Unsupervised classification of non-linear dynamics in optical fiber propagation using intensity clustering

Anastasiia Sheveleva, Andrei Ermolaev, J.M. Dudley, Christophe Finot

HAL Id: hal-04698215
https://hal.science/hal-04698215v1

Submitted on 15 Sep 2024 (v1), last revised 21 Dec 2024 (v2)

Unsupervised classification of non-linear dynamics in optical fiber propagation using intensity clustering

Anastasiia Sheveleva[1], Andrei. V. Ermolaev[2], John M. Dudley[2], Christophe Finot[1,*]

1. Laboratoire Interdisciplinaire Carnot de Bourgogne, UMR 6303 CNRS-Université de Bourgogne, Dijon, France
2. Institut FEMTO-ST, UMR 6174 CNRS-Université de Franche-Comté, Besançon, France

* Corresponding author:
E-mail address: christophe.finot@u-bourgogne.fr

Abstract: We demonstrate that centroid-based clustering of normalized intensity profiles is able to successfully isolate different classes of pulses associated with physically distinct regimes of nonlinear and dispersive fiber propagation. Remarkable for its simplicity, this approach shows how only temporal intensity profiles reveal sufficient similarities to allow physical classification of different propagation behavior.

Keywords: nonlinear fiber optics, nonlinear dynamics, pulse propagation, unsupervised classification

1. Introduction

Machine learning techniques have seen rapid and impressive development in the field of ultrafast photonics in the last 5 years [1-3]. A particular area that has especially benefited from this work is the study of nonlinear dynamics in optical waveguides, where different machine learning techniques have been applied to the study of output waveform properties [4-7], supercontinuum evolution maps [8, 9], extreme event emergence [10] and the dynamics of four-wave mixing [11, 12]. The methods used in this work have been varied, including simple feedforward neural network architectures [4, 10-12], recurrent neural networks [8] convolutional networks [5] and more complex combinations of different deep learning algorithms [5, 7]. Other work has used physics-informed methods to solve the underlying propagation equations [13], inverse-problem methods for prediction [4, 6, 12], sparse regression to discover dynamical models from data [14, 15], and dominant-balance methods to automate detection of dominant physics during nonlinear evolution [16-18].

In this paper, we report a further application of machine learning methods in nonlinear fiber optics, using simple clustering techniques to analyze and classify intensity profiles resulting from a range of propagation scenarios. Specifically, we show that simple centroid-based clustering, such as the widely used K-means algorithm [19, 20], can successfully reveal patterns during higher-order soliton and wavebreaking evolution dynamics, and can also distinguish normal and anomalous dispersion-regime dynamics in specific regimes of propagation. These results reveal that clustering provides a complementary tool to existing analytical and numerical methods to automatically highlight patterns and similarities between different regimes of evolution, potentially yielding new insights into the underlying physics. Moreover, although motivated by the study of dynamics in optical fiber propagation, our analysis is based on an ideal nonlinear Schrödinger equation model, so that our results can be readily extended to other fields such as atom optics, hydrodynamics and plasma physics.

2. Propagation model and methods

We consider propagation in optical fiber described by the dimensional nonlinear Schrödinger equation (NLSE) [21]:

$$i\frac{\partial \psi}{\partial z} - \frac{\beta_2}{2}\frac{\partial^2 \psi}{\partial t^2} + \gamma |\psi|^2 \psi = 0 \tag{1}$$

where $\psi(t,z)$ is the complex pulse envelope, γ and β_2 are respectively the nonlinear and dispersion coefficients, z and t are respectively the propagation distance and comoving time. We define nonlinear and dispersive lengths as $L_{NL} = 1/\gamma P_0$ and $L_D = T_0^2/|\beta_2|$ respectively, where P_0 and T_0 are the peak power and temporal width of a hyperbolic secant input pulse $\psi(t,0) = P_0^{1/2} \operatorname{sech}(t/T_0)$. Loss and noise are neglected. The NLSE in normalized form is then given by:

$$i\frac{\partial u}{\partial \xi} - \frac{\text{sgn}(\beta_2)}{2}\frac{\partial^2 u}{\partial \tau^2} + N^2 |u|^2 u = 0 \qquad (2)$$

where $\xi = z/L_D$, $\tau = t/T_0$, and $N^2 = L_D/L_{NL}$ defines the soliton number N. The normalized input pulse is $u(\tau,\xi) = \text{sech}(\tau)$.

Our aim here is to use centroid-based clustering to differentiate different regimes of nonlinear and dispersive pulse propagation, based solely on analyzing temporal intensity profiles $I(\tau,\xi) = |u(\tau,\xi)|^2$. We stress that this is a severe constraint because we have no access to the temporal phase of the pulse or any spectral information. To generate large data sets of intensity profiles, we perform numerical simulations to solve the NLSE [21] for different initial conditions by randomly scanning N over the range 1–5 (continuously i.e. not only integer values), and at different propagation distances by extracting $I(\tau,\xi)$ at distances ξ over the range 0.0–2.5. We generated 5000 independent numerical simulations for both normal ($\beta_2 > 0$) and anomalous ($\beta_2 < 0$) dispersion regime propagation. Given the temporal symmetry of our initial conditions and the fact we are in an ideal NLSE model, the temporal intensity profiles are always symmetric, and so we can restrict ourselves to performing clustering for only positive-valued τ i.e. for $I(\tau > 0, \xi)$. Considering the intensity profile in this way allows us to characterize the structure of the intensity profiles in terms of their computed kurtosis as we describe below. Our simulations used 8192 temporal grid points, but for improved clustering performance, we downsample the 4096 positive time points to 171 points, using a nonuniform sampling procedure to ensure that we capture both broad envelope and short compressed features in the intensity profiles. And as we describe below, we introduce a normalized intensity profile $I_n(\tau,\xi) = I(\tau,\xi)/I_{\max}(\xi)$ with $I_{\max}$ is the maximum intensity at a given ξ.

We first present some general results in Fig. 1, showing three typical examples of normalized intensity profiles under three different propagation conditions: (blue) $\xi = 0.8$, $N = 4.1$, $\text{sgn}(\beta_2) = -1$; (red) $\xi = 1$, $N = 2.6$, $\text{sgn}(\beta_2) = +1$; (yellow) $\xi = 1.5$, $N = 2.9$, $\text{sgn}(\beta_2) = -1$. The profiles are plotted for positive times ($\tau > 0$) and illustrate the large qualitative differences between pulse profiles arising from propagation under different conditions, with shapes ranging from highly compressed pulses with subpulses associated with soliton evolution (blue), to highly flattened pulse envelopes associated with high power normal dispersion propagation dynamics (red) [21, 22].

The two sets of 5000 temporal intensity profiles sampled over 171 points serve as unlabeled data of the clustering algorithm. Note that we have also calculated from the intensity profiles the temporal moments of order two and four to characterize the intensity profile through its root mean square

(rms) temporal width σ and its kurtosis κ [23]. We recall here that the rms duration and the kurtosis of a hyperbolic secant pulse are 0.907 and 4.2 respectively.
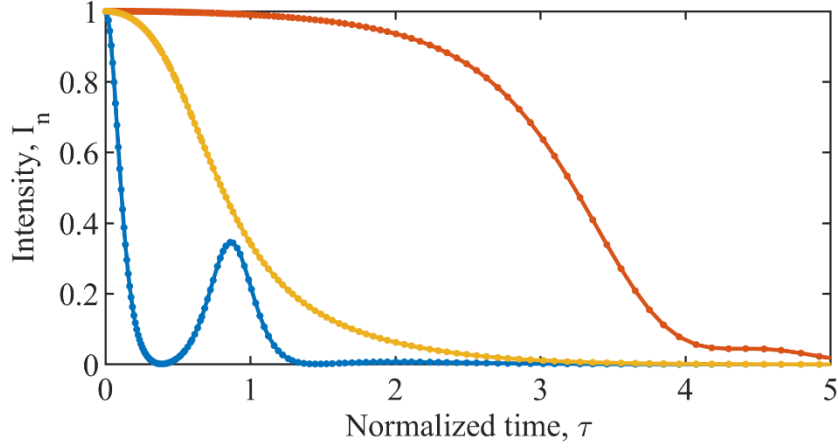
Figure 1: Examples of three normalized intensity profiles used as inputs for clustering. These profiles result from the propagation of a hyperbolic secant pulse under the following conditions: (blue) $\xi = 0.8$, $N = 4.1$, sgn(β_2) = -1; (red) $\xi = 1$, $N = 2.6$, sgn(β_2) = $+1$; (yellow) $\xi = 1.5$, $N = 2.9$, sgn(β_2) = -1

The clustering method we used is the K-means approach, which is often regarded as the most popular and widely used algorithm. Whereas K-means clustering has been implemented in the context of optical telecommunication to mitigate non-linear distortions of phase-coherent transmissions [24, 25], to the best of our knowledge it has never been exploited to get insights on the nonlinear dynamics. K-means is a centroid-based algorithm where the user must first define the required number of clusters. In our case, we choose between 2-4 clusters as we describe below. Then, centroids are randomly created based on the number of clusters and the distance between data points, and each centroid is calculated so that each data point can be assigned to the nearest centroid. The mean of the centroid is recalculated based on all the assigned data points, and this will change the position of the centroid. The process is iterated until it converges and each data is linked to a single cluster.

3. Clustering for anomalous dispersion regime dynamics

We begin by considering the clustering behavior for intensity profiles generated from soliton dynamics in the fiber anomalous dispersion regime. We consider 5000 intensity profiles scanning N over the range 1–5 and ξ over the range 0.0–2.5 as input to the K-means algorithm. To illustrate how simple clustering can identify different dynamical characteristics, we first present results when partitioning into only 2 clusters and considering as input the non-normalized profiles $I(\tau)$. Fig. 2(a1) displays these results by plotting the $\left(N^2, \xi\right)$ pair associated with each profile, but assigning a different color depending on the cluster into which the profile is sorted. This clearly shows a periodic band structure in the $\left(N^2, \xi\right)$ plane, but to gain more insight, we plot in Fig. 2(a2) a three-dimensional plot where the clustered profiles are plotted in terms of their root-mean-squared (rms) duration σ, their kurtosis κ and their peak intensity $I_{\max}$. It is clear from this plot that the main separation between the clusters arises from the difference in profile intensities, allowing us to conclude that the bands in Fig. 2(a1) are identifying different stages of periodic temporal expansion and compression.

To consider how clustering can identify more general pulse shape characteristics in addition to peak intensity, Fig. 2(b1) and Fig. 2(b2) show results using normalized intensity profiles $I_n(\tau)$ as input. Fig. 2(b1) shows a similar band structure in the $\left(N^2, \xi\right)$ plane although the bands are wider. Again, the clusters can be readily attributed to different stages of pulse evolution, with the red cluster linked to stages of noticeable compression and significant deviation from the input profile, and the blue color to stages where there is only moderate change with respect to the input profile. This is particularly the case for short propagation distances where the dominance of the nonlinear regime (self-phase modulation) does not affect the temporal intensity profile [18]. Similarly, the blue regime dominates for areas with low nonlinearity and for values of $N \sim 1$ where the pulse preserves its shape or can even be stationary in the case of the fundamental soliton [21]. We also recognize the special conditions corresponding to integers values of N and propagation distances of $\xi = \pi/2$ (the black crosses in the $\left(N^2, \xi\right)$ plane) associated with perfect soliton recurrence to its initial state [26]. Fig. 2(b2) separates the clusters in terms of computed temporal duration and kurtosis, with the projection in the rms/kurtosis (σ, κ) plane showing that the different regions strongly overlap. This highlights how an approach using the normalized profile is not equivalent to an approach considering only the temporal moments of the pulses.
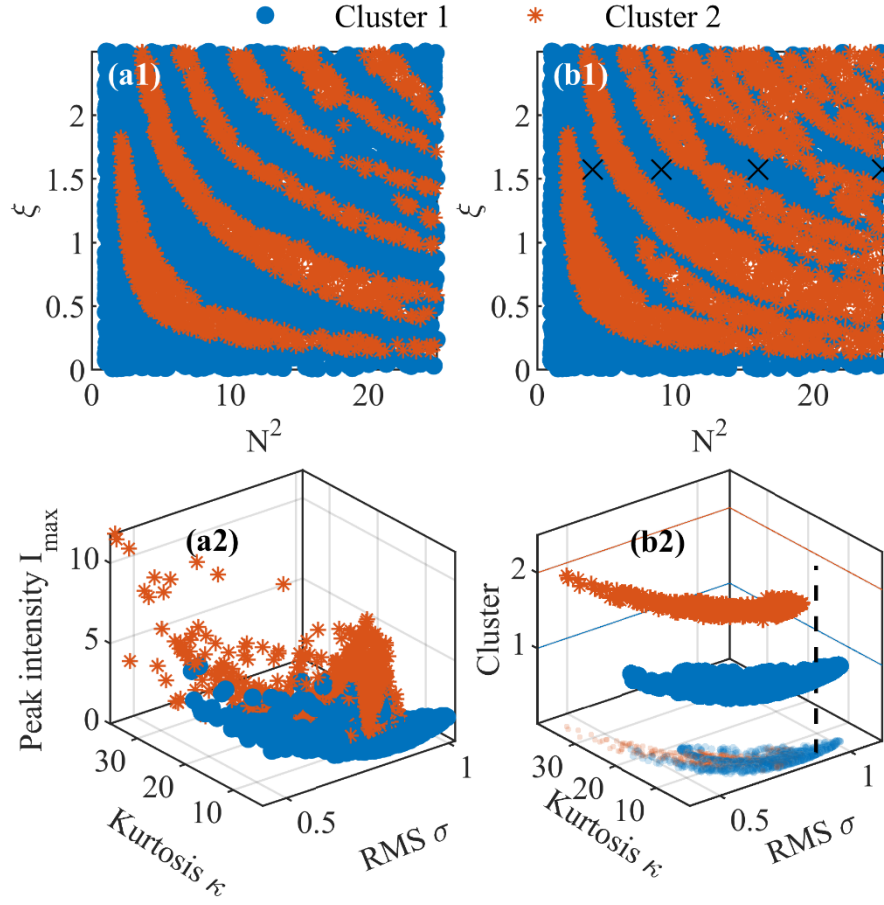
Figure 2 : Clustering into two sets of temporal intensity profiles generated for propagation in the anomalous dispersion regime. Panels (a) and (b) show results using non-normalized profiles $I(\tau)$ and normalized profiles $I_n(\tau)$ respectively. Panels a1 and b1 display clusters according to normalized propagation distance ξ and the soliton number N. The crosses indicate the parameters leading to ideal periodic recovery of the initial intensity profile. Panels a2 and b2 display clusters according to their rms duration σ and kurtosis κ as well as their peak power I_{max} for the non-normalized dataset. The black dashed line is a visual guideline for the properties of the input pulse.

We now discuss how this clustering behaviour changes when the number of clusters for partitioning is increased to 3. Results of this new clustering for the normalized profiles $I_n(\tau)$ are shown in Fig. 3(a1). We again see a band structure in the (N^2, ξ) plane, with the blue regions again corresponding to intensity profiles that have experienced little variation with propagation. This cluster region is similar to that in the 2 cluster results in Fig. 2(b1). However, when selecting 3 clusters, the single cluster in Fig. 2(b1) (corresponding to pulses that underwent significant evolution in their intensity profiles) now splits into two subsets. And we see in Fig. 3(a2) that in the rms/kurtosis (σ, κ) plane, these clusters are superimposed to a large degree so that clustering in terms of these features is insufficient to separate the different behaviours.

To better understand the physical characteristics that are being selected and associated with these different clusters, Fig. 3(b1) and (b2) plots the longitudinal evolution of the intensity profile of pulses of soliton order $N = 4$ and $N = 5$, respectively, accompanied by the variations of the rms duration and kurtosis in panels (c1) and (c2) respectively. We now see that the red bands in the (N^2, ξ) plane in Fig. 3(b1) correspond to the zones undergoing maximum temporal compression, while the yellow bands correspond to the zone where the pulse has temporally-spit into several subpulses with intensity at $\tau = 0$ that tends to vanish. This is confirmed by also showing on panel (a1) with the dashed curve, the empirical equation giving the distance ξ_c leading to the maximum compression as a function of the soliton number [27, 28]:

$$\xi_c \sim 1/N \tag{3}$$

Using the recurrence conditions and the longitudinal symmetry of the evolution achieved for integer values of N, one can note that that the corresponding points (black triangles) are well contained within the second cluster.
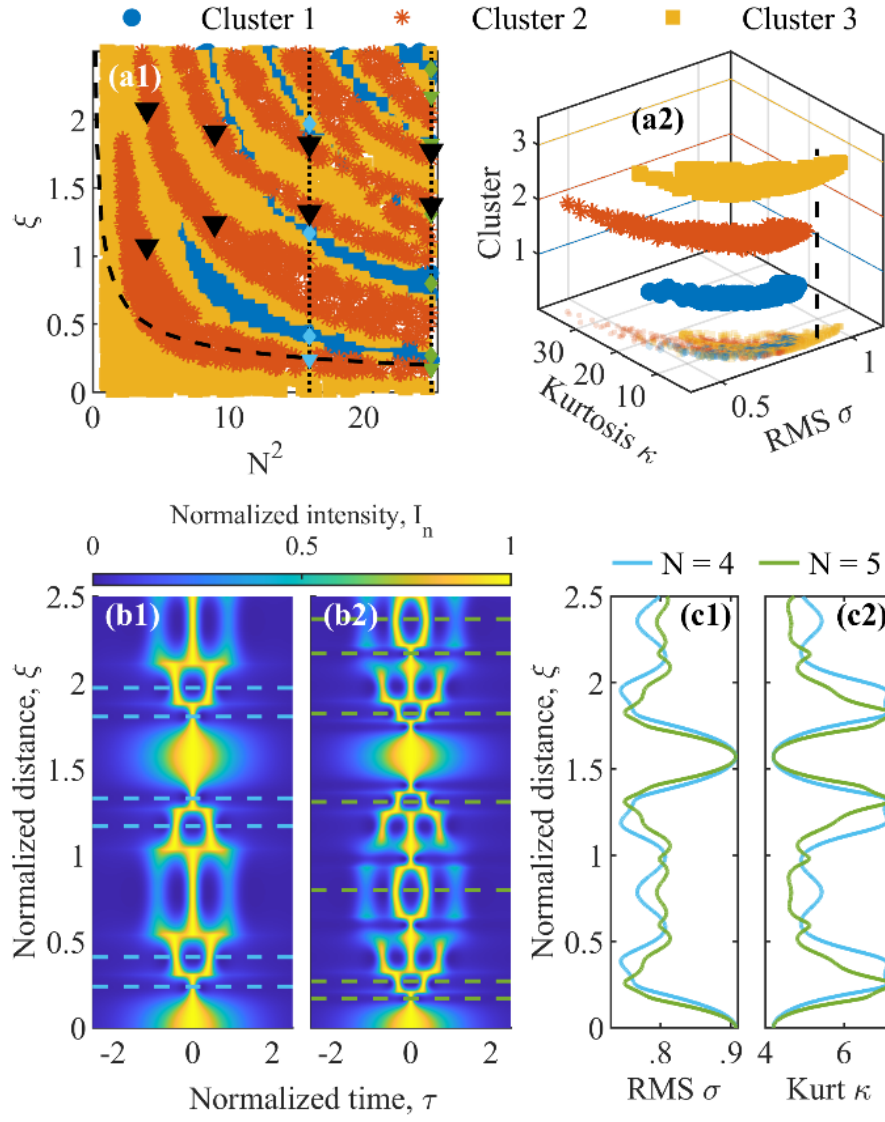
Figure 3 : Clustering into three sets of the normalized temporal intensity profiles generated upon propagation in the anomalous regime of dispersion. (a1) Clusters according to the normalized propagation distance ξ and the solitonic number N. The dashed black line marks the positions of the compression stages approximated by Eq. (3) and the black triangles highlight the parameters leading to other marked compression for an integer value of N. Dashed blue and green lines label $N = 4$ and 5, respectively. The blue and green triangles mark positions of the compressed pulses, while the blue and green diamonds – the split pulses' positions. (a2) Clusters according to their rms duration σ and their kurtosis κ. The black dashed line is a visual guideline for the properties of the input pulse. (b) Longitudinal evolution of the normalized intensity profile for N = 4 and 5 (panels b1 and b2 respectively). (c) Longitudinal evolution of c1: the corresponding rms duration σ and c2:kurtosis κ.

4. Clustering for normal dispersion regime dynamics

Pulse evolution in the normal dispersion regime differs drastically from the anomalous case. Instead of periodic evolution dynamics and temporal localization associated with soliton effects, the interaction between the Kerr nonlinearity and normal dispersion is marked by effects such as optical wavebreaking, self-similar propagation, and temporal broadening [29, 30]. We study here the clustering of intensity profiles associated with normal dispersion propagation using the same approach as above, with 5000 intensity profiles scanning N over the range 1–5 and ξ over the range 0.0–2.5, except that now we take $\text{sgn}(\beta_2) > 0$, and the input parameter N is no longer interpreted as a soliton number.

Clustering into three clusters as shown in Fig. 4(a1) leads to the isolation of three successive stages of dynamics associated with well-known characteristics of nonlinear and dispersive propagation [21]. The blue cluster, where the rms duration of the pulse does not vary significantly, corresponds to the initial propagation regime when the spectral broadening and chirp induced by the nonlinearity has not yet translated into significant temporal broadening. However, the shape of the temporal profile does change in this regime, as can be observed in the significant drop of the kurtosis [31] apparent from Fig. 4(a2). The boundary between the blue and red clusters can be associated with a known analytical result that estimates the distance ξ_{WB} after which wave-breaking occurs for a hyperbolic secant pulse [29, 30]

$$\xi_{WB} = \sqrt{\frac{3}{2}} \frac{1}{\sqrt{N^2 + 1}},$$

(4)

which is plotted as the solid line in Fig. 4(a1). Although strictly speaking accurate only for large N, the trend is nonetheless remarkably consistent with the results of the clustering. In the red cluster, the pulse undergoes both temporal broadening and change in shape, representing a transition stage before the pulse shape tends to evolve much less, with intensity profiles in the yellow cluster showing little variation in kurtosis while still undergoing temporal broadening. This stage of evolution is well-understood, associated with the temporal profile taking on the same shape as the pulse spectrum (the dispersive Fourier transform or "spectron" regime [22, 32-34]).

Finally in this section, we note that in contrast to the case of anomalous dispersion regime propagation where the clusters are mixed in the (σ, κ) plane, the clusters are much betted isolated for normal dispersion regime propagation. In this context, it is interesting to consider whether direct clustering according to the rms duration and kurtosis could reveal a similar picture. These results are plotted in panels (b) of Fig. 4. However, whilst the trends in Fig. 4(b2) are consistent with the discussion of the dynamics from (N^2, ξ) clustering, the borders between the clusters cannot be readily associated with the wave-breaking condition which is again plotted as the solid line in Fig. 4(b1).
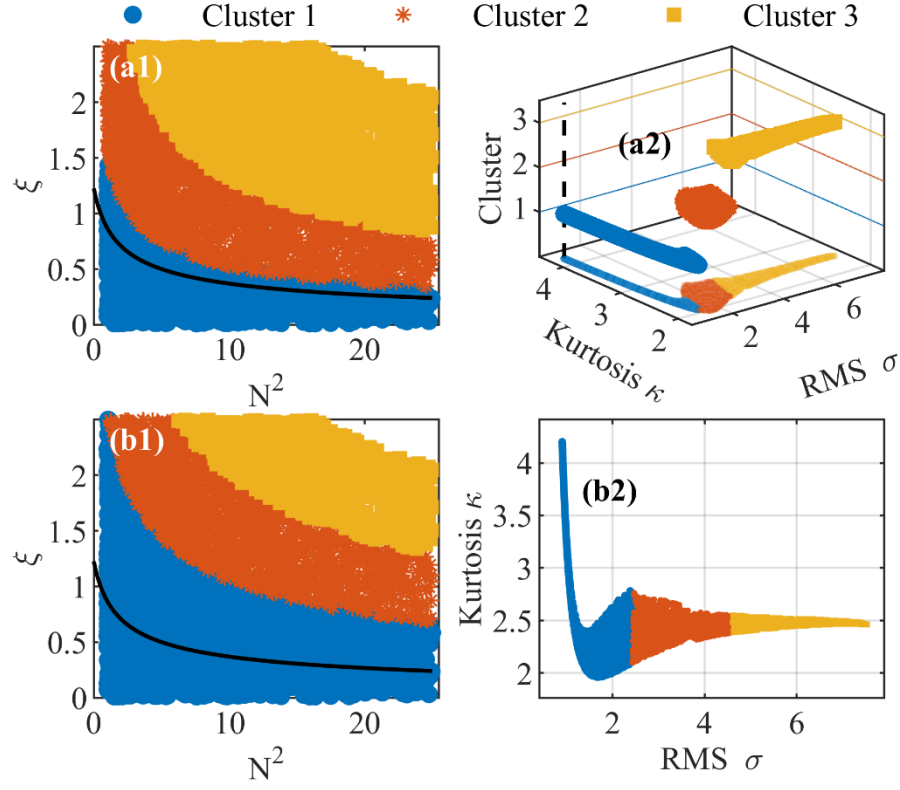
Figure 4: Clustering into three sets for pulses generated during propagation in the normal dispersion regime. Fig. 4(a) plots the clustering results in terms of the normalized intensity profile whilst Fig. 4(b) plots the results in terms of σ and κ properties. Panels a1 and b1 plot the clusters according to the normalized propagation distance ξ and the number N. The black line marks the wave-breaking condition (Eq. (4)). Panels a2 and b2 plot the clusters according to their rms duration σ and their kurtosis κ.

5. Clustering profiles for both anomalous and normal dispersion regime dynamics

We now evaluate the performance of the clustering algorithm when input data combines intensity profiles arising from both anomalous and normal dispersion regime propagation. That is, we now input 10,000 profiles into the algorithm combining intensity profiles arising from soliton dynamics as well as normal dispersion broadening and wave breaking. In Fig. 5 panels (a), (b), (c), we show results of this clustering into two, three, and four clusters respectively.

We first consider the results in Fig. 5(a1) where we plot how the two clusters (red and blue, plotted in the $\left(N^2, \xi\right)$ plane) sort the intensity profiles relative to the known values of dispersion $\text{sgn}(\beta_2) = \pm 1$ in which they were generated. The classification shows that based only on the output normalized intensity profile, it is possible to infer to a large extent the dispersion regime in which pulse propagation took place. However, the classification is not perfect and for short propagation distances, some profiles known to arise from normal dispersion regime propagation are sorted into the same blue cluster associated primarily with soliton-like profiles in the anomalous dispersion regime. However, this can be readily understood because these are profiles associated with the initial stages of normal dispersion propagation (below the wave breaking distance) where the combination of dispersion and nonlinearity is not yet sufficient to lead to major variation in the temporal pulse duration. In fact, we can see from Fig. 5(a2) where the clusters are plotted in the (σ, κ) plane that the algorithm selects different characteristics of the pulse profiles that follow distinct branches: one in which the duration changes little but the pulse kurtosis is significantly modified (blue); and another where the pulse shape changes little but its duration significantly changes (red).

When classifying into three and four clusters, it is essentially within the anomalous dispersion regime that we see additional cluster structure appear. Indeed, as can be seen in Fig. 5(b1) and (b2), moving to three clusters leads to the identification of intensity profiles linked to periodic soliton compression dynamics, as discussed in section 3.2. For four clusters as shown in Fig. 5(c1) and (c2), it is once again in the anomalous dispersion regime that new structure emerges, which in this case corresponds to the typical splitting of the higher soliton pulse into multiple subpulses. In this regard, it is also of interest to explicitly plot the centroids of the four clusters in Fig. 6(a), and because the centroid does not actually result from any physical propagation, we also plot in Fig. 6(b) the normalized intensity profiles of the pulse in the dataset that is the closest to each centroid. This provides additional insights into the typical profiles each cluster and confirms our physical discussion above. Cluster 1 is associated with profiles close to the initial pulse, typical of low propagation distances or parameters leading to a recurrence of the initial higher-order solitonic condition. The second cluster corresponds to pulse evolution beyond the wave-breaking distance where the pulse has experienced strong temporal broadening and reshaping towards a flattened profile. The third centroid corresponds to a significantly

compressed profile whereas the last centroid is typical of the splitting of the higher order soliton into two subpulses.
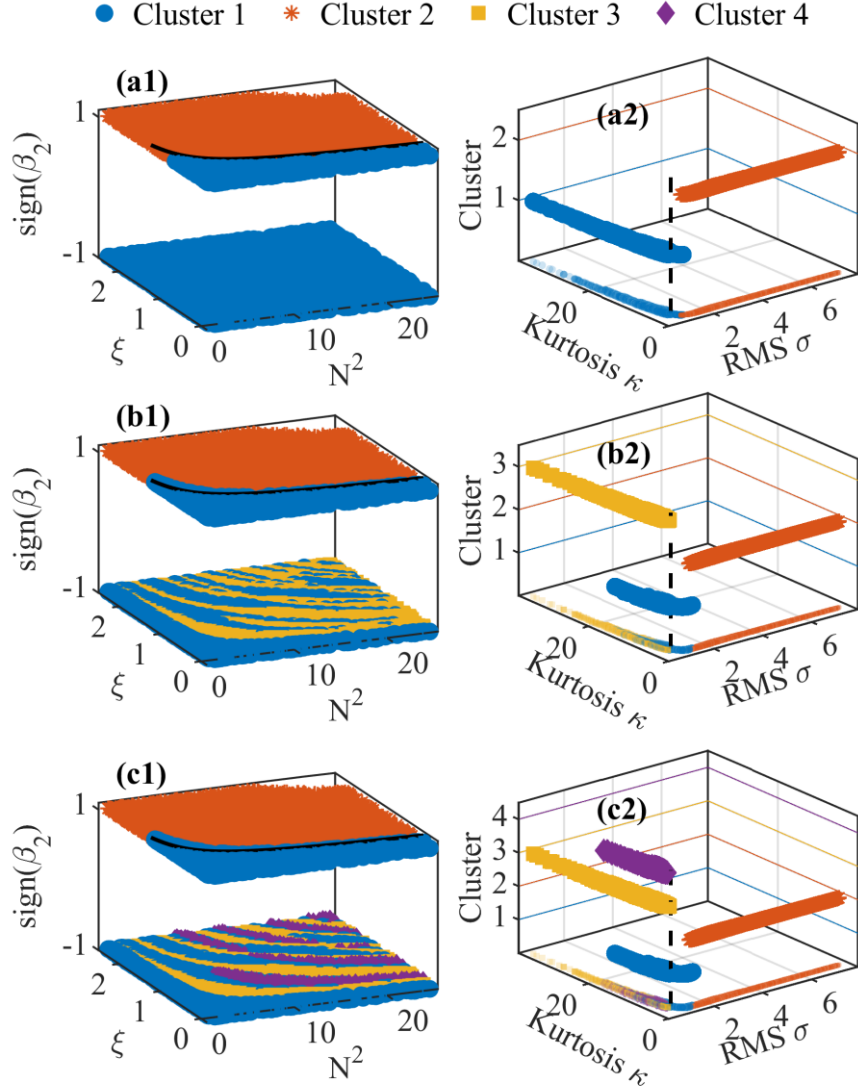

Figure 5: Clustering of the temporal intensity profiles generated upon propagation in the normal and anomalous regimes of dispersion. Figures 5(a), (b), (c) show results according to the different number of clusters: 2, 3 and 4 respectively. Panels a1,b1,c1 are clusters displayed according to the normalized propagation distance ξ, the number N and the dispersion regime ($\mathrm{sign}(\beta_2)$). The black line marks the wave-breaking condition (Eq. (4)). Panels a2, b2, c2 are the clusters plotted according to their rms duration σ and their kurtosis κ. The black dashed line is a visual guideline for the properties of the input pulse.
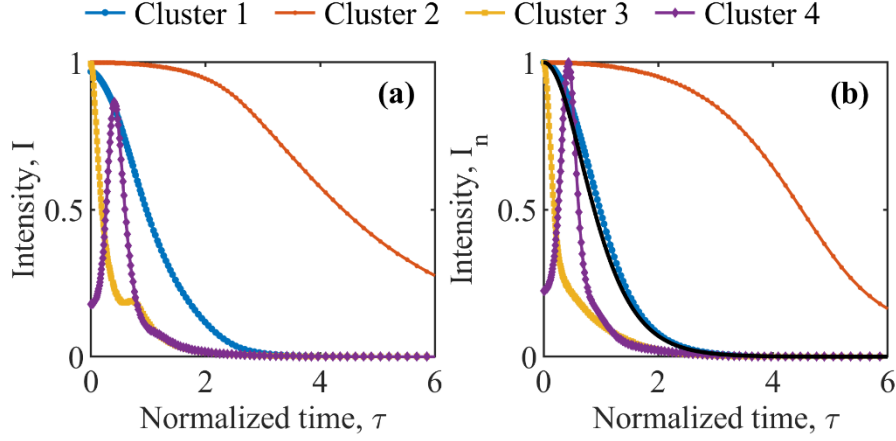
Figure 6 : (a) Centroids of the four clusters. (b) Normalized temporal intensity profiles in our dataset that are the closest to each centroid. The black curve is the input condition.

6. Conclusions

The major result of this work has been to show that centroid-based clustering of normalized intensity profiles can successfully isolate different classes of pulses associated with physically distinct regimes of nonlinear and dispersive fiber propagation. Remarkable for its simplicity, this approach shows how only temporal intensity profiles (i.e. without spectral intensity or phase information) nonetheless reveal sufficient similarities to allow physical classification of different propagation behavior. Extensions of this work could be in numerous directions, including for example the use of simultaneous spectral intensity profiles to generalize the technique to more difficult cases with initial chirp leading to effects such as spectral compression [35, 36] which would be hard if not impossible to distinguish using temporal profile clustering only. Also, whilst we have considered background-free pulses in this analysis, it could be readily extended to nonlinear and dispersive pulse structures upon a continuous background, opening the possibility to obtain insights and empirical intuition into the properties and emergence of extreme rogue wave events. It is also possible that going beyond the simple NLSE and including additional higher order linear or non-linear terms will allow a useful cluster analysis of the highly complex process of supercontinuum generation. And in addition to NLSE-related problems, the process could readily be adapted to handle propagation in cavities typical of lasers or resonators supporting dissipative or cavity solitons. And of course, extensions of the clustering algorithm beyond the use of K-means may also prove useful, although this is beyond the scope of this present paper.

Funding

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Anastasiia Sheveleva: Conceptualization, Formal analysis, Methodology, Investigation, Software, Visualization, Writing - review & editing. **Andrei Ermolaev:** Validation, Writing - review & editing. **John M. Dudley:** Validation, Writing-original draft, Funding acquisition. **Christophe Finot:** Conceptualization, Methodology, Validation, Writing-original draft, Supervision, Funding acquisition.

Data availability

The data that support the findings of this study are available from the corresponding author, CF, upon reasonable request.

References

[1] G. Genty, L. Salmela, J.M. Dudley, D. Brunner, A. Kokhanovskiy, S. Kobtsev, S.K. Turitsyn, Machine learning and applications in ultrafast photonics, Nat. Photon., 15 (2021) 91-101.

[2] P. Freire, E. Manuylovich, J.E. Prilepsky, S.K. Turitsyn, Artificial neural networks for photonic applications—from algorithms to implementation: tutorial, Adv. Opt. Photon., 15 (2023) 739-834.

[3] J.W. Nevin, S. Nallaperuma, N.A. Shevchenko, X. Li, M.S. Faruk, S.J. Savory, Machine learning for optical fiber communication systems: An introduction and overview, APL Photonics, 6 (2021) 121101.

[4] S. Boscolo, C. Finot, Artificial neural networks for nonlinear pulse shaping in optical fibers, Opt. Laser Technol., 131 (2020) 106439.

[5] N. Gautam, A. Choudhary, B. Lall, Comparative study of neural network architectures for modelling nonlinear optical pulse propagation, Opt. Fiber Technol., 64 (2021) 102540.

[6] S. Boscolo, J.M. Dudley, C. Finot, Modelling self-similar parabolic pulses in optical fibres with a neural network, Results in Optics, 3 (2021) 100066.

[7] H. Sui, H. Zhu, L. Cheng, B. Luo, S. Taccheo, X. Zou, L. Yan, Deep learning based pulse prediction of nonlinear dynamics in fiber optics, Opt. Express, 29 (2021) 44080-44092.

[8] L. Salmela, N. Tsipinakis, A. Foi, C. Billet, J.M. Dudley, G. Genty, Predicting ultrafast nonlinear dynamics in fibre optics with a recurrent neural network, Nature Machine Intelligence, 3 (2021) 344-354.

[9] L.C.B. Silva, M.E.V. Segatto, Nonlinear autoregressive with external input neural network for predicting the nonlinear dynamics of supercontinuum generation in optical fibers, J. Opt. Soc. Am. B, 40 (2023) 1292-1298.

[10] M. Närhi, L. Salmela, J. Toivonen, C. Billet, J.M. Dudley, G. Genty, Machine learning analysis of extreme events in optical fibre modulation instability, Nat. Commun., 9 (2018) 4923.

[11] A. Sheveleva, P. Colman, J.M. Dudley, C. Finot, Phase space topology of four-wave mixing reconstructed by a neural network, Opt. Lett., 47 (2022) 6317-6320.

[12] S. Boscolo, J.M. Dudley, C. Finot, Predicting nonlinear reshaping of periodic signals in optical fibre with a neural network, Opt. Commun., 542 (2023) 129563.

[13] X. Jiang, D. Wang, Q. Fan, M. Zhang, C. Lu, A.P.T. Lau, Physics-Informed Neural Network for Nonlinear Dynamics in Fiber Optics, Laser & Photonics Rev., (2022) 2100483.

[14] A.V. Ermolaev, A. Sheveleva, G. Genty, C. Finot, J.M. Dudley, Data-driven model discovery of ideal four-wave mixing in nonlinear fibre optics, Sci. Rep., 12 (2022) 1-11.

[15] M. Sorokina, S. Sygletos, S.K. Turitsyn, Sparse identification for nonlinear optical communication systems: SINO method, Opt. Express, 24 (2016) 30433-30443.

[16] J.L. Callaham, J.V. Koch, B.W. Brunton, J.N. Kutz, S.L. Brunton, Learning dominant physical processes with data-driven balance models, Nat. Commun., 12 (2021) 1016.

[17] A.V. Ermolaev, M. Mabed, C. Finot, G. Genty, J.M. Dudley, Analysis of interaction dynamics and rogue wave localization in modulation instability using data-driven dominant balance, Sci. Rep., 13 (2023) 10462.

[18] A.V. Ermolaev, C. Finot, G. Genty, J.M. Dudley, Automating physical intuition in nonlinear fiber optics with unsupervised dominant balance search, Opt. Lett., 49 (2024) 4202-4205.

[19] S. Lloyd, Least squares quantization in PCM, IEEE transactions on information theory, 28 (1982) 129-137.

[20] G. Seber, Multivariate observations, John Wiley & Sons 2009.

[21] G.P. Agrawal, Nonlinear Fiber Optics, Sixth Edition, Academic Press, San Francisco, CA, 2019.

[22] S.A. Akhmanov, V.A. Vysloukh, A.S. Chirkin, Self-action of wave packets in a nonlinear medium and femtosecond laser pulse generation, Soviet Physics Uspekhi, 29 (1986) 642.

[23] L.T. DeCarlo, On the meaning and use of kurtosis, Psychol. Methods, 2 (1997) 292-307.

[24] N.G. Gonzalez, D. Zibar, A. Caballero, I.T. Monroy, Experimental 2.5-Gb/s QPSK WDM Phase-Modulated Radio-Over-Fiber Link With Digital Demodulation by a K-Means Algorithm, IEEE Photon. Technol. Lett., 22 (2010) 335-337.

[25] J. Zhang, W. Chen, M. Gao, G. Shen, K-means-clustering-based fiber nonlinearity equalization techniques for 64-QAM coherent optical communication system, Opt. Express, 25 (2017) 27570-27580.

[26] R.H. Stolen, L.F. Mollenauer, W.J. Tomlinson, Observation of pulse restoration at the soliton period in optical fibers, Opt. Lett., 8 (1983) 187-189.

[27] C.-M. Chen, P.L. Kelley, Nonlinear pulse compression in optical fibers: scaling laws and numerical analysis, J. Opt. Soc. Am. B, 19 (2002) 1961-1967.

[28] G. Genty, S. Coen, J.M. Dudley, Fiber supercontinuum sources, J. Opt. Soc. Am. B, 24 (2007) 1771-1785.

[29] D. Anderson, M. Desaix, M. Lisak, M.L. Quiroga-Teixeiro, Wave-breaking in nonlinear optical fibers, J. Opt. Soc. Am. B, 9 (1992) 1358-1361.

[30] C. Finot, B. Kibler, L. Provost, S. Wabnitz, Beneficial impact of wave-breaking on coherent continuum formation in normally dispersive nonlinear fibers, J. Opt. Soc. Am. B, 25 (2008) 1938-1948.

[31] C.-K. Rosenberg, D. Anderson, M. Desaix, P. Johannisson, M. Lisak, Evolution of optical pulses towards wave breaking in highly nonlinear fibres, Opt. Commun., 273 (2007) 272-277.

[32] A. Zeytunyan, G. Yesayan, L. Mouradian, P. Kockaert, P. Emplit, F. Louradour, A. Barthélémy, Nonlinear-dispersive similariton of passive fiber, J. Europ. Opt. Soc. Rap. Public., 4 (2009) 09009.

[33] S.O. Iakushev, O.V. Shulika, I.A. Sukhoivanov, Passive nonlinear reshaping towards parabolic pulses in the steady-state regime in optical fibers, Opt. Commun., 285 (2012) 4493-4499.

[34] T. Jannson, Real-time Fourier transformation in dispersive optical fibers, Opt. Lett., 8 (1983) 232-234.

[35] A.V. Zohrabian, L.K. Mouradian, Compression of the spectrum of picosecond ultrashort pulses, Quantum Electron., 25 (1995) 1076.

[36] C. Finot, S. Boscolo, Design rules for nonlinear spectral compression in optical fibers, J. Opt. Soc. Am. B, 33 (2016) 760-767.