

PAPER • OPEN ACCESS

An iterative deep learning procedure for determining electron scattering cross-sections from transport coefficients

To cite this article: Dale L Muccignat *et al* 2024 *Mach. Learn.: Sci. Technol.* **5** 015047

View the [article online](#) for updates and enhancements.

You may also like

- [ECoPANN: A Framework for Estimating Cosmological Parameters Using Artificial Neural Networks](#)
Guo-Jian Wang, Si-Yao Li and Jun-Qing Xia
- [A novel ANN adaptive Riemannian-based kernel classification for motor imagery](#)
Fodil Yacine, Haddab Salah, Kachenoura Amar et al.
- [CoLFI: Cosmological Likelihood-free Inference with Neural Density Estimators](#)
Guo-Jian Wang, Cheng Cheng, Yin-Zhe Ma et al.



PAPER

OPEN ACCESS

RECEIVED

21 November 2023

REVISED

4 February 2024

ACCEPTED FOR PUBLICATION

4 March 2024

PUBLISHED

14 March 2024

Original Content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.



An iterative deep learning procedure for determining electron scattering cross-sections from transport coefficients

Dale L Muccignat¹ , Gregory G Boyle¹ , Nathan A Garland^{2,3,*} , Peter W Stokes^{1,4}
and Ronald D White¹

¹ College of Science & Engineering, James Cook University, Townsville, QLD 4814, Australia

² Centre for Quantum Dynamics, Griffith University, Nathan, QLD 4111, Australia

³ School of Environment and Science, Griffith University, Nathan, QLD 4111, Australia

⁴ Department of Medical Physics, Townsville University Hospital, Townsville, QLD 4814, Australia

* Author to whom any correspondence should be addressed.

E-mail: n.garland@griffith.edu.au

Keywords: swarm analysis, inverse problem, Boltzmann equation, machine learning

Abstract

We propose improvements to the artificial neural network (ANN) method of determining electron scattering cross-sections from swarm data proposed by coauthors. A limitation inherent to this problem, known as the inverse swarm problem, is the non-unique nature of its solutions, particularly when there exists multiple cross-sections that each describe similar scattering processes. Considering this, prior methods leveraged existing knowledge of a particular cross-section set to reduce the solution space of the problem. To reduce the need for prior knowledge, we propose the following modifications to the ANN method. First, we propose a multi-branch ANN (MBANN) that assigns an independent branch of hidden layers to each cross-section output. We show that in comparison with an equivalent conventional ANN, the MBANN architecture enables an efficient and physics informed feature map of each cross-section. Additionally, we show that the MBANN solution can be improved upon by successive networks that are each trained using perturbations of the previous regression. Crucially, the method requires much less input data and fewer restrictive assumptions, and only assumes knowledge of energy loss thresholds and the number of cross-sections present.

1. Introduction

Electron transport models are crucial to enable the predictive control of low-temperature plasma systems [1]. Underpinning these techniques are the use of accurate and complete electron scattering cross-section sets. The derivation of scattering cross-sections is typically conducted through experimental and theoretical techniques, coupled with verification through swarm scattering experiments to ensure their validity [2]. In regions where these techniques are limited, ‘educated guesses’ and numerical techniques are often used to bridge the gap. This knowledge gap motivates the need for reliable and benchmarked numerical techniques to aid in the development of accurate and complete cross-section sets.

Here, we focus on the determination of cross-section sets from swarm data, otherwise known as the inverse swarm problem [3]. Presently, two primary numerical techniques exist that aim to solve the inverse swarm problem: the iterative swarm technique and, more recently, the application of artificial neural networks (ANNs). The iterative swarm technique first used approximate distributions, such as a Maxwellian or Druyvesteyn distribution, of the electron energy distribution function to calculate transport coefficients that are compared to experimental transport coefficients and improved iteratively [4–6]. The accuracy of the iterative swarm technique was then improved with the inclusion of an accurate electron energy distribution functions derived from the solution of the Boltzmann equation [7–10].

A substantial limitation to solutions of the inverse swarm problem lies in its ill-posed nature. In particular, the existence of multiple cross-sections that describe similar scattering processes, such as similar

threshold vibrational modes, results in substantial degeneracy of transport data for a given species [3, 11]. In this limit, the iterative swarm technique relies on the intuition and experience of an expert, which, along with the trial and error nature of the approach, results in an inefficient procedure that is difficult to reproduce. Several methods that attempt to automate this methodology have been proposed [12–17] to address this issue. Of interest to this study, is use of ANNs trained on existing cross-section data to determine scattering cross-sections for electron transport in gases.

In the early 1990s, Morgan [18] first demonstrated a solution to the inverse swarm problem through an ANN trained on example cross-sections and their associated transport coefficients. Recently, Stokes *et al* [3] revisited this problem utilising advances in network architecture, model size and available cross-section data to improve the network's predictive power. Since then, the method has been successfully applied to the improvement of tetrahydrofuran, α -tetrahydrofurfuryl and nitric oxide electron scattering cross-sections [19–21]. Jetly and Chaudhury [22] evaluated the performance of three network architectures for the regression of single electron-scattering cross-section and found that a DenseNet architecture resulted in the highest regression accuracy.

While the application of ANNs to the inverse swarm problem shows great promise, the method is constrained due to a number of key limitations. This study aims to address the following two limitations. First, we demonstrate that a conventional ANN limits the ability of the network to reliably and self-consistently model multiple independent cross-section regressions when compared to an equivalent network that is designed in a physics-informed manner to contain multiple parallel branches of densely connected layers. In this design, each possible scattering process that may occur for a given target is provided a separate branch, which assures the user that the total of all parallel branches reproduces the required total scattering cross-section. Additionally, as the prediction of multiple cross-sections is limited due to the degenerate nature of the inverse swarm problem, we propose an iterative procedure to enable the network to incrementally explore the solution space of the problem.

Reducing the impact of these limitations will extend the applications for the ANN method to complex bio-molecules critical for applications in fields such as plasma medicine [23] and dielectric insulators [24]. In previous work, Stokes *et al* [3], used ANNs to approximate argon and helium's cross-sections. However, argon and helium do not exhibit vibration and attachment collisions that are generally important in bio-molecules. Here, we use methane as a case study for the influence of vibration and attachment processes, as has been widely studied in the literature [25–29]. For example, the strong excitation processes demonstrated by methane lead to a breakdown of the two-term approximation that is often employed in the solution of Boltzmann's equations [29–31].

In section 2, we outline each modification to the network architecture and methodology before evaluating their performance using methane as a case study. We then summarise the results in section 3.

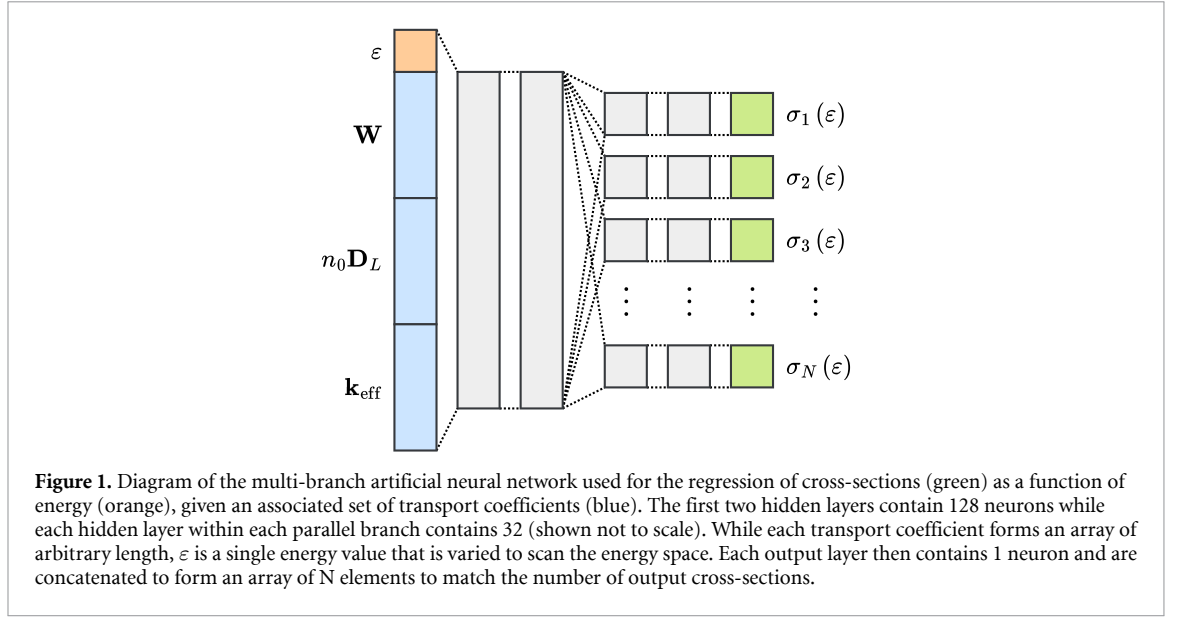
2. ANN regression of cross-section sets

The application of ANNs towards determining complete cross-section sets through the inversion of macroscopic experimental data has been the focus of a recent project at James Cook University [3, 19–21, 32]. While the technique has predominately been used to improve existing cross-section sets [19–21], the determination of complete cross-section sets for complex targets remains elusive due to the ill-posed nature of the problem. In this section, we present two modifications to the methodology that aim to improve the ability of the network to determine complete cross-sections from transport data.

First, to aid the network in representing the independent nature of each cross-section, we propose a multi-branch ANN (MBANN) and compare its performance through a regression of the cross-section set for methane recommended by Song and Bouwman [29, 33]. To reduce the impact of the non-unique nature of determining cross-section sets with multiple similar scattering processes, we then propose an iterative procedure that incrementally explores the solution space by using perturbations of the previous regression. To demonstrate the iterative procedure, we compare the initial regression and the best regression found for methane's cross-section set.

2.1. Multi-branch neural network regression

Stokes *et al* [3] proposed an ANN where each element of the output corresponds to a single cross-section. The simultaneous prediction of each cross-section ensures that the full set of cross-sections are self-consistent, which ensures an accurate replication of the target swarm transport data. In their investigation of various ANN architectures, Jetly and Chaudhury used separately trained networks to enforce independent feature maps for each cross-section. The authors state that the simultaneous prediction of multiple cross-sections would force the network to share feature maps across different types of cross-sections and thus severely inhibit its predictive capability.



Here, we propose a MBANN to bridge the gap between the requirement for self-consistency and the desire for independent feature maps for each cross-section. That is, for each cross-section, there exists an independent block of dense layers that each extend from a single block of dense layers. Each parallel branch is then allowed to develop a feature set specific to a single cross-section while still ensuring each regression is conducted in context of the full cross-section set.

We utilise a MBANN of the form,

$$\sigma(\mathbf{x}) = \left[\mathbf{A}_5^{[n]} \circ \text{mish} \circ \mathbf{A}_4^{[n]} \circ \text{mish} \circ \mathbf{A}_3^{[n]} \circ \text{mish} \circ \mathbf{A}_2 \circ \text{mish} \circ \mathbf{A}_1 \right] (\mathbf{x}), \quad (1)$$

where $\mathbf{A}_i(\mathbf{x}) \equiv \mathbf{W}_i \mathbf{x} + \mathbf{b}_i$ are affine mappings defined by dense *weight* matrices \mathbf{W}_i and *bias* vectors \mathbf{b}_i , and $\text{mish}(x) = x \tanh(\ln(1 + e^x))$ is a nonlinear activation function [34] that is applied element-wise. The final output, $\sigma^{[n]}$, then represents the n th cross-section of interest within a set of N cross-sections. $\mathbf{A}_3^{[n]}$, $\mathbf{A}_4^{[n]}$ and $\mathbf{A}_5^{[n]}$ form an array of N parallel branches that each utilise the output of \mathbf{A}_2 to independently represent the n th cross-section. $\mathbf{b}_3^{[n]}$ and $\mathbf{b}_4^{[n]}$ contain 32 elements each, $\mathbf{b}_5^{[n]}$ contains 1 for each output n , while those in the initial two layers contain 128. The *weight* matrices are sized accordingly.

From previous investigations and a simple hyper-parameter optimization procedure outlined in appendix B, we found that approximately 32 neurons in each parallel layer is required for a suitable regression of each cross-section with more neurons resulting in modest improvements. Here, we choose this minimum to isolate and demonstrate the differences between a MBANN and an equivalent ANN architecture. Each other parameter, such as the activation function and number of hidden layers, was chosen from a set of reasonable values using a comparison of validation accuracy and prior experience. A schematic representation of the MBANN architecture is shown in figure 1.

For comparison, we use an ANN of the form,

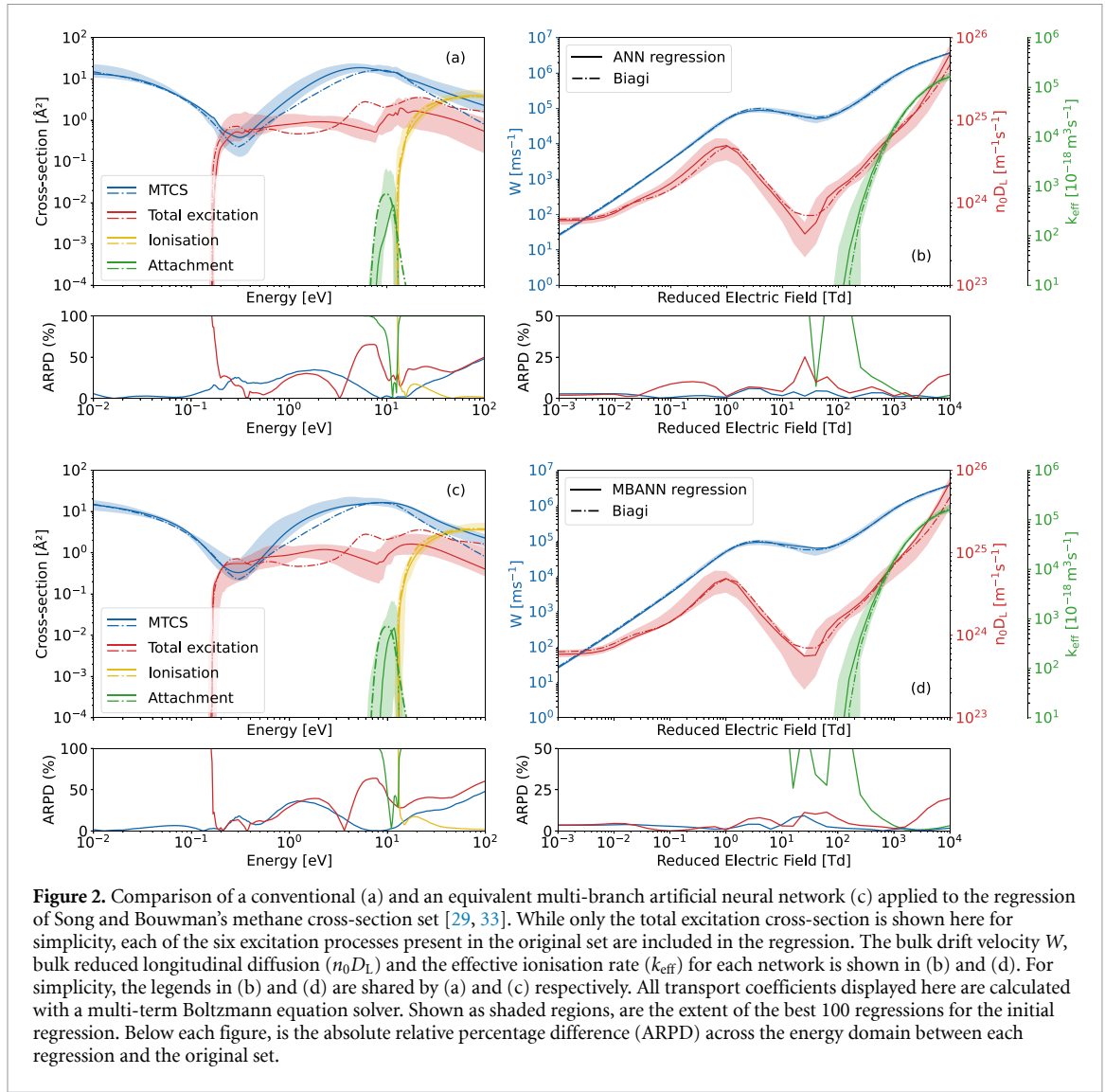
$$\sigma(\mathbf{x}) = [\mathbf{A}_5 \circ \text{mish} \circ \mathbf{A}_4 \circ \text{mish} \circ \mathbf{A}_3 \circ \text{mish} \circ \mathbf{A}_2 \circ \text{mish} \circ \mathbf{A}_1] (\mathbf{x}), \quad (2)$$

where \mathbf{b}_3 and \mathbf{b}_4 now contain $32 \times N$ elements while \mathbf{b}_5 contains N elements to match the number of outputs. The ANN thus contains the same number of neurons per layer as the MBANN network outlined above. We note however that the number of trainable parameters is larger than that of the MBANN.

Each cross-section is a function of the incoming projectile electron kinetic energy, ε , which, alongside the available swarm data, forms the input to the neural network,

$$\mathbf{x} = \begin{bmatrix} \varepsilon \\ \mathbf{W}(E/n_0) \\ n_0 \mathbf{D}_L(E/n_0) \\ \mathbf{k}_{\text{eff}}(E/n_0) \end{bmatrix}, \quad (3)$$

where n_0 , \mathbf{W} , $n_0 \mathbf{D}_L$ and \mathbf{k}_{eff} are the neutral density, bulk drift velocity, reduced bulk longitudinal diffusion and effective ionisation rate of the electron swarm evaluated at a number of reduced electric fields E/n_0 .

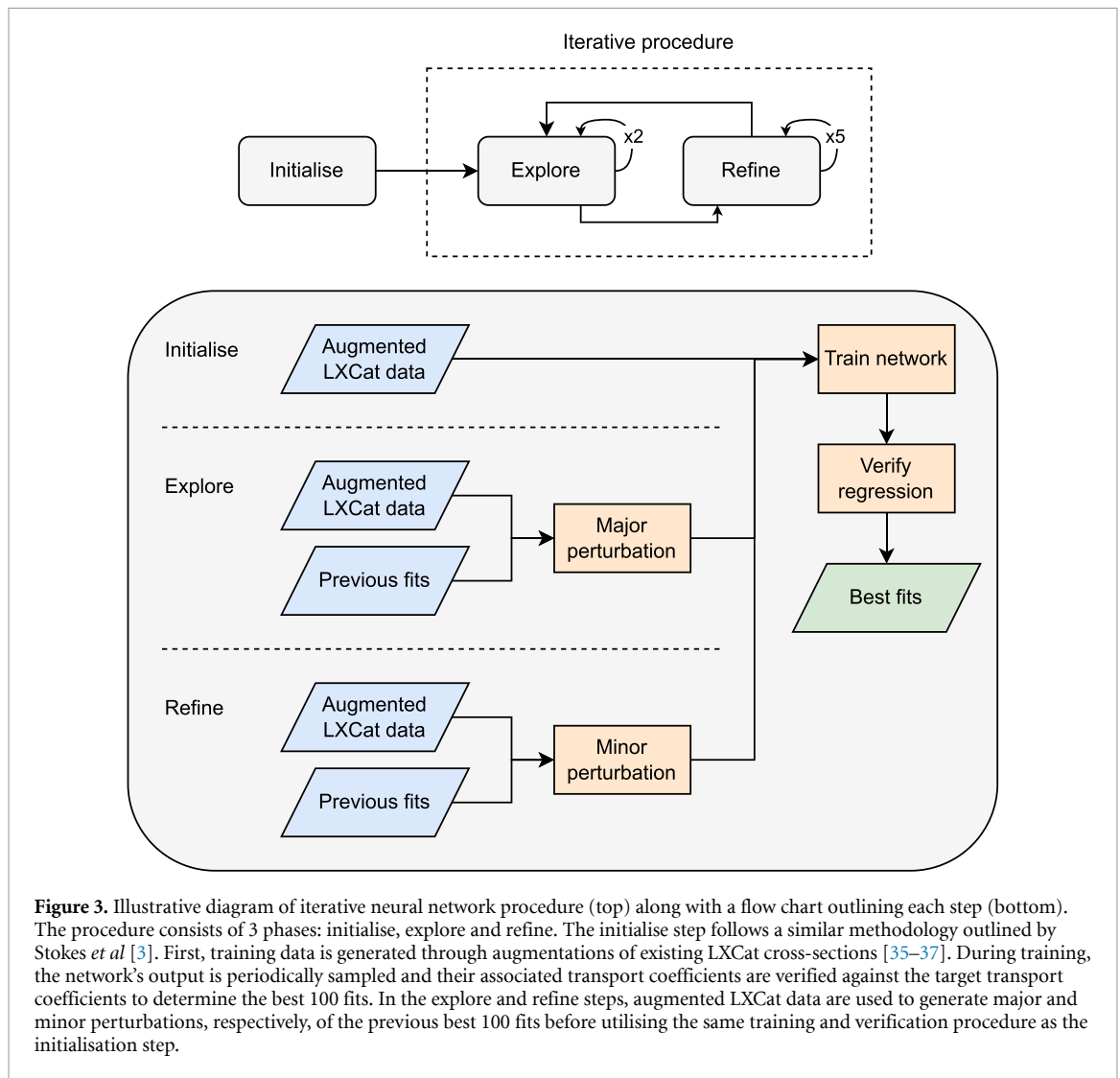


To train each neural network, we generate an appropriate set of physically plausible example swarm transport data using augmentations of cross-sections from the LXCat project [35–37]. The data generation process and subsequent training procedure follows the method outlined by Stokes *et al* [3] with some modifications that are described in detail in appendix A.

A total of 10^5 training iterations are performed, each consisting of a mini-batch of 32 cross-section sampled at 128 energies from a total of 2×10^4 training examples for each cross-section. For each training set, a multi-term Boltzmann equation solver was used to calculate W , $n_0 D_L$ and k_{eff} at 80 log-spaced reduced electric fields between 10^{-3} and 10^4 Td (1 Td = 1 Townsend = 10^{-21} Vm²) while the cross-section regression was conducted between 0.01 and 200 eV. A detailed outline of the training procedure can be found in appendix B.

To demonstrate the improvements offered by the proposed architecture, we present a regression of methane's cross-section set for both the MBANN and an equivalent ANN network. The cross-section set was retrieved from the LXCat database [35–37] and originates from the work of Song and Bouwman [29, 33]. In this work, we perform a regression of the elastic momentum transfer cross-section, total ionisation, total attachment, and each of the 6 excitation cross-sections. In addition, while it has been shown that an ANN can determine some energy loss thresholds [3], any target cross-section set that exhibits multiple similar threshold processes introduces a high degree of degeneracy. In this work, we assume knowledge of each energy loss threshold and leave their determination for future investigations.

A comparison between the resulting regression for both the ANN and MBANN architectures is shown in figure 2. In each, the extent of the 100 best fits sampled during the training process is shown as a shaded region to provide an indication of the network's variability. The ANN regression resulted in a mean absolute relative percentage difference (MARPD, see equation (B.2)) of 2.2, 6.3 and 21% for W , $n_0 D_L$ and k_{eff} ,



respectively. The MBANN regression resulted in a comparable MARPD of 2.7, 5.2 and 21% for W , $n_0 D_L$ and k_{eff} , respectively.

While each network exhibited a similar global accuracy in the replication of transport coefficients, non-physical fluctuations are present in the ANN regression of the elastic and excitation cross-sections between 0.1–0.4 eV and 7–20 eV. It is clear that the large gradients present in these regions, due to the energy loss thresholds of 0.162, 0.363, 7.5, 9.1, 12.36, 15.5 and 15.5 eV, resulted in the ANN being unable to independently represent each cross-section’s feature map when compared to the MBANN fit, despite the same number of neurons available to each architecture. Similar fluctuations were also observed in the best 100 regressions represented by the shaded regions in figure 2. While a larger ANN would mitigate this effect, the MBANN is able to leverage its independent branches to ensure each cross-section’s feature map is physics-informed and intuitive. The branching architecture also offers a simple linear scaling for variations in the number of target cross-sections through the addition, or removal of, individual branches.

As demonstrated by both networks presented here, there remains much room for improvement in the regression of methane’s electron-scattering cross-section set. While additional improvements of the network architecture may be available, such as those seen in the work of Jetly and Chaudhury [22], the ill-posed nature of the inverse swarm problem places an inherent limit on the accuracy of methods that seek to learn the feature map between transport data and cross-section sets. In what follows, we aim to mitigate this restriction through a new procedure that uses a sequence of MBANNs to incrementally explore the solution space.

2.2. An iterative approach to neural network regression

The regression of numerous similar cross-sections poses a substantial challenge for solutions to the inverse swarm problem. As shown previously by Stokes *et al* [3, 20, 21], the predictive power of the neural network can be improved by restricting the training data to perturbations around a reference cross-section set. We extend this work and propose an iterative procedure in which a sequence of networks are trained using a

weighted mixing of the previous solution with example cross-section data. As illustrated in figure 3, the procedure consists of three phases; initialise, explore and refine.

In the initialise phase we follow the procedure outlined in appendix B. In this phase, no prior information of the target cross-section set is given to the network other than energy loss thresholds and the number of processes present. Each energy loss cross-section in the training set is thus energy shifted to their respective target threshold energy. The resulting best 100 regressions then form an array of current fits σ_c . During the following two phases, we seek to improve the regression by generating stochastic perturbations around each current fit through a weighted mixing with example cross-section data.

To train each subsequent network, we use augmented LXCat cross-sections to generate perturbations around each current fit. First, σ_s , is generated with the same method used in the initialisation phase. We then use σ_s to generate a perturbation around the i th current fit $\sigma_{c,i}$ using a weighted sum in log space,

$$\sigma_s(\varepsilon) = \sigma_s^{1-r}(\varepsilon - \varepsilon_s + \varepsilon_t) \sigma_{c,i}^r(\varepsilon - \varepsilon_{c,i} + \varepsilon_t), \quad (4)$$

where i is a uniformly distributed random number and r is a pseudo-random number sampled from a scaled Laplace distribution. ε_s and $\varepsilon_{c,i}$ represent the respective threshold energies for each cross-section while ε_t is the target threshold energy. The parameter r then defines how similar each training sample σ_s is to $\sigma_{c,i}$, where values close to 1 results in minor perturbation around $\sigma_{c,i}$ while values close to 0 results in major perturbations. Values of r greater than 1 can be used to produce accentuated perturbations around $\sigma_{c,i}$ to extend the solution space beyond the available data. The extent of these perturbations then define the network's ability to either explore the solution space or refine the existing solution. If the training data is restricted to minor perturbations, the solution may become trapped in a local minimum. Conversely, major perturbations may result in the network being unable to determine a sufficiently accurate cross-section set. The explore and refine phases of the procedure aim to strike a balance between these two regimes. In the explore phase, major perturbations of σ_c are made to assist the network in traversing the solution space beyond the current fits while in the refine phase, minor perturbations are used to further refine σ_c .

Through a trial and error process, we found the following parameters to be suitable for each training phase. In the explore phase, we conduct two iterations while in the refine phase we conduct five to help ensure sufficient refinement of a particular solution is conducted after each exploratory phase. For high energy (>10 eV) processes, such as electronic excitation and ionisation, we sample r from the domain $[0.5, 0.8]$ for each iteration in the explore phase while during the refine phase we set $r = 0.8$. For low energy processes, such as vibration and elastic, r is sampled from $[0.5, 1.5]$ in the explore phase and $[0.8, 1.2]$ in the refine phase. In the case of low energy processes, r values greater than 1 are used to generate training examples that accentuate low energy cross-section features that are present in the sample data.

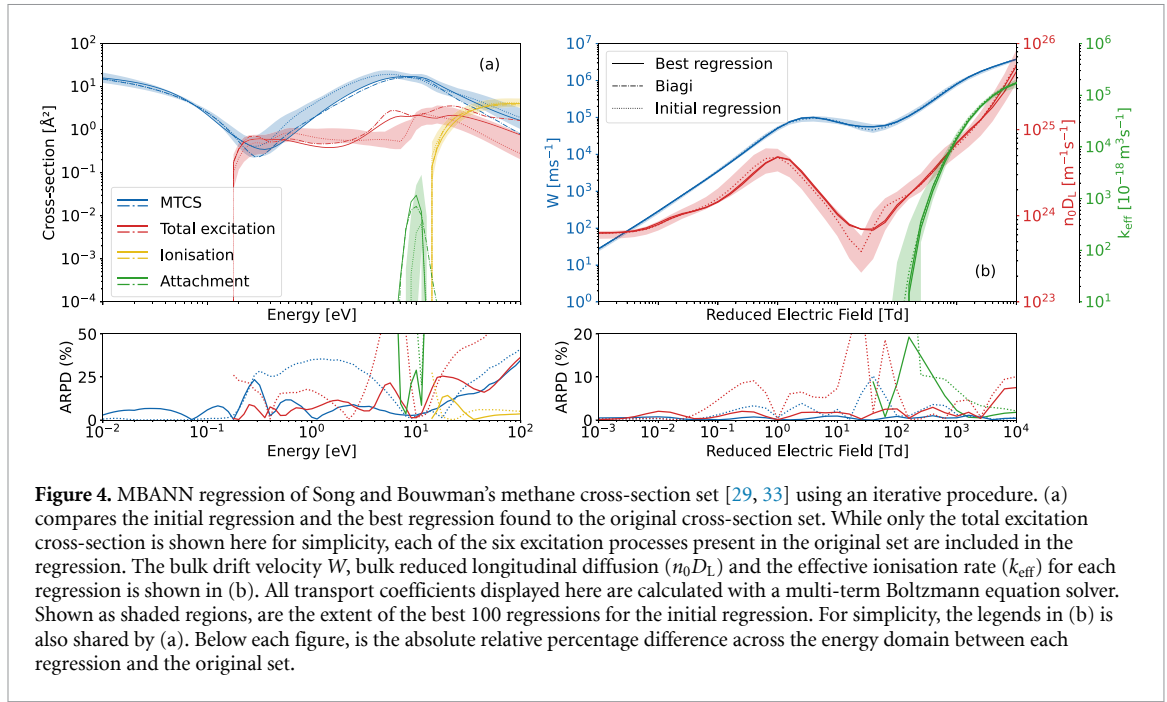
Direct parallels can be drawn between the iterative MBANN technique and the well known iterative swarm technique. In each, an informed supervisor guides the procedure towards both a physical and accurate solution to the inverse problem. In the iterative swarm technique, this is generally the role of an expert in the field who may make adjustments to the solution or procedure where necessary. In the iterative MBANN technique this role is, in the ideal case, automated by the neural network. Depending on the application, the guidance of an expert may still be required to choose suitable parameters and monitor its performance.

We demonstrate the proposed iterative procedure through a regression of methane's cross-section set presented in section 2.1. While 32 neurons was chosen for the hidden layers in each parallel branch as the limiting case in the previous section, we increase this to 64 in what follows due to modest improvements in the validation accuracy. In figure 4, we compare both the initial and the best regression found of methane's cross-section set during the procedure, along with their associated transport coefficients.

The associated transport coefficients of the initial fit of methane's cross-section set results in substantial discrepancies to the original set. The initial MBANN regression resulted in a MARPD of 1.8, 5.1 and 27% for W , $n_0 D_L$ and k_{eff} , respectively. After 40 iterations, the procedure was then able to substantially improve upon the initial regression with the best iteration resulting in a MARPD of 0.48, 1.69, and 5.84% for W , $n_0 D_L$ and k_{eff} , respectively.

Crucially, we also find substantial improvements in the agreement between the total cross-sections for each collision type in the set and the target cross-section set. Provided as shaded regions in figure 4, is the extent of the best 100 regressions found in the initial fit. Both the total excitation and the attachment cross-section of the best regression exist, in part, outside of the initial extent of σ_c . The network was therefore able to effectively explore where necessary to improve the resulting fit of the target transport coefficients.

In its current form, the procedure assumes the prior knowledge of threshold energies. This assumption is particularly important when groupings of similar thresholds are present. If instead, effective excitation cross-sections are utilised in the network to represent groupings of similar threshold energies, this assumption could be avoided at the expense of physical threshold energies in the resulting cross-section set.



In this investigation, we utilise only calculated transport coefficients over a large range of reduced electric fields. In reality, such a range is not often available. While tailoring the energy domain of the cross-section regression to the transport data available will alleviate the limitation in part, this has limited returns. We thus encourage the measurement of swarm coefficients over a broad range of electric field domains where possible. Finally, while we have made a concerted effort to develop a robust iterative procedure, depending on the particular problem and the extent of available transport data, the network may still produce non-physical cross-sections or become trapped in a local minimum. The parameters utilised here should thus only serve as a guide for future applications to be modified as needed.

3. Conclusion

In this work, we demonstrate a new iterative procedure that uses a MBANN to solve the inverse swarm problem. Building upon the foundations outlined by Stokes *et al* [3, 19–21] and Jetly and Chaudhury [22], we address two key limitations of an ANN solution to the inverse swarm problem.

We first evaluate the use of a MBANN that includes an independent branch of dense layers for each output that each stem from common feature map of the input. We then compare the MBANN to an equivalent conventional ANN using Song and Bouwman's methane cross-section set [29, 33] and demonstrate that the use of parallel layers can improve the resulting regression as the network is able to efficiently and independently represent multiple distinct cross-sections.

In addition, taking inspiration from the iterative swarm technique, we propose an iterative MBANN procedure to that incrementally explores the solution space to reduce the ill-posed nature of the problem. After an initial regression is found, we use a sequence of MBANNs that are each trained using perturbations around the previous regression. The iterative MBANN procedure then converges towards a particular solution of the inverse swarm problem.

To demonstrate the iterative MBANN procedure, we evaluate its performance using Song and Bouwman's methane cross-section set [29, 33]. In the 40 iterations that were conducted, the MARPD of the initial regression's resulting transport coefficients was substantially decreased from 1.8, 5.1 and 27%, to 0.48, 1.69, and 5.84% for W , $n_0 D_L$ and k_{eff} , respectively. Additionally, the total cross-section for each collision type within the best set found exhibited good agreement with the original set, in contrast to the initial regression.

Overall, we have demonstrated an improved artificial neural network solution to the inverse swarm problem that utilises both an iterative procedure and parallel branches of densely connected layers that represent each cross-section. In conjunction, these additions improve the ability of the network to generate both self-consistent and physical cross-sections, particularly when large degeneracies may exist for a particular species.

In future work, we aim to apply this procedure to derive complete cross-section sets for complex targets while also investigating the use of convolutional architectures.

Data availability statement

The data cannot be made publicly available upon publication because they are not available in a format that is sufficiently accessible or reusable by other researchers. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

The authors gratefully acknowledge the financial support of the Australian Research Council (ARC) through the Discovery Projects Scheme (DP220101480 and DP190100696).

Appendix A. LXCat data augmentation

To provide the network with sufficient training samples, we implement the data augmentation method first presented by Stokes *et al* [3]. First, we collate an appropriate set of physically plausible example swarm transport data using cross-sections from the LXCat project [35–37]. Following the method outlined in the work of Jetly and Chaudhury [22], from a total of 397 elastic and effective cross-sections available, 3 groups of similar features were created by visual inspection. The total number of cross-sections contained in each group were 60, 130, and 197. In addition, 6763, 451 and 332 cross-sections in total were present for excitation, ionisation and attachment processes respectively.

In each iteration, we then generate a database of example cross-sections through a randomised mixing of cross-section pairs using a weighted sum in log space, [3]:

$$\sigma_s(\varepsilon) = \sigma_1^{1-r'} \left(\varepsilon - \varepsilon_1 + \varepsilon_1^{1-r'} \varepsilon_2^{r'} \right) \sigma_2^{r'} \left(\varepsilon - \varepsilon_2 + \varepsilon_1^{1-r'} \varepsilon_2^{r'} \right), \quad (\text{A.1})$$

where σ_1 and σ_2 are a random pair of LXCat electron scattering cross-sections, sampled from the available targets, for a given process type (e.g. excitation, ionisation, etc), while ε_1 and ε_2 are their respective threshold energies. The parameter r' is a pseudo-random mixing ratio. To help ensure a physical representation of cross-sections within the training set, a scaled Laplace distribution truncated to the domain $r \in [0, 1.5]$ was chosen to bias r' towards small perturbations around each example cross-section. Any decaying distribution may be sufficient for this purpose however. Ratios greater than 1 are used here to accentuate cross-section features found within the sample set to reduce the extend of outliers within the set. Note that for the elastic cross-section, we sample two cross-sections from the three groups so that each group is equally represented in the training set.

In addition, due to the limited nature of the available data, the solution may exist at the extremes of the available training data which can introduce unwanted bias in the data augmentation process. To alleviate this, equation (A.1) is modified such that the energy domain and magnitude of each cross-section are multiplied by the scaling factors 10^a and 10^b respectively. Each factor is a pseudo-random number uniformly distributed within a defined range. Here, we set $a \in [-0.5, 0.5]$, $b \in [-0.5, 0.5]$ for elastic cross-sections, $a = 0$, $b \in [0, 2]$ for excitation cross-sections, $a = 0$, $b \in [0, 1]$ for ionisation cross-sections and $a \in [-1, 1]$, $b \in [-1, 1]$ for attachment cross-sections. Each was chosen to reasonably extend the extent of the available training data. Finally, we log transform then scale each cross-section between -1 and 1 . If a cross-section magnitude is below $\delta = 10^{-6}$, it is replaced by 10^{-7} before applying the transform.

For each generated cross-section set, the resulting transport coefficients are calculated through a multi-term Boltzmann equation solver according to the target experimental transport coefficients. If a particular set results in non-physical transport coefficients, the set is removed from the training set and a new set is found until the condition is satisfied. For this investigation, physical transport coefficients are defined as $\mathbf{W} > 0$ and $n_0 \mathbf{D}_L > 0$, where the electric field is directed along the negative z -axis. In addition, we apply a logarithmic transformation to ensure that all inputs and outputs of the network are dimensionless and lie within $[-1, 1]$, with special consideration given to \mathbf{k}_{eff} due to the presence of negative values. The vectors $\mathbf{k}_{\text{eff}}^+$ and $\mathbf{k}_{\text{eff}}^-$ are created to represent the positive and negative portions of the original input respectively. For $\mathbf{k}_{\text{eff}}^+$, each negative value is set to a sufficiently small positive value while for $\mathbf{k}_{\text{eff}}^-$, each positive value is set to a sufficiently small negative value before taking the absolute value of each.

Appendix B. Training procedure

Training of the network is conducted using a mini-batch of 32 cross-section sets evaluated at 128 energies for a particular iteration. Each *weight* and *bias* is then updated using the ‘NAdam’ optimiser [38] with a learning

rate of 0.001 and 0.9 and 0.999 as the exponential decay for the first and the second momentum estimate respectively. For each batch, random noise is applied to each transport coefficient. Noise is sampled from a lognormal distribution with a standard deviation chosen to replicate experimental uncertainties typically observed for each transport coefficient.

The prediction of multiple cross-sections of the same collision type has previously been shown to be a highly degenerate problem, particularly in the case of excitation collisions [3]. To emphasise the importance of the total cross-section, we use a loss function that includes a penalty for the total cross-section for each process type, in addition to each individual cross-section. The loss function is defined as follows,

$$L = \frac{1}{I} \sum_i \frac{1}{M_i + 1} \left[\sum_{\sigma \in \sigma_i} |\sigma - \hat{\sigma}| + \left| \sum_{\sigma \in \sigma_i} \sigma - \sum_{\sigma \in \sigma_i} \hat{\sigma} \right| \right], \quad (\text{B.1})$$

where σ_i is the set of M_i cross-sections for the collision type i , I is the number of collision types present and $\hat{\sigma}$ is the predicted cross-section. Note that the calculation of the total cross-sections is conducted in linear space before scaling is re-applied.

To validate the network, we set aside 5% of the available cross-sections for each collision type to form the basis of our validation set. In the case of the elastic cross-sections, 5% of each of the feature groups is selected. The chosen validation cross-sections then undergo the same data augmentation as the training data to produce 100 samples. Every 100 iterations we test the network on the validation set and calculate the MARPD over the energy domain,

$$\text{MARPD} = \frac{1}{I} \sum_i \left| 100 \times \frac{\sigma_{t,i} - \hat{\sigma}_{t,i}}{|\sigma_{t,i}| + |\hat{\sigma}_{t,i}|} \right|, \quad (\text{B.2})$$

where $\sigma_{t,i}$ is the total cross-section for the collision type i . During training, the validation loss is used to both compare network parameters and aid in preventing overfitting.

In addition, every 10 training iterations we store the neural network's prediction of the target transport data and compare the resulting transport coefficients. Due to the presence of multiple distinct fit parameters, we repeatedly select one random parameter and then remove the worst predicted cross-section set until only one remains. This process is then replicated 1000 times before the cross-section set which appears most frequently is chosen as the best overall fit. As desired, the best overall fit can be removed, and the processes repeated to find the next best fit.

To select each hyperparameter, a simple tuning process was conducted that compared the validation loss between reasonable values for batch size (32, 64, 128), number of hidden layers (2, 3, 4, 5), hidden layer size (32, 64, 128, 256) and the optimiser used (NAdam, Adam). For simplicity, this processes was conducted with a conventional ANN in which each layer contained an equal number of neurons and the output consisted of an elastic and ionisation cross-section along with three electronic excitation cross-sections. The tuning processes indicated that 4–5 hidden layers with 128–256 neurons resulted in the smallest validation error. Due to computational limitations, a more extensive tuning that includes branching hidden layers and a greater variation of parameters was not conducted.

ORCID iDs

Dale L Muccignat  <https://orcid.org/0000-0002-6693-6851>

Gregory G Boyle  <https://orcid.org/0000-0002-8581-4307>

Nathan A Garland  <https://orcid.org/0000-0003-0343-0199>

Peter W Stokes  <https://orcid.org/0000-0002-0956-5927>

Ronald D White  <https://orcid.org/0000-0001-5353-7440>

References

- [1] Adamovich I *et al* 2017 *J. Phys. D: Appl. Phys.* **50** 323001
- [2] Brunger M J and Buckman S J 2002 *Phys. Rep.* **357** 215–458
- [3] Stokes P W, Casey M J E, Cocks D G, de Urquijo J, García G, Brunger M J and White R D 2020 *Plasma Sources Sci. Technol.* **29** 105008
- [4] Mayer H F 1921 *Ann. Phys., Lpz.* **369** 451–80
- [5] Ramsauer C 1921 *Ann. Phys., Lpz.* **369** 513–40
- [6] Townsend J S and Bailey V A 1922 *London, Edinburgh Dublin Phil. Mag. J. Sci.* **43** 593–600
- [7] Frost L S and Phelps A V 1962 *Phys. Rev.* **127** 1621–33
- [8] Engelhardt A G and Phelps A V 1963 *Phys. Rev.* **131** 2115–28
- [9] Engelhardt A G, Phelps A V and Risk C G 1964 *Phys. Rev.* **135** A1566–74

- [10] Hake R D and Phelps A V 1967 *Phys. Rev.* **158** 70–84
- [11] Brunger M J, Buckman S J and Ratnavelu K 2017 *J. Phys. Chem. Ref. Data* **46** 023102
- [12] Duncan C W and Walker I C 1972 *J. Chem. Soc. Faraday Trans. 2* **68** 1514
- [13] O'Malley T F and Crompton R W 1980 *J. Phys. B: At. Mol. Phys.* **13** 3451–64
- [14] Taniguchi T, Suzuki M, Kawamura K, Noto F and Tagashira H 1987 *J. Phys. D: Appl. Phys.* **20** 1085–7
- [15] Morgan W L 1991 *IEEE Trans. Plasma Sci.* **19** 250–5
- [16] Morgan W L 1993 *J. Phys. D: Appl. Phys.* **26** 209–14
- [17] Brennan M J and Ness K F 1993 *Aust. J. Phys.* **46** 249
- [18] Morgan W L 1991 *Phys. Rev. A* **44** 1677–81
- [19] Stokes P W, Cocks D G, Brunger M J and White R D 2020 *Plasma Sources Sci. Technol.* **29** 055009
- [20] Stokes P W, White R D, Campbell L and Brunger M J 2021 *J. Chem. Phys.* **155** 084305
- [21] Stokes P W, Foster S P, Casey M J E, Cocks D G, González-Magaña O, de Urquijo J, García G, Brunger M J and White R D 2021 *J. Chem. Phys.* **154** 084306
- [22] Jetly V and Chaudhury B 2021 *Mach. Learn.: Sci. Technol.* **2** 035025
- [23] Laroussi M 2018 *Plasma* **1** 47–60
- [24] Chachereau A, Rabie M and Franck C M 2016 *Plasma Sources Sci. Technol.* **25** 045005
- [25] Cho H, Park Y S, y Castro E A, de Souza G L C, Iga I, Machado L E, Brescansin L M and Lee M-T 2008 *J. Phys. B: At. Mol. Opt. Phys.* **41** 045203
- [26] Işık N, Doğan M and Bahçeli S 2016 *J. Phys. B: At. Mol. Opt. Phys.* **49** 065203
- [27] Gadoum A and Benyoucef D 2019 *IEEE Trans. Plasma Sci.* **47** 1505–13
- [28] Harvey M, Sakaamini A, Patel M, Amami S, Madison D and Murray A J 2019 *J. Chem. Phys.* **151** 194305
- [29] Bouwman D, Martinez A, Braams B J and Ebert U 2021 *Plasma Sources Sci. Technol.* **30** 075012
- [30] Ness K F 1994 *J. Phys. D: Appl. Phys.* **27** 1848
- [31] White R D, Ness K F and Robson R E 1999 *J. Phys. D: Appl. Phys.* **32** 1842
- [32] Muccignat D L, Stokes P W, Cocks D G, Gascooke J R, Jones D B, Brunger M J and White R D 2022 *Int. J. Mol. Sci.* **23** 3354
- [33] Song M, Yoon J, Cho H, Itikawa Y, Karwasz G P, Kokkoulouline V, Nakamura Y and Tennyson J 2015 Cross Sections for Electron Collisions with Methane *J. Phys. Chem. Ref. Data.* **44**
- [34] Misra D 2019 arXiv:1908.08681
- [35] Pancheshnyi S, Biagi S, Bordage M C, Hagelaar G J, Morgan W L, Phelps A V and Pitchford L C 2012 *Chem. Phys.* **398** 148–53
- [36] Pitchford L C et al 2017 *Plasma Process. Polym.* **14** 1600098
- [37] Carbone E, Graef W, Hagelaar G, Boer D, Hopkins M M, Stephens J C, Yee B T, Pancheshnyi S, van Dijk J and Pitchford L 2021 *Atoms* **9** 16
- [38] Dozat T 2016 Incorporating Nesterov momentum into Adam *Proc. 4th Int. Conf. on Learning Representations, Workshop Track*