ANN-MoC Method for Inverse Transient Transport Problems in One-Dimensional Geometry

ARTICLE HISTORY

Received 5 March 202
Accepted 19 April 2024

Nelson Garcia Roman
Universidade Federal de Rio Grande do Sul (UFRGS)
Porto Alegre, Brazil
ngroman1992@gmail.com
ORCID: 0009-0006-8794-9500

Pedro Costas dos Santos
Universidade Federal de Rio Grande do Sul (UFRGS)
Porto Alegre, Brazil
pedro.costa4137@gmail.com
ORCID: 0009-0001-9927-2860

Pedro Henrique de Almeida Konzen
Universidade Federal de Rio Grande do Sul (UFRGS)
Porto Alegre, Brazil
pedro.konzen@ufrgs.br
ORCID: 0000-0002-0411-1563

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2024

"ANN-MoC Method for Inverse Transient Transport Problems in One-Dimensional Geometry",
Latin-American Journal of Computing (LAJC), vol. 11, no. 2, 2024.
N. Roman, P. dos Santos and P. Konzen,

ANN-MoC Method for Inverse Transient Transport Problems in One-Dimensional Geometry

Nelson Garcia Roman
Instituto de Matemática e Estatística (IME)
Universidade Federal de Rio Grande do Sul (UFRGS)
Porto Alegre, Brazil
ngroman1992@gmail.com
ORCID: 0009-0006-8794-9500

Pedro Costas dos Santos
Instituto de Matemática e Estatística (IME)
Universidade Federal de Rio Grande do Sul (UFRGS)
Porto Alegre, Brazil
pedro.costa4137@gmail.com
ORCID: 0009-0001-9927-2860

Pedro Henrique de Almeida Konzen
Instituto de Matemática e Estatística (IME)
Universidade Federal de Rio Grande do Sul (UFRGS)
Porto Alegre, Brazil
pedro.konzen@ufrgs.br
ORCID: 0000-0002-0411-1563

***Abstract*—**Transport problems of neutral particles have important applications in engineering and medical fields, from safety and quality protocols to optical medical procedures. In this paper, the ANN-MoC approach is proposed to solve the inverse transient transport problem of estimating the absorption coefficient from scalar flux measurements at the boundaries of the model domain. The central idea is to fit an Artificial Neural Network (ANN) using samples generated by direct solutions computed by a Method of Characteristics (MoC) solver. The direct solver validation is performed on a manufactured solution problem. Two inverse problems are then presented for testing the ANN-MoC method. In the first, a homogeneous medium is assumed, and, in the second, the medium is heterogeneous with a piecewise constant absorption coefficient. We show that the method can achieve good estimates, with accuracy depending on that of the direct solver. We also include a test of sensibility by studying the propagation of noise on the input data. The results highlight the potential of the proposed method to be applied to a broader range of inverse transport problems.

Keywords*—*artificial neural network, method of characteristics, particle neutral transport, inverse problem

I. Introduction

Neutral particle transport problems have many important applications in engineering and medical fields. The main fields of radiative heat transfer and neutron transport share the fundamental model based on the linear Boltzmann equation [1], [2]. Applications include engineering at high temperatures, such as glass and ceramic manufactures [3], combustion chambers [4], solar energy production [5], nuclear energy production [6], and optical medicine [7], [8]. Related inverse problem solutions can enhance the development of safety protocols, quality control procedures, and technological innovations.

We consider the time-dependent linear Boltzmann equation with initial and boundary conditions and with isotropic scattering

$$\forall \mu: \frac{1}{c}\frac{\partial}{\partial t} I(t,x,\mu) + \mu \frac{\partial}{\partial x} I(t,x,\mu) + \sigma_t I(t,x,\mu) = \sigma_s \Psi(t,x) + q(t,x,\mu),\ (t,x) \in \left(0, t_f\right] \times D, \tag{1.1}$$

$$\forall \mu: I(0,x,\mu) = I_0(x,\mu),\ x \in D, \tag{1.2}$$

$$\forall \mu > 0: I(t,a,\mu) = I_{in,a}(t,\mu),\ t \in \left(0, t_f\right], \tag{1.3}$$

$$\forall \mu < 0: I(t,b,\mu) = I_{in,b}(t,\mu),\ t \in \left(0, t_f\right], \tag{1.4}$$

$I(t,x,\mu)\ (W/sr)$ denotes the radiation intensity at time $0 \leq t \leq t_f\ (s)$ at point $x \in D = [a,b]\ (m)$, and in the direction $-1 \leq \mu \leq 1, \mu \neq 0$. The average speed of light in the medium is denoted by $c\ (m/s)$. The total absorption coefficient is denoted by $\sigma_t = \kappa + \sigma_s$, while $(1/m)$ and $\sigma_s\ (1/m)$ are, respectively, the absorption and scattering coefficients. The sources are denoted by $q(t,x,\mu)\ (W/(msr))$ in the domain and $I_{in,a}(t,\mu),\ I_{in,b}(t,\mu)\ (W/sr)$ at boundaries. At $t = 0$, initial condition $I = I_0(x,\mu)\ \ (W/sr)$ is assumed. The average scalar flux (W/sr) is defined as

$$\Psi(t,x) := \frac{1}{2}\int_{-1}^{1} I(t,x,\mu) d\mu. \tag{2}$$

Inverse transport problems have been the subject of important research for many decades. The books of [9] and [10] discuss the fundamental methods applied to the solution of inverse problems. Concerning the problems of parameter estimation, the main approaches consist of estimating parameters as solutions to an associated minimization problem. The problem can then be solved by optimization methods, which usually require a good initial approximation of the solution. When this is not known, meta-heuristic algorithms can be applied to this end (see, for instance [11]). Alternatively, Deep Learning [12] techniques are also applied [13], [14]. A well-known approach is to fit an Artificial Neural Network (ANN, [15]) with samples built from solutions to the associated direct problem.

In this context, we introduce the ANN-MoC approach to the inverse transport problem of the absorption coefficient estimation from the scalar flux measured at the boundaries of the model domain. The core concept is to fit an ANN using data derived from direct solutions of Eq. (1) computed by a solver based on the Method of Characteristics (MoC) [16]. The designed methodology is here presented together with

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2024
DOI: 10.5281/zenodo.12191947

selected test cases. After testing the direct solver, two inverse problems are considered. The first is a transport problem in a homogeneous medium. In the second, the medium has two regions with different absorption coefficients.

In the following, the methodology of the MoC direct solver and the ANN model are presented. Numerical experiments with the proposed approach are then presented. They include the selection of ANN architectures, data preprocessing, and model sensibility tests. Conclusions are then presented.

II. The ANN-MOC Method

The ANN-MoC approach consists of solving the inverse transport problem by an Artificial Neural Network (ANN) trained from samples generated by directly solving a set of transport problems by the Method of Characteristics (MoC).

A. MoC direct solver

The MoC direct solver computes an approximation of Eq. (1) built with the Discrete Ordinates Method (DOM) [1] followed by an implicit Euler time discretization [17]. The raised system of ordinary differential equations is decoupled by a Source Iteration (SI, [1]) scheme and then, solved with the Method of Characteristics (MoC, [15]).

Discrete ordinates formulation. The following DOM form of Eq. (1) is obtained by assuming the Gauss-Legendre quadrature $\{(\mu_j, w_j)\}_{j=1}^{n_q}$, with even $n_q > 1$,

$$\forall \mu: \frac{1}{c}\frac{\partial}{\partial t} I_j(t,x) + \mu_j \frac{\partial}{\partial x} I_j(t,x) + \sigma_t I_j(t,x) = \sigma_s \Psi(t,x) + q_j(t,x), \ (t,x) \times D, \tag{3.1}$$

$$\forall \mu: I_j(0,x) = I_{j,0}(x), \ x \in D, \tag{3.2}$$

$$\forall \mu > 0: I_j(t,a) = I_{j,in,a}, \ \forall t \in (0, t_f], \tag{3.3}$$

$$\forall \mu < 0: I_j(t,b) = I_{j,in,b}, \ \forall t \in (0, t_f], \tag{3.4}$$

where the notation $I_j(t,x) \approx I(t, x, \mu_j)$ (analogous to the others) is assumed with $j = 1,2,\dots,n_q$. The scalar flux is approximated by

$$\Psi(t,x) \approx \frac{1}{2}\sum_{j=1}^{n_q} I_j \, w_j. \tag{4}$$

Time discretization. For the time discretization, it is assumed that $t^{(k)} = kh_t, k = 0,1,2,\dots,n_t, h_t = t_f/n_t$ (see Fig. 1). The implicit Euler formulation of Eq. (3) gives an iterative procedure with initialization

$$\forall \mu_j: I_j^{(0)}(x) = I_{j,0}(x), \ x \in D, \tag{5}$$

$j = 1,2,\dots,n_q$, and the following steps

$$\forall \mu_j: \frac{1}{c}\frac{I_j^{(k+1)}(x) - I_j^{(k)}(x)}{h_t} + \mu_j \frac{\partial I_j^{(k+1)}(x)}{\partial x} + \sigma_t I_j^{(k+1)}(x) = \sigma_s \Psi^{(k+1)}(x) + q_j^{(k+1)}(x), \tag{6.1}$$

$$\forall \mu_j > 0: I_j^{(k+1)}(a) = I_{j,in,a}^{(k+1)}, \tag{6.2}$$

$$\forall \mu_j < 0: I_j^{(k+1)}(b) = I_{j,in,b}^{(k+1)}, \tag{6.3}$$

where the notation $I_j^{(k)}(x) \approx I(t^{(k)}, x, \mu_j)$ (analogous to the others) is assumed with $k = 0,1,2,\dots,n_t - 1$ and $j = 1,2,\dots,n_q$. For the sake of simplicity, in the following the index k will be suppressed, with $I_j^{(1)}$ denoting $I_j^{(k+1)}$ and $I_j^{(0)} = I_j^{(k)}$ (analogous to the others).

Source iteration. The decoupling of system Eq. (6) is performed with the Source Iteration (SI) technique. From a given initial scalar flux $\Psi^{(0,0)}(x)$, successive approximations $\Psi^{(1,l)}(x)$are iteratively computed from

$$\forall \mu_j: \frac{1}{c}\frac{I_j^{(1,l+1)}(x) - I_j^{(0)}(x)}{h_t} + \mu_j \frac{\partial I_j^{(1,l+1)}(x)}{\partial x} + \sigma_t I_j^{(1,l+1)}(x) = \sigma_s \Psi^{(1,l)}(x) + q_j^{(1)}(x), \tag{7.1}$$

$$\forall \mu_j > 0: I_j^{(1,l+1)}(a) = I_{j,in,a}^{(1,l+1)}, \tag{7.2}$$

$$\forall \mu_j < 0: I_j^{(1,l+1)}(b) = I_{j,in,b}^{(1,l+1)}, \tag{7.3}$$

where

$$\Psi^{(1,l)}(x) := \frac{1}{2}\sum_{j=1}^{n_q} I_j^{(1,l)}(x) \, w_j, \tag{8}$$

for $j = 1,2,\dots,n_q$, and $l = 0, 1, 2, \dots$ until some given stop criteria are fulfilled.

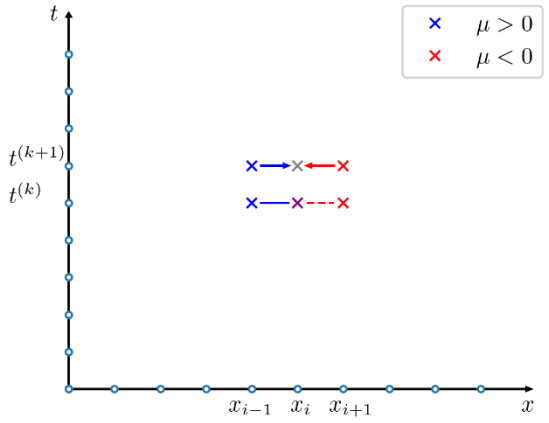

Fig. 1. Scheme of the space-time discretization. Points (×) and intervals (lines and sets) for directions μ > 0 (blue) and μ < 0 (red)

Method of characteristics. At each time step and each source iteration, Eq. (7) is solved by the Method of Characteristics (MoC). First, it is observed that Eq. (7a) can be rewritten as

$$\mu_j \frac{\partial I_j^{(1,l+1)}(x)}{\partial x} + \left(\sigma_t + \frac{1}{ch_t}\right) I_j^{(1,l+1)}(x) = \sigma_s \Psi^{(1,l)}(x) + q_j^{(1)}(x) + \frac{1}{ch_t} I_j^{(0)}(x), \tag{9}$$

$j = 1,2,\dots,n_q$ and $l = 0, 1, 2, \dots$. Again, for the sake of simplicity, the index j is suppressed in the following.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2024
DOI: 10.5281/zenodo.12191947

The MoC form of Eq. (9) is obtained by assuming $x(s) = x_0 + s\mu, s \in \mathbb{R}$, from where Eq. (9) is rewritten as

$$\frac{d}{ds}I^{(1,l+1)}(s) + \left(\sigma_t + \frac{1}{ch_t}\right)I^{(1,l+1)}(s) = \sigma_s \Psi^{(1,l)}(s) + q^{(1)}(s) + \frac{1}{ch_t}I^{(0)}(s), \quad (10)$$

$l = 0, 1, 2, \ldots$. This linear first-order differential equation can now be solved using an integration factor, which gives the solution from

$$I^{(1,l+1)}(s) = I^{(1,l+1)}(0)e^{-\int_0^s \tilde{\sigma}_t ds'} + \int_0^s S^{(l)}(s')e^{-\int_{s'}^s \tilde{\sigma}_t ds''} ds', \quad (11)$$

where $\tilde{\sigma}_t := \sigma_t + \frac{1}{ch_t}$ and

$$S^{(l)}(s) := \sigma_s \Psi^{(1,l)}(s) + q^{(1)}(s) + \frac{1}{ch_t}I^{(0)}(s), \quad (12)$$

$l = 0, 1, 2, \ldots$.

One observes that choosing $x_0 = a$, Eq. (11) gives the particle intensity $I^{(1,l+1)}\big(x(s)\big)$ at each domain $x(s)$ for a given direction $\mu > 0$. Analogously, by choosing $x_0 = b$, one obtains the particle intensity point for a given direction $\mu < 0$.

Direct solver algorithm. Assuming a spatial mesh with n_x nodes $x_i = a + ih_x$, and mesh size $h_x = (b-a)/n_x, i = 0,1,2,\ldots,n_x$, see Fig. 1, the direct solver algorithm can be summarized as follows:

1. Set time, mesh and quadrature parameters
2. From initial condition, set

$$I_{i,j}^{(0)} \leftarrow I(0, x_i, \mu_i), \ \forall i, j, \quad (13.1)$$

$$\Psi_i^{(0)} \leftarrow \frac{1}{2}\sum_{j=1}^{n_q} I_{i,j}^{(0)} w_j, \ \forall i. \quad (13.2)$$

3. (Time loop). For $k = 0,1,2,\ldots,n_t$:

a. (SI loop) For $l = 0,1,2,\ldots,n_{s.i.}$:

a.1. For $j = 0,1,2,\ldots,n_q$ and $\mu_j > 0$:

For $i = 0,1,2,\ldots,n_x - 1$:

$$I_{i+1,j}^{(1,l+1)} \leftarrow I_{i,j}^{(1,l+1)}e^{-\int_0^s \tilde{\sigma}_t ds'} + \int_0^s S^{(l)}(s')e^{-\int_{s'}^s \tilde{\sigma}_t ds''} ds'. \quad (14)$$

a.2. For $j = 1,2,\ldots,n_q$ and $\mu_j < 0$:

For $i = n_x, n_x - 1, \ldots, 1$:

$$I_{i-1,j}^{(1,l+1)} \leftarrow I_{i,j}^{(1,l+1)}e^{-\int_0^s \tilde{\sigma}_t ds'} + \int_0^s S^{(l)}(s')e^{-\int_{s'}^s \tilde{\sigma}_t ds''} ds'. \quad (15)$$

a.3. Compute new scalar flux

$$\Psi_i^{(1+l)} \leftarrow \frac{1}{2}\sum_{j=1}^{n_q} I_{i,j}^{(1,l+1)} w_j, \ \forall i. \quad (16)$$

a.4. SI stop criterion

B. ANN inverse model

The inverse problem is solved by fitting a Multilayer Perceptron network (MLP, [14]) from a data set $\left\{\left(\boldsymbol{\psi}_{train}^{(s)}, \boldsymbol{\kappa}_{train}^{(s)}\right)\right\}_{s=1}^{n_{train}}$ generated from computed solutions of the direct problem for several values of the absorption coefficient. The MLP of $n_h + 2$ layers is written as

$$\tilde{\boldsymbol{\kappa}} = \mathcal{N}\left(\boldsymbol{\psi}; \left\{\left(\boldsymbol{f}^{(l)}, \boldsymbol{b}^{(l)}, \boldsymbol{W}^{(l)}\right)\right\}_{l=1}^{n_h+1}\right), \quad (17)$$

where, in the l-th network layer with $n_n^{(l)}$ neuron units, $\left(\boldsymbol{f}^{(l)}, \boldsymbol{b}^{(l)}, \boldsymbol{W}^{(l)}\right)$ denotes the triple of the activation function, the bias $n_n^{(l)}$-vector, and the weights $n_n^{(l)} \times n_n^{(l+1)}$-matrix. By denoting the input $\boldsymbol{y}^{(\boldsymbol{0})} = \boldsymbol{\psi}$ of detector measurements, its forward propagation through the network layers $l = 1, 2, \ldots n_h + 1$ is given by

$$\boldsymbol{y}^{(l)} = \boldsymbol{f}^{(l)}\left(W^{(l)}\boldsymbol{y}^{(l-1)} + \boldsymbol{b}^{(l)}\right), \quad (18)$$

and the output is the estimated absorption coefficient $\tilde{\boldsymbol{\kappa}} = \boldsymbol{y}^{(\boldsymbol{n_h+1})}$ (see Fig. 2).

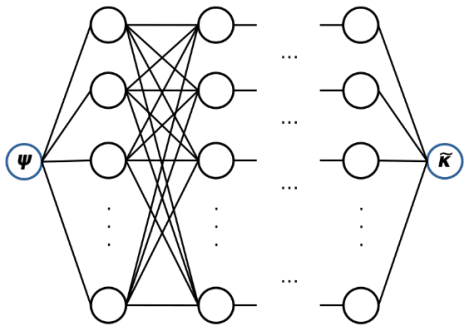

Fig. 2. Architecture of a MLP neural network with n_n neurons on each $n_h$hidden layer

Basic training algorithm. The basic training algorithm can be summarized as follows:

1. Set the MLP architecture.

Sets n_h, n_n, $\boldsymbol{f}^{(l)}$, and initial $b^{(l)}$, $W^{(l)}$ and a global learning rate $l_r > 0$.

2. Loop over epochs $e \leftarrow 1,2,\ldots,n_e$:

2.a. Forward the training set.

$$\tilde{\boldsymbol{\kappa}}_{train} \leftarrow \mathcal{N}(\boldsymbol{\psi}_{train}). \quad (19)$$

2.b. Compute the loss function.

$$\mathcal{L} \leftarrow \frac{1}{n_{train}}\sum_{s=1}^{n_{train}}\left|\tilde{\boldsymbol{\kappa}}_{train}^{(s)} - \boldsymbol{\kappa}_{train}^{(s)}\right|^2. \quad (20)$$

2.c. Backward the loss function to compute the gradients

$$\frac{\partial \mathcal{L}}{\partial W^{(l)}}, \frac{\partial \mathcal{L}}{\partial b^{(l)}}, l = 1, 2, \ \ldots n_l. \quad (21)$$

2.d. Perform an optimizer gradient based step.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2024
DOI: 10.5281/zenodo.12191947

$$(W^{(l)}, \boldsymbol{b}^{(l)}) \leftarrow (W^{(l)}, \boldsymbol{b}^{(l)}) - l_r \frac{\partial \mathcal{L}}{\partial (W^{(l)}, \boldsymbol{b}^{(l)})}, \quad (22)$$

where $l = 1, 2, \ldots n_l$, and l_r is a given learning rate.

The MLPs reported in this paper have been implemented with the help of the machine learning package PyTorch [18], and trained with the Adam method [19]. The learning rate has been set to $l_r = 10^{-2}$.

ANN model test. The test of the trained neural network model consists of verifying its performance for a new data set $\left\{\left(\boldsymbol{\psi}_{test}^{(0)}, \boldsymbol{\kappa}_{test}^{(s)}\right)\right\}_{s=1}^{n_{test}}$ which has not been used for training. The test data set has also been computed by solving the direct problem for several values of the absorption coefficient. The accuracy of the network estimated values $\tilde{\boldsymbol{\kappa}}_{test}^{(s)}$ can be measured by the squared error $\mathcal{L}_{test}$ and the coefficient of determination.

C. Data preprocessing

Data preprocessing for deep learning may reduce generalization errors and reduce the size of the model needed to fit the training set [12]. There are many available techniques [20], and we have chosen to work with the preprocessing now as Standard Scaler. This function transforms the features to have zero mean and unit standard deviation. The general formula for the transformation is:

$$X_{scaled} = \frac{X - mean(X)}{std(X)}, \quad (23)$$

where X is the original value of the feature, $mean(X)$ is the mean and $std(X)$ the standard deviation over the data set X. It ensures that features have comparable scales, which is known to enhance training gradient-based methods.

III. Results

Numerical experiments with the proposed ANN-MoC approach are presented. First, the direct solver validation is presented on a manufactured solution problem. Two inverse problems are then discussed. In the first, a homogeneous medium is assumed, and, in the second, it is considered a two-region heterogeneous medium.

A. Direct solver test

In order to test the direct solver, we have considered the manufactured solution

$$\hat{I}(t, x, \mu) := e^{-\sigma_t |x-t|^2}, \; x \in (0, t_f] \times [0,1]. \quad (24)$$

By substituting Eq. (24) into Eq. (1.1), the source is found to be

$$q(t, x, \mu) = [2\sigma_t(1-\mu)(x-t) + \kappa] e^{-\sigma_t |x-t|^2}, \quad (25)$$

and from the definition of the scalar flux Eq. (2), one has $\hat{\Psi} = \hat{I}$.

After numerical tests, we have chosen the solver parameters $h_t = 0.01, n_x = n_q = 100$, and $tol = 1.49 \times 10^{-8}$ as the absolute L^2-norm tolerance for the SI stopping criterion. Table I shows a comparison between the direct solver approximations and the exact scalar flux solutions at $t_f = 1.0$ for different absorption coefficients. The relative L^2-error is denoted by ε_{rel} and indicates that the chosen parameters were enough for the direct solver to produce an accurate solution with $\varepsilon_{rel} < 10^{-2}$.

TABLE I. COMPARISON BETWEEN THE DIRECT SOLVER APPROXIMATIONS AND THE EXACT SOLUTION AT $t_f = 1.0$

κ	$\Psi(0.0)$	$\Psi(0.5)$	$\Psi(1.0)$	ε_{rel}
0.9	$3.667e-1$	$7.748e-1$	$9.974e-1$	$4.5e-3$
0.5	$3.664e-1$	$7.740e-1$	$9.971e-1$	$5.3e-3$
0.1	$3.660e-1$	$7.730e-1$	$9.968e-1$	$6.4e-3$
Exact	$3.679e-1$	$7.788e-1$	$1.000e+0$	--x--

B. Inverse problem 1 – homogeneous medium

In the inverse problem 1, we assume a homogeneous medium with a constant absorption coefficient. The problem consists of estimating $0.1 < \kappa < 0.9$ from detectors measurements of the scalar fluxes at $x_{d,0} = 0, x_{d,1} = 1$ and at time $t_{d,3} = 3.0$. Boundary conditions are taken as $I(t, 0, \mu) = 1$, for all $\mu > 0$, and $I(t, 1, \mu) = 0$, for all $\mu < 0$. The source is considered null, and the initial condition is $I(0, 0, \mu) = 1, \mu > 0$, and $I(0, x, \mu) = 0$ for all $x > 0$.

The ANN inverse model has the detectors' measurements $d_0 = \Psi(t_{d,3}, 0)$, $d_1 = \Psi(t_{d,3}, 1)$ as inputs and outputs the estimated absorption coefficient $\tilde{\kappa}$. For its training, we have used the direct solver to build a training set $\left\{\left(\boldsymbol{d}_{train}^{(s)}, \kappa_{train}^{(s)}\right)\right\}_{s=1}^{n_{train}}$ of $n_{train} = 17$ samples (patterns) with $\kappa^{(s)} = 0.1 + (s-1)h_s$, $h_s = 0.05$. The test set $\left\{\left(\boldsymbol{d}_{test}^{(s)}, \kappa_{test}^{(s)}\right)\right\}_{s=1}^{n_{test}}$ has been generated with $n_{test} = 32$ with uniformly distributed random choices $0.1 < \kappa^{(s)} < 0.9$ (see Fig. 3).

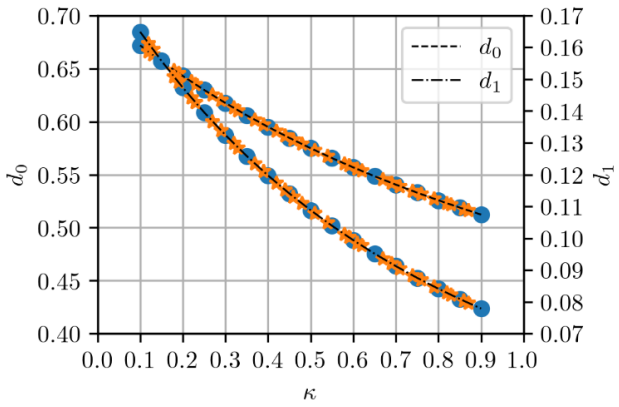

Fig. 3. Inverse problem 1. Training (*circles*) and test (*stars*) samples

We have performed several numerical tests to choose an adequate MLP architecture. Here, we tried architectures $2 - n_n \times n_h - 1$ (2 inputs, n_n neurons on each n_h hidden layer, and 1 output). Training has been stopped when the loss function $\mathcal{L} < 10^{-5}$. Due to the stochasticity of the training method, each test has been repeated three times. Table II presents the results with the hyperbolic tangent ($tanh$) and the identity as activation functions in the hidden and in the output layers, respectively. The demanded averaged total

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2024
DOI: 10.5281/zenodo.12191947

number of epochs n_e and computational time t_c are tabulated. Table III presents results for similar numerical test, but with the $ReLU$ as activation function in hidden layers. We observe that, if MLP, with $tanh$ have demanded last resources to train with small architectures, the $ReLU$ in an $2-30\times4-1$ MLP was even better.

TABLE II. INVERSE PROBLEM 1. TRAINING TESTS FOR MLP ARCHITECTURES WITH TANH AS ACTIVATION FUNCTION

n_h/n_n	10	15	20	25	30
1	3807/ 5,52 s	3281/ 3.66 s	3160/ 3.66s	4711/ 6.55s	9070/ 5.01 s
2	986/ 2.3 s	818/ 1.68 s	920/ 1.27 s	764/ 1.40 s	895/ 1.24 s
3	1294/ 0.95 s	628/ 1.05 s	737/ 2.03 s	885/ 1.01 s	710/ 0.77 s
4	1603/ 1.41 s	1500/ 3.18 s	1041/ 1.94 s	1696/ 2.51 s	634/ 0.76 s

To enhance the training, we have then performed trials with data preprocessing. Inputs of the training samples have been scaled with the Standard Scaler. Setting the $ReLU$ as activation function in hidden layers, several MLP architectures have been tested, and the results can be found in Table IV. The enhancement with preprocessing is notable, with the $2-30\times4-1$ providing the best results.

TABLE III. INVERSE PROBLEM 1. TRAINING TESTS FOR MLP ARCHITECTURES WITH RELU AS ACTIVATION FUNCTION

n_h/n_n	10	15	20	25	30
1	15670/ 21.09 s	11368/ 12.83 s	11884/ 10.9 s	12697/ 15.40 s	16679/ 9.11 s
2	8040/ 4.64 s	14567/ 8.5 s	956/ 0.56 s	2398/ 1.46 s	1577/ 0.99 s
3	2140/ 1.42 s	521/ 0.27 ss	859/ 0.55 s	698/ 0.45 s	226/ 0.15 s
4	1992/ 1.37 s	1015/ 0.83 s	912/ 0.64 s	396/ 0.28 s	196/ 0.24 s

TABLE IV. INVERSE PROBLEM 1. TRAINING TESTS OF MLP ARCHITECTURES WITH DATA PREPROCESSING

n_h/n_n	10	15	20	25	30
1	3280/ 1.48 s	1842/ 0.85 s	374/ 0.18 s	237/ 0.13 s	230/ 0.1 s
2	647/ 0.38 s	238/ 0.2 s	333/ 0.21 s	170/ 0.1 s	93/ 0.06 s
3	553/ 0.34 s	230/ 0.14 s	214/ 0.14 s	145/ 0.1 s	113/ 0.07 s
4	507/ 0.33 s	227/ 0.13 s	212/ 0.14 s	174/ 0.13 s	83/ 0.06 s

Following the previous numerical tests, we have chosen to work with an $2-30\times4-1$ MLP model (two inputs, four hidden layers with 30 neurons each, and one output neuron), the $ReLU$ and the identity as activation functions in the hidden and in the output layers, respectively. With approximately $n_e = 83$, the model reaches a mean squared error $\mathcal{L}_{train} < 10^{-5}$ and coefficient of determination $R^2_{train} = 0.9998$. The application of the trained model to the test data gave results with $\mathcal{L}_{test} < 10^{-5}$ and $R^2_{test} = 0.9998$ (see Fig. 4).

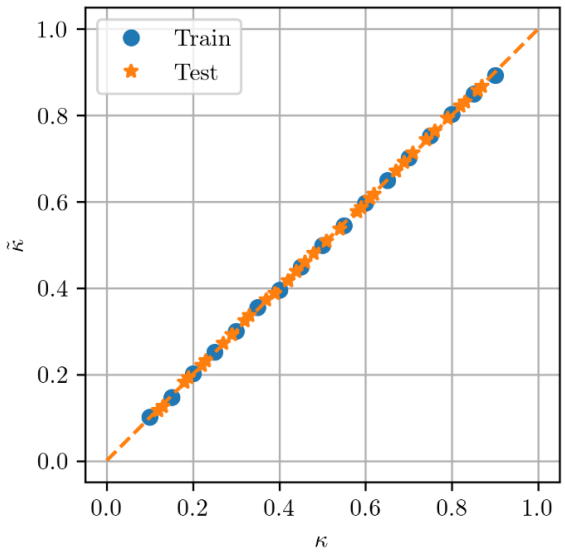

Fig. 4. Inverse problem 1. Expected κ versus estimated $\tilde{\kappa}$. Train: circles. Test: starts. Line fitted to test data results: dashed line

C. *Inverse problem 2 – heterogeneous medium*

In the inverse problem 2, we assume a heterogeneous medium with piecewise constant absorption coefficients

$$\kappa(x) = \begin{cases} \kappa_1 \ , & 0 \le x \le 0.5, \\ \kappa_2 \ , & 0.5 < x \ \le 1. \end{cases} \tag{26}$$

The inverse problem consists of estimating $0.1 \le \kappa_1, \kappa_2 \le 0.9$ from detectors measurements of the scalar fluxes at $x_{d,0} = 0$, $x_{d,1} = 1$ and at the times $t_{d,2} = 2.0$ and $t_{d,3} = 3.0$. The initial and boundary conditions, as well as the source, are the same as for inverse problem 1.

The ANN inverse model has the detector measurements $\boldsymbol{d}_0 = \big(\Psi(t_{d,2},0),\Psi(t_{d,3},0)\big)$, $\boldsymbol{d}_1 = \big(\Psi(t_{d,2},1),\Psi(t_{d,3},1)\big)$ as inputs and outputs the estimated absorption coefficients $\tilde{\kappa}_1$ and $\tilde{\kappa}_2$. For its training, we have used the direct solver to compute the training set $\left\{\left(\boldsymbol{d}^{(s)}_{train}, \boldsymbol{\kappa}^{(s)}_{train}\right)\right\}_{s=1}^{n_{train}}$ of $n_{train} = 81$ samples (patterns) with $\kappa^{(s)}_{1,2} = 0.1 + (s-1)h_s$, $h_s = 0.1$. The test set $\left\{\left(\boldsymbol{d}^{(s)}_{test}, \boldsymbol{\kappa}^{(s)}_{test}\right)\right\}_{s=1}^{n_{test}}$ has been generated with $n_{test} = 64$ uniformly distributed random choices $0.1 < \kappa^{(s)}_{1,2} < 0.9$.

For this inverse problem, we tested MLP architectures $4-n_n\times n_h-2$ (4 inputs, n_n neurons in each hidden layer n_h, and 2 outputs) with Standard Scaler preprocessing the input data. The training was stopped when the loss function $\mathcal{L} < 10^{-5}$. Due to the stochasticity of the training method, each test has been repeated three times. Table V presents the results with the $ReLU$ and the identity as activation functions in the hidden and in the output layers, respectively. It is tabulated the required average total number of epochs n_e and computational time t_c. Like the inverse problem 1, the MLP architecture $4-30\times4-2$ provided the best results, which we now set to report the results to follow.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2024
DOI: 10.5281/zenodo.12191947

TABLE V. INVERSE PROBLEM 2. TRAINING TESTS FOR MLP ARCHITECTURES WITH DATA PREPROCESSING

n_h/n_n	10	15	20	25	30
1	7516/ 8.60 s	2662/ 3.86 s	2628/ 3.93 s	1480/ 1.79 s	1292/ 1.83 s
2	1753/ 3.30 s	1336/ 2.29 s	1062/ 1.47 s	467/ 0.85 s	448/ 0.54 s
3	3581/ 6.76 s	1177/ 2.06 s	752/ 1.72 s	513/ 1.07 s	417/ 0.49 s
4	3434/ 7.05 s	1748/ 3.81 s	565/ 0.99 s	554/ 0.92 s	266/ 0.40 s

With approximately $n_e = 266$, the model reaches a mean squared error $\mathcal{L}_{train} < 10^{-5}$ and coefficient of determination $R^2_{train} = 0.9998$. The application of the trained model to the test data gave results with $\mathcal{L}_{test} < 10^{-5}$ and $R^2_{test} = 0.9998$. Figs. 5 and 6 show the expected versus estimated absorption coefficients for the training and test samples. The fitted least square line is also shown for the test data.

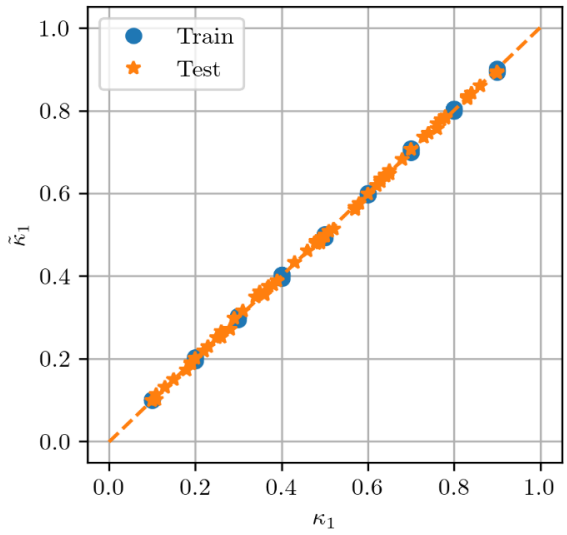

Fig. 5. Inverse problem 2. Expected κ versus estimated $\tilde{\kappa}_1$. Train: circles. Test: starts. Line fitted to test data results: dashed line

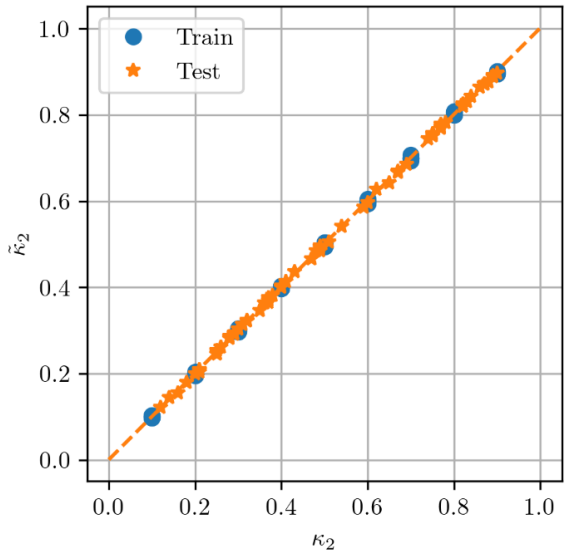

Fig. 6. Inverse problem 2. Expected κ versus estimated $\tilde{\kappa}_2$. Train: circles. Test: starts. Line fitted to test data results: dashed line

Sensitivity test. To validate the robustness and stability of the proposed MLP model in this problem, a sensitivity test was applied, which involves adding uniformly distributed noise into the input data, more specifically in detectors $\boldsymbol{d}_0$ and $\boldsymbol{d}_1$. Table VI shows the results of the mean squared error R^2 and the mean absolute squared error ($MAPE$) for different levels of noise.

The results indicate that the MLP model is relatively robust to low and moderate levels of noise in the input data. The noise is propagated to the output by a factor of 3.4 times. The $R^2 > 0.85$ is reached even with a noise level up to 5%. Figs. 7 and 8 show the expected *versus* estimated κ_1 and κ_2 of the test data set with noise levels of 2%, 3% and 4%. In the figures, the identify line is plotted as a dashed line as a guide. We observe the absence of outliers, which also indicates a good generalization of the ANN-MoC method.

TABLE VI. INVERSE PROBLEM 2. SENSITIVITY TESTS

Noise (%)	R^2	$MAPE$ (%)
1	0.994	3.84
2	0.981	7.10
3	0.948	11.06
4	0.929	13.27
5	0.873	16.95
6	0.809	20.43
7	0.726	25.26
8	0.758	23.15
9	0.483	35.52
10	0.658	33.31

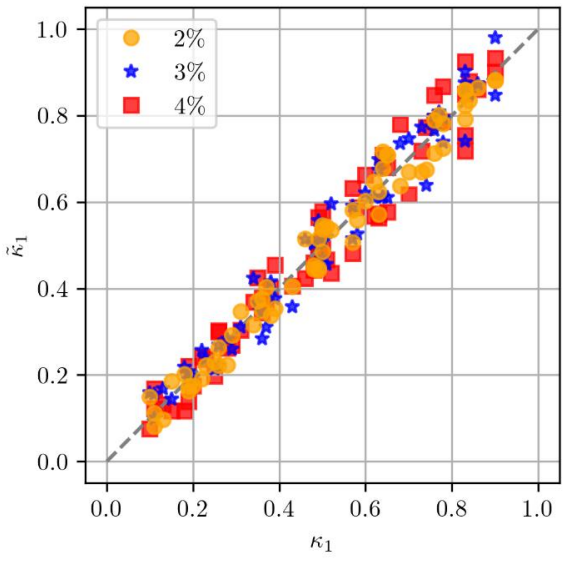

Fig. 7. Sensibility test for inverse problem 2. Comparison between expected κ_1 *versus* estimated $\tilde{\kappa}_1$ for different levels of input data noise

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2024
DOI: 10.5281/zenodo.12191947

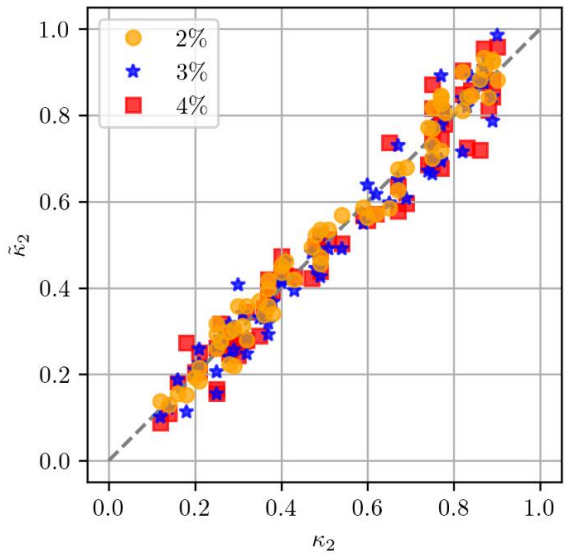

Fig. 8. Sensibility test for inverse problem 2. Comparison between expected κ_2 *versus* estimated $\tilde{\kappa}_2$ for different levels of input data noise

IV. Conclusions

In this paper, the ANN-MoC approach has been proposed to solve the inverse transient transport problem of estimating the absorption coefficient from scalar flux measurements at the boundaries of the model domain. The central idea is to fit an Artificial Neural Network (ANN) using samples generated by direct solutions computed by a Method of Characteristics (MoC) solver.

Applications of two different inverse transport problems were reported, one with homogenous medium and the other two region medium with piecewise constant absorption coefficient. After several numerical tests, we found that small MLPs could provide good estimations. Better results were reached by preprocessing the input data with the Standard Scaler. A sensitivity test was also reported for the second problem. The results highlight the potential of the proposed method to be applied to a broader range of inverse transport problems.

Further developments should aim to improve the direct solver. Improvements in the solution accuracy and, primarily, in computational performance are important to provide the ANN model with a higher-quality dataset. Solutions to more complex inverse transport problems could also benefit from the proposed approach, but once again, it will require additional improvements in the direct solver. Finally, the use of the proposed methodology for realistic problems depends on how good the direct transport model is for the intended application.

Acknowledgment

The authors thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) for partially financing this research (Finance Code 001).

References

[1] M. F. Modest, *Radiative Heat Transfer*, 3rd. New York: Elsevier, 2013.

[2] E. E. Lewis and W. F. Miller, *Computational Methods of Neutron Transport*. New York: John Wiley & Sons, Inc., 1984.

[3] E. W. Larsen, G. Thömmes, A. Klar, M. Seaïd, and T. Götz, "Simplified P_N approximations to the equations of radiative heat transfer and applications," *Journal of Computational Physics*, vol. 183, pp. 652–675, 2002.

[4] M. Frank, M. Seaïd, A. Klar, R. Pinnau, G. Thömmes, and J. Janicka, "A comparison of approximate models for radiation in gas turbines," *Progress in Computational Fluid Dynamics, an International Journal*, vol. 4, pp. 191–197, 2004.

[5] W. Fuqiang, M. Lanxin, C. Ziming, T. Jianyu, H. Xing, and L. Linhua, "Radiative heat transfer in solar thermochemical particle reactor: A comprehensive review," *Renewable and Sustainable Energy Reviews*, vol. 73, pp. 935–949, 2017.

[6] W. M. Stacey, *Nuclear Reactor Physics*, 2nd ed. Weinheim: Wiley-VCH, 2007.

[7] L. V. Wang and H. Wu, *Biomedical Optics: Principles and Imaging*. John Wiley & Sons, 2007.

[8] G. Bal, "Inverse transport theory and applications," *Inverse Problems*, vol. 25, p. 053001, 2009.

[9] M. N. Özisik and H. R. B. Orlande, *Inverse Heat Transfer. Fundamentals and Applications*, 2nd ed. CRC press, 2021.

[10] F. D. Moura Neto and A. J. Silva Neto, *An Introduction to Inverse Problems with Applications*, 1st ed. Heidelberg: Springer, 2014.

[11] F. S. Lobato, Jr. V. Steffen, and A. J. Silva Neto, "A comparative study of the application of differential evolution and simulated annealing in radiative transfer problems," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 32, pp. 518–526, 2010.

[12] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, 1st ed. London: MIT Press, 2016.

[13] J. C. Bokar, "The estimation of spatially varying albedo and optical thickness in a radiating slab using artificial neural networks," *International Communications in Heat and Mass Transfer*, vol. 26, pp. 359–367, 1999.

[14] Jr. J. Lugon, A. J. Silva Neto, and C. C. Santana, "A hybrid approach with artificial neural networks, Levenberg–Marquardt and simulated annealing methods for the solution of gas–liquid adsorption inverse problems," *Inverse Problems in Science and Engineering*, vol. 17, pp. 85–96, 2009.

[15] S. Haykin, *Neural Networks and Learning Machines*, 3rd ed. New Jersey: Pearson, 2009.

[16] L. C. Evans, *Partial Differential Equations*, 1st ed. Providence: AMS (American Mathematical Society), 1997.

[17] J. Stoer, R. Bulirsch, R. Bartels, W. Gautschi, and C. Witzgall, *Introduction to Numerical Analysis*, 3rd ed. New York: Springer, 1980.

[18] PyTorch Developers, "PyTorch." [Online]. Available: https://pytorch.org/

[19] D. P. Kingma and J. L. Ba, "Adam: a method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
[20] A. Zheng and A. Casari, *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*, 1st ed. O'Reilly Media, Inc.

LATIN-AMERICAN JOURNAL OF COMPUTING (LAJC), Vol XI, Issue 2, July 2024

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2024
DOI: 10.5281/zenodo.12191947

AUTHORS

Nelson Garcia Roman

Nelson García Román was born on September 20, 1992, in Pinar del Río, Cuba. He graduated as a Mechanical Engineer from the University of Pinar del Rio (2011-2016). During his studies, he was involved as a student assistant in Calculus and received two Honorable Mention awards for the presentation of two papers on Applied Mathematics in Engineering at the Scientific Conferences. After graduation, he became a professor at the same university from 2016 to 2018, and subsequently as a mathematics professor at the José Antonio Echeverría Technological University (CUJAE) until 2022. Currently, he is pursuing a Master's degree in the Graduate Program in Applied Mathematics (PPGMAp) at the Federal University of Rio Grande do Sul (UFRGS), where he holds a scholarship from CAPES, focusing on numerical methods, computational modeling, and deep learning for the numerical solution of particle neutral transport inverse problems.

Pedro Costas dos Santos

Pedro Costa dos Santos was born on September 23, 1998, in Rio de Janeiro - RJ, Brazil. During his secondary education, he successively received Honorable Mentions in the Brazilian Public School Mathematics Olympiad (OBMEP), and in 2015, he was awarded the Silver Medal in the competition. From 2016 to 2018, he attended an undergraduate course in Industrial Chemistry at the Federal University of Rio Grande do Sul(UFRGS). Since 2018, he has been a student in the undergraduate course of Applied Mathematics at UFRGS. Since 2022, he has been granted a scholarship for research initiation in Applied Mathematics, with an aim on Deep Learning applications to the numerical solution of Inverse Particle Neutral Transport problems. In 2023, he received the Honorable Mention for his research developments presented in the Scientific Initiation Week (SIC) at UFRGS.

Pedro Henrique de Almeida Konzen

Pedro Henrique de Almeida Konzen was born on June 12, 1981, in Santa Cruz do Sul - RS, Brazil. Doctor in Applied Mathematics from the Federal University of Rio Grande do Sul (UFRGS, 2010), having conducted doctoral research at Ruprecht-Karls-Universität Heidelberg/Germany (Uni-HD, 2008-2010). Currently, Adjunct Professor at the Department of Pure and Applied Mathematics (DMPA), Institute of Mathematics and Statistics (IME), Federal University of Rio Grande do Sul (UFRGS, since 2014). Permanent member of the Graduate Program in Applied Mathematics (PPGMAp-UFRGS, since 2022). Has experience in the field of applied mathematics, with emphasis on numerical methods, computational simulation, mathematical modeling and deep learning.