Journal of Fusion Energy (2024) 43:7
https://doi.org/10.1007/s10894-024-00398-8

RESEARCH

Causality Detection and Quantification by Ensembles of Time Delay Neural Networks for Application to Nuclear Fusion Reactors

Michela Gelfusa[1] · **Riccardo Rossi**[1] · **Andrea Murari**[2,3]

Accepted: 4 March 2024 / Published online: 3 April 2024

Abstract

The understanding and control of complex systems in general, and thermonuclear plasmas in particular, require analysis tools, which can detect not the simple correlations but can also provide information about the actual mutual influence between quantities. Indeed, time series, the typical signals collected in many systems, carry more information than can be extracted with simple correlation analysis. The objective of the present work consists of showing how the technology of Time Delay Neural Networks (TDNNs) can extract robust indications about the actual mutual influence between time indexed signals. A series of numerical tests with synthetic data prove the potential of TDNN ensembles to analyse complex nonlinear interactions, including feedback loops. The developed techniques can not only determine the direction of causality between time series but can also quantify the strength of their mutual influences. An important application to thermonuclear fusion, the determination of the additional heating deposition profile, illustrates the capability of the approach to address also spatially distributed problems.

Keywords Causality detection · Time delay neural networks · Nuclear fusion · Additional heating systems

Identification and Quantification of Causal Relations

The observation of regularities between events is an essential cognitive task to understand the physical world. In the time domain, these regularities consist of constant sequences of phenomena, which nowadays are investigated by various mathematical disciplines such as statistics and machine learning [1, 2]. In the last years, artificial intelligence has driven impressive progress in the analysis of time series. An incredible number of tasks, from customer advice to investment decisions and medical diagnosis, is performed entirely or crucially supported by statistical and machine learning tools. And the range of human activities, can benefit from information processing tools is constantly increasing as confirmed recently by ChatGPT success [3].

However traditional artificial intelligence and machine learning tools are not conceived to distinguish between association and causation [4, 5]. On the other hand, the real objective of scientific and engineering studies consists often of determining the causal relations between phenomena. Causality requires a level of analysis deeper than the assessment of statistical dependence. Indeed from a mathematical standpoint, the objective of dependence investigations is to obtain the actual probability density function (pdf) of the data. This is not sufficient for causal analysis, which must guide operational behaviour and in particular provide guidance about the effects of interventions. Basically, the difference between statistical correlation and causality is the one between observing and acting.

Even if determining causal relationships is more difficult than simply calculating correlations, nowadays modern societies and the big physics experiments tend to produce huge amounts of data. Consequently, a lot of potentially very useful information is becoming available to support deeper

Michela Gelfusa and Riccardo Rossi have contributed equally to the paper.

✉ Michela Gelfusa
gelfusa@ing.uniroma2.it

[1] Department of Industrial Engineering, University of Rome "Tor Vergata", Via del Politecnico 1, 00133 Rome, Italy

[2] Consorzio RFX (CNR, ENEA, INFN, University of Padova, Acciaierie Venete SpA), C.so Stati Uniti 4, 35127 Padua, Italy

[3] Istituto per la Scienza e la Tecnologia dei Plasmi, CNR, Padua, Italy

level causality studies. These considerations have motivated the development of techniques aimed at extracting as much knowledge as possible about causal influences from data, a field called Observational Causality Detection (OCD). The developed theories, tools and software packages constitute today a very significant body of knowledge for cross sectional data as reported in various overview papers. On the other hand, techniques for the analysis of the causal interactions between time series have not reached the same level of maturity. Some approaches have been proposed but an organic, general view is not consolidated [6–10].

The goal of the present work consists of showing the potential of Time Delay Neural Networks (TDNNs) and their ensembles to quantify the causal relation between time series [11, 12], and to show how they can be applied to nuclear fusion problems, where time series are a crucial for both physics understanding and control. Their potential is extremely high, because they cannot only shed light on the mutual influences between signals but can also quantify the strength of their causal interactions. The developed TDNN ensembles are very competitive with all the other techniques available and can outperform them significantly, particularly in the most difficult applications.

The use of TDNN ensembles can play a relevant role in several nuclear fusion problems, being thermonuclear plasmas a typical unsteady complex physical system. High temperature plasmas are complex systems that have to be kept well out of equilibrium by the injection of matter and energy [13–15]. Indeed to produce the required rate of fusion reactions they have to reach temperatures higher than the core of the sun. This energy content cannot be achieved with simple ohmic heating and therefore sophisticated additional heating schemes have been developed in the course of the years. The calculation of the local effects of this injected power is a difficult task that requires very sophisticated simulation codes [16–19]. The ensembles of TDNNs could be very useful not only to confirm the results of these simulations but also to provide rapid answers for intershot analysis.

With regard to the organisation of the paper, next section introduces the technology of Time Delay Neural Networks (TDNNs), their ensembles and how they can be refined to investigate causal relationships between time series. In Sect. 3 the potential of TDNN ensembles is substantiated with the help of various numerical tests. The application to the deposition profile in fusion is the subject of Sect. "Validation of the Method with Synthetic Cases" before the conclusions.

Causality Detection with TDNN Ensembles

Causality detection techniques are based on different assumptions. One of the most used class of algorithms for causality detection is based on the definition of causality proposed by Granger [6, 9, 10]: a variable X(t) evolving in time "Granger-causes" another time-evolving variable Y(t) if the past values of X(t) (t-1, t-2, etc.) allows for a better prediction of Y(t) than the prediction without the use of past X(t). So basically Granger causality is a simple statistical hypothesis test for determining whether one time series provides useful information to forecast another. The methods reported in the present work, already analysed with various synthetic cases in [11, 12], are based on this concept.

Adopted Approach to Causality Detection

Consider the following experimental time series: a set of causes $X_1(t), X_2(t), \ldots, X_M(t)$, one new candidate cause $X_C(t)$, and one possible effect $Y(t)$. Given two functions predicting $Y(t)$, one $Y_f(t)$ using also past values of $X_C(t)$ and one $Y_g(t)$ not relying on the information contained in $X_C(t)$:

$$\begin{aligned} Y_f(t) = & f\big(X_1\big(t-1,\ldots,t-N_d\big),\ldots,X_M\big(t-1,\ldots,t-N_d\big), \\ & Y\big(t-1,\ldots,t-N_d\big),X_C\big(t-1,\ldots,t-N_d\big)\big) \end{aligned} \tag{1}$$

$$\begin{aligned} Y_g(t) = & g\big(X_1\big(t-1,\ldots,t-N_d\big),\ldots, \\ & X_M\big(t-1,\ldots,t-N_d\big), \\ & Y\big(t-1,\ldots,t-N_d\big)\big) \end{aligned} \tag{2}$$

The prediction error of the two functions can be written as:

$$\mathrm{E}_f = \frac{1}{N}\sum_{i=1}^{N}\big(Y\big(t_i\big) - Y_f\big(t_i\big)\big)^2 \tag{3}$$

$$\mathrm{E}_g = \frac{1}{N}\sum_{i=1}^{N}\big(Y\big(t_i\big) - Y_g\big(t_i\big)\big)^2 \tag{4}$$

where N is the number of samples in the time series. One may conclude that $X_C(t)$ causes $Y(t)$ if $\mathrm{E}_g > \mathrm{E}_f + \mathrm{E}_{threshold}$, where $\mathrm{E}_{threshold}$ is an uncertainty threshold determined by the error and nature of the time series. The $\mathrm{E}_{threshold}$ can be determined with statistical techniques such as the method of the surrogate data.

A possible and more robust alternative to the just described approach is based on fitting several f_k and g_k functions by using slightly different data ("training set"). So, the

average and the standard deviation of the uncertainties of f_k and g_k are calculated:

$$\overline{\mathrm{E}_f} = \frac{1}{K}\sum_{k=1}^{K}\mathrm{E}_{f_k}; \sigma^2_{E_f} = \frac{1}{K}\sum_{k=1}^{K}\left(\mathrm{E}_{f_k} - \overline{\mathrm{E}_f}\right)^2 \tag{5}$$

$$\overline{\mathrm{E}_g} = \frac{1}{K}\sum_{k=1}^{K}\mathrm{E}_{g_k}; \sigma^2_{E_g} = \frac{1}{K}\sum_{k=1}^{K}\left(\mathrm{E}_{g_k} - \overline{\mathrm{E}_g}\right)^2 \tag{6}$$

In such a situation, a statistical hypothesis test can be performed (based on the difference of the means) and one can also provide the confidence interval of the results (see Fig. 1 for an overall pictorial description of the entire methodology).

Traditional implementations of Granger causality, the basis of the approach followed in the present work, are affected by a couple of quite significant limitations. The most important is that many algorithms assume a priori the functional for of the dependence between the effect and the potential causes. Typically, some sort of autoregressive models are implemented, which work satisfactorily only for linear dependencies [8]. Secondly, the threshold is usually a parameter tuned to maximise the sensitivity and specificity of the causality detection performances. Unfortunately, such a tuning is not fully general. In this work, the use of TDNN allows for a fully general training independent of any a priori assumption, while the ensemble methodology allows for a statistical threshold that does not need any tuning (it just depends of the confidence intervals required by the specific application).

Brief Description of Time Delay Neural Networks (TDNNs)

In the perspective of deriving information about the time evolution of systems and their mutual influence, Time Delay Neural Networks (TDNNs), which constitute a natural extension of traditional feed-forward neural networks [24], have proved to be very useful tools. Indeed, they are explicitly designed to learn from the past. Consequently, the topology of TDNNs is also well suited to the investigation of causal relations between time series.

The architecture of Time Delay Neural Networks is explicitly devised to predict the future of time series on the

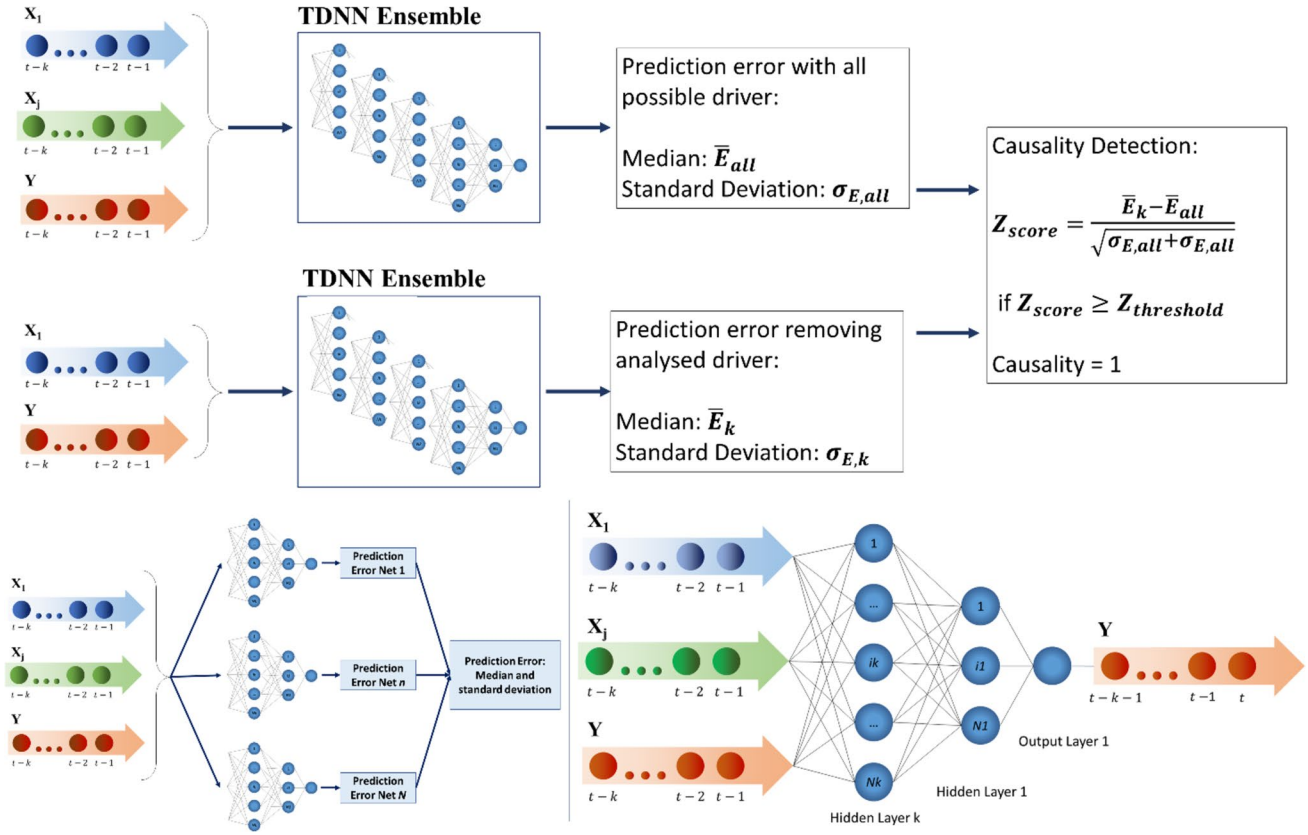

Fig. 1 Top: overview of the entire procedure to investigate the causal coupling between systems. Bottom-left: the architecture of the ensembles. Bottom-right: topology of an individual time delay neural network (particularized for the trivariate case)

basis of their past evolution. This is achieved by providing as inputs to this type of networks sequences of subsequent time points and not single time slices of data. Therefore, the inputs to TDNNs can be conceptualised as windows of length p into the past. In mathematical terms, this type of network implements a non-linear autoregressive model of order p. Given their simple architecture, the training of traditional feed-forward networks can be easily adapted to the topology of TDNNs.

However, the simple architecture of traditional TDNNs is not completely adequate to learn the temporal structure of the data with the aim of assessing their causal relationships. A slightly modified version of TDNNs, shown in the bottom panel of Fig. 1, has therefore been devised and it is the one adopted to perform the studies reported in the rest of the paper. Moreover, individual TDNNs, even if very performing, are not immune to bias when deployed to address involved cases. Therefore, the final technology implemented consist of Ensembles of DNNs, with the topology depicted in the middle panel of Fig. 1. Other approaches to reduce bias, such as repeating the training with different seeds or dropout learning [25], are of course viable alternatives. However, the proposed ensembles provide additional opportunities, such as the use of bagging noise-based ensembles or adaption of random forests [26, 27], which can be very useful in many practical applications.

These ensembles of TDNNs, deployed to obtain the results described in the following sections, have been implemented with the MATLAB toolbox. The training technique is backpropagation using the Levenberg–Marquardt algorithm. About 5000 epochs have proved to be normally sufficient to solve the numerical examples reported (in any case, convergence has also been achieved for less than the maximum limit set to 10,000 epochs and with 200 validation checks).

As mentioned, the concept of causality adopted in the present work is the one of improved predictivity. Consequently, the TDNNs ensembles have been trained and deployed to predict the future evolution of the signal targets. The quality of the predictions is then determined by the residuals, the differences between the predictions and the actual values of the target time series. The residuals of the target quantity are calculated first providing all the variables as inputs to the TDNN ensembles. Then the residuals are computed again but after removing a candidate driver from the input set (see top panel of Fig. 1). The variance of the residuals, the one obtained considering all the inputs, is then compared with the one obtained after removing the specific driver. If the variance, obtained after removing the potential driver, is statistically significantly higher, than the one calculated when the candidate driver was included, then the removed quantity is considered to have a causal influence on the target.

Causality Quantification

In many applications, it is relevant to determine not only whether one variable causes another, but also to quantify the level of influence exerted by one system to another. A very useful indicator to evaluate such information is the error ratio, defined in the previous work as $R_\sigma\left(X_c, Y\right)$:

$$R_\sigma\left(X_c, Y\right) = \frac{\overline{\mathrm{E}_g}}{\overline{\mathrm{E}_f}} \tag{7}$$

Its uncertainty can be easily calculated with simple error propagation analysis:

$$\sigma_{R_\sigma}\left(X_c, Y\right) = R_\sigma\left(X_c, Y\right)\sqrt{\frac{\sigma^2_{E_f}}{\overline{\mathrm{E}_f}^2} + \frac{\sigma^2_{E_g}}{\overline{\mathrm{E}_g}^2}} \tag{8}$$

By using the average value ($R_\sigma\left(X_c, Y\right)$) and its uncertainty ($\sigma_{R_\sigma}\left(X_c, Y\right)$) one may perform a statistical test to determine if $R_\sigma\left(X_c, Y\right)$ is statistically larger than one (and therefore if there is causality between X_c and Y).

The interpretability of such indicator can be easily understood with the help of two examples.

In the first case, let us suppose that the equation governing $Y(t)$ is:

$$Y(t) = h\left(X_1(t-1), \ldots, X_M(t-1), Y(t-1)\right) + wX_C(t-1) + \epsilon \tag{9}$$

where w is constant and ϵ is term associated to noise or random fluctuations. A statistical analysis allows writing:

$$\overline{\mathrm{E}_f} = \epsilon^2; \overline{\mathrm{E}_g} = w^2\sigma^2_{X_c} + \epsilon^2; R_\sigma\left(X_c, Y\right) = \frac{w^2\sigma^2_{X_c}}{\epsilon^2} + 1 \tag{10}$$

Equation (10) lends itself to a clear interpretation: $R_\sigma\left(X_c, Y\right) - 1$ quantifies the amount of influence that X_c exerts on Y, in terms of the variance that flows from X_c to Y (normalised by the variance of the noise affecting Y).

A slightly more general case may be:

$$Y(t) = h\left(X_1(t-1), \ldots, X_M(t-1), Y(t-1)\right) + q\left(X_C(t-1)\right) + \epsilon \tag{11}$$

In this case, simple statistical analysis leads to:

$$R_\sigma\left(X_c, Y\right) - 1 = \frac{\sigma^2_{q(X_c)}}{\epsilon^2} \tag{12}$$

The most general case is expressed by the following equation:

$$\begin{aligned} Y(t) =& h\big(X_1(t-1),\dots,X_M(t-1),\\ & Y(t-1)) + q\big(X_1(t-1),\dots,X_M(t-1),\\ & Y(t-1),X_C(t-1)\big) + \epsilon \end{aligned} \tag{13}$$

Now the indicator $R_\sigma\left(X_c, Y\right)$ can be written as:

$$R_\sigma\left(X_c, Y\right) - 1 = \frac{\sigma^2_{q\left(X_1(t-1),\dots,X_M(t-1),Y(t-1),X_C(t-1)\right)}}{\epsilon^2} \tag{14}$$

In this case, the cause of $X_C(t-1)$ depends on other variables (for example, one may have something like $q\big(X_1(t-1),\dots,X_M(t-1),Y(t-1),X_C(t-1)\big) = wY(t-1)X_C(t-1)$). Since there is not a decoupling between the variables (because causes are coupled), this is a more delicate problem in terms of both analysis and interpretation, and different methodologies based on modelling or knowledge of a priori information is needed. A possible approach is the use of the genetic programming via symbolic regression for time series [5].

Validation of the Method with Synthetic Cases

In this section, some numerical models are introduced to investigate the capabilities of new methodology for both causality detection and quantification. The training, validation and test sets are in the proportion of 75%, 15% and 15% of the overall database.

Analysed Systems

Three different systems, all based on three coupled autoregressive equations, are investigated as case studies.

$$System\,A: \begin{cases} x(t) = 0.95x(t-1) + \sigma(0,0.1) \\ y(t) = 0.95y(t-1) + \sigma(0,0.1) \\ z(t) = 0.95z(t-1) + w_x x(t-1) + w_y y(t-1) + \sigma(0,0.1) \end{cases} \tag{15}$$

In system A, both $x(t)$ and $y(t)$ are independent from external causes and both influence $z(t)$ with an intensity equal to w_x and w_y respectively. The causality is linear and independent, and therefore the amount of influence (variance) of x on z is independent of the influence (variance) of y on z. $\sigma(0,0.1)$ is a random number sampled from a normal distribution with mean equal to zero and standard deviation equal to 0.1.

$$System\,B: \begin{cases} x(t) = 0.5x(t-1) + \sigma(0,0.1) \\ y(t) = 0.5y(t-1) + \sigma(0,0.1) \\ z(t) = 0.5z(t-1) + wx(t-1)y(t-1) + \sigma(0,0.1) \end{cases} \tag{16}$$

In system B, again, both $x(t)$ and $y(t)$ are independent and both exert an influence on $z(t)$. However, in this case their effects are not independent, because their influence depends on their mutual product. Consequently, they have no individual causal influence but they affect y (t) only if they are both present.

$$System C: \begin{cases} x(t) = 0.5x(t-1) + \sigma(0,0.1) \\ y(t) = 0.5y(t-1) + 0.5z(t-1) + \sigma(0,0.1) \\ z(t) = 0.5z(t-1) + w_x x(t-1) + w_y y(t-1) + \sigma(0,0.1) \end{cases} \tag{17}$$

This last system C replicates the same of case A, with the difference that there is a feedback loop between $y(t)$ and $z(t)$. Being able to handled causal relationships with feedback is particularly important, because in nature many complex systems present this type of interactions. It will be shown than in this situation, even if the same coefficient (0.5) is applied to $z(t-1)$ in the equation of $y(t)$, the value of $R_\sigma(y,z)$ will change because of w_y, that in the feedback loop causes a variation of z. The draw and stock diagrams of the three systems are shown in Fig. 2.

Results

The coefficients of causality R_σ as a function of both W_x and W_y are shown in Fig. 3 for system A.

For what concerns the time series x(t) and y(t), they result to be caused only by their past values. In fact, only $R_\sigma(x,x) - 1$ and $R_\sigma(y,y) - 1$ are significantly larger than zero (see $R_\sigma(y,x)$, $R_\sigma(z,x)$, $R_\sigma(y,y)$, $R_\sigma(z,y)$). Moreover, their values are independent of both coefficients W_x and W_y, which model the strength of the causal influence between the two time series.

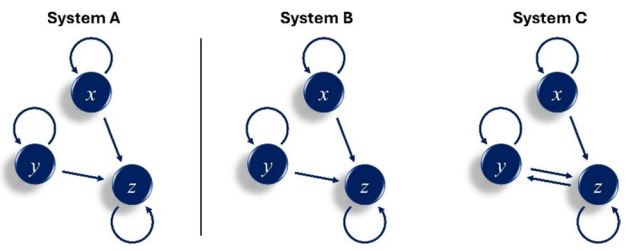

Fig. 2 Pictorial view of the causal relationships for the three numerical cases *A*, *B* and *C*. All systems present inertia whereas only the last one contains also a feedback loop

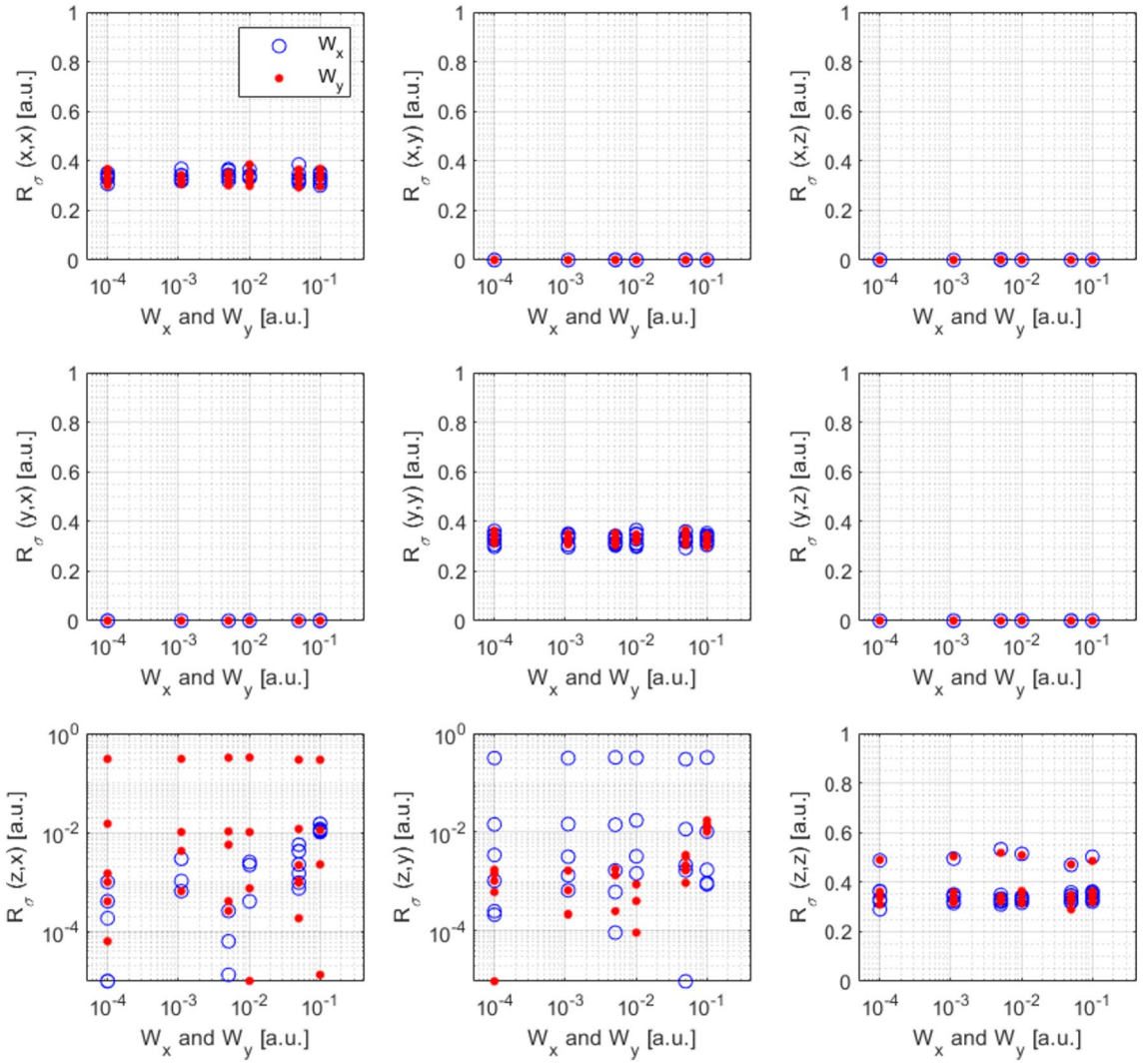

Fig. 3 Coefficients of causality R_σ as a function of W_x and W_y for System A

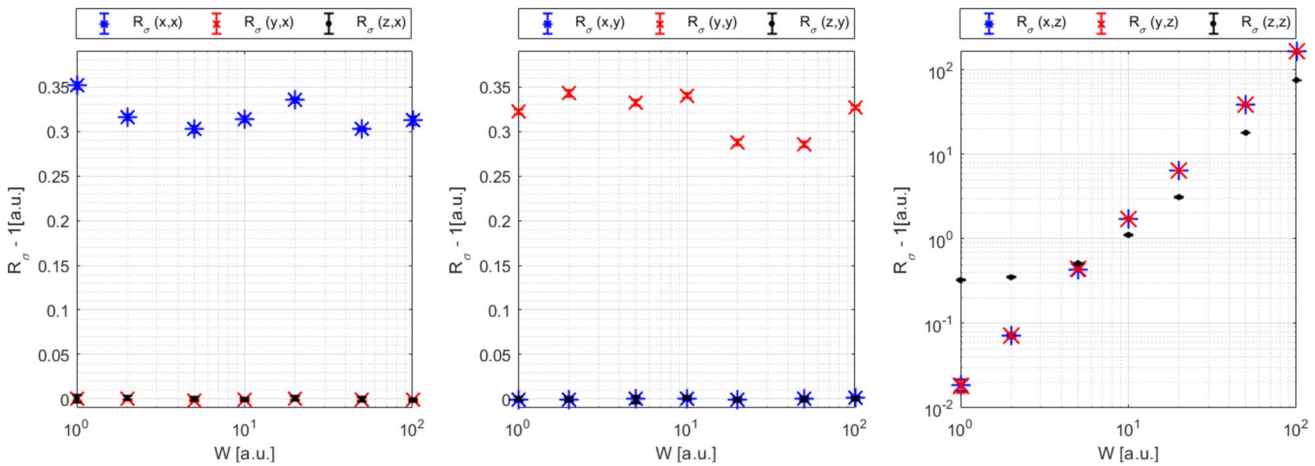

Fig. 4 Coefficients of causality R_σ as a function of W for System B

In the case of time-series z(t), it is observed what expected: $R_\sigma(x,z)$ increases proportionally with W_x, while $R_\sigma(y,z)$ with W_y. No dependences between $R_\sigma(x,z)$ and W_y or between $R_\sigma(y,z)$ and W_x are detected. $R_\sigma(z,z)$ is large and almost independent of the two causal intensities determined by the coefficients W_x and W_y.

The results for system B are summarised in Fig. 4. Again, x(t) and y(t) depend just on their past values respectively. The time series z(t) depends on all values. In this case, both $R_\sigma(x,z)$ and $R_\sigma(y,z)$ increase with W and the causal effects of x(t) and y(t) on z(t) are not separable.

Figure 5 shows the results for System C. The time series x(t) is independent of other observations. High causality values for y(t) are found for both y(t) and z(t). As for system A, $R_\sigma(x,z)$ increases proportionally with W_x, while $R_\sigma(y,z)$ with W_y. However, since the system C is characterised by a feedback loop between y and z, a variation of W_x and W_y directly involves a different causality intensity $R_\sigma(z,y)$, as it is shown

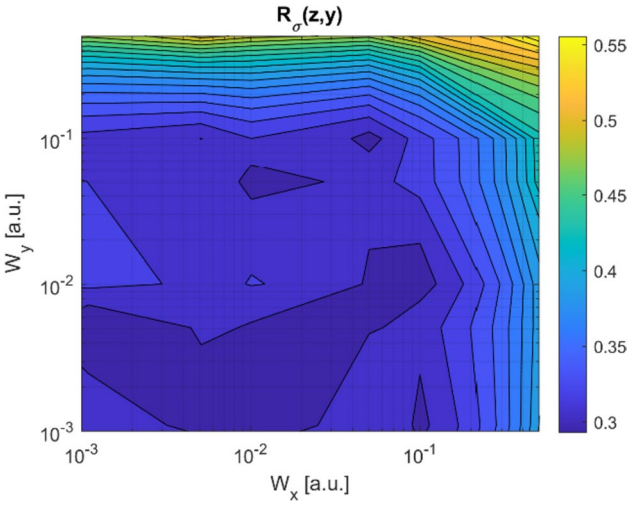

Fig. 6 $R_\sigma(z,y)$ as a function of both W_x and W_y

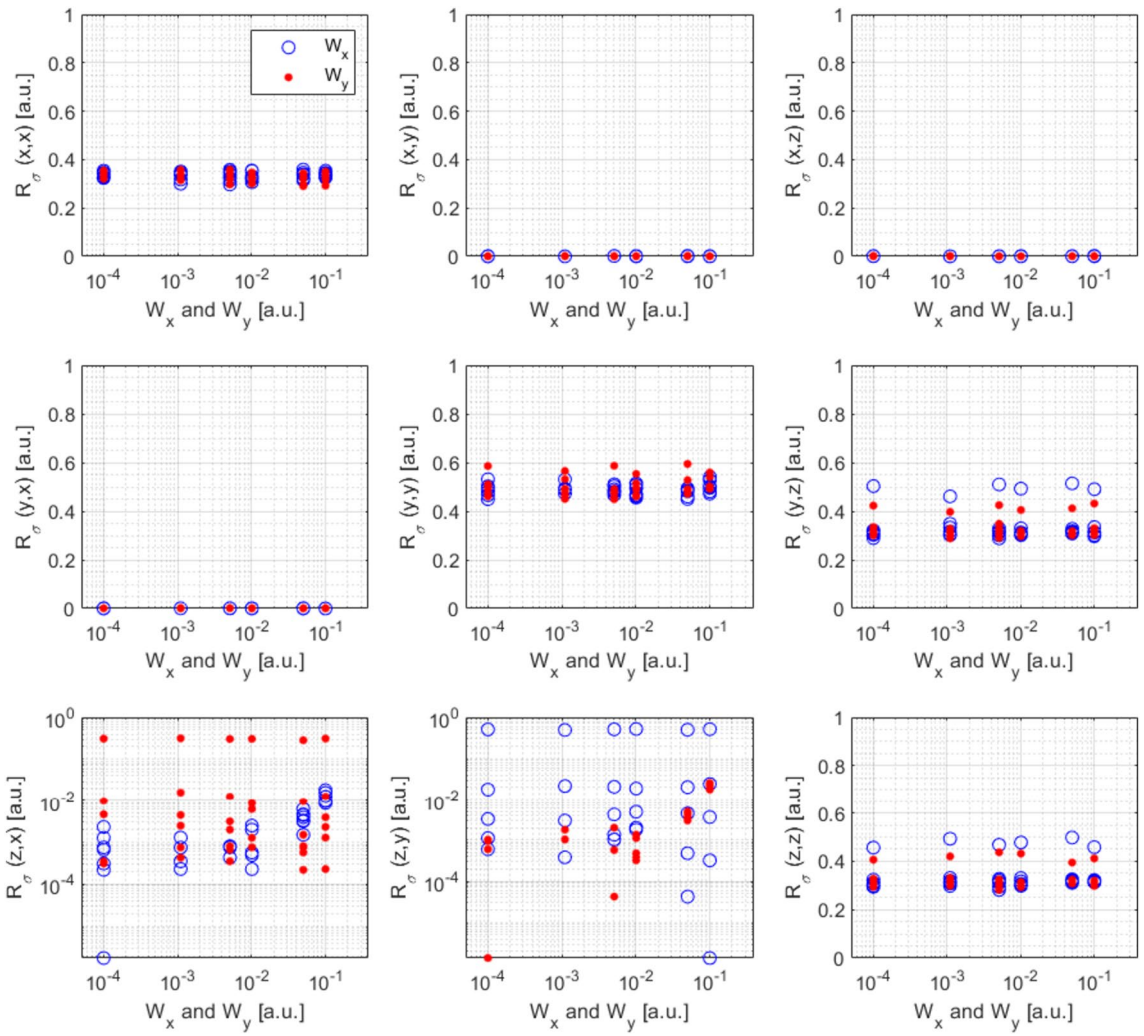

Fig. 5 Coefficients of causality R_σ as a function of W_x and W_y for System C

in Fig. 6. Indeed, even if the past value of z linearly causes y, the feedback loop implies that the cause is not linear.

Nuclear Fusion Case Study: Additional Heating Profile Estimation

An important real-life example of a very complex system, in which the maze of causal relationships is very difficult to disentangle, is the tokamak. In the devices implementing this magnetic configuration, high temperature plasmas are confined for the research on thermonuclear fusion [11]. The main objective consists of achieving conditions, in which the production of energy from the coalescence of light nuclei becomes economically viable [20]. A reactor implementing this concept constitutes one of the most ambitious engineering challenges of present-day societies (see the building of the next generation experiment ITER in the south of France). One particularly delicate aspect is the heating of the plasmas confined by the tokomak field configuration. Indeed, at the temperatures of hundreds of millions of degrees ohmic heating becomes ineffective and other schemes have to be deployed. The research in the field of additional heating techniques for tokamak devices is a very active area [21–23]. Among other aspects, the determination of the power deposition profile is a particularly delicate task, which has significant implications both operational and interpretative. For example, in many studies the properties of the transport are investigated, e.g. using modulated ECRH, while assuming the theoretical deposition profile. This highly problematic, as typically, the experimentally deposition profiles are significantly broader than the theoretical profiles [24]. As the transport partial differential equation contains source, sinks and transport, a wrong assumption on the deposition, will lead to miss-interpretation of the transport. The details of the power deposition inside the plasma is typically investigated with very sophisticated codes, which require human supervision and are computationally quite demanding. There is therefore scope for fast but sufficiently accurate algorithms, capable of determining the effects of the additional heating power on the temperature profile in reasonably short time. The potential of the TDNN ensembles to contribute to this aspect is discussed in the rest of this section. It should be emphasised that the case reported in the following is not a realistic physics study but just a numerical example, to motivate the use of the TDNN technology in more detailed and realistic investigations of the subject.

Model

Let us consider a one-dimensional case where both the temperature (T) and the density (n) are a function of the normalised minor radius coordinate $r_n = r/a$ (a is the minor radius). Moreover, suppose that the density profile is given by the following equation:

$$n(r_n) = n_0\left(1 - {r_n}^{\alpha}\right)^{\beta} \tag{18}$$

where n_0, α, and β are free parameters defining the shape and the value of the electron density.

One may assume that the temperature changes according with the energy conservation law:

$$n(r_n)\frac{\partial T(r_n)}{\partial t} = \Gamma_{heat}(r_n) + \Gamma_{diffusion}(r_n) + \Gamma_{rad}(r_n) \tag{19}$$

where $\Gamma_{heat}(r_n)$, $\Gamma_{diffusion}(r_n)$ and $\Gamma_{rad}(r_n)$ are the heating, diffusive, and radiative terms respectively.

The heating term is assumed to be completely determined by the additional heating systems with a function like:

$$\Gamma_{heat}(r_n) = H_p(r_n)g(t) \tag{20}$$

where $g(t)$ describes how the total input power varies as a function of time, while $H_p(r_n)$ is directly responsible for the local deposition along the minor radius. One of the objectives of the following analysis is the demonstration that the causality score is proportional to the heat deposition function $H_p(r_n)$.

The diffusion term is simply defined as.

$$\Gamma_{diffusion}(r_n) = D_{diff}\frac{\partial^2 T(r_n)}{\partial r^2} \tag{21}$$

where D_{diff} is a diffusion coefficient, which is assumed constant for simplicity sake in the present treatment.

The radiative term is assumed to be dominated the Bremsstrahlung radiation:

$$\Gamma_{rad}(r_n) = C_{Brem}Z^2(r_n)n^2(r_n)T(r_n)^{1/2} \tag{22}$$

where C_{Brem} is a constant and $Z(r_n)$ is the equivalent charge number.

Boundary conditions at the edge just imposed that the temperature at $T(r_n = -1) = T(r_n = 1) = 1eV$.

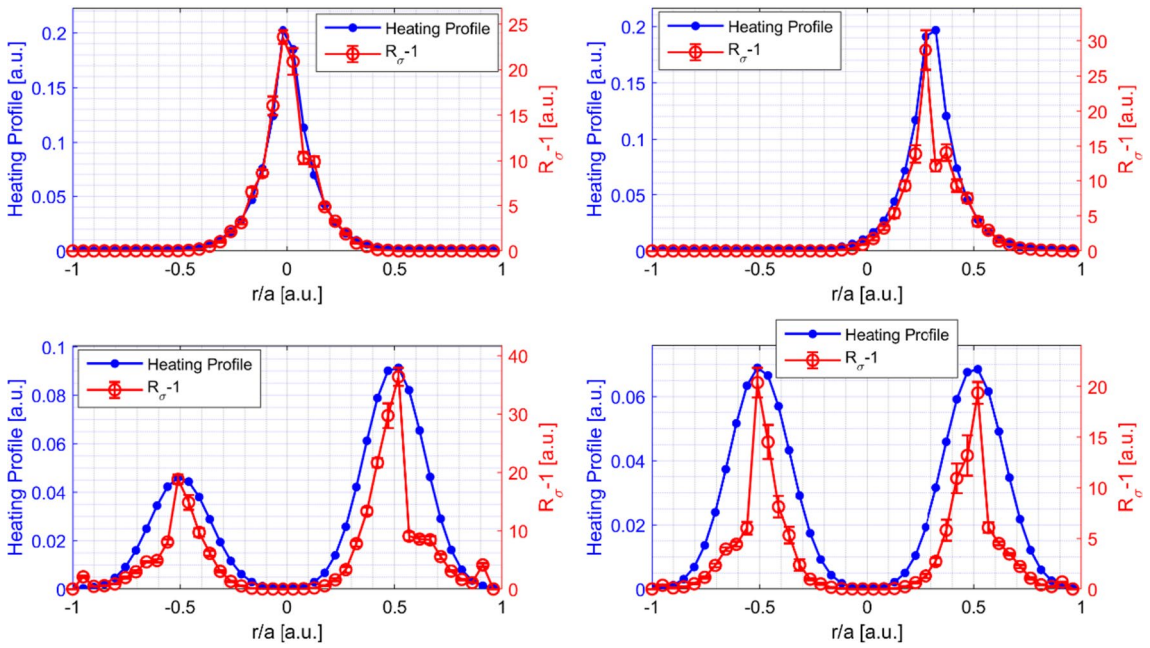

Fig. 7 Power deposition profiles in blue and the indicator $R_\sigma(r_i/a) - 1$ in blue

Results

Figure 7 shows in blue the heating profiles for the four cases analysed. In red the causality intensity indicator $R_\sigma - 1$, quantifying the effect of the input power on the temperature, is reported. The causality intensity is resolved for each position r_i/a, by comparing the two TDNN models:

$$T_f\left(\frac{r_i}{a},t\right) = f\left(T\left(\frac{r_{i-1}}{a},t-1\right),T\left(\frac{r_i}{a},t-1\right),T\left(\frac{r_{i+1}}{a},t-1\right),P_{heating}(t-1)\right) \tag{23}$$

$$T_g\left(\frac{r_i}{a},t\right) = g\left(T\left(\frac{r_{i-1}}{a},t-1\right),T\left(\frac{r_i}{a},t-1\right),T\left(\frac{r_{i+1}}{a},t-1\right)\right) \tag{24}$$

So, $R_\sigma(r_i/a) - 1$ should quantify how much the input power is directly causing a local variation of the local temperature. Four different power deposition profiles have been simulated. They are shown as blue curves in Fig. 7. At first glance, it can be observed that there is a clear proportionality between the causality intensity and the heating profile, which perfectly overlap for case 1 (plot on the top-left). For the other cases, there is a good match between the causality intensity coefficients and the respective heating profiles, even if not perfect.

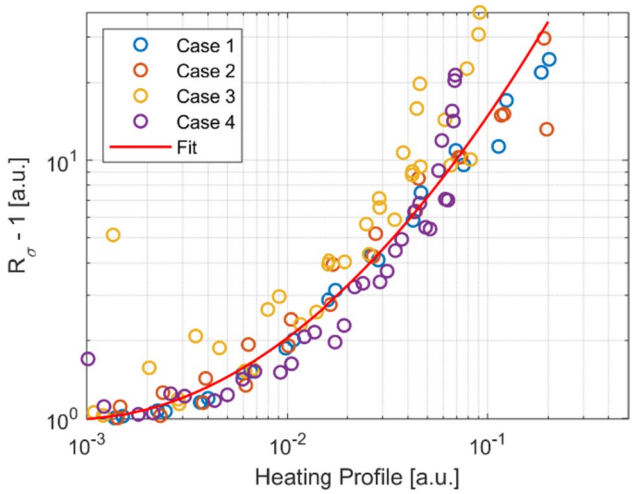

Fig. 8 $R_\sigma - 1$ vs the true value of the heating profile for the four cases. The red line is the result from a 2nd order polynomial function

In Fig. 8, the correlation between the heating profile and the causality intensity is analysed for the four cases. A regression analysis has been performed:

$$\log(R_\sigma - 1) = 0.118\log(H_p)^2 + 1.648\log(H_p) + 5.928 \tag{25}$$

and the result of the fitting is reported in the figure. The adjusted R^2 of the fit is 98.2% (calculated in logarithmic scale, as the rest of the figure).

These differences between causality intensity and heating profiles are due, mostly, to the non-linearity of the phenomenon described here. In fact, a local value of the heating profile can have completely different causality values on the temperature since both diffusion and radiation depend on the temperature themselves. This clearly implies that a measurement of the heating profile can be investigated by using classical methodologies (e.g. numerical simulations) or by using the new branch of artificial intelligence that combines physics with data-driven techniques, the so-called Physics-Informed Machine Leaning [25], which is gaining increasing attention in most branches of the applied sciences.

Conclusions and Future Work

The investigation of complex system requires the development of new analysis tools capable of providing information about the causal relationships between time series and not only about their correlations. Ensembles of TDNNs have proved to be very performing in this respect. Various numerical tests have proved that they can handle quite complex situations, even including feedback loops between the quantities involved. Moreover, they tend to outperform all the main techniques reported in the literature in terms of both flexibility and accuracy. In addition, they are easy to implement and fast to run (Computational times are of course dependent of sample size. For time-series of 1000 samples, the computational time is around 30 min on a common laptop). Another competitive advantage is the capability of TDNNs to quantify the level of mutual influence between time series not only the directionality of the causal relationships.

Application to arguably the most complex laboratory system in big physics, magnetic confinement thermonuclear fusion, has proven the capability of the approach to handle real life data without any major issue. The prospects in this field are particularly promising for both the tokamak [15, 26] and the RFP configuration [27]. Indeed, the approach could be used to address various crucial issues from the control of the total radiation [28, 29] to the prediction of disruptions [30–35]. With regard to the specific aspect of investigating the heating system, very interesting would be the simultaneous analysis of the deposition and the power transport [36]. From a methodological perspective, further work would be necessary to couple TDNNs with symbolic regression in order to extract from the data also the mathematical equations governing the interactions between the systems involved [37, 38].

Author contributions A. M. wrote the main manuscriptR.R and A.M. visualizationR.R. and M.G. building the experimental DBM.G: supervision, verification numerical testR.R. writing the softwareAll authors reviewed the manuscript.

Funding Open access funding provided by Università degli Studi di Roma Tor Vergata within the CRUI-CARE Agreement.

Data availability statement No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. J. Runge, A. Gerhardus, G. Varando, V. Eyring, G. Camps-Valls, Causal inference for time series. Nat. Rev. Earth Environ. **4**(7), 487–505 (2023). https://doi.org/10.1038/s43017-023-00431-y
2. R. Moraffah et al., Causal inference for time series analysis: problems, methods and evaluation. Knowl. Inf. Syst. **63**(12), 3041–3085 (2021). https://doi.org/10.1007/s10115-021-01621-0
3. Q. Wen et al., Transformers in time series: a survey. In: Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, California: International Joint Conferences on Artificial Intelligence Organization, Aug. 2023, pp. 6778–6786. https://doi.org/10.24963/ijcai.2023/759.
4. L.R. Goldberg, The book of why: the new science of cause and effect. Quant Financ. **19**(12), 1945–1949 (2019). https://doi.org/10.1080/14697688.2019.1655928
5. A. Murari, R. Rossi, M. Gelfusa, Combining neural computation and genetic programming for observational causality detection and causal modelling. Artif. Intell. Rev. **56**(7), 6365–6401 (2023). https://doi.org/10.1007/s10462-022-10320-3
6. A. Shojaie, E.B. Fox, Granger causality: a review and recent advances. Ann. Rev. Stat. Appl. **9**(1), 289–319 (2022). https://doi.org/10.1146/annurev-statistics-040120-010930
7. T. Edinburgh, S.J. Eglen, A. Ercole, Causality indices for bivariate time series data: a comparative review of performance. Chaos Interdiscip J Nonlinear Sci (2021). https://doi.org/10.1063/5.0053519
8. A. Krakovská, J. Jakubík, M. Chvosteková, D. Coufal, N. Jajcay, M. Paluš, Comparison of six methods for the detection of causality in a bivariate time series. Phys. Rev. E **97**(4), 042207 (2018). https://doi.org/10.1103/PhysRevE.97.042207
9. D. Marinazzo, M. Pellicoro, S. Stramaglia, Kernel method for nonlinear granger causality. Phys. Rev. Lett. **100**(14), 144103 (2008). https://doi.org/10.1103/PhysRevLett.100.144103

10. C.W.J. Granger, Investigating causal relations by econometric models and cross-spectral methods. Econometrica **37**(3), 424 (1969). https://doi.org/10.2307/1912791
11. R. Rossi, A. Murari, P. Gaudio, On the potential of time delay neural networks to detect indirect coupling between time series. Entropy **22**(5), 584 (2020). https://doi.org/10.3390/e22050584
12. R. Rossi, A. Murari, L. Martellucci, P. Gaudio, NetCausality: a time-delayed neural network tool for causality detection and analysis. SoftwareX **15**, 100773 (2021). https://doi.org/10.1016/j.softx.2021.100773
13. J. Wesson, Tokamaks. Oxford University Press.
14. E. Joffrin et al., Overview of the JET preparation for deuterium–tritium operation with the ITER like-wall. Nucl. Fusion vol. 59, no. 11, (2019), https://doi.org/10.1088/1741-4326/ab2276
15. J. Mailloux et al., Overview of JET results for optimising ITER operation. Nucl. Fusion **62**(4), 042026 (2022). https://doi.org/10.1088/1741-4326/ac47b4
16. M.J. Singh, D. Boilson, A.R. Polevoi, T. Oikawa, R. Mitteau, Heating neutral beams for ITER: negative ion sources to tune fusion plasmas. New J. Phys. **19**(5), 055004 (2017). https://doi.org/10.1088/1367-2630/aa639d
17. E. Lerche et al., Sawtooth pacing with on-axis ICRH modulation in JET-ILW. Nucl. Fusion **57**(3), 036027 (2017). https://doi.org/10.1088/1741-4326/aa53b6
18. D. Gallart et al., Modelling of JET hybrid plasmas with emphasis on performance of combined ICRF and NBI heating. Nucl. Fusion **58**(10), 106037 (2018). https://doi.org/10.1088/1741-4326/aad9ad
19. A. Murari, T. Craciunescu, E. Peluso, E. Lerche, M. Gelfusa, On efficiency and interpretation of sawteeth pacing with on-axis ICRH modulation in JET. Nucl. Fusion **57**(12), 126057 (2017). https://doi.org/10.1088/1741-4326/aa87e7
20. F.F. Chen, *An Indispensable Truth* (Springer, New York, NY, 2011). https://doi.org/10.1007/978-1-4419-7820-2
21. D. Van Eester, E. Lerche, R. Ragona, A. Messiaen, T. Wauters, Ion cyclotron resonance heating scenarios for DEMO. Nucl. Fusion **59**(10), 106051 (2019). https://doi.org/10.1088/1741-4326/ab318b
22. B. Na et al., Experimental and numerical evaluation of the neutral beam deposition profile in KSTAR. Fusion Eng. Des. **185**, 113320 (2022). https://doi.org/10.1016/j.fusengdes.2022.113320
23. J.H. Slief, R.J.R. van Kampen, M.W. Brookman, J. van Dijk, E. Westerhof, M. van Berkel, Quantifying electron cyclotron power deposition broadening in DIII-D and the potential consequences for the ITER EC system. Nucl. Fusion **63**(2), 026029 (2023). https://doi.org/10.1088/1741-4326/acaedc
24. M.W. Brookman, Resolving ECRH deposition broadening due to edge turbulence in DIII-D. Phys. Plasmas (2021). https://doi.org/10.1063/1.5140992
25. R. Rossi, M. Gelfusa, A. Murari, On the potential of physics-informed neural networks to solve inverse problems in tokamaks. Nucl. Fusion **63**(12), 126059 (2023). https://doi.org/10.1088/1741-4326/ad067c
26. M.E. Puiatti et al., Radiation pattern and impurity transport in argon seeded ELMy H-mode discharges in JET. Plasma Phys. Control Fusion **44**(9), 1863–1878 (2002). https://doi.org/10.1088/0741-3335/44/9/305
27. S. Martini et al., Active MHD control at high currents in RFX-mod. Nucl. Fusion **47**(8), 783–791 (2007). https://doi.org/10.1088/0029-5515/47/8/008
28. M. Odstrcil, J. Mlynar, T. Odstrcil, B. Alper, A. Murari, Modern numerical methods for plasma tomography optimisation. Nucl. Instrum. Methods Phys. Res. A **686**, 156–161 (2012). https://doi.org/10.1016/j.nima.2012.05.063
29. A. Murari et al., Investigating the thermal stability of highly radiative discharges on JET with a new tomographic method. Nucl. Fusion **60**(4), 046030 (2020). https://doi.org/10.1088/1741-4326/ab7536
30. J. Vega et al., Disruption prediction with artificial intelligence techniques in tokamak plasmas. Nat. Phys. **18**(7), 741–750 (2022). https://doi.org/10.1038/s41567-022-01602-2
31. A. Murari, M. Lungaroni, M. Gelfusa, E. Peluso, J. Vega, Adaptive learning for disruption prediction in non-stationary conditions. Nucl. Fusion 59(8): 086037(2019)
32. A. Murari et al., Adaptive predictors based on probabilistic SVM for real time disruption mitigation on JET. Nucl. Fusion **58**(5), 056002 (2018). https://doi.org/10.1088/1741-4326/aaaf9c
33. A. Murari et al., On the transfer of adaptive predictors between different devices for both mitigation and prevention of disruptions. Nucl. Fusion **60**(5), 056003 (2020). https://doi.org/10.1088/1741-4326/ab77a6
34. A. Pau, A machine learning approach based on generative topographic mapping for disruption prevention and avoidance at JET. Nucl. Fusion **59**(10), 106017 (2019). https://doi.org/10.1088/1741-4326/ab2ea9
35. R. Rossi, A systematic investigation of radiation collapse for disruption avoidance and prevention on JET tokamak. Matter Radiat. Extr. (2023). https://doi.org/10.1063/5.0143193
36. R. van Kampen, J. de Vries, S. Weiland, M. de Baar, M. van Berkel, Fast simultaneous estimation of nD transport coefficients and source function in perturbation experiments. Sci. Rep. **13**(1), 3241 (2023). https://doi.org/10.1038/s41598-023-30337-0
37. A. Murari, E. Peluso, M. Lungaroni, P. Gaudio, J. Vega, M. Gelfusa, Data driven theory for knowledge discovery in the exact sciences with applications to thermonuclear fusion. Sci. Rep. **10**(1), 19858 (2020). https://doi.org/10.1038/s41598-020-76826-4
38. A. Murari, M. Lungaroni, E. Peluso, T. Craciunescu, M. Gelfusa, A model falsification approach to learning in non-stationary environments for experimental design. Sci. Rep. **9**(1), 17880 (2019). https://doi.org/10.1038/s41598-019-54145-7