# Supporting information for

# Using generative adversarial networks to match experimental and simulated inelastic neutron scattering data

*Andy S. Anker*[1*], *Keith T. Butler*[2,4*], *Manh Duc Le*[3], *Toby G. Perring*[3], *Jeyan Thiyagalingam*[2]

*Correspondence to andy@chem.ku.dk (ASA) and k.butler@qmul.ac.uk (KTB)

1: Nano-Science Center and Department of Chemistry, University of Copenhagen, Denmark

2: Scientific Computing Department, Rutherford Appleton Laboratory, England

3: ISIS Neutron and Muon Source, Rutherford Appleton Laboratory, England

4: Current affiliation: School of Engineering and Materials Science, Queen Mary University of London, Mile End Rd, London E1 4NS, England

**Table of Contents**

A: Distribution of simulated inelastic neutron scattering (INS) 2D data ........................................... 3

B: Applying Sim2ExpGAN for resolution convolution of a simulated 2D INS spectra .................... 4

C: Similarity between datasets can be calculated using the Wasserstein distance in the $\mathrm{Exp2Sim}_{\mathrm{Featurespace}}$ ........................................................................................................ 5

D: Applying $\mathrm{Sim2Exp}_{\mathrm{network}}$ on a range of simulated and experimental 2D INS spectra ................... 6

E: Distribution of simulated inelastic neutron scattering (INS) 3D data ............................................ 7

F: Evaluating the Exp2SimGAN on simulated 3D INS data from the test set using the Goodenough spin wave model ........................................................................................................ 8

**A: Distribution of simulated inelastic neutron scattering (INS) 2D data**

| | Training set | | Testing set | |
|---|---|---|---|---|
| | Resolution convolved | Resolution unconvolved | Resolution convolved | Resolution unconvolved |
| Goodenough | 2627 (2622) | 2627 (2622) | 656 (656) | 656 (656) |
| Dimer | 2622 (2622) | 2622 (2622) | 654 (654) | 654 (654) |

**Table S1 | Distribution of simulated inelastic neutron scattering (INS) 2D data.** The numbers denote how many datasets were used to train the Exp2SimGAN algorithm and DUQ classifier (in parenthesis).

## B: Applying Sim2ExpGAN for resolution convolution of a simulated 2D INS spectra

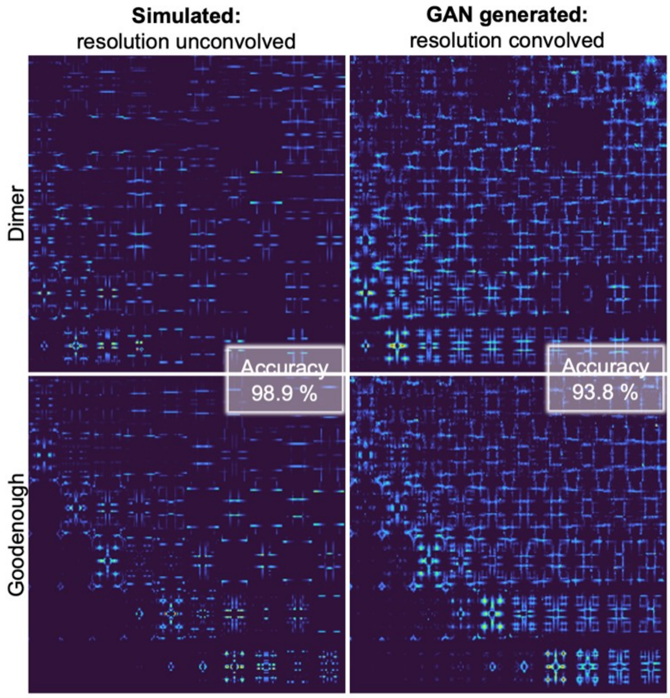


**Fig. S1 | Evaluating the Sim2Exp$_{\mathbf{network}}$ on simulated 2D INS spectra from the test set.** The INS data is split into 80 % training set and 20 % test set. After the network has trained on the training set, we apply it on the data in the test set. Here are shown an example of performing resolution convolution on 2D INS spectra simulated with the Dimer spin wave model and an example using the Goodenough spin wave model. Note that the experimental axis is the same as in Figure 1a. The highlighted accuracies are the performance of the DUQ classifier[44], trained on simulated INS spectra without resolution convolution, on the test set. If the DUQ classifier is trained on GAN resolution convoluted INS spectra the accuracies are 98.6 % applied to simulated data with resolution convolution and 75.1 % applied to GAN-deconvolved data.

**C: Similarity between datasets can be calculated using the Wasserstein distance in the $\mathrm{Exp2Sim}_{\mathrm{Featurespace}}$**

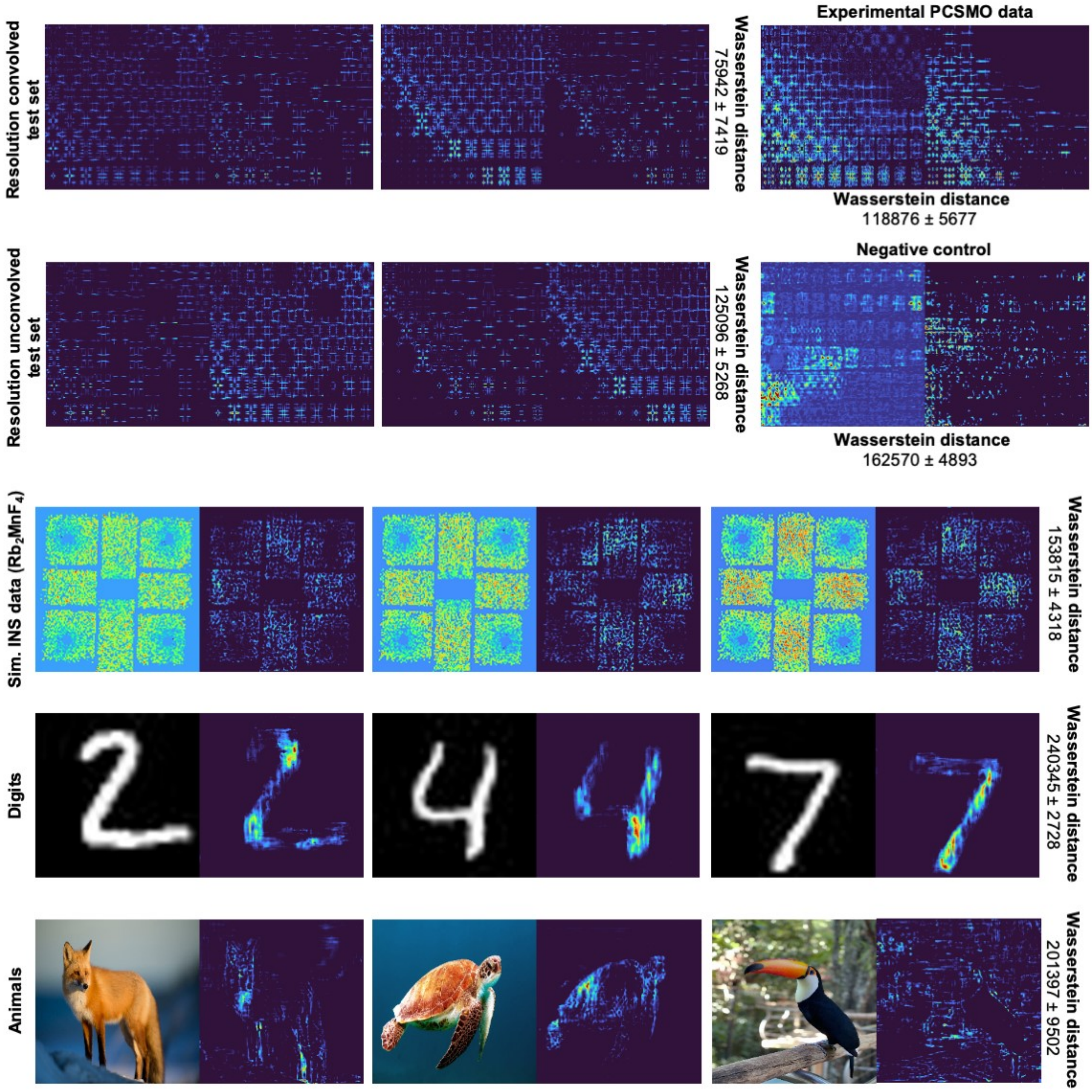


**Figure S2 | $\mathrm{Exp2Sim}_{\mathrm{network}}$ translation of various target distributions to resolution deconvolved INS data and their Wasserstein distances to the training set.** We approximate the Wasserstein distance distribution using sinkhorn distances between various datasets (target distributions) and 20 randomly chosen points from the training set. This process is repeated 1000 times. The Wasserstein distance between the training set distribution to itself is 73168 ± 7803.

## D: Applying Sim2Exp$_{network}$ on a range of simulated and experimental 2D INS spectra

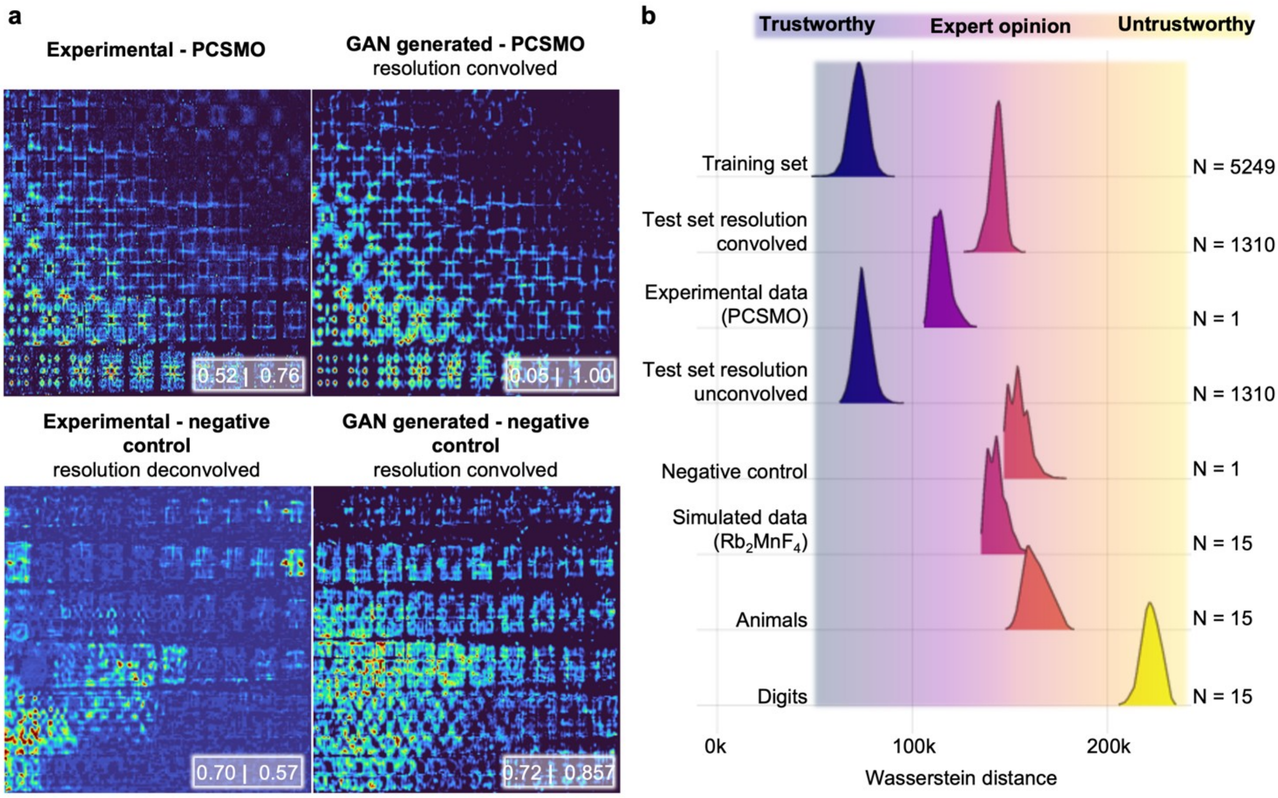


**Figure S3 | Applying Sim2Exp$_{network}$ on a range of simulated and experimental 2D INS spectra.** A) After the network has been trained, it is used to compute a resolution convolution of the experimental INS spectra measured at 4 K on PCSMO (upper panels) and on a dataset that is used as a negative control (lower panels). The negative control dataset is composed of experimental INS spectra measured with the same instrumental settings (and hence instrumental resolution) on various different materials. Note that the experimental axis is the same as in Figure 1a. The insets show the DUQ classifications when trained on GAN-generated resolution convolved INS data.[44] B) The Wasserstein distance of the Sim2Exp$_{Featurespace}$ position has been calculated between various datasets (target distributions) and the Sim2Exp$_{Featurespace}$ position of 20 randomly chosen points from the training set. This process was repeated 1000 times to sample distributions of Wasserstein distances from the target distributions to the training set distribution.

**E: Distribution of simulated inelastic neutron scattering (INS) 3D data**

| | Training set | | Testing set | |
|---|---|---|---|---|
| | Resolution convolved | Resolution unconvolved | Resolution convolved | Resolution unconvolved |
| Goodenough | 768 | 768 | 190 | 190 |
| Dimer | 782 | 782 | 194 | 194 |

**Table S3 | Distribution of simulated inelastic neutron scattering (INS) 3D data.** The numbers denote how many datasets were used for the Exp2SimGAN algorithm.

**F: Evaluating the Exp2SimGAN on simulated 3D INS data from the test set using the Goodenough spin wave model**

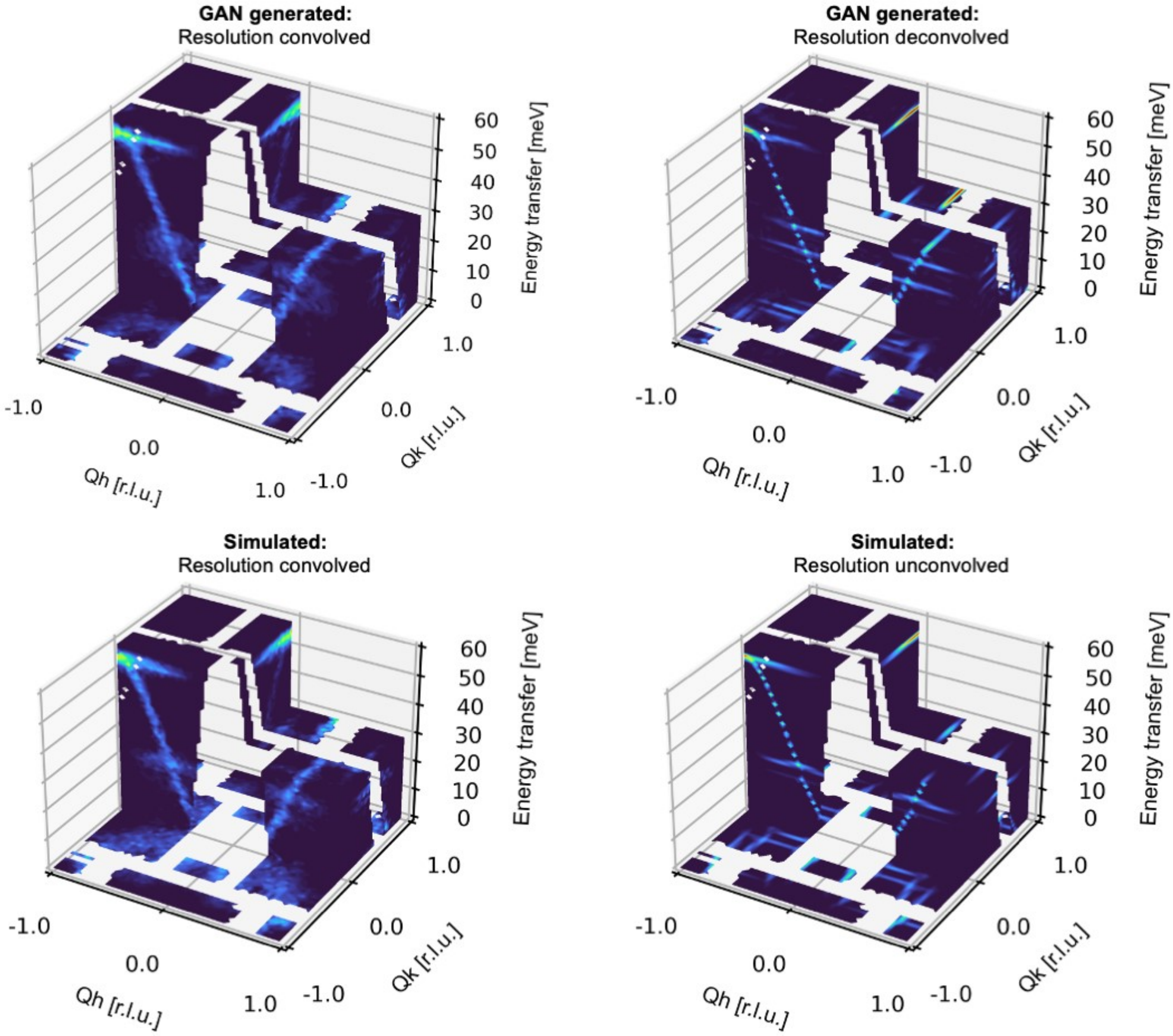


**Fig. S4 | Evaluating the Exp2SimGAN on simulated 3D INS data from the test set.** The INS data is split into 80 % training set and 20 % test set. After the network has trained on the training set, we apply it on the data in the test set. Here is shown an example of performing resolution deconvolution and convolution on 3D INS spectra simulated with the Goodenough spin wave model.