ARTICLES
https://doi.org/10.1038/s42256-021-00297-z

nature machine intelligence

Predicting ultrafast nonlinear dynamics in fibre optics with a recurrent neural network

Lauri Salmela[1], Nikolaos Tsipinakis[2], Alessandro Foi[2], Cyril Billet[3], John M. Dudley[3] and Goëry Genty[1]✉

The propagation of ultrashort pulses in optical fibre plays a central role in the development of light sources and photonic technologies, with applications from fundamental studies of light–matter interactions to high-resolution imaging and remote sensing. However, short pulse dynamics are highly nonlinear, and optimizing pulse propagation for application purposes requires extensive and computationally demanding numerical simulations. This creates a severe bottleneck in designing and optimizing experiments in real time. Here, we present a solution to this problem using a recurrent neural network to model and predict complex nonlinear propagation in optical fibre, solely from the input pulse intensity profile. We highlight particular examples in pulse compression and ultra-broadband supercontinuum generation, and compare neural network predictions with experimental data. We also show how the approach can be generalized to model other propagation scenarios for a wider range of input conditions and fibre systems, including multimode propagation. These results open up novel perspectives in the modelling of nonlinear systems, for the development of future photonic technologies and more generally in physics for studies in Bose–Einstein condensates, plasma physics and hydrodynamics.

The past decade has seen major developments in the field of machine learning, and societal applications in healthcare, autonomous vehicles and language processing are becoming commonplace[1]. The impact of machine learning on basic research has been just as noteworthy, and the use of advanced algorithmic tools in data analysis has resulted in new insights into many areas of science. In physics, there has been particular interest applying the tools of machine learning to study dynamical complex systems that evolve in time. These systems exhibit extreme sensitivity to small variations of the governing parameters, and the use of conventional numerical methods to understand and potentially control these dynamics is challenging.

Nonlinear pulse propagation in optical-fibre waveguides is known to exhibit highly complex evolution, and machine learning methods have been applied in a variety of ways to both optimize and analyse their spectrum or temporal intensity profile at the fibre output. For example, from a feedback and control perspective, evolutionary algorithms (which are typically slow to converge) have been used in experiments optimizing particular characteristics of supercontinuum (SC) sources in single-mode[2,3] or multimode fibres[4,5], as well as the experimental control of mode-locked fibre lasers[6–9]. Machine learning using neural networks has also been applied to ultrashort pulse characterization[10,11], in particular to reduce the complexity of real-time measurements of extreme events, and for the classification of different localization regimes in nonlinear modulation instability[12]. Applications to the control of mode locking[10,13] and pulse shaping[14] have also been demonstrated numerically. Yet, all these applications have been restricted either to (slow) genetic algorithms or to feedforward neural network architectures limited to determine the correspondence between a given input and some single output parameter.

More generally, experiments in optical fibres are of very wide interest in nonlinear science as they provide a convenient means of studying the nonlinear dynamics common to many nonlinear Schrödinger equation (NLSE) systems, including hydrodynamics, plasmas and Bose–Einstein condensates. However, because propagation in an NLSE system depends sensitively on both the input pulse and fibre characteristics, the design and analysis of experiments require extensive numerical simulations based on the numerical integration of the NLSE or its extensions. This is computationally demanding and creates a severe bottleneck in using numerical techniques to design or optimize experiments in real time.

In this Article, we present a solution to this problem using machine learning to predict complex nonlinear propagation in optical fibres with a recurrent neural network (RNN), bypassing the need for direct numerical solution of a governing propagation model. The general context of our work is the recent development of machine-learning approaches exploiting knowledge-based and model-free methods to forecast and thus control complex evolving dynamics. Knowledge-based (or physics-informed) methods rely on some a priori knowledge of the mathematical model governing the physical system, and they perform especially well in capturing nonlinear dynamics[15–17]. By contrast, model-free forecasting is a purely data-driven approach where a neural-network structure will learn the system dynamical behaviour from a set of training data, without any prior knowledge of the physics of the system or any underlying governing equation(s). Model-free methods have been particularly successful in forecasting spatiotemporal dynamics of physical systems exhibiting high-dimensional chaos, instabilities and turbulence[18–20], as well as reproducing the propagation dynamics of certain analytical solutions of the NLSE[21].

Our objective here is to extend the use of model-free methods in nonlinear physics by showing how a long short-term memory (LSTM) RNN can accurately reproduce the complex nonlinear dynamics of ultrashort pulse evolution in optical fibre governed by an NLSE system. With such dynamics, the temporal (and spectral) intensity profiles at a particular distance depend on the intensity profile at earlier distance, and this motivates the use of an RNN

[1]Photonics Laboratory, Physics Unit, Tampere University, Tampere, Finland. [2]Laboratory of Signal Processing, Tampere University, Tampere, Finland. [3]Institut FEMTO-ST, Université Bourgogne Franche-Comté CNRS UMR 6174, Besançon, France. ✉e-mail: goery.genty@tuni.fi

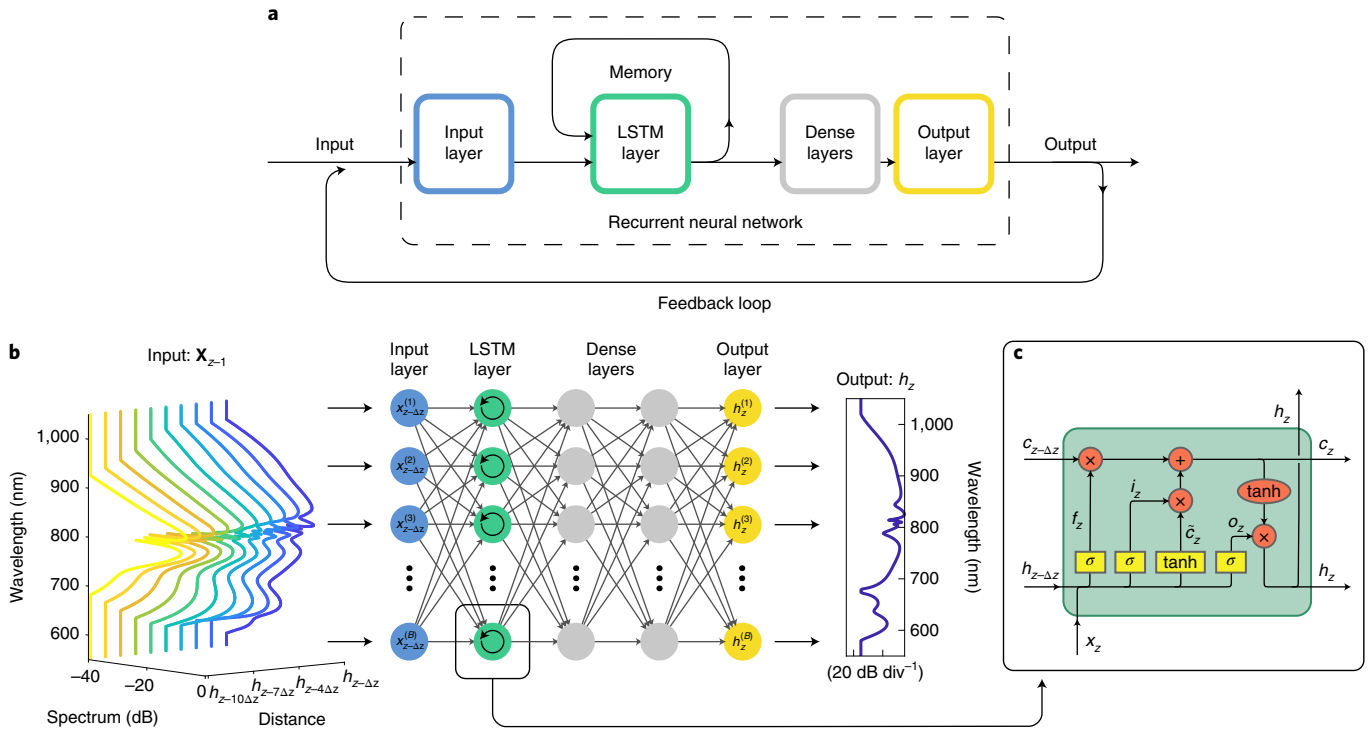


Fig. 1 | Recurrent neural networks. a, Schematic of the RNN architecture used, showing the input layer, the LSTM recurrent layer, two hidden (dense) layers and the output layer. **b**, The neural network uses the spectral (or temporal) intensity profiles $\mathbf{X}_{z-1}$ from the 10 previous intensity profiles $h_{z-10\Delta z}..h_{z-\Delta z}$ in the evolution to yield the subsequent spectrum h_z. Each intensity profile h consists of B intensity bins denoted as x^k, where k indicates the bin number. **c**, The LSTM cell receives the cell input, hidden and cell states from the previous step as an input, and the output of the cell is the new hidden state that is also passed on to the next prediction step along with the new cell state. x_i is the cell input, where $i = z, z-\Delta z$, h_i denotes the hidden state and c_i is the cell state. The yellow rectangles denote layer operations and the orange circles denote pointwise operations. See Methods for more details on the number of nodes used per layer, activation functions and so on. More details and a definition of the different cell elements are provided in the Methods.

with internal memory, an architecture particularly well suited to modelling sequential dependencies. This approach contrasts with previous machine-learning studies of nonlinear pulse propagation, which used only a relatively simple feedforward network architecture to map broadband modulation instability spectra to the local intensity maximum of the corresponding temporal field[12].

Exploiting the memory capacity of RNNs allows modelling of the full nonlinear evolution map of injected short pulses in an optical waveguide in both temporal and spectral domains. We study two particular cases of practical importance: high-power pulse compression associated with the generation of Peregrine-soliton structures, and broadband optical SC generation. In the first case, we demonstrate that the network accurately models the temporal and spectral evolution of higher-order solitons and the appearance of the Peregrine soliton from a transform-limited intensity profile, and we also show how the predicted results agree with reported experimental measurements[22]. We then expand our analysis to even more complex dynamics and show how the network can also predict the full development of an octave-spanning SC with fine details in the spectral and temporal domains.

Finally, we show how the RNN model can be generalized to different input-pulse characteristics and optical-fibre systems to reproduce a wider range of propagation scenarios, including nonlinear dynamics in multimode fibres. These results represent a major extension of the model-free methods applied to nonlinear optics, with potential important impact for high-field physics, nonlinear spectroscopy and precision frequency comb metrology. Moreover, we anticipate that our results will stimulate similar studies in all areas of physics where NLSE-like dynamics play a governing role.

Model-free modelling of nonlinear propagation dynamics

The propagation of light in an optical fibre can be represented as a sequence of electric-field complex amplitude distributions (spectral or temporal) at different points along the propagation path in the fibre. The amplitude at any specific propagation distance is naturally determined by the evolution that precedes it, and modelling this evolution is conventionally carried out by numerically integrating a governing NLSE model over a large number of elementary steps[23]. Unfortunately, this conventional approach can be extremely time-consuming.

Here, we show that such a direct numerical approach can in fact be replaced with model-free forecasting using an RNN. RNNs are a particular class of neural network that possess internal memory, allowing them to account for long-term dependencies and thus to robustly identify patterns in sequential data[24]. The fact that RNNs intrinsically allow modelling of dynamic behaviour makes them particularly adapted to the processing and predictions of time series with applications in speech recognition, predictive texting, handwriting recognition, natural language processing or stock market analysis. They are equally a natural choice to predict the evolution of nonlinear propagation dynamics as a high-power optical field propagates in an optical fibre.

The particular form of RNN we use is the LSTM cell architecture[25]. Although other approaches such as reservoir computing or the gated recurrent unit would also be possible, our choice of LSTM is based on its simplicity of implementation and demonstrated success in various applications[26,27]. We train the network to be able to separately and independently forecast the evolution of temporal and spectral intensity during nonlinear pulse propagation in optical fibre, based only on the initial condition of a transform-limited

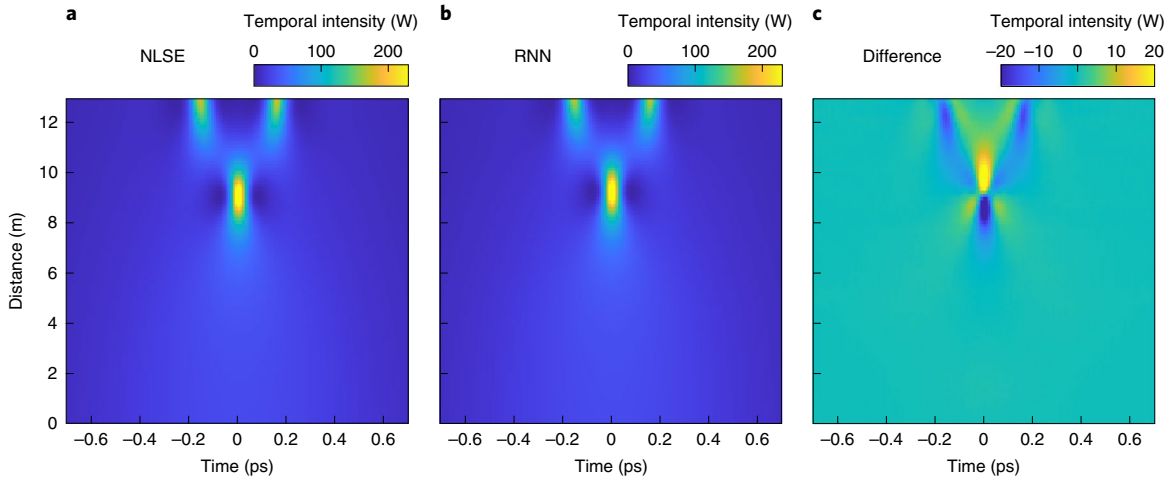


Fig. 2 | Temporal-evolution RNN modelling of higher-order soliton propagation dynamics. a–c, Temporal intensity evolution of a 1.1-ps (FWHM) pulse with peak power of 26.3 W corresponding to an $N=6$ soliton injected into the anomalous dispersion regime of a 13-m-long highly nonlinear fibre: NLSE numerical simulation (**a**), RNN prediction (**b**) and relative difference (**c**). The RNN predictions use only the injected pulse intensity profile as input.

pulse. Of course, physically, the temporal and spectral field characteristics are tightly coupled, and it is therefore remarkable that the network is able to learn, independently, the temporal and spectral evolution dynamics using only intensity data. To teach the network the pulse propagation dynamics, initial training is performed using ensembles of temporal and spectral intensity evolution maps, generated numerically using simulations of the NLSE (or its generalized version, the GNLSE) for a range of input pulse characteristics. To reduce the computational load during training, the simulation profiles are downsampled along both the propagation direction and the temporal and spectral dimensions (Methods).

A general schematic of the RNN is shown in Fig. 1a and an illustration of the training stage is shown in Fig. 1b. Ten consecutive temporal or spectral intensity profiles $h_{z-10\Delta z}..h_{z-\Delta z}$ (that is, the evolution from distance $z-10\Delta z$ to $z-\Delta z$) are fed to the RNN. Here, Δz corresponds to the sampling distance along the propagation direction (Methods). The choice to feed the network with 10 consecutive intensity profiles at propagation interval Δz was found to be a good heuristic compromise between speed and performance (Methods). These intensity profiles are then passed to the LSTM layer consisting of cells (Fig. 1c), governed by a specific algorithm (Methods). The LSTM layer essentially uses three different types of information to predict the (spectral or temporal) intensity profile h_z at distance z: (1) the intensity profile $h_{z-\Delta z}$ at distance $z-\Delta z$, which is the input of the LSTM layer, (2) the hidden state of the layer corresponding to the predicted intensity profile $h_{z-2\Delta z}$ at distance $z-2\Delta z$ and (3) the cell state that contains the long-term dependency information from the intensity profiles $h_{z-10\Delta z}..h_{z-3\Delta z}$ corresponding to the evolution from distance $z-10\Delta z$ to $z-3\Delta z$.

The output of the LSTM layer is subsequently fed to a fully connected feedforward neural network with two hidden (dense) layers whose function is to further improve the predicted intensity at distance z. The prediction made by the RNN (output layer) is compared with the intensity profile from the NLSE (or GNLSE) simulations. The error is backpropagated to the weights and biases of the network nodes (both dense and LSTM layers), which are subsequently adjusted to minimize the prediction error. The RNN cycle is then initiated again with an updated input consisting of the consecutive temporal or spectral intensity profiles $h_{z-9\Delta z}..h_z$ until the full evolution is predicted. Note that the RNN loop is initiated with a 'cold start' where the input sequence contains only the spectral or temporal intensity profile of pulses injected into the fibre (replicated 10 times). In the prediction phase, the RNN model is tested using a separate set of temporal and spectral evolution data that was not used in the training phase.

Results

Higher-order soliton compression. We begin by training the RNN to model the propagation of picosecond pulses in the anomalous dispersion regime of a highly nonlinear fibre. This propagation regime is of particular significance as it is associated with extreme self-focusing dynamics and practical 'higher-order soliton' pulse compression schemes[23]. Moreover, the dynamics of this nonlinear temporal compression have been shown, recently, to be associated with the emergence of the celebrated Peregrine soliton, which appears in the semiclassical limit of the NLSE[22].

The training data were generated by performing 3,000 NLSE numerical simulations of propagation in 13 m of fibre using initial conditions of transform-limited hyperbolic-secant input pulses. The fibre parameters were kept constant between simulations and corresponded to experiments performed around 1,550 nm (ref. [22]). On the other hand, we varied the pulse duration $\Delta\tau$ (full-width at half-maximum, FWHM) and peak power P_0 uniformly over ranges of 0.77–1.43 ps and 18.6–34.2 W, respectively. This yields a variation in soliton number $N=\sqrt{\gamma P_0 T_0^2/|\beta_2|}$ from 3.5 to 8.9 (γ and β_2 are the fibre nonlinear and group velocity dispersion parameters respectively, and $T_0=\Delta\tau/1.763$). Further details are provided in the Methods.

We first illustrate the results obtained when training the network to model the temporal intensity evolution. Figure 2 compares the evolution of the temporal intensity simulated using the NLSE (Fig. 2a) with that predicted by the RNN (Fig. 2b). The particular results shown correspond to an input soliton number $N=6$. One can see the overall excellent visual agreement between the propagation dynamic predicted by the RNN and those simulated from the NLSE. Also notice that the distance of maximum compression and associated temporal intensity profile are particularly well predicted by the RNN. Figure 2c shows the relative difference between the NLSE and RNN evolution maps, with a root-mean-square (r.m.s.) error, computed over the full evolution, of $R=0.04$ (Methods). Comparisons between NLSE and RNN evolution for 100 different input conditions spanning the full range of parameter variation showed similar results, with an r.m.s. error computed over the 100 evolution maps of $R=0.097$ (Methods).

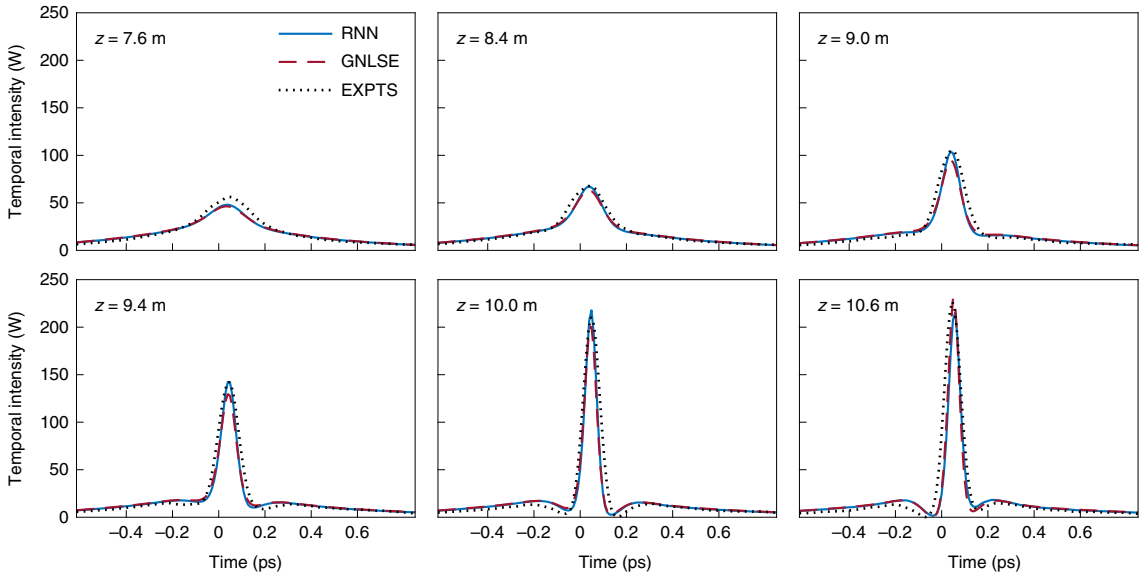


Fig. 3 | Experimental validation of higher-order soliton temporal dynamics RNN modelling. Higher-order soliton ($N=6$) temporal intensity at selected distances z predicted by the neural network (solid blue lines), simulated with the NLSE (dashed red lines), and experimentally measured (black dotted lines). Experimental data are from ref. [22].

A more detailed comparison between the NLSE simulations, RNN prediction and experimental measurements at selected distances is plotted in Fig. 3. For this case, note that third-order dispersion was also included in the training simulations (Methods). The figure shows the intensity profiles predicted by the RNN (solid blue line), the profiles from the NLSE simulations (dashed red line), as well as the experimental measurements (black dotted line) previously reported in ref. [22]. One can see remarkable agreement at all distances between the three sets of results, and we stress in particular that the RNN reproduces both the compressed central portion and the side lobes of the Peregrine soliton associated with maximal compression around 10 m.

We also tested the ability of the RNN to predict the propagation dynamics in the spectral domain from the corresponding input spectrum. Here, we use the same ensemble of NLSE numerical simulations as for the temporal evolution, but this time we train the network by feeding the spectral intensity evolution. Results for input conditions identical to those of Figs. 2 and 3 are shown in Fig. 4. For convenient visualization, the evolution is plotted in logarithmic scale. The spectral evolution consists of an initial stage of spectral broadening dominated by self-phase modulation and corresponding to the compression observed in the time domain. After the point of maximum expansion, we see a breathing phase of narrowing and re-expansion typical of higher-order soliton propagation. One can see excellent agreement between the dynamics predicted from the network and that simulated with the NLSE, with a relative discrepancy of less than 5 dB over the entire evolution (r.m.s. error computed over the full spectral evolution, $R=0.039$). Note, here, that the training data are fed to the network in logarithmic scale, which reduces the noise level in the prediction arising from the minimization of the r.m.s. error function during the training phase (Methods). Of course, it is perfectly possible to train the network using linear spectra, but this yields a higher noise floor in the predicted evolution, as illustrated in the Extended Data Fig. 1.

The excellent correspondence is confirmed in Fig. 5 when plotting a detailed comparison between the RNN predicted (solid blue line), simulated (dashed red line) and experimentally measured spectra (black dotted line) at selected distances z around the maximal temporal compression point, as previously considered and which is naturally also the point of maximum spectral broadening. In particular, one can see the excellent agreement between the NLSE and RNN results over a dynamic range of 40 dB. We also performed a series of tests for 100 different input pulse spectra spanning the full range of parameter variation and found similar network performances in terms of predicted evolution, with an r.m.s. error of $R=0.042$ (computed over the 100 evolution maps tested).

Supercontinuum generation. We next extended our study to even more complex propagation dynamics and the generation of a broadband SC. Here, we focus our attention on a SC generated by injecting femtosecond pulses into the anomalous dispersion of a highly nonlinear fibre. This regime is of particular significance as it has been shown to be associated with high spectral coherence and the generation of stable frequency combs as well as yielding the broadest SC spectra[28].

To test whether a recurrent neural network could learn SC generation dynamics and model their evolution, we generated an ensemble of SC propagation dynamics using the GNLSE, which includes the frequency dependence of dispersion and nonlinearity, as well as the delayed Raman response[28]. Specifically, we simulated the propagation of 100-fs transform-limited pulses at 810 nm injected into the anomalous dispersion regime of a 20-cm-long photonic-crystal fibre with zero dispersion at 750 nm, similar to that used in ref. [12]. Detailed parameter values are provided in the Methods. The ensemble includes simulations for a transform-limited input pulse with peak power uniformly distributed in the 500 W to 2 kW range that yields SC spectra with different characteristics, from isolated dispersive wave generation to a fully developed octave-spanning SC with very fine spectral features. We emphasize that, although the input pulse duration was kept constant for all simulations, the predicted results for other durations show similar agreement with the GNLSE, as in the specific cases discussed below.

We begin by training the network from the temporal intensity evolution. Similarly to the higher-order soliton compression case, the simulation profiles are downsampled along both the propagation direction and the temporal and spectral dimensions (Methods). After training, the RNN model is tested for an input peak power not used in the training stage and the predicted evolution is compared with that directly simulated with the GNLSE for the same input power.

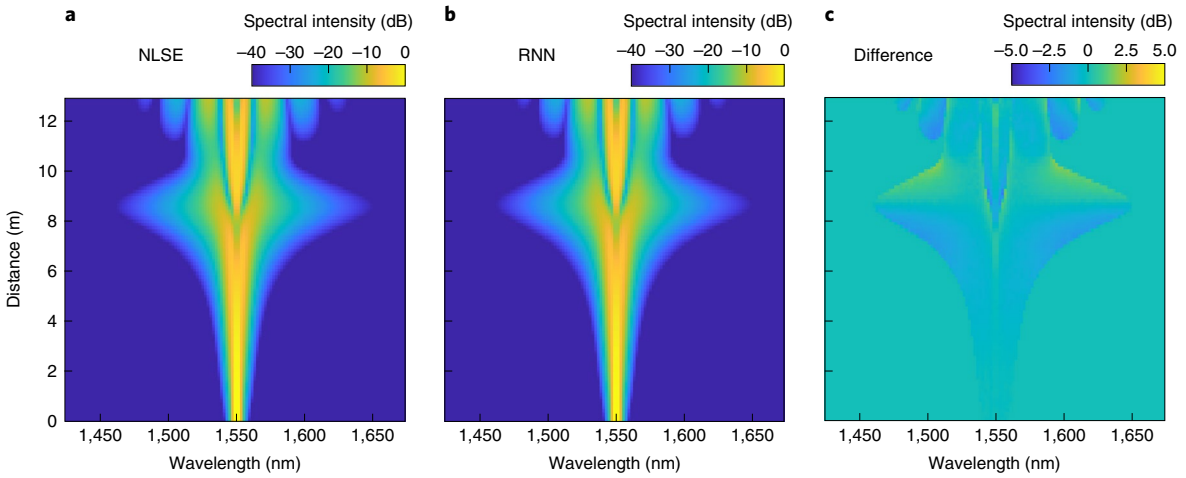


Fig. 4 | Spectral evolution RNN modelling of higher-order soliton propagation dynamics. a–**c**, Spectral intensity evolution of a 1.1-ps (FWHM) pulse with peak power of 26.3 W corresponding to an $N=6$ soliton injected into the anomalous dispersion regime of a 13-m-long highly nonlinear fibre: the results of numerical simulation (**a**), RNN prediction (**b**) and relative difference (**c**). The RNN predictions use only the injected pulse spectrum as input.

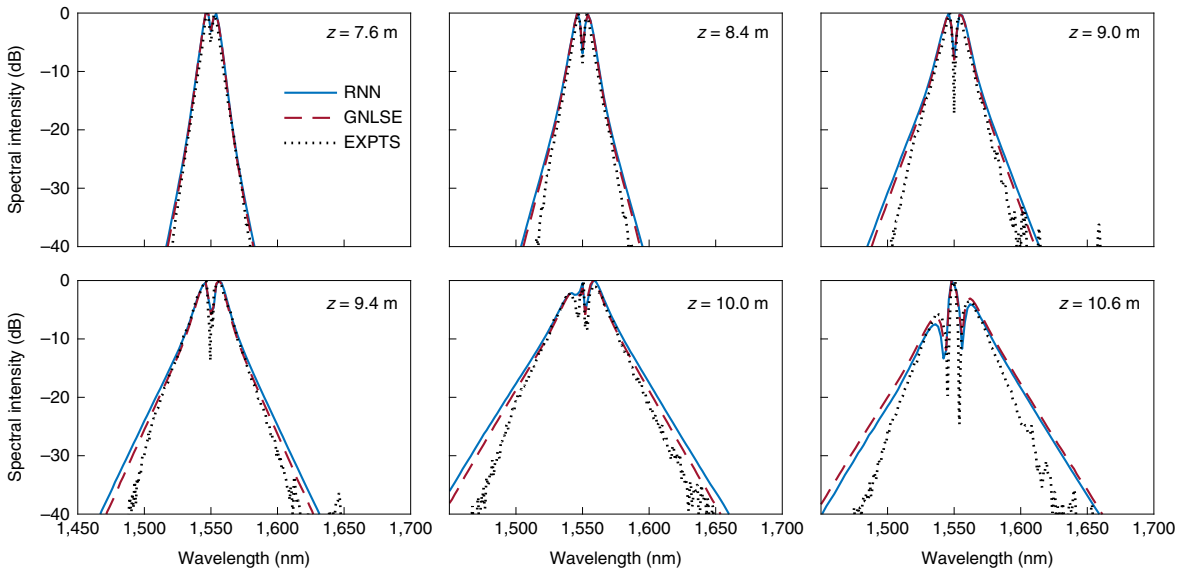


Fig. 5 | Experimental validation of higher-order soliton spectral dynamics RNN modelling. Higher-order soliton ($N=6$) spectral intensity at selected distances z predicted by the neural network (solid blue lines), simulated with the NLSE (dashed red lines), and experimentally measured (black dotted lines).

Results are shown in Fig. 6a,b for input peak powers of 630 W and 1.96 kW, corresponding to input soliton numbers of $N=4.6$ and $N=8.1$, respectively. These values were chosen as they lead to a SC with very distinct characteristics. The left panel shows the temporal intensity evolution from the GNLSE simulation and the central panel shows the evolution predicted by the RNN. The SC generation process arises from soliton dynamics including higher-order soliton compression, soliton fission and dispersive waves emission on the short wavelength side[28]. For longer propagation distances, solitons emerging from the fission experience a Raman self-frequency shift, expanding the SC spectrum towards the long wavelength side[28]. Importantly, in both scenarios, one can see excellent visual agreement between the GNLSE simulations and RNN model. The point of soliton fission and dispersive emission as well as the red-shifting solitons' parabolic trajectories are perfectly reproduced by the network. The r.m.s. errors calculated over the full intensity evolution are $R=0.097$ and $R=0.049$ for Fig. 6a,b, respectively. The remarkable ability of the RNN to predict very complex nonlinear dynamics is further highlighted in the right panels, in which we plot a detailed comparison between the predicted and simulated SC temporal intensity at selected distances along the propagation, where we can see how the amplitude and delay of the dispersive waves and Raman-shifted solitons are also predicted with excellent accuracy at all stages of the propagation. Additional predictions run for 50 different values of pulse peak power (not used in the training phase) also showed very good agreement with the GNLSE simulations (r.m.s. error of $R=0.176$ computed over 50 different evolution maps tested).

We then tested the ability of the RNN model to predict the SC spectral intensity evolution from the input pulse spectrum. The results for an input peak power of 630 W and 1.96 kW are shown in Fig. 6c,d, respectively. For convenient visualization, the evolution is plotted in logarithmic scale. In the case of lower peak power, one can see that the SC spectrum at the fibre output essentially consists

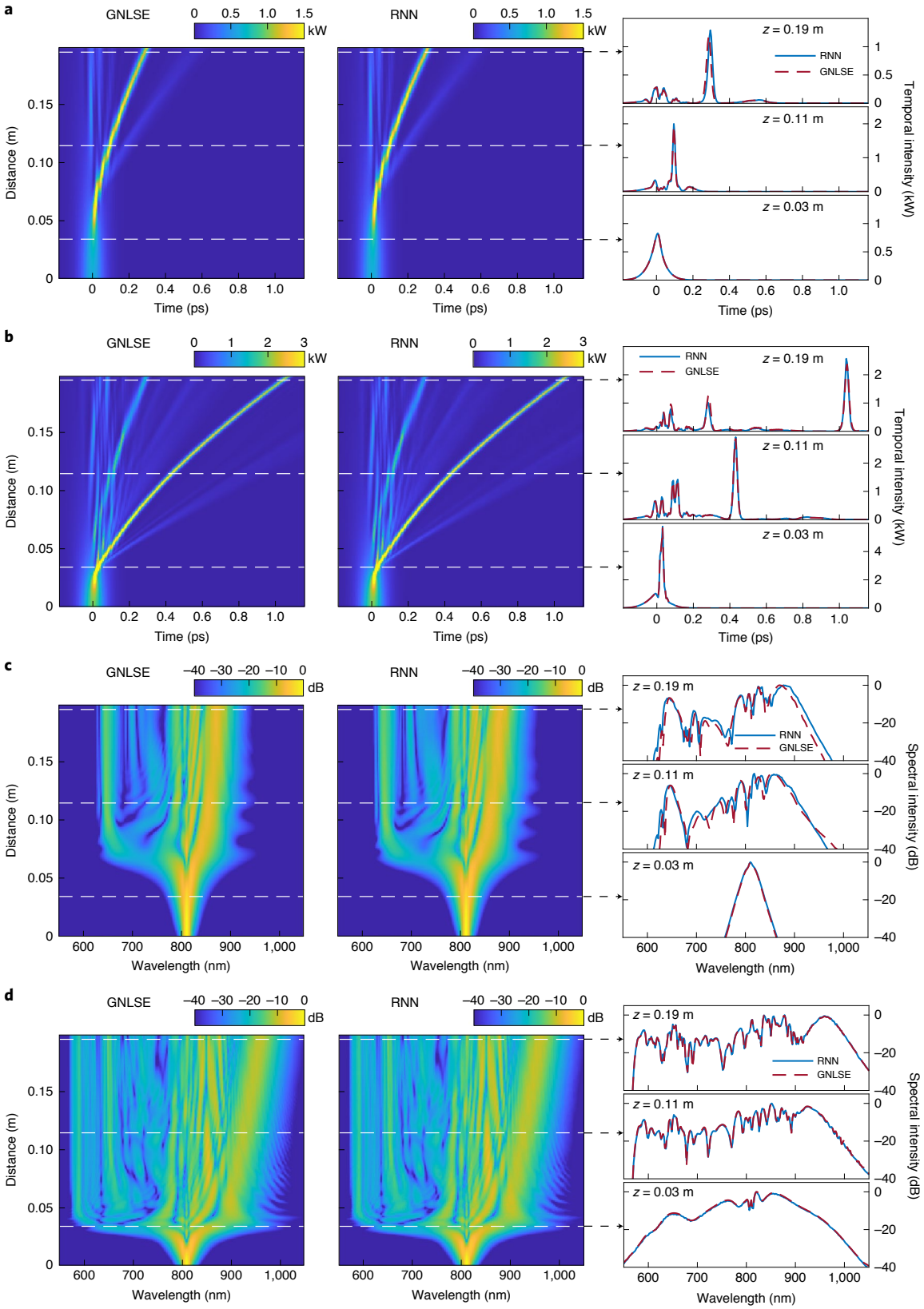


Fig. 6 | Temporal and spectral evolution RNN modelling of a supercontinuum. a–d, Temporal (**a**,**b**) and spectral (**c**,**d**) evolution RNN modelling of the SC. Left panels: numerical simulation (GNLSE) of SC evolution in a 20-cm photonic-crystal fibre for a 100-fs pulse with peak power of 630 W (**a**,**c**) and 1.96 kW (**b**,**d**). See Methods for a full description of the fibre parameters. Middle panels: predicted (RNN) temporal intensity evolution for the same initial temporal intensity profile as in the GNLSE simulations. Right panels: comparison between the predicted (solid blue lines) and simulated (dashed red lines) profiles at selected distances *z* (indicated by white dashed lines in the left and middle panels).

of an isolated dispersive wave and solitons with a limited amount of redshift. For larger input peak power, we see multiple dispersive wave emission and well-separated Raman-shifted solitons resulting in an octave-spanning SC. Again, we can see very good visual agreement between the simulated and predicted evolution maps, and all spectral features—including dispersive waves, Raman-shifted solitons and their interference, which lead to fine spectral features—are perfectly reproduced by the RNN. Additional predictions run for 50 different input pulse peak power values (not used in the training phase) also showed very good agreement with the GNLSE simulations (r.m.s. error, $R = 0.09$).

Generalization

The results above have focused on the specific cases of the Peregrine soliton dynamics that appear during the initial compression stage of transform-limited picosecond pulses, and SC generation in the femtosecond regime. For these examples we have shown that the RNN can learn and model evolution maps for a range of different peak powers. More generally, however, nonlinear dynamics are not only sensitive to input peak power, but also to other characteristics, including, for example, input pulse duration and phase (chirp) as well as the fibre (or waveguide) parameters (dispersion, nonlinearity). This is particularly true for soliton compression and SC generation where, for example, a change in pulse duration or dispersion profile can lead to dramatically different evolution maps and output spectral and temporal profiles.

From the perspective of potential applications, the ability to generalize the use of RNN modelling of evolution dynamics over a wider range of propagation scenarios is especially important. A natural question is then how the neural network architecture trained for a particular system can be 'transferred' to a different system. To develop a more generic model that can reproduce nonlinear dynamics over a large range of input conditions and system parameters, the RNN can be trained from evolution maps simulated using the normalized form of the (G)NLSE (Methods), while adding the particular characteristics of interest (input duration, chirp, dispersion, nonlinearity) as additional input parameters to the RNN. The propagation dynamics for a particular set of characteristics can then be reproduced using normalized variables. Using this approach, the Extended Data figures show a series of examples demonstrating the ability of the network to model dynamics for various additional propagation scenarios including: higher-order soliton propagation of the NLSE for different pulse durations (Extended Data Fig. 2) and chirp values (Extended Data Fig. 3), and the SC dynamics of the GNLSE for input pulses with different duration, peak power and fibre dispersion (Extended Data Fig. 4).

To mimic realistic experimental conditions, we also tested the ability of the network to predict the nonlinear evolution dynamics over a wide range of initial conditions in the presence of multiplicative noise and found that the network still performs well for a noise level in excess of 20% (Methods), as shown in Extended Data Fig. 5.

Finally, we emphasize that the RNN architecture can be adapted to model multidimensional systems such as nonlinear propagation in multimode fibres. For example, one can feed the network with evolution maps corresponding to different injection conditions (Methods), which can be learned and reproduced by the network as illustrated in Extended Data Fig. 6. These particular results could extend the operation range and speed up the optimization and control of spatiotemporal instabilities in multimode systems[4,5]. A similar approach is also adapted to model polarization-dependent propagation in a single-mode fibre.

Discussion

As seen from the results presented above, the propagation dynamics of short pulses are complex and highly nonlinear. For practical application purposes, for example in pulse compression, optical sensing and high-resolution and real-time imaging, optimizing these dynamics generally requires extensive and computationally demanding integration of the GNLSE. From the results above it is clear that, once trained, the RNN is able to substitute itself for the integration of the GNLSE and can predict evolution maps in excellent correspondence with the GNLSE simulations. An important advantage of the RNN is then the greatly reduced computation time required to produce evolution maps, as shown in Extended Data Fig. 7. In particular, unlike the numerical integration of the GNLSE, the RNN computation time is independent of the number of points in the temporal and spectral grid, such that, for a large number of evolution maps, the discrepancy between GNLSE and RNN simulations can exceed two orders of magnitude (9,655 s for 1,000 GNLSE simulations with 8,192 temporal grid points using MATLAB v2020a running on a 2.5 GHz dual-core Intel Core i7 processor versus 36 s for the 1,000 RNN predictions in Python using Keras running on an NVIDIA Quadro K620 graphics processing unit, GPU). The difference is even more marked when considering multimode fibre simulations. For example, 1,000 realizations of multimode propagation simulations for different coupling conditions and including five modes with 4,096 grid points takes nearly 95 h (more than 5 min each), but, once trained, the RNN prediction for 1,000 distinct input coupling conditions only takes ~30 s.

In general, the GPU and central processing unit (CPU) resources needed for training the RNN will vary depending on the size of training data, number of temporal/spectral grid points, the network hyperparameters and characteristics of the computer used (processor and memory), but it is useful to examine them in more detail to illustrate the benefit of the RNN model. For the particular computer used in this work, these resources are summarized in Table 1 for all the cases illustrated in the different figures. One can clearly see that, besides significantly reduced computation time, there are also other advantages associated with the RNN, including reduced resources, dimensionality reduction and the ability to construct dynamics from sparse data, resulting in important memory saving (for example, 1,000 GNLSE evolution maps with 8,192 temporal grid points use over 8 GB of random access memory, whereas the RNN hyperparameters use a mere 5.6 MB). Such time and memory saving shows the potential of RNNs to overcome the bottlenecks encountered in designing and optimizing experiments in real time, allowing us to identify particular boundary conditions to operate in a specific dynamical regime, determine optimum injection conditions for a given system configuration, adjust the input pulse characteristics to maintain stable operation or generate tailored spectral (or temporal) profiles at specific propagation distances.

Conclusions

We have shown that machine-learning techniques can bring new insight into the study and prediction of nonlinear optical systems. Specifically, we have demonstrated that an RNN with LSTM can learn the complex dynamics associated with the nonlinear propagation of short pulses in optical fibres, including higher-order soliton compression and SC generation using solely the pulse intensity profile as input condition. The network is also able to reproduce dynamics in both the temporal and spectral domains, and for the particular case of higher-order soliton compression we have been able to confirm that the predicted evolution maps are also in excellent agreement with experiments. Previous applications of machine learning to ultrafast dynamics have been restricted to slow genetic algorithms or feedforward neural networks designed to establish the transfer function between specific input–output parameters[6,9,10,14,29]. Although useful from a purely application perspective, these approaches are nevertheless not capable of capturing nonlinear dynamics. Our results using an RNN are thus particularly notable considering the extremely complex nature of the temporal and spectral evolution, and they show how a network with advanced

Table 1 | Comparison of GPU/CPU and RAM resources used during the (G)NLSE simulations and RNN training

(G)NLSE	CPU (%)	RAM (GB)[a]	Points	Steps	Real.			
HOS compression	400	4.96	1,024	10,000	3,000			
SC generation	400	8.52	2,048	10,000	1,300			
Normalized NLSE	400	0.84	512	10,000	1,000			
Chirped NLSE	400	19.9	2,048	10,000	6,000			
Normalized GNLSE	400	79.4	8,192	9,000	12,000			
Multimode GNLSE	400	6.58	4,096	5,000	1,000			
RNN	**CPU (%)**	**GPU (GB)**	**RAM (GB)**	**NN (MB)**	**Var.**	**Bins**	**Steps**	**Real.**
HOS (temporal)	192	1.24	4.54	2.25	278k	151	101	3,000
HOS (spectral)	194	1.24	3.98	2.09	258k	126	101	3,000
SC (temporal)	134	1.24	6.70	9.79	1.22M	276	200	1,300
SC (spectral)	157	1.24	6.17	5.56	691k	251	200	1,300
Norm. NLSE								
(temporal)	143	1.24	2.94	7.44	926k	256	101	1,000
(spectral)	165	1.24	1.91	4.32	537k	128	101	1,000
Chirped NLSE	135	1.24	13.9	7.53	938k	256	101	6,000
Norm. GNLSE	161	1.24	7.76	2.49	609k	132	51	12,000
MMGNLSE	116	1.24	3.51	18.5	2.31M	301	50	1,000

[a]The instantaneous memory usage can reduced by running the simulations in batches or by convolving the spectral/temporal intensity profiles during the simulations. Results are shown for higher-order soliton (HOS) compression, supercontinuum (SC) generation, normalized NLSE, chirped NLSE, normalized GNLSE and multimode GNLSE (MMGNLSE) cases. The number of bins, propagation steps and realizations (Real.) used during training as well as the number of variables (Var.) and memory utilized by the RNN hyperparameters (NN) to reproduce the nonlinear dynamics evolution are also indicated. CPU usage of greater than 100% reflects the distribution of tasks among the four cores of the processor.

architecture is able to learn and reproduce the essential physics of nonlinear propagation.

From an experimental perspective, it is well recognized that the study of distance-dependent nonlinear propagation in waveguides is difficult, as it requires extensive cut-back measurements in incremental steps, and it is precisely here where a network trained on simulations and performing accurately within the typical range of experimental parameters would be extremely valuable. We expect that neural networks will very soon become an important and standard tool for analysing complex ultrafast dynamics, for optimizing the generation of broadband spectra and frequency combs, as well as for designing ultrafast optics experiments. Future work may also expand the parameter space of the RNN operation by including additional training variables or extending the predictions to the complex field (amplitude and phase).

Ultrafast laser beams are multidimensional systems and the extent to which their spatial, temporal and spectral properties can be controlled is central to achieving optimum performance in adaptive optics and coherent control. Although there are various pulse-shaping technologies that can be applied to tune these characteristics, the optimization process can be particularly tedious, generally involving multiple parameters that are interconnected in complex ways, and heuristic or manual approaches are often used. By enabling more systematic strategies, this is a particular area of application where the use of an RNN capable of modelling the propagation dynamics associated with a wide range of input parameters could enable an unprecedented level of control, with particular applications, for example, in optimized nonlinear pulse compression schemes, the generation of tailored broadband spectra, or spatial beam profiles with particular temporal (spectral) intensity distributions. For example, one could use the RNN to identify specific initial conditions that yield a desired output (spectrum, temporal and spatial profiles) for a particular fibre system or, given particular input conditions, to determine the optimum propagation distance to generate the output of interest. In another scenario, one could also consider using the injected and measured spectral (or temporal) profiles in a waveguide or fibre to characterize its dispersion and nonlinear coefficient.

From a more fundamental perspective, although, here, we have focused our study on optical pulse propagation, the approach introduced is very generic and can be applied to other physical systems, such as hydrodynamic waves or Bose–Einstein condensates where, under particular conditions, the dynamics are governed by nonlinear Schrödinger equations. For example, one could use an RNN architecture to model the propagation of water waves and identify specific initial conditions that lead to the observation of particular dynamics in a controlled environment such as water tanks. More generally, we believe that the use of RNNs will impact the future design and analysis of nonlinear physics experiments as they represent a natural candidate for exploring and analysing complex operation regimes with long-term dependencies.

Methods

Numerical simulations. The numerical simulations in this work are from the NLSE and its generalized extension (1 + 1D), which describe the propagation of the slowly varying optical field envelope.

Higher-order soliton compression. We model the propagation of short pulses in the anomalous dispersion regime of a 13-m nonlinear optical fibre. The pulses have a hyperbolic-secant intensity profile centred at 1,550 nm, with pulse duration and peak power varying from 0.77 to 1.43 ps and from 18.41 to 34.19 W, respectively. The nonlinear coefficient of the fibre is $\gamma = 18.4 \times 10^{-3}\,\mathrm{W^{-1}\,m^{-1}}$ and the group velocity dispersion coefficient at 1,550 nm is $\beta_2 = -5.23 \times 10^{-27}\,\mathrm{s^2\,m^{-1}}$. When comparing with the experiments, third-order dispersion ($\beta_3 = 4.27 \times 10^{-41}\,\mathrm{s^3\,m^{-1}}$) was also included in the training in addition to a small input pulse asymmetry caused by the experimental implementation[22]. The simulations use 1,024 spectral/temporal grid points with a temporal window size of 10 ps and step size of 0.13 mm (10,000 steps). For completeness, shot noise is added via one-photon-per-mode with random phase in the frequency domain, although noise effects were found to play no major physical role in the regime of coherent propagation studied here.

Supercontinuum generation. We model the propagation of a sech-type pulse centred at 810 nm and with a pulse duration of 100 fs. The peak power of the input pulse is

randomly varied in the range 0.5–2 kW. The pulses are injected in the anomalous dispersion regime of a 20-cm nonlinear optical fibre, including higher-order dispersion terms, self-steepening and the Raman effect. The nonlinear coefficient of the fibre is $\gamma = 0.1\,\mathrm{W^{-1}\,m^{-1}}$ and the Taylor-series expansion coefficients of the dispersion at 810 nm are $\beta_2 = -9.59\times10^{-27}\,\mathrm{s^2\,m^{-1}}$, $\beta_3 = 7.84\times10^{-41}\,\mathrm{s^3\,m^{-1}}$, $\beta_4 = -6.84\times10^{-56}\,\mathrm{s^4\,m^{-1}}$, $\beta_5 = -4.78\times10^{-70}\,\mathrm{s^5\,m^{-1}}$, $\beta_6 = 2.71\times10^{-84}\,\mathrm{s^6\,m^{-1}}$ and $\beta_7 = -5.00\times10^{-99}\,\mathrm{s^7\,m^{-1}}$. The simulations use 2,048 spectral/temporal grid points with a temporal window size of 5 ps and step size of 0.02 mm (10,000 steps). Shot noise is added via one-photon-per-mode with random phase in the frequency domain, but in the coherent propagation regime studied here, noise effects were found to play no major physical role.

Normalized NLSE. To generalize the range of input pulse characteristics and fibre parameters over which the RNN can predict higher-order soliton dynamics, we model the propagation of pulses using the normalized form of the NLSE:

$$\mathrm{i}\frac{\partial\psi}{\partial\xi} - \frac{\mathrm{sgn}\,(\beta_2)}{2}\frac{\partial^2\psi}{\partial\tau^2} + |\psi|^2\psi = 0 \tag{1}$$

where $\tau = T/T_0$, $\xi = z|\beta_2|/T_0^2$ and $\psi(\xi,\tau) = NA(z,T)/\sqrt{P_0}$ are the normalized time, propagation distance and amplitude, respectively, and $A(z,T)$ represents the dimensional temporal envelope of the field. The input pulses are taken to have a hyperbolic-secant intensity profile centred at 1,550 nm and the soliton number N is varied randomly from 1 to 7. The simulations use 512 temporal grid points with temporal window size of 10 (normalized units) and the fibre length is set to $\pi/8$ (corresponding to a quarter of the soliton period $\xi_{\mathrm{sol}} = \pi/2$) with 10,000 steps. The pulses are injected in the anomalous dispersion regime of the NLSE.

We performed another set of simulations where a linear frequency chirp was included as an additional input parameter to the network and corresponding to an input pulse amplitude of the form $\psi(0,\tau) = N\,\mathrm{sech}\,(\tau)\exp(\mathrm{i}\,C\tau^2/2)$, where N is the soliton number, τ is normalized time and C is the chirp parameter. The soliton number N and chirp parameter C were varied randomly in the intervals 1 to 7 and −5 to +5, respectively. The simulations use 2,048 spectral grid points with temporal window size of 40 (normalized units) over a normalized fibre length of $\pi/8$ split into 10,000 steps.

Normalized GNLSE. To extend the applicability of the RNN predictions to a wider range of SC scenarios, we modelled SC generation dynamics with the normalized form of the GNLSE:

$$\mathrm{i}\frac{\partial\psi}{\partial\xi} - \frac{\mathrm{sgn}\,(\beta_2)}{2}\frac{\partial^2\psi}{\partial\tau^2} - \mathrm{i}\frac{q}{6}\frac{\partial^3\psi}{\partial\tau^3} + \left(1 + \mathrm{i}s\frac{\partial}{\partial\tau}\right)\left(\psi\int_{-\infty}^{+\infty} r(\tau')|\psi(\tau-\tau')\xi|^2\,\mathrm{d}\,\tau'\right) = 0 \tag{2}$$

where $\tau = T/T_0$, $\xi = z|\beta_2|/T_0^2$ and $\psi(\xi,\tau) = NA(z,T)/\sqrt{P_0}$, $q = \beta_3/(|\beta_2|T_0)$, $s = 1/(\omega_0 T_0)$ and r are the normalized time, propagation distance, amplitude, third-order dispersion, shock-term and Raman response, respectively. We assume transform-limited hyperbolic-secant pulses centred at 835 nm in the anomalous dispersion regime (that is, $\mathrm{sgn}(\beta_2) = -1$). The soliton number, pulse duration and third-order dispersion parameter were randomly varied in the intervals 2–8, 30–130 fs and 1–9, respectively. The simulations used 8,192 temporal grid points with temporal window size of 350 (normalized units) and normalized distance set to 2 with 9,000 steps.

Multimode GNLSE. We model the nonlinear propagation of hyperbolic-secant pulses with peak power of 2.5 MW and pulse duration of 150 fs (FWHM) at 1,500 nm in a 25-cm multimode step-index silica fibre using the multimode GNLSE[30]. The fibre was assumed to have a 50-μm core radius and core-cladding refractive index difference of $\Delta n = 0.0137$. We consider the propagation in five transverse modes ($\mathrm{LP_{01}}$, $\mathrm{LP_{11}}$, $\mathrm{LP_{02}}$, $\mathrm{LP_{12}}$ and $\mathrm{LP_{03}}$) whose initial coupling conditions are randomly varied while keeping the injected power constant. The simulations use 4,096 spectral/temporal grid points with a temporal window size of 6 ps and a step size of 0.05 mm (5,000 steps).

Recurrent neural networks. *LSTM network operation.* The operation of an LSTM cell can be described at time step t with input $\mathbf{x}_t \in \mathbb{R}^{d_o}$ by a set of equations given by[25]

$$\begin{aligned} \mathbf{f}_t &= \sigma(\mathbf{W}_f[\mathbf{h}_{t-1},\mathbf{x}_t] + \mathbf{b}_f) & \mathbf{i}_t &= \sigma(\mathbf{W}_i[\mathbf{h}_{t-1},\mathbf{x}_t] + \mathbf{b}_i) \\ \tilde{\mathbf{c}}_t &= \tanh(\mathbf{W}_c[\mathbf{h}_{t-1},\mathbf{x}_t] + \mathbf{b}_c) & \mathbf{c}_t &= \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t \\ \mathbf{o}_t &= \sigma(\mathbf{W}_o[\mathbf{h}_{t-1},\mathbf{x}_t] + \mathbf{b}_o) & \mathbf{h}_t &= \mathbf{o}_t \odot \tanh(\mathbf{c}_t) \end{aligned} \tag{3}$$

where $\mathbf{f}_t$, $\mathbf{i}_t$ and $\mathbf{o}_t \in \mathbb{R}^{d_h}$ are the forget, input and output gate vectors, respectively, with d_h denoting the dimensionality of the hidden state (that is the number of hidden units). Vectors $\mathbf{c}_t$ and $\mathbf{h}_t \in \mathbb{R}^{d_h}$ are the updated cell and hidden state, respectively, and $\mathbf{W}_f$, $\mathbf{W}_i$, $\mathbf{W}_c$ and $\mathbf{W}_o \in \mathbb{R}^{d_h\times(d_h+d_o)}$ represent the cell weights and $\mathbf{b}_f$, $\mathbf{b}_i$, $\mathbf{b}_c$ and $\mathbf{b}_o \in \mathbb{R}^{d_h}$ are the biases. The sign $\odot$ denotes pointwise multiplication. The weights and biases of the network are iteratively trained via backpropagation[31].

Feedforward network operation. The operation of the fully connected layers is similar to that in ref. [12]. The codes were written in Python using Keras[32] with the Tensorflow backend[33].

Comparison between RNN prediction and (G)NLSE simulations. A quantitative comparison between the network predicted evolution map and that simulated with the (G)NLSE can be performed using the average (normalized) r.m.s. errors as a metric:

$$R = \sqrt{\frac{\sum_{i,d}(x_{m,i,d} - \hat{x}_{m,i,d})^2}{\sum_{i,d}(x_{m,i,d})^2}} \tag{4}$$

where $\mathbf{x}_m$ and $\hat{\mathbf{x}}_m$ denote the (G)NLSE and RNN predicted intensity profile for realization m. The variables i and d indicate summation over the intensity profiles and propagation steps, respectively. When evaluating the performance of the prediction over an ensemble of M evolution maps, the r.m.s. error is calculated over M distinct realizations.

Higher-order soliton compression. An ensemble of 3,000 numerical simulations was generated. A total of 2,900 realizations are used for the training of the RNN and 100 unseen realizations are used for testing. The simulated intensity evolution maps are uniformly downsampled at a constant propagation step of $\Delta z = 0.13$ m, yielding 101 intensity profiles along propagation for each simulated evolution map. At every one of the 101 steps, the intensity profile is convolved and downsampled with a 10 fs FWHM super-Gaussian temporal filter corresponding to 145 equally spaced bins in the $[-0.7, +0.7]$ ps time interval. The spectral intensity profiles are convolved and downsampled with a 2 nm FWHM super-Gaussian spectral filter resulting in 126 equally spaced intensity bins spanning from 1,425 to 1,675 nm. The temporal and spectral intensity profiles are normalized by the peak intensity over all realizations. From an experimental viewpoint, commonly employed devices to characterize short pulses, such as autocorrelators and spectrometers, only yield measurements related to the temporal and spectral intensities, respectively. Full-field characterization requires the use of more advanced techniques based on some form of phase retrieval or measurement such as frequency-resolved optical gating, and we therefore choose to only use transform-limited intensity profiles during the RNN training while the phase information is omitted. Note that the network is trained with intensity profiles in linear and logarithmic scale for the temporal and spectral intensity intensity evolution, respectively, as these are the scales commonly used for plotting these quantities. See Extended Data Fig. 1 for a comparison between the linear and logarithmic training results. When comparing with the experiments, to account for the slight input pulse asymmetry, the NLSE simulated intensity profiles of every map used in the training phase of the RNN were convolved and downsampled with a 10 fs FWHM super-Gaussian temporal filter corresponding to 151 equally spaced bins in the $[-0.62, +0.85]$ ps time interval. The spectral intensity profiles were convolved and downsampled similarly to the case of ideal higher-order soliton propagation but spanning from 1,450 to 1,700 nm.

The LSTM and two hidden layers consist of 161 nodes, each with rectified linear unit (ReLU) activations $f(x) = \max(0,x)$, and the output layer consists of 151 and 126 nodes for temporal and spectral predictions, respectively, with sigmoid activation $f(x) = 1/[1+\exp(-x)]$. The network is trained for 60 and 120 epochs with RMSprop optimizer[34] and adaptive learning rate for the temporal and spectral intensity predictions, respectively.

The input of the RNN consists of 10 consecutive temporal or spectral intensity profiles $h_{z-10\Delta z}..h_{z-\Delta z}$ at a distance along the fibre of $z-10\Delta z$ to $z-\Delta z$.

A smaller number of intensity profiles was also found to give satisfactory results, but of course this is at the expense of the relative prediction error, which increases from 0.097 to 0.174 (temporal intensity evolution of higher-order soliton) when reducing the number of consecutive intensity profiles from 10 to 5. As the number of consecutive intensity profiles used in the training is increased, the training time also increases and therefore the training process is always a compromise between prediction accuracy and the time required to train the network.

Supercontinuum generation. An ensemble of 1,300 numerical simulations was generated. A total of 1,250 realizations were used for training the RNN and 50 realizations for testing. The simulated intensity evolution maps are uniformly downsampled at a constant propagation step of $\Delta z = 0.2$ mm, yielding 200 intensity profiles along propagation for each simulated evolution. To reduce the computational load, when training the RNN to predict temporal intensity maps, the profiles at each of the 200 steps are convolved and downsampled with a 10 fs FWHM super-Gaussian temporal filter corresponding to 276 equally spaced bins spanning the $[-0.18, +1.16]$ ps time interval. Note that the asymmetry in the modelled time interval is implemented to account for the soliton self-frequency shift effect. When training the RNN from spectral intensity profiles, each spectrum is convolved and downsampled with a 2 nm FWHM super-Gaussian spectral filter such that the wavelength grid consisted of 251 spectral intensity bins spanning from 550 to 1,050 nm. The profiles are normalized by the peak intensity over all realizations and the spectral intensity profiles are converted to logarithmic scale.

For the temporal intensity evolution, the LSTM and two hidden layers consist of 300 nodes each with ReLU activations, and the output layer consists of 276 nodes with sigmoid activation. The network was trained for 120 epochs with RMSprop optimizer and an adaptive learning rate. For the spectral intensity evolution, the LSTM and two hidden layers consist of 250 nodes each with ReLU activations, and the output layer consists of 251 nodes with sigmoid activation. The network is trained for 100 epochs.

Normalized NLSE. Here, we used an ensemble of 1,000 numerical simulations, and 950 and 50 realizations were used for training and testing the RNN, respectively. The simulated intensity evolution maps are uniformly downsampled at a constant propagation step $\Delta\xi = 0.004$ (normalized units), yielding 101 intensity profiles along propagation for each simulated evolution. The temporal intensity profiles consist of 256 equally spaced bins spanning the $[-2.5, +2.5]$ (normalized units) time interval. The spectral intensity profiles consist of 128 frequency bins whose spacing is determined by the pulse duration.The profiles are normalized by the peak intensity over all realizations and the spectral data are transformed to a logarithmic scale. The LSTM and two hidden layers consist of 300 nodes each with ReLU activations, and the output layer consists of 256 nodes with sigmoid activation for the temporal intensity predictions. For the spectral data, the network node numbers are 250 and 128, respectively. The networks were trained for 80 epochs with RMSprop optimizer and adaptive learning rate.

Noise implementation. To study the sensitivity of the RNN prediction to input noise, we included ±10% and up to ±20% random (multiplicative) noise on the temporal intensity profiles generated from the normalized NLSE. This noisy data were then tested with the network trained on the noise-free data to evaluate the robustness of the network against noise.

Linear chirp. An ensemble of 6,000 normalized NLSE numerical simulations was generated with randomly varying linear frequency chirp. A total of 5,800 realizations were used for training the RNN and 200 realizations for testing. Similarly to the unchirped normalized NLSE, the simulated intensity evolution maps are uniformly downsampled to 101 intensity profiles along propagation and the temporal axis consisted of 256 grid points in the range $[-2.5, +2.5]$ (normalized units). Along with the temporal intensity profile, the input chirp parameter is included to the network as an additional parameter. The LSTM and two hidden layers consist of 300 nodes each with ReLU activations, and the output layer consists of 256 nodes with sigmoid activation. The network was trained for 80 epochs with RMSprop optimizer and adaptive learning rate.

Normalized GNLSE. When generalizing the RNN predictions over a wider range of input pulse characteristics and fibre parameters, an ensemble of 12,000 numerical simulations was generated using the normalized form of the GNLSE; 11,800 realizations were used for training the RNN and 200 realizations for testing. The simulated intensity evolution maps are uniformly downsampled at a constant propagation step of $\Delta\xi = 0.02$ (normalized units), yielding 101 spectra along propagation for each simulated evolution. Each spectrum along the fibre is convolved and downsampled with a 8 nm FWHM super-Gaussian spectral filter such that the wavelength grid consisted of 132 spectral intensity bins spanning from 450 to 1,500 nm. The profiles are normalized by the peak intensity over all realizations and the data are transformed to logarithmic scale. A third-order dispersion parameter is included as an additional input parameter to the network. The LSTM and two hidden layers consist of 265 nodes each with ReLU activations, and the output layer consists of 132 nodes with sigmoid activation. The network was trained for 80 epochs with RMSprop optimizer and adaptive learning rate.

Multimode GNLSE. To evaluate the RNN ability to model the propagation of short pulses in a multimode fibre, we generated an ensemble of 1,000 numerical simulations using the multimode GNLSE; 950 realizations were used for training the RNN and 50 realizations for testing. Here, we consider the total field summed over the five simulated propagating modes. The simulated spectral evolution maps are uniformly downsampled at a constant propagation step of $\Delta z = 5$ mm, yielding 50 spectra along propagation for each simulated evolution. Each spectrum along the fibre is convolved and downsampled with an 8 nm FWHM super-Gaussian spectral filter such that the wavelength grid consisted of 301 spectral intensity bins spanning from 600 to 3,000 nm. The profiles are normalized by the peak intensity over all realizations and the data are converted to logarithmic scale. The initial coupling conditions of the individual are included in the network as additional parameters. The LSTM and two hidden layers consist of 500 nodes each with ReLU activations, and the output layer consists of 301 nodes with sigmoid activation. The network was trained for 80 epochs with RMSprop optimizer and adaptive learning rate.

Data availability

The numerical data used in this work and a public version of the codes are available at https://gitlab.com/salmelal/rnnnonlinear and with full simulation datasets at https://doi.org/10.5281/zenodo.4304771 under an MIT licence.

Received: 27 May 2020; Accepted: 11 January 2021;
Published online: 18 February 2021

References

1. Jordan, M. I. & Mitchell, T. M. Machine learning: trends, perspectives and prospects. *Science* **349**, 255–260 (2015).
2. Wetzel, B. et al. Customizing supercontinuum generation via on-chip adaptive temporal pulse-splitting. *Nat. Commun.* **9**, 4884 (2018).
3. Michaeli, L. & Bahabad, A. Genetic algorithm driven spectral shaping of supercontinuum radiation in a photonic crystal fiber. *J. Opt.* **20**, 055501 (2018).
4. Tzang, O., Caravaca-Aguirre, A. M., Wagner, K. & Piestun, R. Adaptive wavefront shaping for controlling nonlinear multimode interactions in optical fibres. *Nat. Photon.* **12**, 368–374 (2018).
5. Teğin, U. et al. Controlling spatiotemporal nonlinearities in multimode fibers with deep neural networks. *APL Photonics* **5**, 030804 (2020).
6. Andral, U. et al. Fiber laser mode locked through an evolutionary algorithm. *Optica* **2**, 275–278 (2015).
7. Pu, G., Yi, L., Zhang, L. & Hu, W. Intelligent programmable mode-locked fiber laser with a human-like algorithm. *Optica* **6**, 362–369 (2019).
8. Dudley, J. M. & Meng, F. Toward a self-driving ultrafast fiber laser. *Light Sci. Appl.* **9**, 26 (2020).
9. Kokhanovskiy, A., Ivanenko, A., Kobtsev, S., Smirnov, S. & Turitsyn, S. Machine learning methods for control of fibre lasers with double gain nonlinear loop mirror. *Sci. Rep.* **9**, 2916 (2019).
10. Kokhanovskiy, A. et al. Machine learning-based pulse characterization in figure-eight mode-locked lasers. *Opt. Lett.* **44**, 3410–3413 (2019).
11. Zahavy, T. et al. Deep learning reconstruction of ultrashort pulses. *Optica* **5**, 666–673 (2018).
12. Närhi, M. et al. Machine learning analysis of extreme events in optical fibre modulation instability. *Nat. Commun.* **9**, 4923 (2018).
13. Baumeister, T., Brunton, S. L. & Kutz, J. N. Deep learning and model predictive control for self-tuning mode-locked lasers. *J. Opt. Soc. Am. B* **35**, 617–626 (2018).
14. Finot, C., Gukov, I., Hammani, K. & Boscolo, S. Nonlinear sculpturing of optical pulses with normally dispersive fiber-based devices. *Opt. Fiber Technol.* **45**, 306–312 (2018).
15. Brunton, S. L., Proctor, J. L. & Kutz, J. N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl Acad. Sci. USA* **113**, 3932–3937 (2016).
16. Raissi, M. Deep hidden physics models: deep learning of nonlinear partial differential equations. *J. Mach. Learn. Res.* **19**, 932–955 (2018).
17. Raissi, M., Perdikaris, P. & Karniadakis, G. E. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378**, 686–707 (2019).
18. Vlachas, P. R., Byeon, W., Wan, Z. Y., Sapsis, T. P. & Koumoutsakos, P. Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **474**, 20170844 (2018).
19. Vlachas, P. R. et al. Backpropagation algorithms and reservoir computing in recurrent neural networks for the forecasting of complex spatiotemporal dynamics. *Neural Netw.* **126**, 191–217 (2020).
20. Pandey, S. & Schumacher, J. Reservoir computing model of two-dimensional turbulent convection. *Phys. Rev. Fluids* **5**, 113506 (2020).
21. Jiang, J. & Lai, Y.-C. Model-free prediction of spatiotemporal dynamical systems with recurrent neural networks: role of network spectral radius. *Phys. Rev. Res.* **1**, 033056 (2019).
22. Tikan, A. et al. Universality of the Peregrine soliton in the focusing dynamics of the cubic nonlinear Schrödinger equation. *Phys. Rev. Lett.* **119**, 033901 (2017).
23. Agrawal, G. P. *Nonlinear Fiber Optics* 5th edn (Academic Press, 2013).
24. Lipton, Z. C. A critical review of recurrent neural networks for sequence learning. Preprint at https://arxiv.org/pdf/1506.00019.pdf (2015).
25. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
26. Carleo, G. et al. Machine learning and the physical sciences. *Rev. Mod. Phys.* **91**, 045002 (2019).
27. Yu, Y., Si, X., Hu, C. & Zhang, J. A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput.* **31**, 1–36 (2019).
28. Dudley, J. M., Genty, G. & Coen, S. Supercontinuum generation in photonic crystal fiber. *Rev. Mod. Phys.* **78**, 1135 (2006).
29. Farfan, C. A., Epstein, J. & Turner, D. B. Femtosecond pulse compression using a neural-network algorithm. *Opt. Lett.* **43**, 5166–5169 (2018).

30. Wright, L. G. et al. Multimode nonlinear fiber optics: massively parallel numerical solver, tutorial and outlook. *IEEE J. Select. Top. Quantum Electron.* **24**, 1–16 (2017).
31. Werbos, P. J. Backpropagation through time: what it does and how to do it. *Proc. IEEE* **78**, 1550–1560 (1990).
32. Chollet, F. et al. Keras https://keras.io (2015).
33. Abadi, M. et al. Tensorflow: a system for large-scale machine learning. In *Proc. 12th USENIX Symposium on Operating Systems Design and Implementation* (*OSDI 16*) 265–283 (UESNIX Association, 2016).
34. Goodfellow, I et al. *Deep Learning* (MIT Press, 2016).

Acknowledgements

L.S. acknowledges support from the Faculty of Engineering and Natural Sciences Graduate School of Tampere University. J.M.D. acknowledges support from the French Agence Nationale de la Recherche (ANR-20-CE30-0004, ANR-15-IDEX-0003 and ANR-17-EURE-0002). G.G. acknowledges support from the Academy of Finland (298463, 318082 and Flagship PREIN 320165). We also thank D. Brunner for useful discussions.

Author contributions

L.S. performed all the numerical simulations. All authors participated in the data analysis reported and in the writing of the manuscript. L.S., J.M.D. and G.G. planned the research project and G.G. provided overall supervision.

Competing interests

The authors declare no competing interests.

Additional information

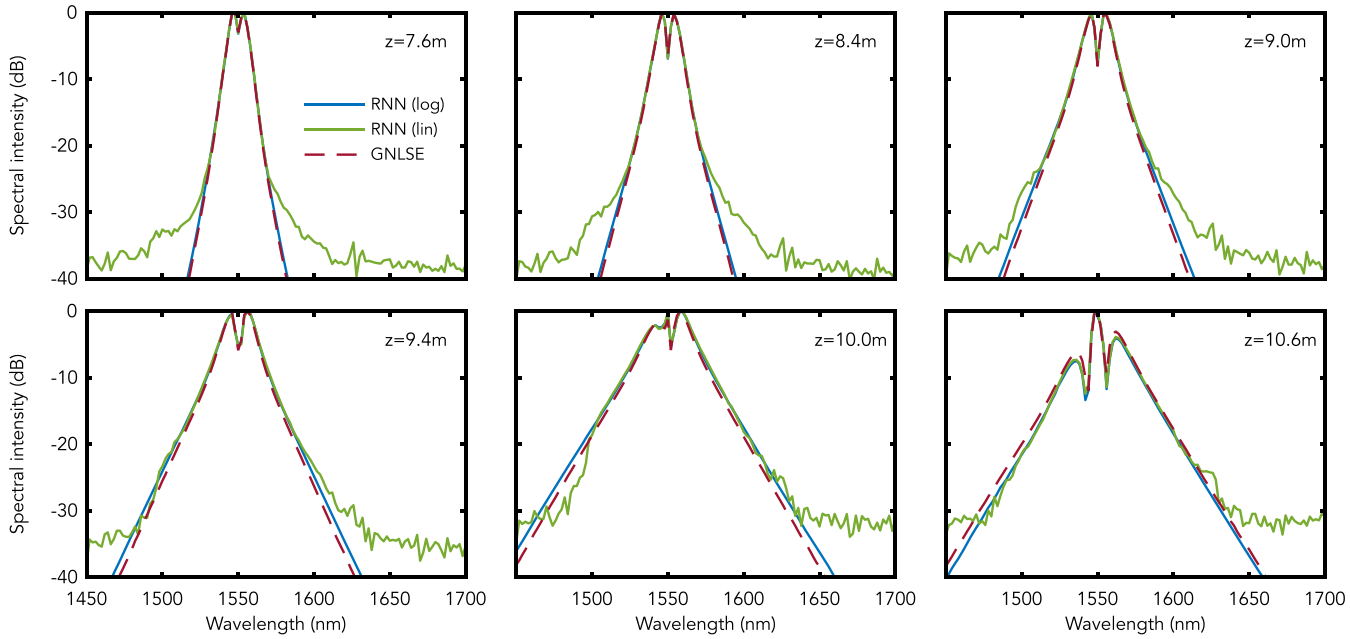
Extended data is available for this paper at https://doi.org/10.1038/s42256-021-00297-z.

Correspondence and requests for materials should be addressed to G.G.

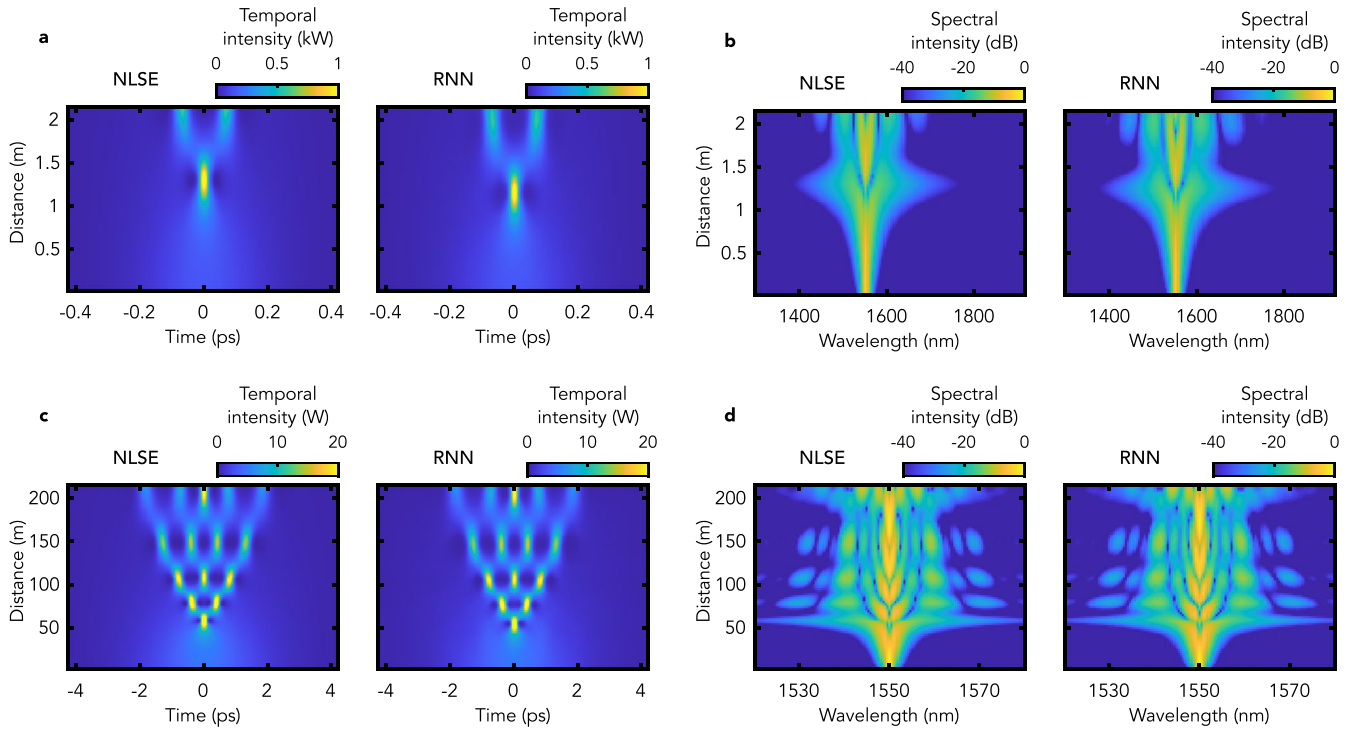
Peer review information *Nature Machine Intelligence* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

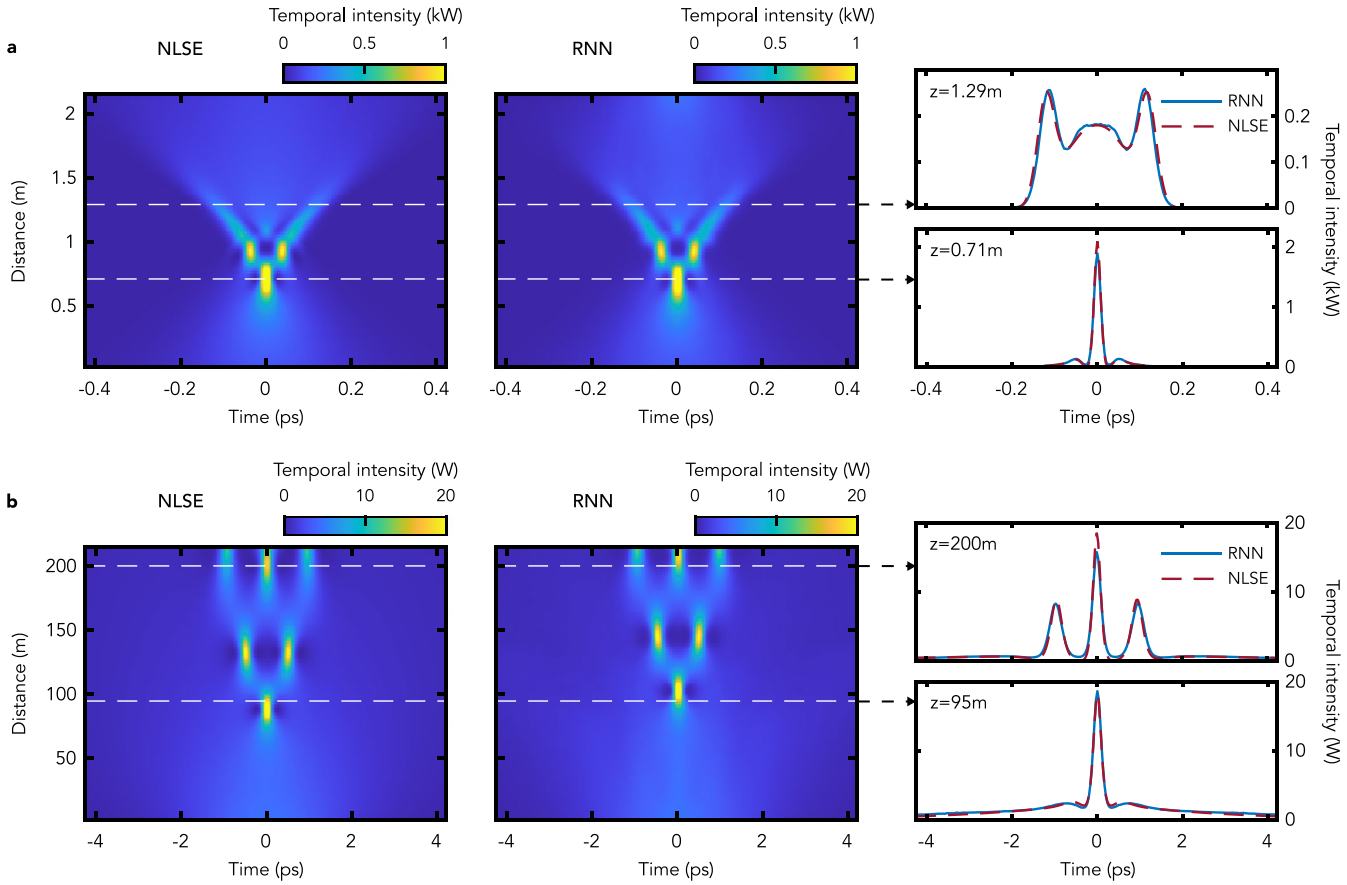
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



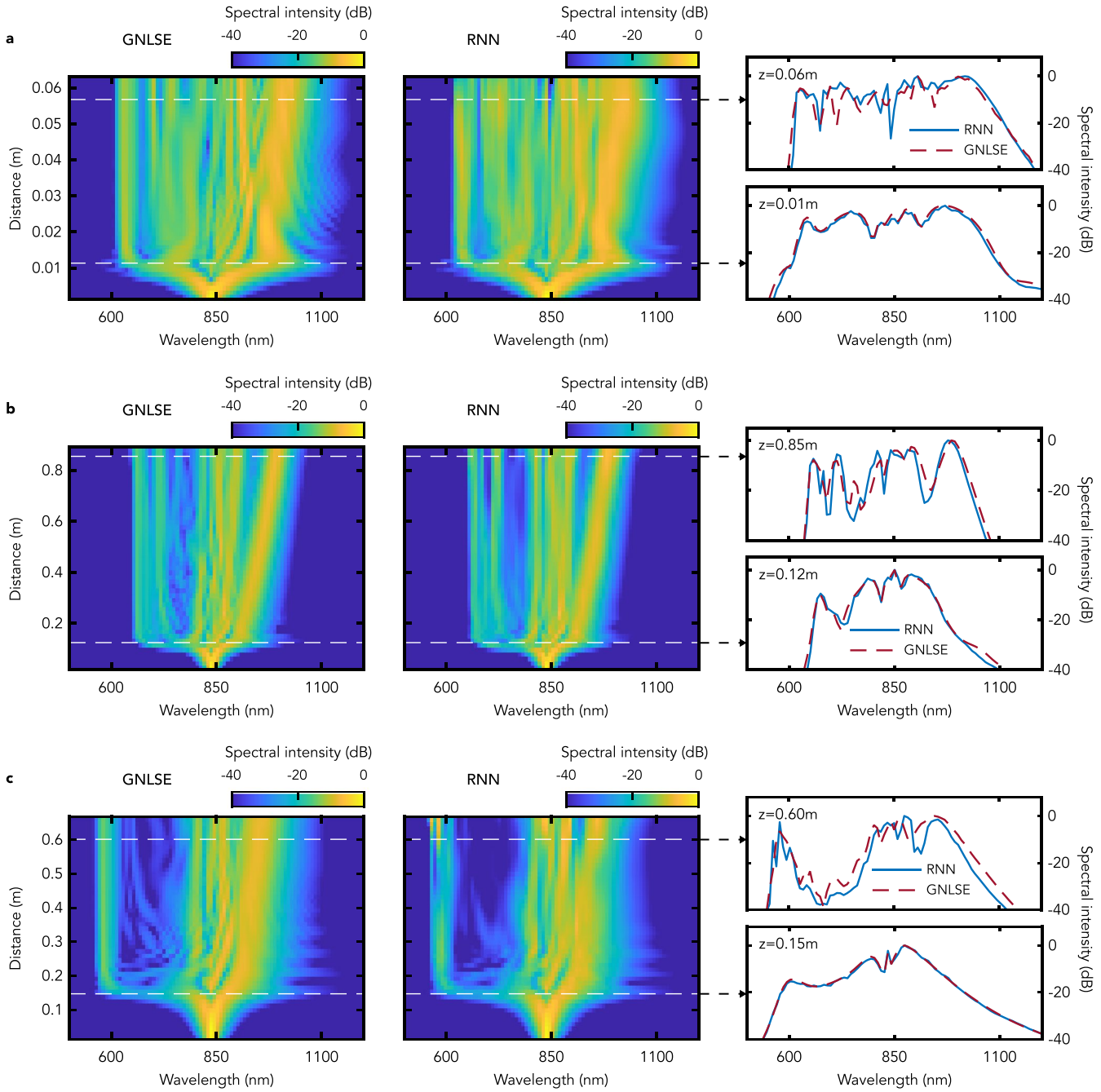
Extended Data Fig. 1 | Comparison of spectral intensity evolution modelling by the RNN when trained using data input in linear and logarithmic (dB) representation. The dashed red line shows the simulated spectra of a higher-order soliton ($N=6$) at selected distances along the fibre similar to those plotted in Fig. 5. The solid blue lines shows the RNN prediction when trained directly with spectral data input in dB. The solid green line shows the RNN prediction (plotted in dB) when trained with spectral data input in linear scale.



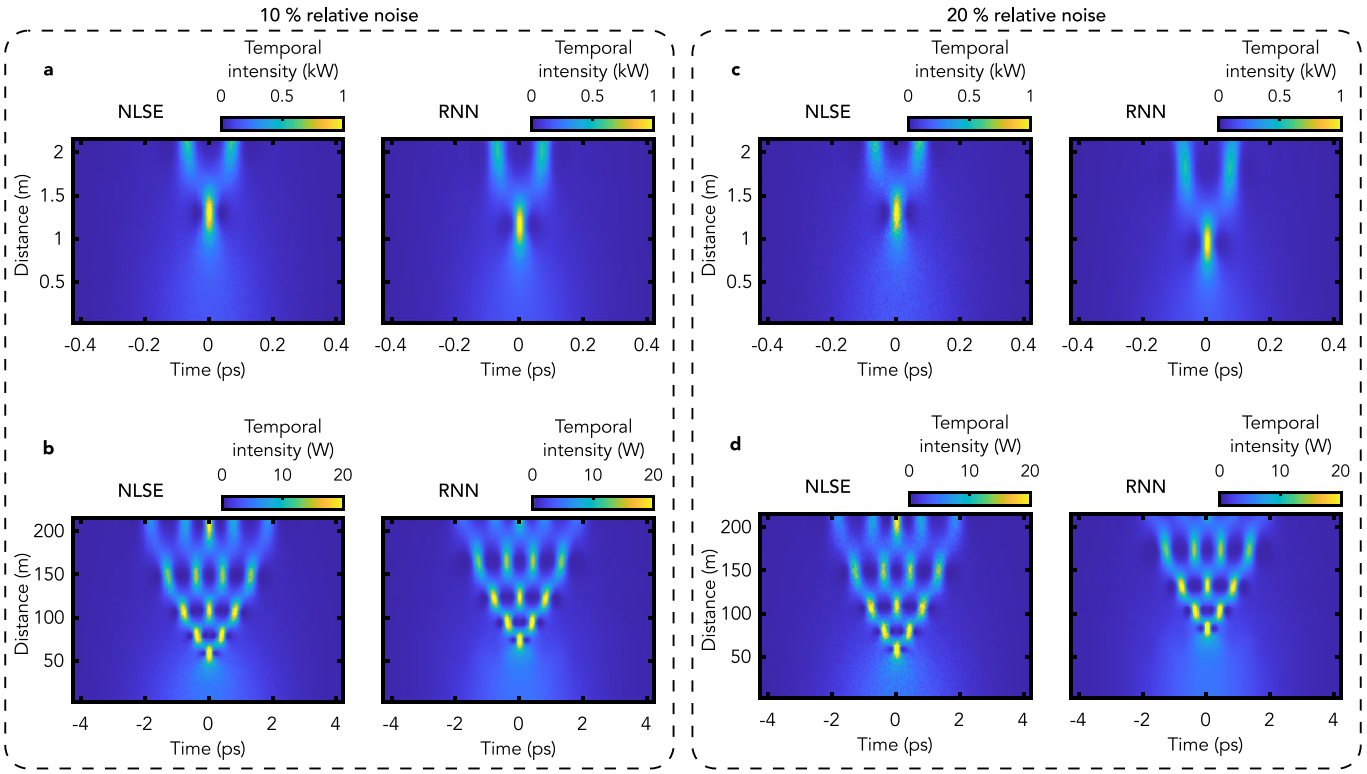
Extended Data Fig. 2 | Generalization of RNN applicability to modelling NLSE dynamics using normalized training simulations. Training data were generated using the normalized form of the NLSE (see Methods). The results are plotted in dimensional units. (a) and (b) shows the temporal and spectral evolution corresponding to the initial stage of higher-order soliton compression and breakup with soliton number $N=4$, pulse duration (FWHM) of 300 fs and peak power of 157 W. (c) and (d) shows the temporal and spectral evolution corresponding to the initial stage of higher-order soliton compression and breakup with soliton number $N=7$, pulse duration (FWHM) of 3 fs and peak power of 4.8 W. In each panel, we show the evolution map directly obtained from the numerical NLSE simulations and that obtained from the RNN network model. The RNN was trained with 950 normalized NLSE realizations where the soliton number was randomly varied in the range 1 to 7. The r.m.s. error (see Eq. (4) in Methods) computed over 50 test realizations was $R=0.152$ and $R=0.077$ for the temporal and spectral intensities, respectively.



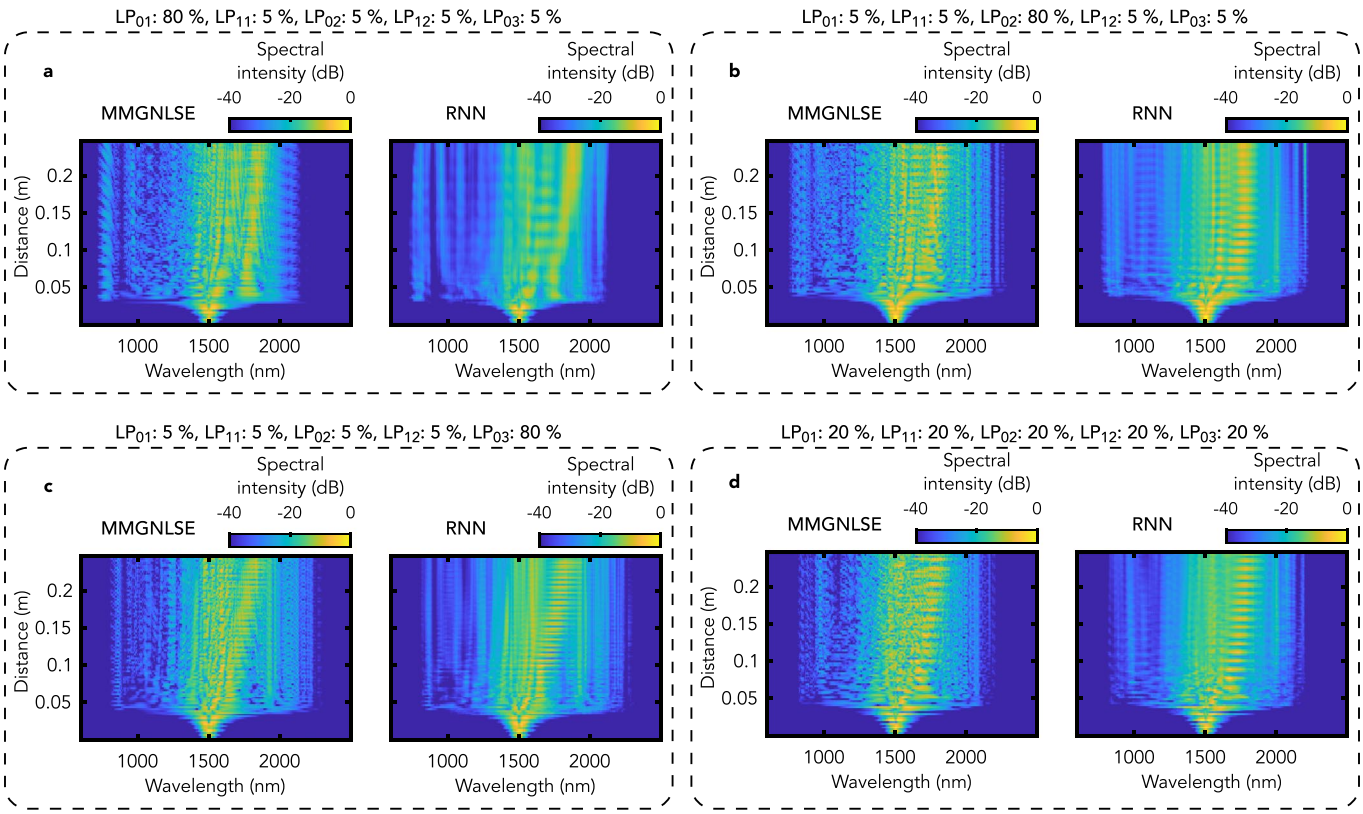
Extended Data Fig. 3 | Inclusion of input pulse chirp. (a) and (b) shows temporal evolution for input pulse parameter similar to those in Extended Data Fig. 2(a) and (c), but when a linear chirp (parabolic phase) of $C=-5$ and $C=5$, respectively, was added to the input pulse (see Methods). The left and middle panels show temporal evolution from NLSE simulations and prediction by the RNN, respectively, and the right panel shows the comparison between these two at selected distances. The network was trained with 5900 NLSE realizations where the soliton number and chirp parameter were varied randomly in the range 1 to 7 and -8 to 8, respectively. The r.m.s. error computed over 100 test realizations is $R=0.156$.



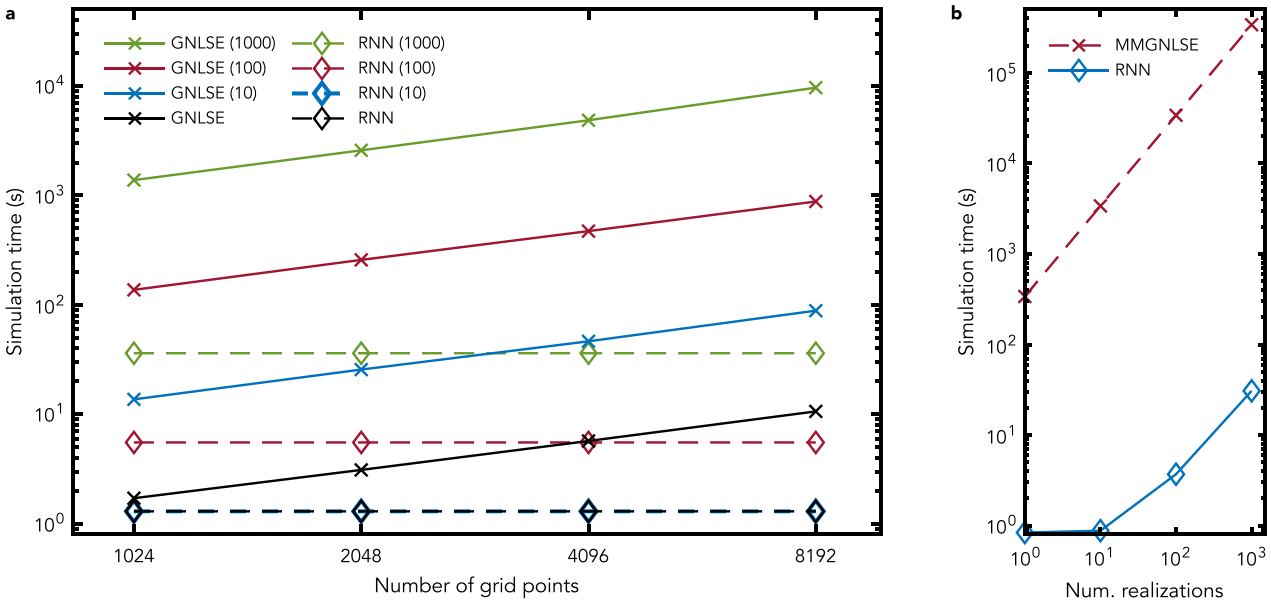
Extended Data Fig. 4 | Generalization of RNN applicability to modelling GNLSE dynamics. Training data were generated using the normalized form of the GNLSE including Raman effect, self-steepening and third-order dispersion (see Methods). The results are plotted in dimensional units. The network was trained using 11800 normalized GNLSE realizations where the soliton number, normalized third-order dispersion parameter, and pulse duration were respectively randomly varied in the range 2 to 8, 1 to 9 and 30 and 130 fs (FWHM). (a) shows the results for a transform-limited $N=4$ input pulse centered at 830 nm with 7.6 kW peak power and 40 fs duration. Corresponding fibre parameters are $\gamma=0.1$ W^{-1}m^{-1}, $\beta_2=-8\times10^{-27}$ s^{2}m^{-1} and $\beta_3=9\times10^{-41}$ s^{3}m^{-1}. (b) shows the results for a transform-limited $N=7$ input pulse with 2.9 kW peak power and 120 fs duration. Corresponding fibre parameters are $\gamma=0.0184$ W^{-1}m^{-1}, $\beta_2=-5.1\times10^{-27}$ s^{2}m^{-1} and $\beta_3=4.3\times10^{-41}$ s^{3}m^{-1}. (c) shows the results for a transform-limited $N=4.5$ input pulse with 3.0 kW peak power and 60 fs duration. Corresponding fibre parameters are $\gamma=0.01$ W^{-1}m^{-1}, $\beta_2=-1.7\times10^{-27}$ s^{2}m^{-1} and $\beta_3=6.5\times10^{-42}$ s^{3}m^{-1}. In each panel, we show the evolution map directly obtained from the numerical GNLSE simulations and that obtained from the RNN network model. The r.m.s. error computed over 200 test realizations is $R=0.092$.



Extended Data Fig. 5 | Effect of input noise on RNN predictions. The left and right panels shows how ±10 and ±20 % relative random multiplicative intensity noise added to the examples shown in Extended Data Fig. 2(a) and (c) affect the RNN predictions (see also Methods). The r.m.s. error computed over 50 test realizations was $R = 0.200$ and $R = 0.271$ for the ±10 and ±20 % cases, respectively ($R = 0.152$ for noise-free data). Although we do note a residual shift in the point of maximum compression, the dynamics of the higher-order soliton compression are overall well reproduced even under noisy conditions.



Extended Data Fig. 6 | Application of RNN to modelling multimode GNLSE dynamics. RNN modelling of nonlinear propagation of an ultrashort pulse of 150 fs duration (FWHM) and 2.5 MW peak power with 1500 nm center wavelength in a 50 μm radius multimode silica fibre. The network was trained with 950 realizations where the energy distribution between the different modes at the fibre input was varied. (a-d) show the spectral evolution integrated over all the modes for input energy distributions as indicated. In each panel, we show the evolution map directly obtained from the numerical multimode GNLSE simulations and that obtained from the RNN network model. The r.m.s. error computed over 50 test realizations is $R = 0.104$.



Extended Data Fig. 7 | Computation time comparison of GNLSE and RNN modelling. (a) Simulation time to compute evolution maps of supercontinuum similar to those presented in Fig. 6 using the GNLSE and RNN, as a function of the number of points N_p in the temporal (or spectral) grid and number of computed maps as indicated in the parentheses. (b) Simulation time to compute evolution maps of multimode simulations similar to those shown in Extended Data Fig. 6 using the multimode GNLSE and RNN, as a function of the number of realizations for a constant number of modes and grid points.