Big)^{\frac{1}{m}}\Big)^m,\\
\omega_{2c}(h) &= \Big(1-\Big(\frac{\theta_\varepsilon(h)-\theta_r}{\theta_s-\theta_r}\Big)^{\frac{1}{m}}\Big)^{m-1}, \quad \omega_{2d}(h) = \Big(\frac{\theta_\varepsilon(h)-\theta_r}{\theta_s-\theta_r}\Big)^{\frac{1}{m}-1}.
\end{aligned}$$

Direct calculations imply that for all $h \in (-\infty, 0]$:

$$\begin{aligned}
&\Big|l\Big(\frac{\theta_\varepsilon(h)-\theta_r}{\theta_s-\theta_r}\Big)^{l-1}\Big| \le l\Big(\frac{\varepsilon}{2(\theta_s-\theta_r)}\Big)^{l-1},\\
&\Big|1-\Big(1-\Big(\frac{\theta_\varepsilon(h)-\theta_r}{\theta_s-\theta_r}\Big)^{\frac{1}{m}}\Big)^m\Big| \le \Big|1-\Big(1-\Big(1-\frac{\varepsilon}{2(\theta_s-\theta_r)}\Big)^{\frac{1}{m}}\Big)^m\Big|,\\
|\omega_{2a}(h)| &\le 2\Big(1-\frac{\varepsilon}{2(\theta_s-\theta_r)}\Big)^{l}, \quad |\omega_{2b}(h)| \le 1-\Big(1-\Big(1-\frac{\varepsilon}{2(\theta_s-\theta_r)}\Big)^{\frac{1}{m}}\Big)^m,\\
|\omega_{2c}(h)| &\le \Big(1-\Big(1-\frac{\varepsilon}{2(\theta_s-\theta_r)}\Big)^{\frac{1}{m}}\Big)^{m-1}, \qquad |\omega_{2d}(h)| \le \Big(1-\frac{\varepsilon}{2(\theta_s-\theta_r)}\Big)^{\frac{1}{m}-1}.
\end{aligned}$$

Combining the results above with the boundedness of θ_ε' shows that $\exists L_{\kappa_\varepsilon} > 0$ such that $|\kappa_\varepsilon'(h)| \le L_{\kappa_\varepsilon}$ for all $h \le 0$, thus showing the Lipschitz continuity of κ_ε for $h \le 0$.
Due to the limiting behaviour of θ_ε, it is the case that for all $h \in (-\infty, 0]$,

$$\begin{aligned}
\kappa_\varepsilon(h) &\le \kappa_{\max} = \Big(1-\frac{\varepsilon}{2(\theta_s-\theta_r)}\Big)^{l}\Big[1-\Big(1-\Big(1-\frac{\varepsilon}{2(\theta_s-\theta_r)}\Big)^{\frac{1}{m}}\Big)^m\Big]^2,\\
\kappa_\varepsilon(h) &\ge \kappa_{\min} = \Big(\frac{\varepsilon}{2(\theta_s-\theta_r)}\Big)^{l}\Big[1-\Big(1-\Big(\frac{\varepsilon}{2(\theta_s-\theta_r)}\Big)^{\frac{1}{m}}\Big)^m\Big]^2.
\end{aligned}$$

5 Properties of the flow-anisotropy matrix H

Proposition 5.1. *For any axial facilitation constant $c_a \ge 1$, the flow-anisotropy matrix $H(c_a; x)$ of any root system $\mathfrak{R}$ is uniformly positive-definite i.e, there exists $\lambda > 0$ such that*

$$\zeta^\top H(c_a; x)\zeta \ge \lambda\,|\zeta|^2 \tag{5.1}$$

for all $x \in \Omega$ and all $\zeta \in \mathbb{R}^3$.

Proof. The flow-anisotropy matrix has the form $H(c_a; x) = \sum_{i=1}^{N_\mathfrak{R}} H_i(c_a; x)$ where, for each segment $\mathfrak{R}_i$ in the underlying root system, the associated matrix is $H_i(c_a; x) = \psi_i(x)R_iA_i(1, c_a)R_i^{-1}$. Given a segment $\mathfrak{R}_i$, let $\vartheta_i \in [0, 2\pi)$ be the anti-clockwise angle of the segment from the positive x_1 axis and $\varphi_i \in [0, \pi]$ be the angle of $\mathfrak{R}_i$ from the positive x_3 axis. In this work it is assumed that $\varphi_i \in \left[\frac{\pi}{2}, \pi\right]$ for each segment in the root system. This means that in proving (5.1) for $H_i(c_a; x)$, it is only necessary to consider case 1: $\varphi_i \in \left[\frac{\pi}{2}, \pi\right)$ and case 2: $\varphi_i = \pi$.

Case 1, $\varphi_i \in \left[\frac{\pi}{2}, \pi\right)$**:** Completing the steps detailed in (1.1) yields the rotation matrix,

$$R_i = \begin{pmatrix}
1-(1-\cos(\varphi_i))\cos^2(\vartheta_i) & -(1-\cos(\varphi_i))\sin(\vartheta_i)\cos(\vartheta_i) & \sin(\varphi_i)\cos(\vartheta_i)\\
-(1-\cos(\varphi_i))\sin(\vartheta_i)\cos(\vartheta_i) & 1-(1-\cos(\varphi_i))\sin^2(\vartheta_i) & \sin(\varphi_i)\sin(\vartheta_i)\\
-\sin(\varphi_i)\cos(\vartheta_i) & -\sin(\varphi_i)\sin(\vartheta_i) & \cos(\varphi_i)
\end{pmatrix}.$$

Using $R_i^{-1} = R_i^\top$, the matrix $H_i(c_a; x) = \psi_i(x) R_i A_i(c_a) R_i^{-1}$ can be computed. For all $\zeta \in \mathbb{R}^3$ it is then the case that

$$\begin{aligned}
\zeta^\top H_i(c_a; x)\zeta =& \psi_i(x)\Bigg[c_a\big(1+\overline{S}_i\big)\Big[\cos(\vartheta_i)\sin(\varphi_i)\zeta_1 + \sin(\vartheta_i)\sin(\varphi_i)\zeta_2 + \cos(\varphi_i)\zeta_3\Big]^2 \\
&+ \big(\sin^2(\vartheta_i) + \cos^2(\vartheta_i)\cos^2(\varphi_i)\big)\zeta_1^2 - 2\sin(\vartheta_i)\cos(\vartheta_i)\sin^2(\varphi_i)\zeta_1\zeta_2 \\
&- 2\cos(\vartheta_i)\sin(\varphi_i)\cos(\varphi_i)\zeta_1\zeta_3 + \big(\cos^2(\vartheta_i) + \sin^2(\vartheta_i)\cos^2(\varphi_i)\big)\zeta_2^2 \\
&- 2\sin(\vartheta_i)\sin(\varphi_i)\cos(\varphi_i)\zeta_2\zeta_3 + \sin^2(\varphi_i)\zeta_3^2\Bigg] \\
\geq& \psi_i(x)\Bigg[\Big[\cos(\vartheta_i)\sin(\varphi_i)\zeta_1 + \sin(\vartheta_i)\sin(\varphi_i)\zeta_2 + \cos(\varphi_i)\zeta_3\Big]^2 \\
&+ \big(\sin^2(\vartheta_i) + \cos^2(\vartheta_i)\cos^2(\varphi_i)\big)\zeta_1^2 - 2\sin(\vartheta_i)\cos(\vartheta_i)\sin^2(\varphi_i)\zeta_1\zeta_2 \\
&- 2\cos(\vartheta_i)\sin(\varphi_i)\cos(\varphi_i)\zeta_1\zeta_3 + \big(\cos^2(\vartheta_i) + \sin^2(\vartheta_i)\cos^2(\varphi_i)\big)\zeta_2^2 \\
&- 2\sin(\vartheta_i)\sin(\varphi_i)\cos(\varphi_i)\zeta_2\zeta_3 + \sin^2(\varphi_i)\zeta_3^2\Bigg] \\
=& \psi_i(x)\,|\zeta|^2 \geq \psi_i^*\,|\zeta|^2 \quad \forall x \in \Omega,
\end{aligned}$$

where $\psi_i^* = \min_{x\in\Omega} \psi_i(x) > 0$.

Case 2, $\varphi_i = \pi$**:** In this scenario the flow-anisotropy matrix associated to the segment is

$$H_i(c_a; x) = \psi_i(x) R_i A_i(c_a) R_i^{-1} = \begin{pmatrix} \psi_i(x) & 0 & 0 \\ 0 & \psi_i(x) & 0 \\ 0 & 0 & \psi_i(x)c_a(1+\overline{S}_i) \end{pmatrix},$$

and as a result $\zeta^\top H_i(c_a; x)\zeta \geq \psi_i^*\,|\zeta|^2 \quad \forall x \in \Omega$ and $\forall \zeta \in \mathbb{R}^3$. For any root segment $\mathfrak{R}_i$ the matrix H_i is therefore uniformly positive-definite. Since the sum of uniformly positive-definite matrices is also uniformly positive-definite, there must then exist $\lambda > 0$ such that $\zeta^\top H(c_a; x)\zeta \geq \lambda\,|\zeta|^2$ for all $x \in \Omega$, $\zeta \in \mathbb{R}^3$ and $\psi_i^* > 0$ with $i = 1, \ldots, N_{\mathfrak{R}}$. □

6 Optimisation

Bayesian optimisation is based on Bayes theorem (Williams, 1991) and uses two components to search for the minimiser of a cost function $v : \mathcal{Z} \to \mathbb{R}$, where $\mathcal{Z} \subset \mathbb{R}^d$ is the parameter space with dimension $d > 0$. The first component is the surrogate probability model $\mathcal{M}$ which is used as the prior distribution for the value of a cost function $v(\mathbf{z})$ evaluated at a parameter value $\mathbf{z} \in \mathbb{R}^d$. Given n observations of the value of the cost function, $\mathcal{D} = \{(\mathbf{z}_1, v(\mathbf{z}_1)) \ldots (\mathbf{z}_n, v(\mathbf{z}_n))\}$, the surrogate $\mathcal{M}$ is conditioned on $\mathcal{D}$ using Bayes theorem. The second component of Bayesian optimisation is the acquisition function $\Lambda : \mathbb{R}^d \to \mathbb{R}$, which is defined by the choice of surrogate $\mathcal{M}$ and depends upon the current contents of the observation set $\mathcal{D}$. This function Λ proposes the best location $\mathbf{z}' \in \mathcal{Z}$ at which to evaluate v, so as to most efficiently find its minimiser $\mathbf{z}^*$. The proposed next evaluation location for any choice of acquisition function Λ is the value $\mathbf{z}'$ which maximises $\Lambda(\mathbf{z}')$ (Brochu et al., 2010). The full Bayesian optimisation routine of N_{fevs} function calls is given by Algorithm 1.

Algorithm 1: Bayesian optimisation (Dewancker et al., 2016)

Inputs : v, $\mathcal{M}$, Λ, N_{fevs}
Output: $\mathbf{z}^*$ optimal value

$\mathcal{D} \leftarrow \{(\mathbf{z}_1, v(\mathbf{z}_1)) ... (\mathbf{z}_n, v(\mathbf{z}_n))\}$ initial data;
for $i \leftarrow 1$ **to** N_{fevs} **do**
  $\mathcal{M}(v|\mathcal{D}) \leftarrow$condition surrogate model;
  $\mathbf{z}' \leftarrow \underset{\mathbf{z}\in\mathcal{Z}}{\operatorname{argmax}}\, \Lambda(\mathbf{z})$ acquisition function proposal location;
  $v(\mathbf{z}')$ evaluate at proposed location;
  $\mathcal{D} \leftarrow \mathcal{D} \cup (\mathbf{z}', v(\mathbf{z}'))$ update observation set;
$\mathbf{z}^* \leftarrow \underset{\mathbf{z}\in\mathcal{D}}{\operatorname{argmin}}\, v(\mathbf{z})$

There are a number of options for the choice of surrogate model $\mathcal{M}$ and acquisition function Λ (Brochu et al., 2010), but in this work we opted for the combination of a Gaussian process $\mathcal{GP}$ surrogate model and an expected improvement acquisition function:

$$\mathcal{M}(v) = \mathcal{GP}(\mu, k), \tag{6.1}$$
$$\Lambda(\mathbf{z}) = \mathbb{E}(w(\mathbf{z})|\mathcal{D}) = \big(v(\mathbf{z}^\dagger) - \mu_{v|\mathcal{D}}(\mathbf{z})\big) F_{v|\mathcal{D}}(v(\mathbf{z}^\dagger)) + \Sigma_{v|\mathcal{D}}(\mathbf{z}) f_{v|\mathcal{D}}(v(\mathbf{z}^\dagger)), \tag{6.2}$$

where $\mathbf{z}^\dagger \in \mathcal{D}$ is the best minimiser of v that has been observed so far and

$$w(\mathbf{z}) = \max(0, v(\mathbf{z}^\dagger) - v(\mathbf{z}))$$

is the so-called improvement function. The functions $F_{v|\mathcal{D}}$ and $f_{v|\mathcal{D}}$ are the cumulative and probability density functions of the conditioned distribution $\mathcal{M}(v|\mathcal{D})$ respectively. By the definition of a Gaussian process, any finite vector of n observations $\mathbf{v}_n = (v(\mathbf{z}_1), ..., v(\mathbf{z}_n))^\top$ follows a multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}_n, \Sigma_n)$ with mean vector $\boldsymbol{\mu}_n = (\mu(\mathbf{z}_1), ..., \mu(\mathbf{z}_n))^\top$ and covariance matrix $\Sigma_n \in \mathbb{R}^{n\times n}$, where $\Sigma_{n,ij} = k(\mathbf{z}_i, \mathbf{z}_j)$ for $i, j = 1, \ldots, n$. For a given set of cost function evaluations $\mathcal{D} = \{(\mathbf{z}_1, v(\mathbf{z}_1)), ..., (\mathbf{z}_n, v(\mathbf{z}_n))\}$, the structure of the Gaussian process allows the formulation of the conditioned surrogate model

$$\mathcal{M}(v|\mathcal{D}) = \mathcal{N}\big(\mu_{v|\mathcal{D}}, \Sigma_{v|\mathcal{D}}\big),$$

where the components of this distribution are

$$\mu_{v|\mathcal{D}}(\mathbf{z}) = \mu(\mathbf{z}) + \mathbf{k}_n(\mathbf{z})\Sigma_n^{-1}(\mathbf{v}_n - \boldsymbol{\mu}_n),$$
$$\Sigma_{v|\mathcal{D}}(\mathbf{z}) = k(\mathbf{z}, \mathbf{z}) - \mathbf{k}_n(\mathbf{z})\Sigma_n^{-1}\mathbf{k}_n(\mathbf{z})^T,$$

with $\mathbf{k}_n(\mathbf{z}) = (k(\mathbf{z}, \mathbf{z}_1), ..., k(\mathbf{z}, \mathbf{z}_n))$.
The first component of the weighted sum in (6.2) corresponds to the exploitation of low mean values and the second component incorporates the exploration of regions of the parameter space with high uncertainty regarding the value of the cost function v. It is common to add a tuning parameter ξ into the acquisition function so that

$$\Lambda(\mathbf{z}) = \big(v(\mathbf{z}^\dagger) - \mu_{v|\mathcal{D}}(\mathbf{z}) - \xi\big) F_{v|\mathcal{D}}(v(\mathbf{z}^\dagger)) + \Sigma_{v|\mathcal{D}} f_{|\mathcal{D}}(v(\mathbf{z}^\dagger)),$$

and increasing values of ξ lead to increased exploration by the algorithm. In our work we use $\xi = 0.01$, which is the common default value (Brochu et al., 2010). For the mean function μ in (6.1) we follow the usual convention of $\mu(\mathbf{z}) \equiv 0$ and for the covariance kernel k we choose the popular Matérn function with $\nu = 1/2$, see (Rasmussen, 2003) for information on many of the available alternatives

References

Alnæs, M. S., J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M. E. Rognes, and G. N. Wells (2015). The fenics project version 1.5. *Archive of Numerical Software 3*(100).

Brochu, E., V. M. Cora, and N. De Freitas (2010). A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv preprint arXiv:1012.2599*.

Dewancker, I., M. McCourt, and S. Clark (2016). Bayesian optimization for machine learning: A practical guidebook. *arXiv preprint arXiv:1612.04858*.

List, F. and F. A. Radu (2016). A study on iterative methods for solving richards' equation. *Computational Geosciences 20*(2), 341–353.

Rasmussen, C. E. (2003). Gaussian processes in machine learning. In *Summer school on machine learning*, pp. 63–71. Springer.

Rodrigues, O. (1840). *Des lois géométriques qui régissent les déplacements d'un système solide dans l'espace: et de la variation des cordonnées provenant de ces déplacements considérés indépendamment des causes qui peuvent les produire.*

Thomée, V. (2006). *Galerkin finite element methods for parabolic problems*, Volume 25 of *Springer Series in Computational Mathematics*. Springer Berlin Heidelberg.

Williams, D. (1991). *Probability with martingales*. Cambridge university press.