

Parameter estimation for land-surface models using Neural Physics

Ruiyue Huang¹, Claire E. Heaney^{2,3}, and Maarten van Reeuwijk¹

¹Department of Civil and Environmental Engineering, Imperial College London, SW7 2AZ London, UK

²Department of Earth Science and Engineering, Imperial College London, SW7 2AZ London, UK

³Imperial-X, Imperial College London, W12 7SL London, UK

Correspondence: Maarten van Reeuwijk (m.vanreeuwijk@imperial.ac.uk)

Abstract. We propose a novel inverse-modelling approach ~~which that~~ estimates the parameters of a simple land-surface model (LSM) by assimilating data into a differentiable, physics-based forward model formulated using convolutional operations. The governing equations are expressed within ~~a machine-learning framework using~~ the Neural Physics ~~approach~~framework, allowing direct gradient-based optimisation of time-dependent parameters without the need to derive and maintain adjoint formulations. The model parameters are ~~updated-estimated~~ by minimising the mismatch between model predictions and synthetic or observational data. Although differentiability is ~~achieved through functionality in-enabled through~~ machine-learning libraries, the forward model itself remains entirely physics-based and ~~no part of either- neither~~ the forward model ~~or-nor~~ the parameter estimation ~~involves-procedure involve~~ training.

~~In-order-to-test-To evaluate~~ the approach, ~~a synthetic dataset is generated-we first generate synthetic observations of soil~~ temperature by running the forward model with known parameter values ~~to create a time series of soil temperature that serves as observations for-and subsequently treat these parameters as unknown in~~ an inverse problem ~~in which the parameters are assumed unknown and subsequently estimated~~. We show that ~~it is not possible to obtain a reliable estimate of the parameters using a time series of soil temperature-observed~~ observations of soil temperature at a single depth. ~~Using measurements are insufficient to reliably constrain the model parameters. Using observations~~ at two depths, ~~however, does yield~~ reliable parameter estimates ~~can be obtained although it is not possible to differentiate between-, although the individual contributions of~~ latent and sensible heat fluxes cannot be distinguished.

We also apply the approach to urban flux tower data from Phoenix, United States, and show that the thermal conductivity, volumetric heat capacity and the combined sensible-latent heat transfer coefficient can be reliably estimated whilst using an observed value for the effective surface albedo. The resulting model accurately predicts the outgoing longwave radiation, conductive soil fluxes and the combined sensible-latent heat fluxes, demonstrating that the Neural Physics framework can be used to accurately determine ~~parameters in LSMs. The approach also provides insight into parameter identifiability and confounding, which are difficult to assess in traditional calibration workflows. Limitations include the need to refactor existing LSMs coded in classical programming languages in order to capitalise on this approach and the lack of~~ the parameters of the particular LSM used here. This model is intentionally simple and does not include, for example, a subsurface moisture model. This simplicity facilitates an exploration of parameter identifiability, confounding and equifinality.

1 Introduction

The exchange of momentum and thermal energy between the land surface and the atmosphere is of great importance to atmospheric processes and thermal comfort (Oke, 2017). At present, such processes are typically modelled with land-surface models (LSMs) using meteorological forcing data and various parameters associated with the site. Many LSMs have been developed over the years, each differing in how they handle the urban surface, and in their modelling of various morphological, vegetation, hydrological and anthropogenic processes. Modern LSMs are capable of closely replicating the heterogeneous surface structure of cities (Lee and Lee, 2020; Krayenhoff and Voogt, 2007), and capturing the effects of various urban processes, such as urban heat storage and vegetation dynamics, using various soil, hydrological or vegetation models (Lipson et al., 2017; Arsenault et al., 2018).

However, larger complexity of LSMs does not always translate to better predictions (Grimmond et al., 2010, 2011; Lipson et al., 2024). The Urban-PLUMBER project (Lipson et al., 2024) evaluated the performance of 30 LSMs in estimating surface fluxes at a suburban site in Melbourne, Australia. The study found that models with highly complex urban schemes often do not perform as well as models with simpler urban schemes, as their performances are penalised by having relatively simple representation of hydrological/vegetation models. Most complex models were also unable to make efficient use of site-specific information.

Besides capturing the complexity of land-surface interactions, another key challenge in modelling urban surface fluxes lies in estimating (or calibrating) the relevant input parameters (Lipson et al., 2017; Chaney et al., 2016). LSMs typically have different sensitivities to changes in different parameters, making it difficult to perform data assimilation or improve model accuracy by manually updating parameter values, as a slight alteration can lead to large and unexpected changes in the final output (Massoud et al., 2019). Grimmond et al. (2011) found that poor estimation of parameter values can result in a significant decline in performance of LSMs. It is hence critical to find a reliable method of parameter estimation.

Raoult et al. (2024) presented an overview of the popular techniques for estimating the parameters of LSMs, which include statistical methods and inverse modelling. Statistical methods involve estimating the probability distributions of model parameters based on prior knowledge and observed data. Some examples include the Markov Chain Monte Carlo (MCMC) method (Hastings, 1970), which draws samples from posterior distributions to determine the likely values of parameters. However, statistical methods can be computationally intensive, and therefore have been mainly used in computationally inexpensive LSMs such as Simplified PnET (SIPNET) ecosystem model (Fer et al., 2018), or to estimate isolated processes (e.g. plant respiration (Jones et al., 2024)). Inverse modelling employs optimisation algorithms, such as the gradient-descent method or genetic algorithms, to determine the optimal set of parameters that minimises the error between the model output and observational data. Such minimisation algorithms have been shown to have better overall optimisation efficiency than statistical methods (Raoult et al., 2024). However, inverse modelling algorithms, especially those based on the gradient-descent method, are prone to become stuck in local minima, leading to non-unique solutions that are dependent on the initial conditions (Kuppel et al., 2014; Raoult et al., 2016).

A number of researchers have begun exploring the coupling of machine-learning models with differentiable physics-based models to construct differentiable hybrid models (Shen et al., 2023). In these approaches, neural networks are embedded within a physics-based framework, which is often trained end-to-end, to infer unknown parameters or represent unknown physical relationships. Several studies focus on learning physical processes or closure relationships. For example, Fang and Gentine (2024) proposed a hybrid model to learn the impact of soil water stress on photosynthesis and evapotranspiration, outperforming traditional terrestrial biosphere models that use empirical soil water stress parametrisations. In a similar vein, ElGhawi et al. (2023) represent leaf surface and aerodynamic resistances using two neural networks within an evapotranspiration model, while Norouzi et al. (2026) use neural networks to separate adsorption and capillary effects on the soil water retention curve. Other work embeds neural networks within land-surface models, learning a snow process representation (Deck et al., 2026) and constitutive relationships for frozen soil (Ren et al., 2026). Hybrid models have also been applied to inverse problems and parameter estimation. For example, Ouyang et al. (2025) re-implemented the Xin'anjiang (XAJ) hydrological model in PyTorch representing the discretised equations for run-off and evapotranspiration by neural networks. An inverse problem is solved to learn the weights of an LSTM model which predicts the 15 parameters of the XAJ model. With limited data, their model gave better streamflow predictions than manually calibrating XAJ. A further example is JAX-CanVeg (Jiang et al., 2025), in which a multi-layer land-surface canopy model is augmented with a neural-network-based closure for leaf relative humidity, improving predictions of latent heat and net ecosystem exchange across four flux tower sites. These studies demonstrate the potential of differentiable hybrid models to improve the performance and calibration of LSMs. However, challenges still remain for these differentiable hybrid methods, including sensitivity to training data; tendency of predictions to become physically inconsistent; and computational cost, arising when physics-based models require many forward simulations for parameter optimisation, particularly in end-to-end differentiable settings (Fang and Gentine, 2024; Dukes et al., 2026).

A related class of differentiable but fully physics-based models ~~have~~has also been developed (Zhu et al., 2021; Bezgin et al., 2023; Chen et al., 2026). One such training-free approach, called Neural Physics, capitalises on the equivalence between discretisation stencils and convolutional operators, allowing partial differential equations and their solvers to be expressed as the convolutional layers of neural networks with analytically determined weights. ~~This formulation enables classical physical models to be implemented in a differentiable framework whilst remaining entirely physics-based. Neural Physics and the associated NN4PDEs code~~While differentiability can also be achieved using general-purpose automatic differentiation frameworks (e.g. Julia, or JAX and NumPy-based implementations), Neural Physics provides an alternative to these by formulating numerical discretisations of physical laws using machine-learning primitives such as convolutional operations. The motivation for using Neural Physics in this work extends beyond differentiability. Among its principal advantages are its established and versatile framework, differentiability, reduced implementation complexity and support for a broad range of numerical methods and hardware architectures. The forward model has been applied to a wide range of problems including single-phase flow benchmarks (Chen et al., 2026), multiphase flow (Chen et al., 2024), shallow water dynamics (Chen et al., 2025) and applications in nuclear engineering (Phillips et al., 2023, 2024). ~~NN4PDEs has several advantages over traditional codes. Firstly, NN4PDEs is platform agnostic as it runs~~(Phillips et al., 2023). It is also platform agnostic, being able to execute on CPUs, GPUs and AI accelerators (Cerebras, 2026; Graphcore, 2026) without modification. ~~Secondly, it has a state-of-the-art~~

~~multigrid solver.~~ The end-to-end differentiability of Neural Physics enables the computation of sensitivities of objective functions with respect to model parameters, allowing standard optimisation tools from machine-learning libraries to be used for parameter estimation. Furthermore, the use of convolutional operations and other functions from machine-learning libraries abstracts away much of the implementation complexity associated with accelerator portability and adjoint development. Neural Physics and the associated NN4PDEs code also supports a broader class of numerical methods, including higher-order discretisations (Phillips et al., 2024); finite volume (Xue et al., 2026), finite element and finite difference discretisations (Chen et al., 2026) and multigrid solvers (Phillips et al., 2023). The latter can be exactly represented as a U-Net ~~architecture (Chen et al., 2026, 2024; Phillips et al., 2023), which requires architectures (Phillips et al., 2023), which require~~ no training and ~~reduces~~ reduce the number of iterations needed for convergence compared with classical Jacobi-based schemes. Although not used in this work, the multigrid solver would reduce the computational burden in more complex LSMs. ~~Thirdly, and most importantly in the present context, end-to-end differentiability of Neural Physics enables the computation of sensitivities of objective functions with respect to model parameters, allowing standard optimisation tools from machine-learning libraries to be used for parameter estimation.~~ We hope to exploit these advantages for a range of inverse problems in future work.

The NN4PDEs code has recently been used to determine the conductivity of the subsurface given a set of steady-state observations (Li et al., 2024). Here, we extend ~~this~~ Li et al.'s methodology to a transient application to determine the parameters of a simple land-surface model, and to study whether a parameter can be uniquely determined from the observations (parameter identifiability), whether the effects of different parameters on the model response can be distinguished from one another (confounding) and whether many different parameter combinations result in similar parameter values (equifinality). Although developed independently, our proposed AI-enabled physics-based approach is similar to the NoahPy model which has been developed to improve the representation of permafrost in LSMs (Tian et al., 2026). In that work, numerical discretisations are formulated as a recurrent neural network, enabling parameter estimation from observations at multiple soil depths. The model estimates four parameters using the Adam optimiser and was shown to converge more efficiently than traditional gradient-free optimisation methods.

The key contribution of this paper is the formulation of a land-surface model as within the Neural Physics framework, an end-to-end, fully differentiable, physics-based system that enables efficient parameter estimation. ~~The LSM includes a full~~ We intentionally chose a simple LSM that still includes a surface energy balance ~~equation and time-dependence, in order to isolate parameter identifiability, confounding and equifinality. The surface energy balance equation is~~ coupled to a soil layer (Section 2.1), but does not incorporate a subsurface moisture model or explicitly represent buildings. By expressing the governing equations as convolutions within a machine-learning framework using Neural Physics (Section 2.2), we eliminate the need to derive and maintain adjoint models, which is a significant practical barrier in traditional LSM calibration workflows. Instead, we demonstrate that our approach can be applied directly using gradient-based optimisation to estimate parameters in time-dependent problems (Section 2.3), a process that can otherwise be tedious (Lipson et al., 2024). Using a synthetic dataset for which the correct input parameters are known, the sensitivity to the initial estimate and dependencies between parameters are investigated (Section 3.1). We show that simultaneous measurements at two depths are needed to obtain reliable estimates for the soil properties independent of the initial estimates (Section 3.2). Finally, we use the model to determine the

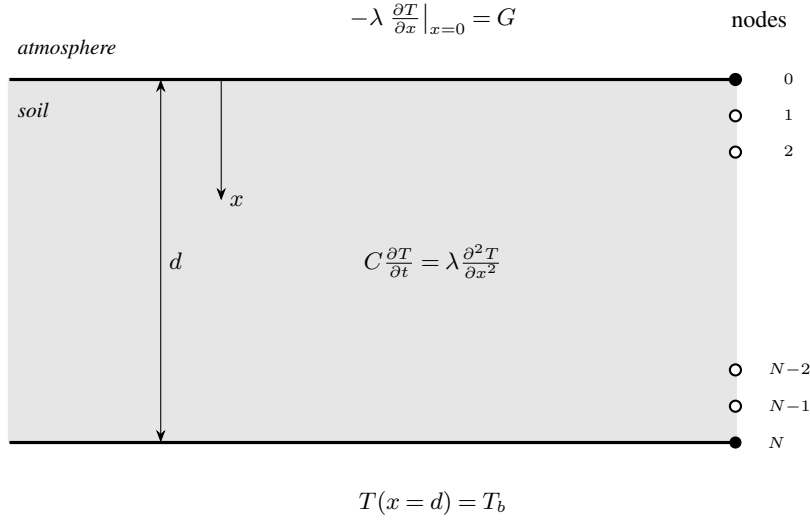


Figure 1. Definition sketch of the domain with the governing equations and boundary conditions. Shown on the right are the nodes of the discretised problem.

site parameters for a 150-hour period of the flux data recorded in the city of Phoenix, United States and evaluate the model on a subsequent 50-hour period (Section 3.3). This approach provides insight into ~~parameter identifiability and confounding effects~~ the degree of underdetermination in the system (e.g., identifiable parameter combinations and the role of observational configuration), which are often difficult to assess using conventional calibration methods. We comment on this in the results and conclusions (Sections 3 and 4).

2 Methods

2.1 Land-surface model

The land-surface model considered comprises a soil layer coupled to a surface energy balance at the soil-atmosphere interface (Figure 1). The surface energy balance equation is given by

$$G(t) = (1 - \alpha)K^\downarrow(t) + L^\downarrow(t) - \epsilon\sigma T_s^4(t) - h(1 + \beta^{-1})(T_s(t) - T_a(t)), \quad (1)$$

where G is the ground heat flux, $K^\downarrow$ and $L^\downarrow$ are the downwelling shortwave and longwave radiation respectively, α is the surface albedo, ϵ is the emissivity, $\sigma = 5.67 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$ is the Stefan-Boltzmann constant, h is the heat transfer coefficient, T_a is the atmospheric temperature and T_s is the absolute surface temperature. The quantity β is the Bowen ratio, defined as

$$\beta = \frac{Q_H}{Q_E}, \quad (2)$$

where Q_H is the sensible heat flux and Q_E is the latent heat flux. The enthalpy equation for the soil is given by

$$C \frac{\partial T}{\partial t} = \lambda \frac{\partial^2 T}{\partial x^2}, \quad (3)$$

145 where x is a downward coordinate starting at the surface, $T(x, t)$ is the soil temperature, C is the volumetric heat capacity and λ is the thermal conductivity. The boundary conditions are given by

$$-\lambda \frac{\partial T}{\partial x}(x=0, t) = G(t), \quad T(x=d, t) = T_b, \quad (4)$$

where d is the vertical extent of the soil layer and T_b is the fixed temperature at the bottom boundary of the soil layer. The initial condition is given by

$$150 \quad T(x, t=0) = T_b. \quad (5)$$

2.2 Forward model

2.2.1 Discretisation

Time and space are discretised uniformly as $t^n = n\Delta t$ and $x_i = i\Delta x$, respectively, where Δt is the time increment, $\Delta x = d/N$ is the grid increment and N is the number of intervals. Using a central finite difference discretisation in space and a backward

155 Euler discretisation in time, (3) becomes

$$-rT_{i+1}^{n+1} + (1+2r)T_i^{n+1} - rT_{i-1}^{n+1} = T_i^n \quad \forall i \in \{1, \dots, N-1\}, \quad (6)$$

where $T_i^n = T(x_i, t^n)$ and $r = \lambda\Delta t/(C\Delta x^2)$. The backward Euler scheme is chosen here because it is absolutely stable and allows for large time-steps, which is useful for the computational efficiency of the inverse model.

The surface boundary condition in (1) is imposed by choosing a second order central discretisation for the surface flux and
160 combining this with the PDE at the boundary (Bradle, 2007), which after discretisation with an Euler backward time-integration method becomes:

$$\frac{T_0^{n+1} - T_0^n}{\Delta t} = \frac{2\lambda}{C\Delta x} \left(\frac{T_1^{n+1} - T_0^{n+1}}{\Delta x} + \frac{G(t^{n+1})}{\lambda} \right), \quad (7)$$

where the ground heat flux $G(t^{n+1})$ uses a mixed discretisation for the outgoing longwave radiation term:

$$G(t^{n+1}) = (1-\alpha)K^\downarrow(t^{n+1}) + L^\downarrow(t^{n+1}) - \epsilon\sigma(T_0^{n+1})^4 - h(1+\beta^{-1})(T_0^{n+1} - T_a(t^{n+1})). \quad (8)$$

165 The discretised bottom boundary condition is given by

$$T_N^{n+1} = T_b. \quad (9)$$

In order to avoid inverting a matrix, the system (6-9) is solved using Jacobi iterations (Greenbaum, 1997). For the PDE discretisation in (6), this implies an iteration scheme of the form

$$T_i^{n+1, (k)} = \frac{T_i^n + r \left(T_{i+1}^{n+1, (k-1)} + T_{i-1}^{n+1, (k-1)} \right)}{1 + 2r}, \quad (10)$$

170 where k denotes the iteration number. The surface boundary condition (8) is solved by linearising the outgoing longwave radiation term (i.e. $\epsilon\sigma \left(T_0^{n+1,(k-1)}\right)^3 T_0^{n+1,(k)}$).

2.2.2 Implementation of Forward Model with Neural Physics

The NN4PDEs code of Neural Physics uses a convolutional layer with pre-determined weights to implement the finite difference scheme described in Section 2.2.1. Equation (10) can be re-written as

$$175 \quad T_i^{n+1,(k)} = \frac{T_i^n}{1+2r} + \mathbf{w} * \mathbf{T}^{n+1,(k-1)} \Big|_i \quad (11)$$

in which $*$ represents the discrete convolution, defined as

$$\mathbf{w} * \mathbf{T} \Big|_i \equiv \sum_{j=-1}^1 w_j T_{i+j}, \quad (12)$$

where $\mathbf{T}^{n+1,(k)}$ is a vector containing the nodal values of temperature at iteration k , and the vector $\mathbf{w}$ has values $[r/(1+2r), 0, r/(1+2r)]$. (11) and (12) apply to all interior nodes $i \in \{1, \dots, N-1\}$. The discretisation in (11) is analogous to a
180 convolutional layer in a neural network where the filter or kernel has weights given by $\mathbf{w}$.

The PyTorch machine-learning library is used to build the forward model. To initialise the neural network, the convolution function and solver for the implicit time-stepping scheme are created. Within the solver, Jacobi iteration for interior nodes is performed by repeated application of the convolution function (11). In each iteration, the boundary temperatures T_0^{n+1} and T_N^{n+1} are updated and appended to the vector for interior nodes to give the full temperature vector. Once the error, which is
185 taken as the maximum absolute difference between consecutive approximations, is smaller than the defined tolerance, the latest temperature vector is saved and used as the new guess for temperature at the next time level. The iteration procedure is repeated for subsequent time steps until the final time is reached. Convergence checks have been performed to ensure that the numerical scheme is stable and consistent (Huang, 2024).

2.3 Inverse model

190 The inverse problem aims to find a set of parameters $\mathbf{p}$ that could have given rise to a set of observations. First, an initial guess is made for the values of the parameters. Second, the forward model is solved through time. Third, a data mismatch functional is calculated, based on the difference between the model output and the observations:

$$J(\mathbf{p}) = \frac{1}{|\mathcal{I}||\mathcal{N}|} \sum_{i \in \mathcal{I}} \sum_{n \in \mathcal{N}} \left(T(x_i, t^n; \mathbf{p}) - \tilde{T}(x_i, t^n) \right)^2, \quad (13)$$

where $T(x_i, t^n; \mathbf{p})$ denotes the model prediction for parameters $\mathbf{p}$, $\tilde{T}(x_i, t^n)$ denotes the observations, $\mathcal{I}$ is a set of indices which
195 indicate where the observations are taken and $\mathcal{N}$ is a set of indices indicating when there are observations. The cardinality or size of these sets is represented by $|\mathcal{I}|$ and $|\mathcal{N}|$. The final stage is to minimise this mismatch with respect to the parameters:

$$\mathbf{p}^* = \arg \min_{\mathbf{p}} J(\mathbf{p}), \quad (14)$$

Table 1. Input parameter ranges and values for the land-surface model. The "Value" column contains the parameter values used to create the synthetic dataset. Data sourced from Brutsaert (2005); Oke (2002); Laloui and Rotta Loria (2020); Oke (2017).

Variable		Typical range	Value	Units
Assumed Constants				
Emissivity (bare soil)	ϵ	0.95–0.98	0.95	—
Total depth	d	0.5–10.0	1.0	m
Time series				
Direct solar irradiance	S_b	—	800	W m^{-2}
Re-emitted longwave radiation	$L^\downarrow$	—	100	W m^{-2}
Atmospheric temperature	T_a	—	296	K
Parameters to be determined				
Albedo	α	0.05 – 0.5	0.2	—
Volumetric Heat capacity	C	Sandy soil: 1.28 – 2.96 Clay soil: 1.42 – 3.10 Peat soil: 0.58 – 4.02	2.2	$\text{MJ m}^{-3} \text{K}^{-1}$
Thermal conductivity	λ	Sandy soil: 0.30 – 2.20 Clay soil: 0.25 – 1.58 Peat soil: 0.06 – 0.50	0.8	$\text{W m}^{-1} \text{K}^{-1}$
Initial / boundary temperature	T_b	Warm: 273 – 303 Snow: 263 – 301 Arid: 268 – 308	293	K
Heat transfer coefficient	h	5 – 30	15	$\text{W m}^{-2} \text{K}^{-1}$
Bowen Ratio	β	0.5 – 10	1.5	—

where $\mathbf{p}^*$ is the optimised parameter vector. Classical optimisation methods such as stochastic gradient descent can be used to minimise the data mismatch and calculate an iterative update of the parameters based on the gradient (Fletcher, 1987; Nocedal and Wright, 2006). We use a modified version of stochastic gradient descent known as Adam, which scales the step length with estimates of the mean and variance of the gradients during the iterative process, in order to improve the speed and stability of the algorithm (Kingma and Ba, 2015). Each parameter has an associated step length, the initial value of which set according to the order of magnitude of the parameter. C has the largest initial step length of 10^4 , T_b has an initial step length of 1, and the remaining parameters have initial step lengths of 0.1. We note that in a machine-learning context, step length is more often referred to as learning rate. However, as we are using an optimisation method to minimise a functional without training a neural network, we use the term step length, in line with classical texts on optimisation (Fletcher, 1987; Nocedal and Wright, 2006).

The optimisation procedure involves solving the forward model, minimising the mismatch between the model output and the observations using the Adam method to give an improved estimation of parameter values that are consistent with the observations. These steps are repeated until the maximum number of iterations (defined as 150) is reached, or the solution has converged. The solution is considered converged at iteration n if both the relative change in the data mismatch functional J is less than 0.01% and the relative changes in all parameters p are less than 0.1%. This procedure is illustrated in Figure 2. The optimisation process is expedited using Python’s multi-processing library and the coding platform Google Colab. Each active session in Google Colab uses 2 CPUs, allowing a maximum of 10 trials to be run simultaneously with 5 active sessions.

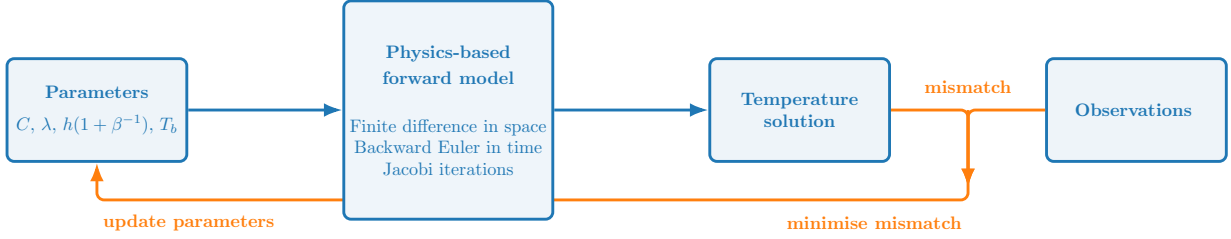


Figure 2. A schematic diagram of the inverse problem. Initial guesses for a set of parameters are fed into the forward model (here, a physics-based model expressed as convolutions using Neural Physics) resulting in a solution for temperature (blue arrows). The mismatch between temperature and the observations is minimised with gradient descent and the parameters are updated (orange arrows). This process continues until convergence.

3 Results

3.1 Observations at a single depth

To investigate the performance of the inverse model, we produce synthetic data using the forward model, so that the correct parameter values are known. We initially pursue a ‘naive’ approach, which is to simply run the inverse model to determine the six main input parameters, namely $p = \{\alpha, \beta, h, C, \lambda, T_b\}$, without consideration of the physics of the underlying model.

Table 1 shows all the inputs that are needed to obtain model output, including the parameter values used to create the synthetic dataset. There is relatively little variation in the emissivity ϵ of longwave radiation, and we will therefore simply assume a fixed value of 0.95. The soil thickness must be chosen such that it is larger than the penetration thickness $d_\omega = \sqrt{8\pi^2\lambda/(C\omega)}$ (Carslaw and Jaeger, 1959), where ω is the angular frequency of a diurnal cycle. For the synthetic dataset, $d_\omega = 0.628$ m, which means that $d = 1$ m is sufficiently deep to assume that the temperature remains unchanged at that depth. The soil layer is discretised with $N = 100$ intervals. A large number of intervals is used to ensure that spatial discretisation errors will not affect the results.

The shortwave radiation varies according to

$$K^\downarrow(t) = S_b \cos(\theta(t)). \quad (15)$$

Here, θ is the solar zenith angle, which is the angle between the direction of the sun's rays and the axis perpendicular to the ground, and S_b is the direct solar irradiance when $\theta = 0$. The solar zenith angle θ is parameterised as

$$\theta(\hat{t}) = \begin{cases} -\frac{\pi}{2} & \text{if } \hat{t} \leq 7; \\ -\frac{\pi}{2} + \frac{\pi}{14}(\hat{t} - 7) & \text{if } 7 < \hat{t} < 21; \\ \frac{\pi}{2} & \text{if } \hat{t} \geq 21 \end{cases}$$

where $\hat{t}$ represents the hour in the day and is given by $t/3600 \bmod 24$. It is assumed that the sun rises at 7am, and sets at 9pm. Before 7am and after 9pm, $\theta = -\frac{\pi}{2}$ which gives zero $K^\downarrow$. Between 7am and 9pm, $K^\downarrow$ varies sinusoidally and peaks at 2pm. The longwave radiation $L^\downarrow$ and atmospheric temperature T_a are assumed constant. The values assumed for S_b , $L^\downarrow$ and T_a are stated in Table 1. The simulation is run for 100 hours and the temperatures recorded over time at a depth of 5 cm are used as

the observations to determine all six parameters with the inverse model.

In order to investigate whether the parameters to be determined are sensitive to the initial parameter values, 50 trials are performed, each starting from different initial parameter values. The parameters for trial i are initialised using $\mathbf{p}_i = \mathbf{p}_i^{min} + R(\mathbf{p}_i^{max} - \mathbf{p}_i^{min})$, where $\mathbf{p}_i^{min}$ and $\mathbf{p}_i^{max}$ are the minimum and maximum parameter values defined such that the range of the initial values is sufficiently large, but still lies within the typical ranges of the parameter (see Table 1), and R is a random number taken from a uniform probability distribution.

The results from the 50 trials are shown in Figure 3 in the blue circles. Scatter plots are shown for each parameter, with its initial value on the x -axis and its final value on the y -axis. Clearly, the ideal outcome is that the same final value is reached regardless of the initial parameter value, but this is not always the case. The results indicate that the algorithm is able to determine T_b and α . However, the value of h is strongly dependent on its initial value. The parameters β , λ and C have a substantial spread in their optimal values, in a seemingly uncorrelated manner with their initial value. Hence, the conclusion from the 'naive' parameter estimation is that it is not possible to obtain a reliable estimation of the parameter values by assuming the parameters are independent and using a single observation of the soil temperature profile.

The reason why this happened is because an optimiser converges to a global minimum only when the optimisation problem is convex. This condition will be violated when two parameters are dependent on each other. In particular, the heat transfer coefficient h and the Bowen ratio β occur only through the product $h(1 + \beta^{-1})$ in (1). As a result, the optimisation problem is non-unique. Figure 4(a) shows the final values for h and β plotted against each other, showing a clear relation between the two quantities. Figure 3(d) shows that the product $h(1 + \beta^{-1})$ has much less spread than h and β individually. Thus, the optimisation should use the parameter $h(1 + \beta^{-1})$ instead of h and β separately.

Furthermore, it is well known that thermal inertia is typically characterised using the thermal admittance $\mu = \lambda C$ (Oke, 2017), which characterises the ability of a material to absorb and release heat. This suggests that there might also be a correlation between C and λ . Figure 4(b) shows the final values for C and λ plotted against each other, once more showing a

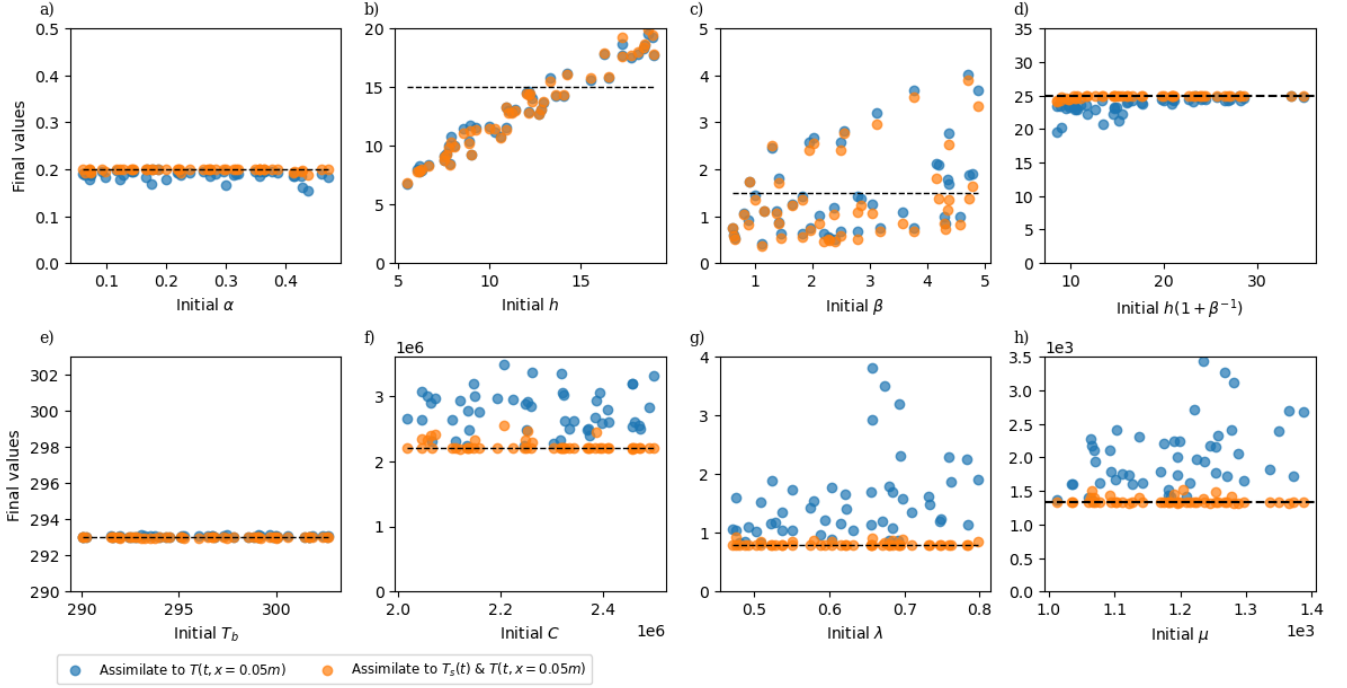


Figure 3. Scatter plots of initial vs final values for the synthetic data set of (a) surface albedo α , (b) heat transfer coefficient h , (c) Bowen ratio β , (d) $h(1 + \beta^{-1})$, (e) Initial/interior temperature T_b , (f) volumetric heat capacity C , g) thermal conductivity λ and h) μ compared with their true values (black dotted line). The blue circles assimilate data from depth of 0.05 m and the orange circles assimilate data from two heights, 0.05 m and 0 m (i.e., the surface).

clear relation between the two quantities. However, Figure 3(h) shows that even though C and λ are clearly correlated, there is considerable uncertainty in μ , and it is not possible to replace C and λ for μ .

To conclude this section, we have learnt that a ‘naive’ parameter estimation approach does not result in reliable parameter values. Firstly, the parameters can be correlated to each other, and we found that it is not possible to determine h and β individually. However, even though it is shown that C and λ are strongly correlated, the quantity μ does not capture this behaviour and we thus conclude that it is impossible to obtain reliable parameter values using temperature observations at a single soil depth.

3.2 Observations at two depths

In this section, parameter estimation using two sets of observations of temperatures is explored, once more using the synthetic data. In addition to $T(t; x=0.05 \text{ m})$, the surface temperature T_s will be used, since this is where the temperature amplitudes are largest. Instead of changing the parameter estimation strategy and using $h(1 + \beta^{-1})$ as a model parameter, we will keep using

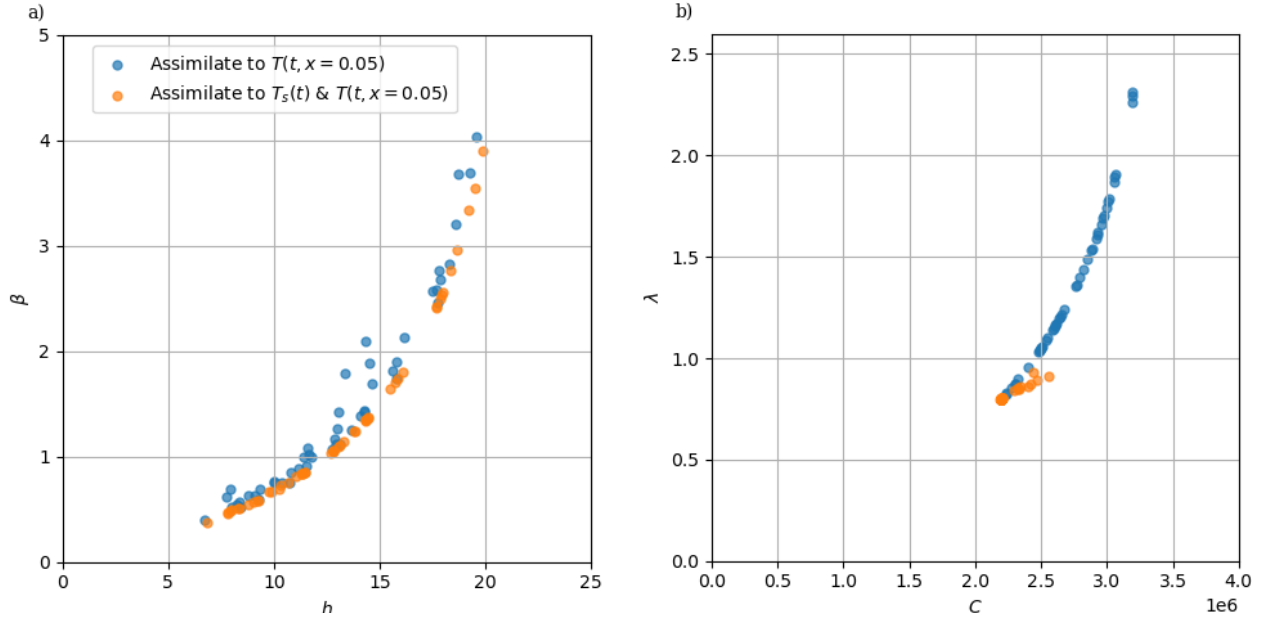


Figure 4. Covariance plots of final parameter values for the synthetic dataset of (a) β against h ; and (b) λ against C

h and β , but with the understanding that we should only look whether the quantity $h(1 + \beta^{-1})$ has a value that is independent of its initial value.

270 Figure 3 shows that by using observations at two depths, it becomes possible to reliably estimate parameter values. The parameter values obtained from the inverse model are $\alpha = 0.199 \pm 0.010$, $T_b = 292.99 \pm 0.05$ K, $\lambda = 0.81 \pm 0.11$ W m⁻¹ K⁻¹, $C = 2.24 \pm 0.32$ MJ m⁻³ K⁻¹ and $h(1 + \beta^{-1}) = 24.90 \pm 0.78$ W m⁻² K⁻¹. Obtained from 50 trials, the mean values of the parameters are within 2% of their real value (see Table 1). Note that C and λ are the only quantities that have outliers up to 15% larger than the mean value.

275 3.3 Application to Phoenix dataset

In order to apply the inverse model to real data, we use data from a flux tower managed by the CAP LTER programme. The tower is located in a residential suburban of Maryvale, in the city of Phoenix, United States. The forcing data (incoming shortwave radiation $K^\downarrow$, re-emitted longwave radiation $L^\downarrow$ and atmospheric temperature T_a) were collected over the calendar year of 2012 (Chow et al., 2014; Chow, 2017). The data was then bias-corrected and gap-filled by the Urban-PLUMBER project (Lipson et al., 2022b, 2024). Phoenix is known for its arid, desert-like climate. This minimises the effect of the moisture content of soil on heat flux exchange, which is not included in the equations of the current model. We extracted 200 hours of flux data from 1 May 2012 0:00 am to 9 May 2012 8:00 am. The first 150 hours of soil temperature measurements are used for data assimilation to estimate the parameters, whilst the remaining 50 hours of data are used for evaluation of the model and its parameters. A longer time series is used than the synthetic data for assimilation to account for the error in assuming a uniform

Table 2. Estimated parameters for the West Phoenix flux tower data set for the period of 1 May 00:00 am – 9 May 08:00 am from assimilating soil temperatures at 0.05 m and 0.15 m.

Parameter	Units	Mean	Standard deviation
C	$\text{MJ m}^{-3} \text{K}^{-1}$	2.44	0.08
λ	$\text{W m}^{-1} \text{K}^{-1}$	0.826	0.043
$h(1 + \beta^{-1})$	$\text{W m}^{-2} \text{K}^{-1}$	19.66	0.11
T_b	K	305.41	0.05

285 initial temperature profile, which is unlikely to occur in real life. For detailed information about the measurement instruments in the West Phoenix flux tower, see Appendix A.

For this dataset, we aim to estimate four parameters, namely T_b , $h(1 + \beta^{-1})$, λ and C . The value of α will not be determined using optimisation, as we found that the parameter values could not be determined reliably without specifying α , presumably since α controls the absorbed shortwave radiation which is the primary driver of the system. However, this value can be inferred
290 directly from measured incoming and outgoing short-wave radiations. In this case, it is taken to be $\alpha = 0.172$, an average mid-day value estimated for the site by Lipson et al. (2022b, 2024). Once more, a total of 50 trials are conducted using initial parameter values that are randomly generated using the procedure discussed in Section 3.1. Table 2 shows the final parameter values from assimilating to temperature measurements at 5 cm and 15 cm, which are taken to be the mean of the optimised values from all 50 trials.

295 Comparing results of inverse analysis using real data to that of synthetic data when assimilated to measurements at two depths (Figure 3), λ has a larger distribution in its optimised values, indicating higher uncertainties. The estimated thermal conductivity λ and volumetric heat capacity C are within the range of typical values for all three types of soils. The estimated boundary temperature T_b is also reasonable for an arid climate during summer (Table 1).

In order to compare the modelled fluxes with the observations, the forward model is run using the mean of the obtained
300 parameter values. Figure 5(a) shows good agreement of the model output for the two temperature time series with the final 50 hours of observations, which are not used for the optimisation. More interesting are comparisons with other observed quantities, such as the outgoing longwave radiation $L^\uparrow$, the combined sensible-latent heat flux $Q_E + Q_H$ and conductive soil fluxes in the soil. Figure 5(b) shows the conductive soil flux measurements in gravel and sandy soil as observed together with the model outputs. For the model, the flux $G = -\lambda \partial T / \partial x$ is estimated using central differences. The modelled ground heat flux lies in
305 between the ground heat flux measured in gravel and sandy soil, as it resembles the flux recorded in sandy soil more in the day, and flux recorded in gravel more at night. Figure 5(c) compares the observed upwelling longwave radiation $L^\uparrow$ with the model output, once more showing good agreement between the two. It is not self-evident that this should be the case, since the long-wave radiation picks up information from buildings as well as the soil, and the surface temperature has a much larger amplitude than the temperatures deeper in the soil, thus allowing for extrapolation errors. Finally the combined sensible-latent
310 heat fluxes $Q_H + Q_E$ are calculated and compared with flux tower measurements in Figure 5(d). Here, the agreement is not

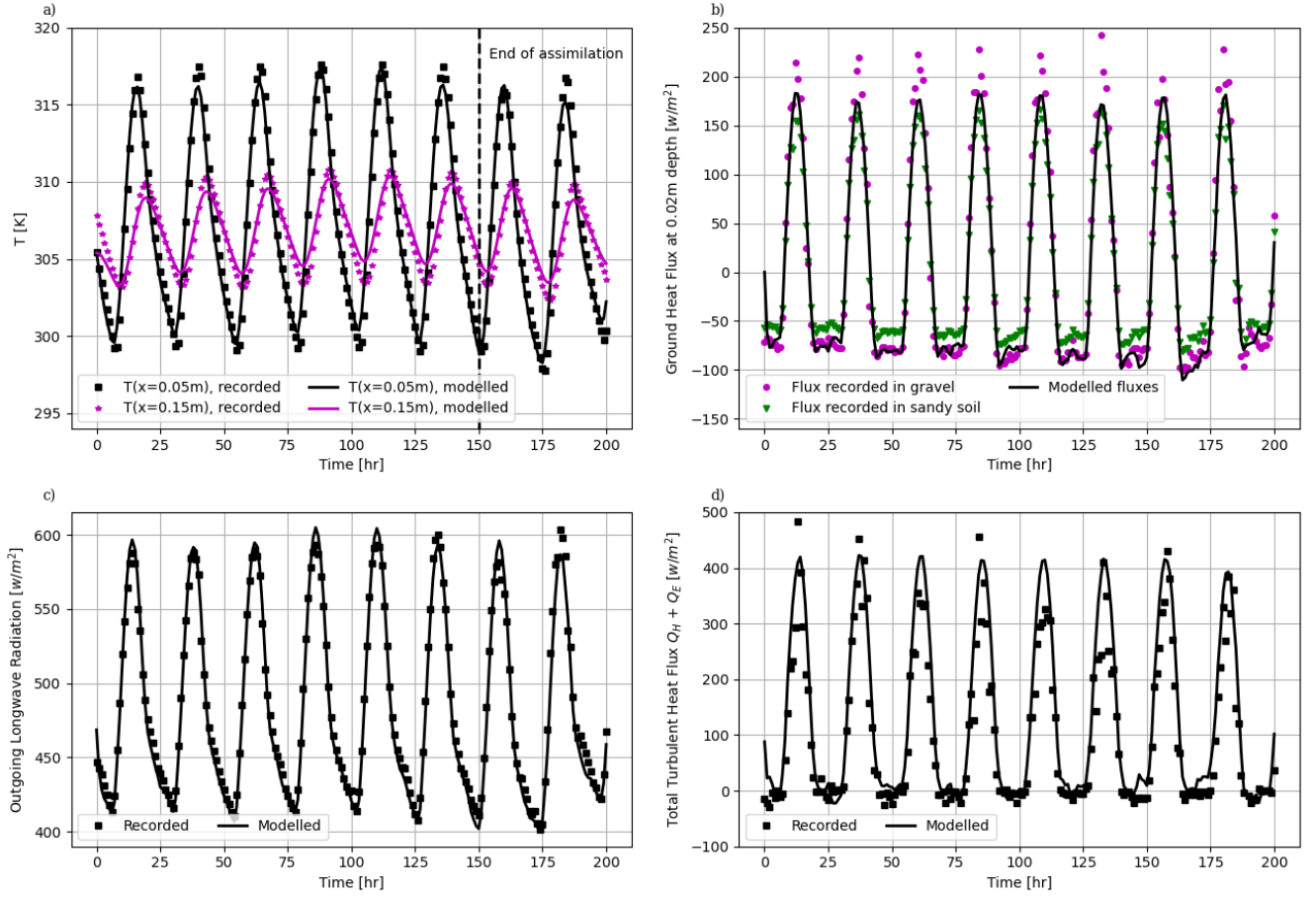


Figure 5. Comparison of modelled and recorded data for the West Phoenix flux tower dataset. a) Temperatures at 0.05 m and 0.15 m depth. b) Ground heat flux at 0.02 m depth. c) Outgoing long-wave radiation. d) Combined sensible-latent heat flux

as good as for the other surface energy balance components. We note that the differences between the observations and the model for the turbulent fluxes could be due to the difference in scale of the forcing data and target temperature measurements. Turbulent fluxes are measured at a height of 22.1 m AGL and represent an averaged value across the source area, which is highly variable as it changes with wind direction and atmospheric stability (Chow et al., 2014; Schmid et al., 1991), whilst soil
 315 temperature measurements are localised (taken only at one location beneath the flux tower).

4 Conclusions

We developed a land-surface model that enables end-to-end parameter estimation using gradient-based optimisation through Neural Physics, an approach which takes advantage of the tools from machine-learning libraries to provide differentiability

whilst remaining entirely physics based. The approach also offers insights into parameter identifiability and confounding, and
320 we provide a concrete, quantitative characterisation of equifinality for this application, identifying specific pathways through which it can be mitigated.

Results from a ‘naive’ approach that assumes parameter independence show that temperature observations at a single depth are not sufficient to obtain reliable parameter values. In fact, the problem is shown to be ill-posed as multiple parameter combinations (e.g., h and β) collapse onto an identifiable product $h(1 + \beta^{-1})$ and soil parameters (λ , C) exhibit strong correlations
325 with a large spread in their individual estimates. We demonstrate that this non-identifiability is not intrinsic to the model alone, but depends on the observational configuration, as adding temperature observations at a second depth resolves this degeneracy and enables reliable recovery of thermal conductivity and heat capacity, independent of initial conditions. We also show that some confounding cannot be resolved without additional physics or data, for example, the sensible and latent heat fluxes remain inseparable without direct flux measurements. The model is then applied to observations of a 150-hour period of the
330 West Phoenix flux tower dataset and evaluated on a subsequent 50-hour period of data. The estimated parameter values for the thermal conductivity, volumetric heat capacity and the combined sensible-latent heat transfer coefficient are physically plausible, and the agreement of the predictions for outgoing longwave radiation, conductive soil fluxes and combined sensible-latent heat fluxes with observations in the evaluation period is good.

In terms of parameter estimation, we have therefore learnt the following:

- 335 1. Data at two depths is required in order to obtain reliable estimates for thermal conductivity and heat capacity.
2. It is not possible to partition sensible and latent heat fluxes without specific measurements of either Q_H or Q_E .
3. It is necessary to specify the effective albedo of the site, as without it parameter values did not become independent of their initial estimate. We surmise that this is because the effective albedo is the primary forcing during day-time and all fluxes will be affected by this. Since this is one of the simplest quantities to measure, this is not considered a major
340 limitation.
4. We recommend performing optimisation with at least 50 random initial parameter estimates to ensure the robustness of the obtained parameter values and to take the mean value over all the trials. However, our results suggest that it is possible to use a lower number.

~~We note several limitations of the present study~~Several challenges remain for future work. First, the LSM used to demon-
345 strate the proposed approach is ~~relatively simple. As no moisture model is included for the subsurface, we focus the analysis intentionally simple. The absence of a subsurface moisture model allows a detailed investigation of parameter identifiability and equifinality, but also motivates future work aimed at extending the framework to more realistic settings. In the present study, we therefore focus~~ on a period from the West Phoenix flux tower dataset during which ~~the latent heat flux is~~ latent heat fluxes are negligible (i.e., no rainfall). ~~Extending this framework to more realistic settings~~Applying the approach to operational
350 LSMs will require methods capable of estimating the larger number of parameters ~~present in operational LSMs~~associated with more complex process representations. In addition, the inverse model is applied over a limited time period, avoiding the need

to assume slowly varying parameter values. This assumption becomes particularly important when considering soil moisture, which is inherently time-dependent and typically requires multiple state variables for accurate representation (Grimmond et al., 2010). For longer time series, more sophisticated approaches may be required, potentially including the incorporation of a learned component. Although it is feasible to augment the current model with such methods, the parameter estimation results indicate that the present physics-based formulation is already capable of reproducing observations with reasonable accuracy.

~~Second, there are practical barriers to adoption-~~

The second challenge concerns practical adoption of the framework. LSMs are typically implemented in legacy languages such as Fortran, whereas machine-learning frameworks are commonly developed in Python, JAX or Julia. ~~Applying~~ Therefore, applying the Neural Physics approach ~~therefore, or, indeed, adapting code to any framework with automatic differentiability,~~ requires some degree of model refactoring, ~~although-~~ Although challenging, this process is becoming increasingly tractable with modern tools such as Github Copilot, Claude Code and ChatGPT (Zhou et al., 2024).

~~Despite these limitations~~ A third challenge is to deal with the non-smooth parameterisations arising from more complex LSMs. Many operational land-surface models include discontinuities arising from lookup tables, threshold-based parameterisations and conditional logic, which can complicate gradient-based optimisation in practice. Although modern automatic differentiation frameworks can handle piecewise-smooth operations, extending our approach to discontinuous or hybrid formulations is an important direction for future work and an active area of research.

Despite these challenges, the proposed framework offers a potential advantage when scaling to more complex systems. Dukes et al. (2026) identify differentiable hybrid physics-machine-learning models as a new and promising direction for parameter estimation in Earth System Models, but highlight the substantial computational burden associated with increasing model complexity. In particular, stability constraints often necessitate smaller time steps, leading to higher iteration counts and increased memory demands in inverse problems. In this context, the Neural Physics formulation provides a natural way of incorporating fully differentiable multigrid solvers through U-Net architectures (Chen et al., 2026, 2024; Phillips et al., 2023). Such approaches are well suited to reducing iteration counts, thereby mitigating a key computational bottleneck. Whilst not required for the simple case considered here, this capability represents a potential advantage of the proposed approach when applied to more complex LSMs.

Appendix A: West Phoenix flux tower data

Information about the West Phoenix flux data measurements is provided in Table A1.

Code availability. The code that supports the findings of this study is available on the Zenodo server under <https://doi.org/10.5281/zenodo.19344692> (Huang, 2026)

Table A1. Summary of West Phoenix flux data

Variable	Description	Units	Instrument	Location	Bias correction	Gap filled
Forcing data¹ (Lipson et al., 2022a, b, 2024)						
$K^\downarrow$	Downward shortwave radiation	W m^{-2}	Net radiometer/NR01	22.1 m AGL	none	1.63 % gap-filled
$L^\downarrow$	Downward longwave radiation	W m^{-2}	Net radiometer/NR01	22.1 m AGL	hourly and daily	1.32 % gap-filled
T_a	Air Temperature	K	Temperature–relative humidity sensor/HMP45AC	22.1 m AGL	hourly and daily	2.95 % gap-filled
Outgoing Radiation & Soil temperature measurements (Chow et al., 2014; Chow, 2017)						
$K^\uparrow$	Upward shortwave radiation	W m^{-2}	Net radiometer/NR01	22.1 m AGL	none	no
$L^\uparrow$	Upward longwave radiation	W m^{-2}	Net radiometer/NR01	22.1 m AGL	none	no
Q_E	Latent heat flux	W m^{-2}	Eddy Covariance	22.1 m AGL	none	no
Q_H	Sensible heat flux	W m^{-2}	Eddy Covariance	22.1 m AGL	none	no
—	Soil temperature ²	K	Soil averaging thermocouple/TCAV	0.05 m & 0.15 m BGL	none	no

¹ Urban-PLUMBER gap-filled the forcing data using a combination of three methods: contemporaneous and nearby flux tower data where available, linear interpolation for small gaps ($\leq 2h$), and bias-corrected ERA5 reanalysis meteorological data for larger gaps. ERA5 combines satellite, atmospheric and ground-based observations to produce globally consistent forcing data at 0.25° spatial and hourly temporal resolutions;

² Soil temperature measurements are only available from March 2012

Author contributions. All authors contributed to the development of methods; RH developed the software, implemented the test cases, analysed the results and prepared the visualisations; RH wrote the paper with contributions from CEH and MvR; CEH and MvR conceptualised and supervised the project.

Competing interests. The contact author has declared that none of the authors has any competing interests.

385 *Acknowledgements.* The authors would like to thank Prof. Sue Grimmond for the insightful conversations about atmospheric temperature readings. CEH gratefully acknowledges support from Imperial-X’s Eric and Wendy Schmidt Centre for AI in Science and funding from the

EPSRC AI-Respire project (grant number EP/Y018680/1). MvR acknowledges funding from the NERC ASSURE project (grant number NE/W002868/1). The authors would also like to thank the reviewers, whose valuable suggestions have helped improve the manuscript.

References

- 390 Arsenault, K. R., Nearing, G. S., Wang, S., Yatheendradas, S., and Peters-Lidard, C. D.: Parameter Sensitivity of the Noah-MP Land Surface Model with Dynamic Vegetation, *Journal of Hydrometeorology*, 19, 815–830, <https://doi.org/10.1175/jhm-d-17-0205.1>, 2018.
- Bezgin, D. A., Buhendwa, A. B., and Adams, N. A.: JAX-Fluids: A fully-differentiable high-order computational fluid dynamics solver for compressible two-phase flows, *Computer Physics Communications*, 282, 108 527, <https://doi.org/10.1016/j.cpc.2022.108527>, 2023.
- Bradle, B.: A friendly introduction to numerical analysis, Pearson Education, 2007.
- 395 Brutsaert, W.: *Hydrology: An Introduction*, Cambridge University Press, 2005.
- Carslaw, H. S. and Jaeger, J. C.: *Conduction of Heat in Solids*, p. 81, Oxford University Press, USA, 1959.
- Cerebras: CS-3 System: Revolutionary AI Infrastructure, <https://www.cerebras.ai/system>, accessed: 03-03-2026, 2026.
- Chaney, N. W., Herman, J. D., Ek, M. B., and Wood, E. F.: Deriving global parameter estimates for the Noah land surface model using FLUXNET and machine learning, *Journal of Geophysical Research: Atmospheres*, 121, 13,218–13,235, <https://doi.org/10.1002/2016JD024821>, 2016.
- 400 Chen, B., Heaney, C. E., Gomes, J. L. M. A., Matar, O. K., and Pain, C. C.: Solving the discretised multiphase flow equations with interface capturing on structured grids using machine learning libraries, *Computer Methods in Applied Mechanics and Engineering*, 426, 116 974, <https://doi.org/10.1016/j.cma.2024.116974>, 2024.
- Chen, B., Nadimy, A., Heaney, C. E., Sharifian, M. K., Estrem, L. V., Nicotina, L., Hilberts, A., and Pain, C.: Solving the Discretised Shallow Water Equations Using Neural Networks, *Advances in Water Resources*, 197, 104 903, <http://dx.doi.org/10.2139/ssrn.4956116>, 2025.
- 405 Chen, B., Heaney, C. E., and Pain, C. C.: Neural Physics: Using AI Libraries to Develop Physics-Based Solvers for Incompressible Computational Fluid Dynamics, *Computers & Fluids*, 308, 106 981, <https://doi.org/j.compfluid.2026.106981>, 2026.
- Chow, W.: Eddy covariance data measured at the CAP LTER flux tower located in the west Phoenix, AZ neighborhood of Maryvale from 2011-12-16 through 2012-12-31 ver 1, Environmental Data Initiative., <https://doi.org/10.6073/pasta/fed17d67583eda16c439216ca40b0669>, 2017.
- 410 Chow, W. T. L., Volo, T. J., Vivoni, E. R., Jenerette, G. D., and Ruddell, B. L.: Seasonal dynamics of a suburban energy balance in Phoenix, Arizona, *International Journal of Climatology*, 34, 3863–3880, <https://doi.org/10.1002/joc.3947>, 2014.
- Deck, K., Braghiere, R. K., Renchon, A. A., Sloan, J., Bozzola, G., Speer, E., Ben Mackay, J., Reddy, T., Phan, K., Gagné-Landmann, A. L., Li, Y., Yatunin, D., Charbonneau, A., Efrat-Henrici, N., Bach, E., Ma, S., Gentine, P., Frankenberg, C., Bloom, A. A., Wang, Y., Longo, M., and Schneider, T.: ClimaLand: A Land Surface Model Designed to Enable Data-Driven Parameterizations, *Journal of Advances in Modeling Earth Systems*, 18, e2025MS005 118, <https://doi.org/10.1029/2025MS005118>, 2026.
- 415 Dukes, J. S., Xu, C., Liao, C., Novick, K. A., Phillips, R. P., Beverly, D. P., Fang, Y., Jacobs, E. M., McAdam, S. A. M., Paudel, I., Rimer, I. M., and Robbins, Z. J.: Improving the representation of plant water stress and water use in Earth System Models, *New Phytologist*, 249, 39–55, <https://doi.org/10.1111/nph.70687>, 2026.
- 420 ElGhawi, R., Kraft, B., Reimers, C., Reichstein, M., Körner, M., Gentine, P., and Winkler, A. J.: Hybrid modeling of evapotranspiration: inferring stomatal and aerodynamic resistances using combined physics-based and machine learning, *Environmental Research Letters*, 18, 034 039, <https://doi.org/10.1088/1748-9326/acbbe0>, 2023.
- Fang, J. and Gentine, P.: Exploring Optimal Complexity for Water Stress Representation in Terrestrial Carbon Models: A Hybrid-Machine Learning Model Approach, *Journal of Advances in Modeling Earth Systems*, 16, e2024MS004 308, <https://doi.org/10.1029/2024MS004308>, 2024.
- 425

- Fer, I., Kelly, R., Moorcroft, P. R., Richardson, A. D., Cowdery, E. M., and Dietze, M. C.: Linking big models to big data: efficient ecosystem model calibration through Bayesian model emulation, *Biogeosciences*, 15, 5801–5830, <https://doi.org/10.5194/bg-15-5801-2018>, 2018.
- Fletcher, R.: *Practical Methods of Optimization*, Wiley & Sons, second edn., <https://doi.org/10.1002/9781118723203>, 1987.
- Graphcore: Designed for AI: Intelligence Processing Unit, <https://www.graphcore.ai/products/ipu>, accessed: 03-03-2026, 2026.
- 430 Greenbaum, A.: *Iterative Methods for Solving Linear Systems*, Society for Industrial and Applied Mathematics, <https://doi.org/10.1137/1.9781611970937>, 1997.
- Grimmond, C. S. B., Blackett, M., Best, M. J., Barlow, J., Baik, J.-J., Belcher, S. E., Bohnenstengel, S. I., Calmet, I., Chen, F., Dandou, A., Fortuniak, K., Gouvea, M. L., Hamdi, R., Hendry, M., Kawai, T., Kawamoto, Y., Kondo, H., Krayenhoff, E. S., Lee, S.-H., Loridan, T., Martilli, A., Masson, V., Miao, S., Oleson, K., Pigeon, G., Porson, A., Ryu, Y.-H., Salamanca, F., Shashua-Bar, L., Steeneveld, G.-J.,
- 435 Tombrou, M., Voogt, J., Young, D., and Zhang, N.: The International Urban Energy Balance Models Comparison Project: First Results from Phase 1, *Journal of Applied Meteorology and Climatology*, 49, 1268–1292, <https://doi.org/10.1175/2010JAMC2354.1>, 2010.
- Grimmond, C. S. B., Blackett, M., Best, M. J., Baik, J.-J., Belcher, S. E., Beringer, J., Bohnenstengel, S. I., Calmet, I., Chen, F., Coutts, A., Dandou, A., Fortuniak, K., Gouvea, M. L., Hamdi, R., Hendry, M., Kanda, M., Kawai, T., Kawamoto, Y., Kondo, H., Krayenhoff, E. S., Lee, S.-H., Loridan, T., Martilli, A., Masson, V., Miao, S., Oleson, K., Ooka, R., Pigeon, G., Porson, A., Ryu, Y.-H., Salamanca, F., Steeneveld, G., Tombrou, M., Voogt, J. A., Young, D. T., and Zhang, N.: Initial results from Phase 2 of the international urban energy balance model comparison, *International Journal of Climatology*, 31, 244–272, <https://doi.org/10.1002/joc.2227>, 2011.
- 440 Hastings, W. K.: *Monte Carlo Sampling Methods Using Markov Chains and Their Applications*, *Biometrika*, 57, 97–109, <http://www.jstor.org/stable/2334940>, 1970.
- Huang, R.: *Inverse Modelling of the Surface Energy Balance using Machine Learning Libraries*, Master’s thesis, Imperial College London,
- 445 2024.
- Huang, R.: *Solving land surface model using Neural Physics*, <https://doi.org/10.5281/zenodo.19344692>, 2026.
- Jiang, P., Kidger, P., Bandai, T., Baldocchi, D., Liu, H., Xiao, Y., Zhang, Q., Wang, C. T., Steefel, C., and Chen, X.: JAX-CanVeg: A Differentiable Land Surface Model, *Water Resources Research*, 61, e2024WR038116, <https://doi.org/10.1029/2024WR038116>, 2025.
- Jones, S., Mercado, L., Bruhn, D., Raoult, N., and Cox, P.: Night-time decline in plant respiration is consistent with substrate depletion, *Communications Earth & Environment*, 5, 148, <https://doi.org/10.1038/s43247-024-01312-y>, 2024.
- 450 Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, in: 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings, edited by Bengio, Y. and LeCun, Y., <https://doi.org/10.48550/arXiv.1412.6980>, 2015.
- Krayenhoff, E. S. and Voogt, J. A.: A microscale three-dimensional urban energy balance model for studying surface temperatures, *Boundary-Layer Meteorology*, 123, 433–461, 2007.
- 455 Kuppel, S., Peylin, P., Maignan, F., Chevallier, F., Kiely, G., Montagnani, L., and Cescatti, A.: Model-data fusion across ecosystems: from multisite optimizations to global simulations, *Geoscientific Model Development*, 7, 2581–2597, <https://doi.org/10.5194/gmd-7-2581-2014>, 2014.
- Laloui, L. and Rotta Loria, A. F.: Chapter 3 - Heat and mass transfers in the context of energy geostructures, in: *Analysis and Design of Energy Geostructures*, edited by Laloui, L. and Rotta Loria, A. F., pp. 69–135, Academic Press, ISBN 978-0-12-820623-2, <https://doi.org/10.1016/B978-0-12-816223-1.00003-5>, 2020.
- 460 Lee, D.-I. and Lee, S.-H.: The Microscale Urban Surface Energy (MUSE) Model for Real Urban Application, *Atmosphere*, 11, <https://doi.org/10.3390/atmos11121347>, 2020.

- Li, Y. et al.: An AI-based Integrated Framework for Anisotropic Electrical Resistivity Imaging, in preparation, 2024.
- 465 Lipson, M., Grimmond, S., Best, M., Chow, W., Christen, A., Chrysoulakis, N., Coutts, A., Crawford, B., Earl, S., Evans, J., Fortuniak, K., Heusinkveld, B. G., Hong, J.-W., Hong, J., Järvi, L., Jo, S., Kim, Y.-H., Kotthaus, S., Lee, K., Masson, V., McFadden, J. P., Michels, O., Pawlak, W., Roth, M., Sugawara, H., Tapper, N., Velasco, E., and Ward, H. C.: Data for "Harmonized gap-filled dataset from 20 urban flux tower sites" for the Urban-PLUMBER project, <https://doi.org/10.5281/zenodo.7104984>, 2022a.
- Lipson, M., Grimmond, S., Best, M., Chow, W. T. L., Christen, A., Chrysoulakis, N., Coutts, A., Crawford, B., Earl, S., Evans, J., Fortuniak, K., Heusinkveld, B. G., Hong, J.-W., Hong, J., Järvi, L., Jo, S., Kim, Y.-H., Kotthaus, S., Lee, K., Masson, V., McFadden, J. P., Michels, O., Pawlak, W., Roth, M., Sugawara, H., Tapper, N., Velasco, E., and Ward, H. C.: Harmonized gap-filled datasets from 20 urban flux tower sites, *Earth System Science Data*, 14, 5157–5178, <https://doi.org/10.5194/essd-14-5157-2022>, 2022b.
- Lipson, M. J., Hart, M. A., and Thatcher, M.: Efficiently modelling urban heat storage: an interface conduction scheme in an urban land surface model (aTEB v2.0), *Geoscientific Model Development*, 10, 991–1007, <https://doi.org/10.5194/gmd-10-991-2017>, 2017.
- 475 Lipson, M. J., Grimmond, S., Best, M., Abramowitz, G., Coutts, A., Tapper, N., Baik, J.-J., Beyers, M., Blunn, L., Boussetta, S., Bou-Zeid, E., De Kauwe, M. G., de Munck, C., Demuzere, M., Fatichi, S., Fortuniak, K., Han, B.-S., Hendry, M. A., Kikegawa, Y., Kondo, H., Lee, D.-I., Lee, S.-H., Lemonsu, A., Machado, T., Manoli, G., Martilli, A., Masson, V., McNorton, J., Meili, N., Meyer, D., Nice, K. A., Oleson, K. W., Park, S.-B., Roth, M., Schoetter, R., Simón-Moral, A., Steeneveld, G.-J., Sun, T., Takane, Y., Thatcher, M., Tsiringakis, A., Varentsov, M., Wang, C., Wang, Z.-H., and Pitman, A. J.: Evaluation of 30 urban land surface models in the Urban-PLUMBER project: Phase 1 results, *Quarterly Journal of the Royal Meteorological Society*, 150, 126–169, <https://doi.org/10.1002/qj.4589>, 2024.
- 480 Massoud, E. C., Xu, C., Fisher, R. A., Knox, R. G., Walker, A. P., Serbin, S. P., Christoffersen, B. O., Holm, J. A., Kueppers, L. M., Ricciuto, D. M., Wei, L., Johnson, D. J., Chambers, J. Q., Koven, C. D., McDowell, N. G., and Vrugt, J. A.: Identification of key parameters controlling demographically structured vegetation dynamics in a land surface model: CLM4.5(FATES), *Geoscientific Model Development*, 12, 4133–4164, <https://doi.org/10.5194/gmd-12-4133-2019>, 2019.
- 485 Nocedal, J. and Wright, S. J.: *Numerical Optimization*, Springer, second edn., <https://doi.org/10.1007/978-0-387-40065-5>, 2006.
- Norouzi, S., Moldrup, P., Moseley, B., Robinson, D., Or, D., Hohenbrink, T. L., Minasny, B., Sadeghi, M., Arthur, E., Tuller, M., Greve, M. H., and de Jonge, L. W.: A differentiable hybrid modeling approach for learning soil water retention mechanisms from partial knowledge and data, *Journal of Hydrology*, 668, 135 008, <https://doi.org/10.1016/j.jhydrol.2026.135008>, 2026.
- Oke, T. R.: *Boundary layer climates*, Routledge, 2002.
- 490 Oke, T. R.: *Urban climates*, Cambridge University Press, 2017.
- Ouyang, W., Ye, L., Chai, Y., Ma, H., Chu, J., Peng, Y., and Zhang, C.: A differentiable, physics-based hydrological model and its evaluation for data-limited basins, *Journal of Hydrology*, 649, 132 471, <https://doi.org/10.1016/j.jhydrol.2024.132471>, 2025.
- Phillips, T. R. F., Heaney, C. E., Chen, B., Buchan, A. G., and Pain, C. C.: Solving the Discretised Neutron Diffusion Equations Using Neural Networks, *International Journal for Numerical Methods in Engineering*, 124, 4659–4686, <https://doi.org/10.1002/nme.7321>, 2023.
- 495 Phillips, T. R. F., Heaney, C. E., Chen, B., Buchan, A. G., and Pain, C. C.: Solving the Discretised Boltzmann Transport Equations Using Neural Networks: Applications in Neutron Transport, *arXiv preprint*, 2301.09991, <https://doi.org/10.48550/arXiv.2301.09991>, 2024.
- Raoult, N., Douglas, N., MacBean, N., Kolassa, J., Quaife, T., Roberts, A., Rosie, F., Fer, I., Bacour, C., Dagon, K., Hawkins, L., Carvalhais, N., Cooper, E., Dietze, M., Gentine, P., Kaminski, T., Kennedy, D., Liddy, H., Moore, D., and Zobitz, J.: Parameter Estimation in Land Surface Models: Challenges and Opportunities with Data Assimilation and Machine Learning, *ESS Open Archive*, 2024, <https://doi.org/10.22541/essoar.172838640.01153603/v1>, 2024.
- 500

- Raoult, N. M., Jupp, T. E., Cox, P. M., and Luke, C. M.: Land-surface parameter optimisation using data assimilation techniques: the adJULES system V1.0, *Geoscientific Model Development*, 9, 2833–2852, <https://doi.org/10.5194/gmd-9-2833-2016>, 2016.
- Ren, Y., Gou, L., Xiao, M., Liu, Z. L., and Shen, C.: DMFS: differentiable modeling for frozen soil thermodynamic characteristics, *Canadian Geotechnical Journal*, 63, 1–19, <https://doi.org/10.1139/cgj-2025-0364>, 2026.
- 505 Schmid, H., Cleugh, H., Grimmond, S., and Oke, T.: Spatial variability of energy fluxes in suburban terrain, *Boundary-Layer Meteorology*, 54, 249–276, <https://doi.org/10.1007/BF00183956>, 1991.
- Shen, C., Appling, A. P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., Baity-Jesi, M., Fenicia, F., Kifer, D., Li, L., Liu, X., Ren, W., Zheng, Y., Harman, C. J., Clark, M., Farthing, M., Feng, D., Kumar, P., Aboelyazeed, D., Rahmani, F., Song, Y., Beck, H. E., Bindas, T., Dwivedi, D., Fang, K., Höge, M., Rackauckas, C., Mohanty, B., Roy, T., C., X., and Lawson, K.: Differentiable modelling to unify
- 510 machine learning and physical models for geosciences, *Nature Reviews Earth & Environment*, 4, 552–567, <https://doi.org/10.1038/s43017-023-00450-9>, 2023.
- Tian, W., Yu, H., Zhao, S., Cao, Y., Yi, W., Xu, J., and Nan, Z.: NoahPy: a differentiable Noah land surface model for simulating permafrost thermo-hydrology, *Geoscientific Model Development*, 19, 57–72, <https://doi.org/10.5194/gmd-19-57-2026>, 2026.
- Xue, T., Jiao, Y., Ba, T., Wang, J., Yang, J., See, S., Chen, B., Heaney, C. E., Pain, C. C., Kang, C. W., Mohamed, M. A. B., and Li,
- 515 H.: NeuralFVM: Neural-physics-based Finite Volume Method for Turbulent Flows Using the k - ω Model, *arXiv preprint*, 2603.21869, <https://doi.org/10.48550/arXiv.2603.21869>, 2026.
- Zhou, A., Hawkins, L., and Gentine, P.: Proof-of-concept: Using ChatGPT to Translate and Modernize an Earth System Model from Fortran to Python/JAX, *arXiv preprint*, 2405.00018, <https://doi.org/10.48550/arXiv.2405.00018>, 2024.
- Zhu, W., Xu, K., Darve, E., and Beroza, G. C.: A general approach to seismic inversion with automatic differentiation, *Computers &*
- 520 *Geosciences*, 151, 104 751, <https://doi.org/10.1016/j.cageo.2021.104751>, 2021.