Supplemental Material for "Data-driven model reconstruction for nonlinear wave dynamics"

I. METHODOLOGY FOR DYNAMIC MODEL IDENTIFICATION

This section provides a description of the machine learning method employed to extract the underlying equations governing the system's dynamics from the simulated data. The method uses regression techniques to identify a sparse set of terms from a comprehensive library of potential functions that best describe the system's temporal or spatial evolution. This approach allows us to discover the governing equations and their coefficients.

We first collect data obtained from numerical simulations of the system using Eq. (1) in the main text, considering the geometry presented in Fig. 1 (leftmost column). These data are structured as a series of field profiles that depend on the coordinates (z, x). Their first-order derivatives along the z-axis are denoted as $\mathbf{Y}$ in Fig. 1 of the main text. To perform regression, we construct a library of functions, forming the matrix $\mathbf{X}$. This library consists of various functional terms, including spatial derivatives of different orders along the x-axis and nonlinear terms involving the field values (see also the next sections in the Supplemental Material). These terms serve as potential building blocks of the governing equation.

With the input data and the library of functions defined, we proceed to solve an optimization problem to identify the coefficients that best represent the underlying dynamics. The goal is to find a sparse vector of coefficients, $\mathbf{c}$, that minimizes the error between the observed data and the model predicted by the library functions. Mathematically, this is formulated as finding the coefficients such that $\mathbf{Xc}$ approximates $\mathbf{Y}$ in the least-squares sense, expressed by the minimization problem: $\min_{\mathbf{c}} \|\mathbf{Xc} - \mathbf{Y}\|^2$. To solve this optimization problem, we use standard linear regression, implemented via the `LinearRegression` function from the `sklearn` library in Python.

This standard linear regression returns the coefficient vector $\mathbf{c}$, which represents the weights of the terms in the library that describe the system's dynamics. The goal of our method is to approximate the dynamical equations, even when the relevant terms are not known a priori. This implies that $\mathbf{c}$ should be sparse, meaning many of its entries should be zero. The sparsity of $\mathbf{c}$ determines the relevance of the terms in the library: when an entry in $\mathbf{c}$ is zero (or very small), the corresponding term in $\mathbf{X}$ is considered irrelevant for reconstructing the system's dynamics.

II. PROCEDURE DETAILS FOR THE LINEAR LOW-INTENSITY REGIME

In the linear regime, the governing equation for the envelope amplitude $\tilde{\mathcal{A}}$,

$$i\frac{\partial \tilde{\mathcal{A}}}{\partial z} = (\beta_0 + i\beta_{0I})\tilde{\mathcal{A}} + i(v + iv_I)\frac{\partial \tilde{\mathcal{A}}}{\partial x} - (\eta + i\eta_I)\frac{\partial^2 \tilde{\mathcal{A}}}{\partial x^2} + (\eta' + i\eta_I')\frac{\partial^3 \tilde{\mathcal{A}}}{\partial x^3}\,, \tag{S1}$$

can be rewritten in terms of real and imaginary parts $\tilde{\mathcal{A}} \equiv \tilde{\mathcal{A}}_R + i\tilde{\mathcal{A}}_I$,

$$\begin{aligned}
\frac{\partial \tilde{\mathcal{A}}_R}{\partial z} &= \beta_0\tilde{\mathcal{A}}_I + \beta_{0I}\tilde{\mathcal{A}}_R + v\frac{\partial \tilde{\mathcal{A}}_R}{\partial x} - v_I\frac{\partial \tilde{\mathcal{A}}_I}{\partial x} - \eta\frac{\partial^2 \tilde{\mathcal{A}}_I}{\partial x^2} - \eta_I\frac{\partial^2 \tilde{\mathcal{A}}_R}{\partial x^2} + \eta'\frac{\partial^3 \tilde{\mathcal{A}}_I}{\partial x^3} + \eta_I'\frac{\partial^3 \tilde{\mathcal{A}}_R}{\partial x^3}\,,\\
\frac{\partial \tilde{\mathcal{A}}_I}{\partial z} &= -\beta_0\tilde{\mathcal{A}}_R + \beta_{0I}\tilde{\mathcal{A}}_I + v\frac{\partial \tilde{\mathcal{A}}_I}{\partial x} + v_I\frac{\partial \tilde{\mathcal{A}}_R}{\partial x} + \eta\frac{\partial^2 \tilde{\mathcal{A}}_R}{\partial x^2} - \eta_I\frac{\partial^2 \tilde{\mathcal{A}}_I}{\partial x^2} - \eta'\frac{\partial^3 \tilde{\mathcal{A}}_R}{\partial x^3} + \eta_I'\frac{\partial^3 \tilde{\mathcal{A}}_I}{\partial x^3}.
\end{aligned}$$

The aim of using a machine learning (ML) approach with linear regression is to identify the structure of the governing equations (relevant terms) and determine the exact values of the coefficients.

First, we prepare datasets with three beams of different widths as the initial condition: $\mathcal{L}_1 = 3a, \mathcal{L}_2 = 3a/2, \mathcal{L}_3 = 3a/2 \times 1.5$. These datasets were generated through solving the paraxial equation using beam propagation method for different widths for the same input intensity and then mixing the data for all three widths. While constructing $\mathbf{X}$, we multiply spatial derivatives on $\mathcal{L}_3$, such as $\partial_x \propto \mathcal{L}_3$. Then we randomly select the data points where the intensity exceeds the threshold value of 0.05.

The training set is divided into 100 validation folds, each consisting of approximately 40,000-60,000 points. For every dataset within these folds, the coefficients are extracted, followed by the calculation of their mean value, denoted as $\langle \mathbf{c}_i \rangle$, and standard deviation, $\delta\mathbf{c}_i$. This rigorous process ensures a robust statistical analysis of the model coefficients across the various validation sets.

dw	set	k	$\langle\beta_0\rangle + \delta\beta_0$ (1/mm)	$\langle v\rangle + \delta v$ $(\cdot 10^{-2})$	$\langle\eta\rangle + \delta\eta$ $(\cdot 10^{-4}$mm)	$\langle\eta_I'\rangle + \delta\eta_I'$ $(\cdot 10^{-6}$mm$^2)$
bearded	II	K_0	-0.18±5·10^{-5}	0.61 ± 0.0006	0.43 ± 0.002	-0.24 ± 0.01
bearded	II	K_1	-0.25±4·10^{-5}	0.52 ± 0.0004	0.43 ± 0.002	0.09 ± 0.009
bearded	II	K_2	-0.09±1·10^{-5}	0.7 ± 0.001	0.52 ± 0.005	-0.53 ± 0.02
bearded	I	K_0	-0.77 ± 0.0001	0.86 ± 0.002	0.54 ± 0.008	0.13 ± 0.03
bearded	I	K_1	-0.87 ± 0.0001	0.72 ± 0.002	0.61 ± 0.007	0.21 ± 0.03
bearded	I	K_2	-0.66 ± 0.0001	0.96 ± 0.001	0.49 ± 0.008	-0.11 ± 0.03
zig-zag	I	K_0	-0.82 ± 0.0001	0.79 ± 0.002	0.09 ± 0.01	0.96 ± 0.06
zig-zag	I	K_1	-0.9 ± 0.0001	0.71 ± 0.002	0.42 ± 0.009	0.79 ± 0.04
zig-zag	I	K_2	-0.73 ± 0.0001	0.69 ± 0.002	-0.33 ± 0.01	-0.33 ± 0.07
zig-zag	II	K_0	-0.67±4·10^{-5}	0.23 ± 0.0005	-0.15 ± 0.002	0.42 ± 0.01
zig-zag	II	K_1	-0.64±4·10^{-5}	0.17±0.0005	-0.29 ±0.002	0.31 ± 0.01
zig-zag	II	K_2	-0.7±5·10^{-5}	0.24 ± 0.0007	0.01 ± 0.0027	0.35 ± 0.01
zig-zag	II	K_3	-0.77±7·10^{-5}	$5 \cdot 10^{-5}$±0.0007	0.31±0.003	-0.0002±0.01

TABLE SI. Linear coefficients $\langle\mathbf{c}_i\rangle$ and their possible ranges, $\langle\mathbf{c}_i\rangle \pm \delta\mathbf{c}_i$, where $\delta\mathbf{c}_i$ is defined as the standard deviation, extracted from linear datasets at different wave numbers k along the edge-state dispersion curve, denoted as $K_0 = 4\pi/(3a)$, $K_1 = (4\pi/3 + 0.4)/a$, $K_2 = (4\pi/3 - 0.4)/a$, $K_3 = 2\pi/a$. "dw" stands for the domain wall, and the column "set" refers to the parameter sets listed in Table I of the main text.

We conduct a ML analysis for three different library function choices. This approach enables a comprehensive evaluation of model performance and helps identify the most suitable library function for the dataset, avoiding both overfitting and underfitting.

In the first library, we analyze each term for the real and imaginary components independently. This step enables us to recognize potential combinations between the real and imaginary parts.

In the second phase, we analyze the coefficients for suitable combinations of these real and imaginary terms collectively, as $[\partial_z\tilde{\mathcal{A}}_R, \partial_z\tilde{\mathcal{A}}_I]$, $[\tilde{\mathcal{A}}_I, -\tilde{\mathcal{A}}_R]$, $[\tilde{\mathcal{A}}_R, \tilde{\mathcal{A}}_I]$, $[\partial_x\tilde{\mathcal{A}}_R, \partial_x\tilde{\mathcal{A}}_I]$, $[-\partial_x\tilde{\mathcal{A}}_I, \partial_x\tilde{\mathcal{A}}_R]$, $[-\partial_{xx}\tilde{\mathcal{A}}_I, \partial_{xx}\tilde{\mathcal{A}}_R]$, $[-\partial_{xx}\tilde{\mathcal{A}}_R, \partial_{xx}\tilde{\mathcal{A}}_I]$, $[\partial_{xxx}\tilde{\mathcal{A}}_I, \partial_{xxx} - \tilde{\mathcal{A}}_R]$, $[\partial_{xxx}\tilde{\mathcal{A}}_R, -\partial_{xxx}\tilde{\mathcal{A}}_I]$.

In the final phase, we focus solely on four largest specific terms $[\partial_z\tilde{\mathcal{A}}_R, \partial_z\tilde{\mathcal{A}}_I]$, $[\tilde{\mathcal{A}}_I, -\tilde{\mathcal{A}}_R]$, $[\partial_x\tilde{\mathcal{A}}_R, \partial_x\tilde{\mathcal{A}}_I]$, $[-\partial_{xx}\tilde{\mathcal{A}}_I, \partial_{xx}\tilde{\mathcal{A}}_R]$.

We rigorously validate the extracted coefficients not only by assessing their performance on test datasets but also by comparing the paraxial data against the numerical solutions of the extracted equations. Additionally, we enhance the analysis by fitting a dispersion curve by $\omega = \beta_0 - v(k - k_0) + \eta(k - k_0)^2 + \eta_I'(k - k_0)^3$ as well.

III. PROCEDURE DETAILS FOR THE NONLINEAR HIGH-INTENSITY REGIME

We prepared data with three intensity values at the input. We then truncate the propagation paths from $z_0 \approx 7.5$ mm to $z_1 \approx 23$ mm. To ensure accurate analysis, we select these propagation segments where the nonlinear edge wave has fully formed (given that at the initial $z = 0$ we set a linear transverse profile of the edge mode, we exclude the initial transitory stage of propagation). These segments are also constrained to regions where the assumptions of the weakly nonlinear wave (the nonlinear frequency shift is smaller than the gap size) and smooth profile hold. Additionally, they are short enough to minimize interactions between the formed nonlinear edge wave and the bulk modes.

Subsequently, in the preparation of $\mathbf{X}$, the data is normalized to $\rho_n = 0.4$, and only points with intensity exceeding 0.1 are selected. The dataset is then segmented into 100 validation samples, similar to the linear case, each containing around $30,000$ points, where the coefficients are calculated and averaged.

Further, we explore various function libraries and assess the error on the respective test datasets using the mean absolute percentage error function for both phase and intensity equations. Our findings indicate that incorporating all three terms with the nonlinear group velocity results in a significant standard deviation in coefficient determination, suggesting overfitting. Hence, sets of functions (2-4) are more favorable for achieving a more robust model. Our analysis revealed that library sets 2 and 3 exhibited the lowest error values on the test dataset.

Function library 1: $[\mathcal{B}\cos\theta; \mathcal{B}\sin\theta] = [v_{g1}\rho^1\frac{\partial\rho^2}{\partial x}, v_{g2}\rho^3\frac{\partial\rho^2}{\partial x}, v_{g3}\rho^5\frac{\partial\rho^2}{\partial x}, \gamma_1\rho^3, \gamma_2\rho^5; G_1\rho^3, G_2\rho^5]$;

Function library 2: $[\mathcal{B}\cos\theta; \mathcal{B}\sin\theta] = [v_{g1}\rho^1\frac{\partial\rho^2}{\partial x}, \gamma_1\rho^3, \gamma_2\rho^5; G_1\rho^3, G_2\rho^5]$;

Function library 3: $[\mathcal{B}\cos\theta; \mathcal{B}\sin\theta] = [v_{g2}\rho^3\frac{\partial\rho^2}{\partial x}, \gamma_1\rho^3, \gamma_2\rho^5; G_1\rho^3, G_2\rho^5]$;

Function library 4: $[\mathcal{B}\cos\theta; \mathcal{B}\sin\theta] = [v_{g3}\rho^5\frac{\partial\rho^2}{\partial x}, \gamma_1\rho^3, \gamma_2\rho^5; G_1\rho^3, G_2\rho^5]$;

Function library 5: $[\mathcal{B}\cos\theta; \mathcal{B}\sin\theta] = [v_{g1}\rho^1\frac{\partial\rho^2}{\partial x}, v_{g2}\rho^3\frac{\partial\rho^2}{\partial x}, \gamma_1\rho^3, \gamma_2\rho^5; G_1\rho^3, G_2\rho^5]$;
Function library 6: $[\mathcal{B}\cos\theta; \mathcal{B}\sin\theta] = [\gamma_1\rho^3, \gamma_2\rho^5; G_1\rho^3, G_2\rho^5]$;
Function library 7: $[\mathcal{B}\cos\theta; \mathcal{B}\sin\theta] = [v_{g2}\rho^3\frac{\partial\rho^2}{\partial x}; G_1\rho^3]$.

library set/dw	$\langle G_1\rangle \pm \delta G_1$ $[(0.4)^2\text{mm}^{-1}$ $(\cdot\text{m}^2/10^{16}\text{W})^2]$	$\langle G_2\rangle \pm \delta G_2$ $[(0.4)^4\text{mm}^{-1}$ $(\cdot\text{m}^2/10^{16}\text{W})^4]$	$\langle v_{g1}\rangle \pm \delta v_{g1}$ $[(0.4)^2\cdot$ $(\cdot\text{m}^2/10^{16}\text{W})^2]$	$\langle v_{g2}\rangle \pm \delta v_{g2}$ $[(0.4)^4\cdot$ $(\cdot\text{m}^2/10^{16}\text{W})^4]$	$\langle v_{g3}\rangle \pm \delta v_{g3}$ $[(0.4)^6\cdot$ $(\cdot\text{m}^2/10^{16}\text{W})^6]$	$\langle\gamma_1\rangle \pm \delta\gamma_1$ $[(0.4)^2\text{mm}^{-1}$ $(\cdot\text{m}^2/10^{16}\text{W})^2]$	$\langle\gamma_2\rangle \pm \delta\gamma_2$, $[(0.4)^4\text{mm}^{-1}$ $(\cdot\text{m}^2/10^{16}\text{W})^4]$
lib 1/ bearded	0.13 ± 0.0002	-0.0007 ± 0.0001	0.03 ± 0.004	0.003 ± 0.007	-0.003 ± 0.003	0.003 ± 0.0002	-0.002 ± 0.0001
lib 2/ bearded	0.13 ± 0.0002	-0.0007 ± 0.0001	0.03 ± 0.001	x	x	0.003 ± 0.0002	-0.002 ± 0.0001
lib 3/ bearded	0.13 ± 0.0002	-0.0007 ± 0.0001	x	0.02 ± 0.001	x	0.003 ± 0.0002	-0.002 ± 0.0001
lib 4/ bearded	0.13 ± 0.0002	-0.0007 ± 0.0001	x	x	0.007 ± 0.0004	0.003 ± 0.0002	-0.002 ± 0.0001
lib 5/ bearded	0.13 ± 0.0002	-0.0007 ± 0.0001	0.03 ± 0.004	0.003 ± 0.007	x	0.003 ± 0.0002	-0.002 ± 0.0001
lib 6/ bearded	0.13 ± 0.0002	-0.0007 ± 0.0001	x	x	x	0.03 ± 0.004	0.003 ± 0.007
lib 7/ bearded	0.13 ± $\cdot 10^{-5}$	x	x	0.03 ± 0.004	x	x	x
lib 1/ zig-zag	0.015 ± 0.0002	0.013 ± 0.0002	0.32 ± 0.006	-0.21 ± 0.013	0.096 ± 0.007	0.003 ± 0.0002	-0.004 ± 0.0001
lib 2/ zig-zag	0.015 ± 0.0002	0.013 ± 0.0002	0.22 ± 0.001	x	x	0.003 ± 0.0002	-0.004 ± 0.0001
lib 3/ zig-zag	0.015 ± 0.0002	0.013 ± 0.0002	x	0.2 ± 0.001	x	0.003 ± 0.0002	-0.004 ± 0.0001
lib 4/ zig-zag	0.015 ± 0.0002	0.013 ± 0.0002	x	x	0.14 ± 0.001	0.003 ± 0.0002	-0.004 ± 0.0002
lib 5/ zig-zag	0.015 ± 0.0002	0.013 ± 0.0002	0.32 ± 0.006	-0.21 ± 0.01	x	0.003 ± 0.0002	-0.004 ± 0.0001
lib 6/ zig-zag	0.015 ± 0.0002	0.013 ± 0.0002	x	x	x	0.3 ± 0.006	-0.2 ± 0.01
lib 7/ zig-zag	0.03 ± $6\cdot 10^{-5}$	x	x	0.32 ± 0.006	x	x	x

TABLE SII. Nonlinear coefficients $\langle\mathbf{c}_i\rangle \pm \delta\mathbf{c}_i$, representing the mean value ± standard deviation. As in Table SI, "dw" stands for domain wall, and the symbol "x" denotes terms that are not included in the corresponding library.

Next, we investigated the effect of small variations in accuracy, calculated using the mean absolute percentage error, when selecting libraries 2, 3, and 4. To do so, we analyzed the reconstruction of the Hopf equation itself. This equation was numerically solved in the form of library 3,

$$i\frac{\partial\tilde{\mathcal{A}}}{\partial z} = iv\frac{\partial\tilde{\mathcal{A}}}{\partial x} - \eta\frac{\partial^2\tilde{\mathcal{A}}}{\partial x^2} + i\eta_I'\frac{\partial^3\tilde{\mathcal{A}}}{\partial x^3} + \beta_0\tilde{\mathcal{A}} - G_1|\tilde{\mathcal{A}}|^2\tilde{\mathcal{A}} - iv_{g2}|\tilde{\mathcal{A}}|^2\tilde{\mathcal{A}}\frac{\partial|\tilde{\mathcal{A}}|^2}{\partial x}. \tag{S2}$$

Subsequently, we fitted all the aforementioned libraries. We compared the results of the evolution from the original simulation with the evolution from the extracted equations. In the initial stages of propagation, the difference is minimal, and it is possible to fit a coefficient equally well for any type of nonlinear group velocity. However, over longer distances, the coefficients of library 3 provide the best fit. Therefore, we conclude that for additional verification, it is necessary to observe the evolution over an extended distance, particularly in its final stages.

To further differentiate between libraries 2 and 3, we conducted numerical simulations using the extracted coefficients. Although both libraries produced solutions exhibiting minor discrepancies from the paraxial data, library 3 demonstrated greater consistency and appears to be a more accurate representation.

Our findings demonstrate that in the nonlinear regime, the dominant factor shaping the propagation of a Gaussian pulse depends on the dispersion strength. For weak dispersion, characteristic of the zig-zag domain wall, the nonlinear group velocity is the primary distorting factor, resulting in an asymmetric pulse profile. Conversely, for strong second-order dispersion, typical of the bearded domain wall, the nonlinear phase shift dominates, causing symmetrical compression of the pulse and an increase in its peak intensity.