

On the Galerkin/Finite-Element Method for the Serre Equations

Dimitrios Mitsotakis · Boaz Ilan · Denys Dutykh

Received: 13 June 2013 / Revised: 12 November 2013 / Accepted: 19 January 2014 /
Published online: 5 February 2014
© Springer Science+Business Media New York 2014

Abstract A highly accurate numerical scheme is presented for the Serre system of partial differential equations, which models the propagation of dispersive shallow water waves in the fully-nonlinear regime. The fully-discrete scheme utilizes the Galerkin / finite-element method based on smooth periodic splines in space, and an explicit fourth-order Runge–Kutta method in time. Computations compared with exact solitary and cnoidal wave solutions show that the scheme achieves the optimal orders of accuracy in space and time. These computations also show that the stability of this scheme does not impose very restrictive conditions on the temporal stepsize. In addition, solitary, cnoidal, and dispersive shock waves are studied in detail using this numerical scheme for the Serre system and compared with the ‘classical’ Boussinesq system for small-amplitude shallow water waves. The results show that the interaction of solitary waves in the Serre system is more inelastic. The efficacy of the numerical scheme for modeling dispersive shocks is shown by comparison with asymptotic results. These results have application to the modeling of shallow water waves of intermediate or large amplitude.

D. Mitsotakis (✉) · B. Ilan
School of Natural Sciences, University of California, Merced, 5200 North Lake Road,
Merced, CA 95343, USA
e-mail: dmitsot@gmail.com
URL: <http://dmitsot.googlepages.com>

B. Ilan
e-mail: bilan@ucmerced.edu
URL: <http://faculty.ucmerced.edu/bilan>

D. Dutykh
School of Mathematical Sciences, University College Dublin, Belfield, Dublin 4, Ireland
e-mail: Denys.Dutykh@ucd.ie
URL: <http://www.denys-dutykh.com>

D. Dutykh
LAMA, UMR 5127 CNRS, Université de Savoie, Campus Scientifique,
73376 Le Bourget-du-Lac Cedex, France

Keywords Green–Naghdi equations · Traveling waves · Undular bores

Mathematics Subject Classification 76B15 · 76B25 · 65M30

1 Introduction

The propagation of waves on the free surface of an ideal irrotational fluid under the force of gravity are governed by the Euler’s equations [42]. Solving the Euler’s equations is very difficult, because of the free surface. There is a hierarchy of asymptotic approximations of the Euler’s equations that do not depend on a free surface. In particular, the propagation of waves in shallow water is governed by the Serre equations, also known as the Su–Gardner equations or Green–Naghdi equations (cf. [25,32,37–39]), which we shall refer to as the Serre system. In dimensionless variables it reads

$$\eta_t + u_x + \varepsilon(\eta u)_x = 0, \tag{1a}$$

$$u_t + \eta_x + \varepsilon u u_x - \frac{\sigma^2}{3h} [h^3(u_{xt} + \varepsilon u u_{xx} - \varepsilon(u_x)^2)]_x = 0, \tag{1b}$$

where

$$h(x, t) \doteq 1 + \varepsilon \eta. \tag{2}$$

Here x is the spatial variable, t is the time, $u(x, t)$ is the depth-averaged horizontal velocity of the fluid, $\varepsilon \eta(x, t)$ is the wave height above an undisturbed level of zero elevation, $h(x, t)$ is the total depth of the fluid with respect to a horizontal bottom (at a normalized elevation of -1 from the undisturbed water level), σ is the ratio between the typical depth d and wavelength λ , and ε is the ratio between the typical amplitude a and bottom depth d (i.e. $\sigma = d/\lambda$ and $\varepsilon = a/d$). See sketch in Fig. 1.

The Serre system can be derived as an asymptotic approximation of the Euler’s equations under the assumption of shallow water waves, or long wavelength or weakly dispersive regime, i.e., $\sigma \ll 1$. Importantly, System (1) does not assume small amplitude waves, i.e., ε can be large.

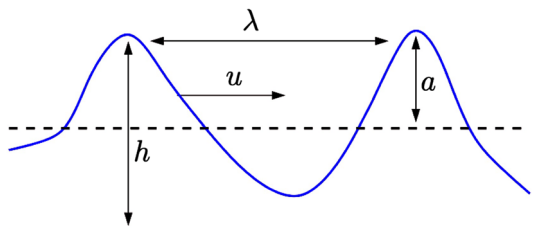
For *small-amplitude* or weakly nonlinear shallow water waves, i.e., when $\varepsilon \ll 1$, $\sigma \ll 1$, and $\varepsilon/\sigma^2 = O(1)$, the Serre system reduces to the ‘**classical**’ Boussinesq (cB) system (cf. [7] for related equations),

$$\eta_t + u_x + \varepsilon(\eta u)_x = 0, \tag{3a}$$

$$u_t + \eta_x + \varepsilon u u_x - \frac{\sigma^2}{3} u_{xxt} = 0. \tag{3b}$$

Comparing the cB and Serre systems, Eqs. (3a) and (1a) are the same. This equation is known as the mass conservation equation. However, Eq. (1b) contains higher order nonlinear-

Fig. 1 Sketch of surface water waves for the Serre (1) and cB (3) systems. u is depth-averaged horizontal velocity, $\varepsilon \eta$ is wave height above an undisturbed level (dashes), h is bottom depth, and λ a typical wavelength



dispersive terms compared with (3b). For this reason, the Serre system is often called **fully-nonlinear** shallow-water equations. Though the two systems are similar, the small amplitude assumption underlying the cB system appears to be too restrictive for model waves of large, or even intermediate, amplitude. Physically, when water waves approach regions of small depth, it is common that their amplitude increases. Therefore, the Serre system is potentially more appropriate for the approximation of long waves in shallow water and also in the nearshore zone.

There are many studies of the cB system and related Boussinesq-type systems. However, there are much fewer studies of the Serre system, in part because it is considerably harder to solve numerically. As a result, many properties of the solutions of the Serre system are unknown or remain unclear. In particular, few numerical schemes have been developed for the Serre system, based on either the finite difference (cf. [36]), hybrid finite-difference finite-volume (cf. [9, 12]), and pseudospectral (cf. [20]) methods. All of these methods can be **formally** very accurate. However, they suffer from aliasing or spurious dissipative effects, due to the approximation of the nonlinear terms. This is usually not a major handicap when solving the cB system, but it becomes debilitating when solving the Serre system for large amplitude and / or rapidly oscillating waves, as in such cases the actual error is large, even when using a fine grid. Even worse, such schemes often fail to converge.

In this study, we design and implement a computational scheme based on the standard **Galerkin / Finite-Element Method** (FEM). One of the main advantages of this method is that it is **non-dissipative**. Another advantage is the sparsity of the resulting linear systems. For this reason, as we show, the FEM scheme for (1) achieves the optimal (formal) order of accuracy even for large amplitude solitary, cnoidal and dispersive shock waves. To achieve this, we use cubic splines for the semi-discretization in space and the classical fourth order **Runge–Kutta** (RK) method for time integration. A similar scheme was studied for a variety of Boussinesq-like systems, cf. [18], and appears to be highly accurate and efficient in a measure that makes the method to be ideal for the study of the dynamics of solitary waves. However, for the Serre system, the dependence of the dispersive terms in (1b) on the unknown function $h(x, t)$ makes the numerical integration considerably more difficult compared with Boussinesq-like systems. In particular, for time integration, a **mass matrix** needs to be assembled at each time step. At every intermediate RK step, two linear systems based on the mass matrices need to be solved. This is a costly computation, yet, in spite of this drawback, the high accuracy of this scheme makes it a strong candidate for computational modeling of the Serre system.

The paper is organized as follows. Section 2 recaps some of the analytical results for solitary, cnoidal, and dispersive shock wave solutions of the Serre system, which serve to validate the computational scheme. Section 3 presents the fully-discrete schemes for the Serre and cB systems. Section 4 validates the convergence, accuracy, and numerical stability of the method. Section 5 presents computational studies of interacting solitary waves and dispersive shock waves.

1.1 Remarks

System (3) was originally derived (in a slightly different form) by Boussinesq [10] and it is a special case of a class of Boussinesq-type models derived by Bona, Chen and Saut, [7]. This system is often used for studying two-way propagation of small amplitude, long waves [34].

The Serre system (1) was originally derived from the Euler's equations in one spatial dimension by Serre [37, 38]. It was re-derived several times later, including independently by Su and Gardner [39]. See also the review by Barthélemy [5]. Green and Naghdi [25]

generalized this system to two spatial dimensions with an uneven bottom, see also [36,40]. Recently, Lannes and Bonneton [30] derived and justified several asymptotic models of surface waves including (1). It is worth mentioning that other systems of a similar ilk have been derived, for example, with improved dispersion characteristics [30,43], with moving bottom [29] and with surface tension [17]. In principle, all these models can be discretized by numerical methods similar to the methods presented below. However, further studies will be required to test the efficacy of the ensuing schemes.

2 Analytical Properties of the Serre System

Below we recap several analytical properties of System (1), which serve as benchmarks for our computational scheme.

2.1 Special Solutions of the Serre System

System (1) admits solitary and cnoidal wave solutions in closed form (cf. [11,22,37,38]). Below we recap these solutions for arbitrary ε and σ . It is convenient to express the solutions of (1) in terms of (h, u) rather than (η, u) . The two ways are equivalent in light of (2).

The two-parameter family of solitary wave solutions of (1) that travel with a constant speed c_s can then be written in the moving frame $\xi = x - c_s t$ as

$$h_{sol}(\xi) = \frac{1}{\sigma} [a_0 + a_1 \operatorname{sech}^2(K_s \xi)] , \tag{4a}$$

$$u_{sol}(\xi) = \frac{c_s}{\varepsilon} \left[1 - \frac{a_0}{\sigma h_{sol}(\xi)} \right] , \tag{4b}$$

where

$$K_s = \sqrt{\frac{3a_1}{4\sigma a_0^2 c_s^2}} , \quad c_s = \sqrt{\frac{a_0 + a_1}{\sigma}} ,$$

and a_0, a_1 are positive (but otherwise arbitrary) constants. Choosing $a_0 = \sigma$ gives the solitary waves that decay to the background average water depth. For the simulations, we choose $\varepsilon = a_0 = \sigma = 1$ and various values of the speed c_s . Then a_1 and K_s can be determined from the above formulae.

Similarly, the three-parameter cnoidal wave solutions of (1) can be written as

$$h_c(\xi) = \frac{1}{\sigma} [a_0 + a_1 \operatorname{dn}^2(K_c \xi, k)] , \tag{5a}$$

$$u_c(\xi) = \frac{c_s}{\varepsilon} \left[1 - \frac{h_0}{\sigma h_c(\xi)} \right] , \tag{5b}$$

where dn denotes the Jacobi elliptic function and

$$h_0 = a_0 + a_1 \frac{E(m)}{K(m)} ,$$

$$K_c = \frac{\sqrt{3a_1}}{2\sqrt{a_0(a_0 + a_1)(a_0 + (1 - k^2)a_1)}} ,$$

$$c_s = \sqrt{\frac{a_0(a_0 + a_1)(a_0 + (1 - k^2)a_1)}{\sigma h_0^2}} .$$

$K(m)$ and $E(m)$ denote the complete elliptic integrals of the first and second kind, respectively, $k \in [0, 1]$, $m = k^2$, and a_0, a_1 are positive constants. Here, a_0, a_1 and m (or k) are arbitrary.

It is remarkable that such closed form solutions of the Serre system have been found and even more remarkable that such closed form solutions *have not* been found for the cB system (3). Nevertheless, it has been proven that the cB system admits solitary and cnoidal wave solutions (cf. [13, 14]).

2.2 Dispersive Shock Waves

When the dispersive terms in the Serre or cB systems are neglected, the resulting non-dispersive equations can give rise to shock waves, i.e., a discontinuous solution. When such shocks are regularized by dissipative effects, this gives rise to classical or viscous shocks, which are characterized by a rapid and monotonic change in the flow properties. On the other hand, in systems where dissipation is negligible compared with dispersion, the dispersive effects give rise to **dispersive shock waves** (DSWs). DSWs are characterized by an expanding train of rapidly-oscillating waves (see sketch in Fig. 2). The leading edge of a DSW possesses large amplitude waves, which decay to linear waves at the trailing edge. DSWs have been studied for several decades, originally in the context of the KdV equation (cf. [6, 26, 41] for some of the early works). In particular, these studies show that, using Whitham’s averaging method, the largest-amplitude oscillation in the leading edge is well-approximated with a solitary wave.

Recently, DSWs were studied analytically and computationally in fully-nonlinear dispersive shallow water systems (cf. [22–24, 33]). In particular, the asymptotic dynamics of the leading edge solitary wave of a simple DSW were studied in [22, 23]. Below we recap some of those results. We use these results to test the non-dissipativity of the numerical schemes.

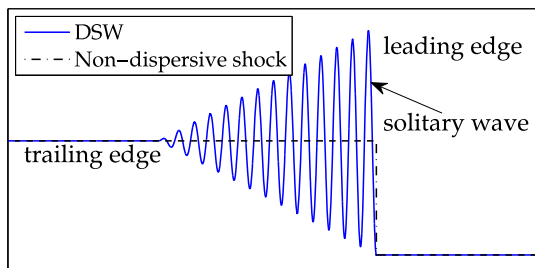
Consider the Serre system (1) with $\varepsilon = \sigma = 1$. As above, writing the solution of the Serre system in terms of h and u , a **simple DSW** traveling to the right is generated using the Riemann initial data

$$h(x, 0) = \begin{cases} h^-, & \text{for } x < 0 \\ h^+, & \text{for } x > 0 \end{cases}, \quad u(x, 0) = \begin{cases} u^-, & \text{for } x < 0 \\ u^+, & \text{for } x > 0 \end{cases}, \tag{6}$$

with the compatibility condition (Riemann invariant)

$$\frac{u^-}{2} - \sqrt{h^-} = \frac{u^+}{2} - \sqrt{h^+}. \tag{7}$$

Fig. 2 Sketch of dispersive shock wave (DSW, *solid*) and non-dispersive shock (*dashes*)



Following [22,23], we assume that the initial jump or total depth variation is small, i.e.,

$$0 < \delta \doteq \frac{h^-}{h^+} - 1 \ll 1. \tag{8}$$

Then, to leading order in δ and for a large propagation time, the leading-edge of the DSW is well approximated with a solitary wave given by (4) with $a_0 = 1$ and amplitude (relative to the constant elevation) and speed

$$a_1^{\text{simple}} \sim 2\delta + \frac{1}{6}\delta^2, \tag{9a}$$

$$c_s^{\text{simple}} \sim 1 + \delta - \frac{5}{12}\delta^2, \tag{9b}$$

respectively.

We also consider the **dam-break problem**. In this case, the initial data for $h(x, 0)$ are the same (6). However, there is no flow at $t = 0$, i.e., $u(x, 0) = 0$. As shown in [22], this generates two counter-propagating DSWs, one on each side of the “dam”, and two rarefaction waves that travel towards the center. Similarly to a simple DSW, the asymptotic amplitude and speed for the leading-edge solitary wave in each DSW is

$$a_1^{\text{dam}} \sim \delta - \frac{1}{12}\delta^2, \tag{10a}$$

$$c_s^{\text{dam}} \sim 1 + \frac{\delta}{2} - \frac{1}{6}\delta^2, \tag{10b}$$

respectively.

2.3 Hamiltonian Conservation

A fundamental property of the Serre system is its Hamiltonian structure, cf. [28,31]. For any solution (h, u) , the energy functional (or Hamiltonian)

$$\mathcal{H}(t) \equiv \mathcal{H}[h, u] = \int_{-\infty}^{\infty} \left[\varepsilon h u^2 + \frac{\varepsilon \sigma^2}{3} h^3 u_x^2 + \frac{1}{\varepsilon} (h - 1)^2 \right] dx, \tag{11}$$

is conserved in the sense that $\mathcal{H}(t) = \mathcal{H}(0)$ for all $t > 0$ and up to the maximal time T of the existence of the solution. In contrast, the cB system (3) does not possess a Hamiltonian structure and its solutions do not conserve an energy functional [7].

3 The FEM Scheme

In this section we present a FEM for the initial-boundary value problem (IBVP) comprised of System (1) subject to periodic boundary conditions. Here and in the computations we choose $\varepsilon = \sigma = 1$. We make this choice in order to simulate large amplitude waves, so as to test the scheme to its limit. For this reason, the ε and σ are dropped from the equations below. It is also convenient to rewrite (1a) in terms of (h, u) rather than (η, u) . This is done using (2) and yields the IBVP

$$h_t + (hu)_x = 0, \tag{12a}$$

$$u_t + h_x + uu_x - \frac{1}{3h} [h^3(u_{xt} + uu_{xx} - (u_x)^2)]_x = 0, \tag{12b}$$

$$\partial_x^i h(a, t) = \partial_x^i h(b, t), \quad i = 0, 1, 2, \dots, \tag{12c}$$

$$\partial_x^i u(a, t) = \partial_x^i u(b, t), \quad i = 0, 1, 2, \dots, \tag{12d}$$

$$h(x, 0) = h_0(x), \tag{12e}$$

$$u(x, 0) = u_0(x), \tag{12f}$$

where $x \in (a, b) \subset \mathbb{R}$ and $t \in [0, T]$. We shall assume that (12) possesses a unique solution, such that h and u are sufficiently smooth and, for any $t \in [0, T]$, in a suitable Sobolev space with periodic boundary conditions, i.e.,

$$h(x, \cdot) \in H^s, \quad u(x, \cdot) \in H^{s+1}, \quad H^s \equiv H_{per}^s(a, b),$$

where $s \geq 1$ (see [27] for sharper results). Here and below, $\|\cdot\|_s$ denotes the standard norm in H^s . We also use the inner product in $L^2 \equiv H^0$, denoted by (\cdot, \cdot) .

3.1 Spatial Discretization

We denote the spatial grid by $x_i = a + i \Delta x$, where $i = 0, 1, \dots, N$, Δx is the grid size, and $N \in \mathbb{N}$, such that $\Delta x = (b - a)/N$. Let (\tilde{h}, \tilde{u}) be the corresponding spatially discretized solution. The **Galerkin / Finite-Element Method** (FEM) seeks a weak solution of (12), i.e., $\tilde{h}(x, t)$ and $\tilde{u}(x, t)$ in $C^1(0, T; S)$, for a suitable finite-dimensional space S . We shall consider the space of the smooth periodic splines

$$S^r = \{ \phi \in C_{per}^{r-2}[a, b] \mid \phi|_{[x_i, x_{i+1}]} \in \mathbb{P}^{r-1}, \quad 0 \leq i \leq N - 1 \},$$

where $r \in \mathbb{N}$,

$$C_{per}^r \doteq \{ f \in C^r[a, b] \mid f^{(k)}(a) = f^{(k)}(b), \quad 0 \leq k \leq r \},$$

and \mathbb{P}^k are the polynomials of degree k . In particular, we shall use cubic splines, which correspond to S^4 , i.e., $r = 4$. Henceforth, we shall denote $S \equiv S^4$. Note that the periodic boundary conditions (12c)–(12d) are satisfied automatically by this choice.

To state the associated weak problem, let $\phi \in S$ be an arbitrary test function. It turns out to be convenient to multiply (12b) by \tilde{h} and group together the first term in this equation with the first term in the brackets. Using integration by parts, gives the **semi-discrete problem**

$$(\tilde{h}_t, \phi) + ((\tilde{h}\tilde{u})_x, \phi) = 0, \tag{13a}$$

$$\mathcal{B}_{Serre}(\tilde{u}_t, \phi; \tilde{h}) + (\tilde{h}(\tilde{h}_x + \tilde{u}\tilde{u}_x), \phi) + \frac{1}{3} (\tilde{h}^3(\tilde{u}\tilde{u}_{xx} - (\tilde{u}_x)^2), \phi_x) = 0, \tag{13b}$$

where the bilinear form is defined for a fixed \tilde{h} (and $\phi, \psi \in S$) as

$$\mathcal{B}_{Serre}(\psi, \phi; \tilde{h}) \doteq (\tilde{h}\psi, \phi) + \frac{1}{3} (\tilde{h}^3\psi_x, \phi_x), \tag{13c}$$

and the initial conditions are

$$\tilde{h}(x, 0) = \mathcal{P}\{h_0(x)\}, \quad \tilde{u}(x, 0) = \mathcal{P}\{u_0(x)\}, \tag{13d}$$

where \mathcal{P} is the L^2 projection onto the S defined by $(\mathcal{P}v, \phi) = (v, \phi)$, for all $\phi \in S$.

The bilinear form (13c) plays a key role in the FEM. Assuming that h is bounded as $h_0 \leq h(x, t) \leq \infty$ for some constant $h_0 > 0$ (the so called “wet bottom” assumption), this bilinear form is bounded and coercive. Specifically, for all $\phi, \psi \in H^1$, there exist c_1, c_2 , such that

$$\mathcal{B}_{Serre}(\psi, \phi; \tilde{h}) \leq c_1 \|\phi\|_1 \|\psi\|_1, \quad \mathcal{B}_{Serre}(\phi, \phi; \tilde{h}) \geq c_2 \|\phi\|_1^2, \tag{14}$$

where $\|\cdot\|_1$ denotes the usual norm of the Sobolev space H^1 . (In general $\|\cdot\|_r$ denotes the usual norm of the Sobolev space H^r and $\|\cdot\|$ the norm of the space L^2 .) These two properties are of fundamental importance for the FEM scheme. In particular, it follows from (14) that the corresponding linear systems are not singular.

3.2 Temporal Discretization

Upon choosing appropriate basis functions for S , System (13) represents a system of ordinary differential equations (ODEs). For time integration, we employ the classical, explicit, four-stage, fourth-order **Runge–Kutta** (RK) method, which is described by the following *tableau*:

$$\frac{A}{\tau} \mid b = \begin{array}{cccc|c} & 0 & 0 & 0 & 0 & 1/6 \\ & 1/2 & 0 & 0 & 0 & 1/3 \\ & 0 & 1/2 & 0 & 0 & 1/3 \\ & 0 & 0 & 1/2 & 0 & 1/6 \\ \hline & 0 & 1/2 & 1/2 & 1 & \end{array} . \tag{15}$$

We use a uniform time-step Δt , such that $\Delta t = T/K$ for a suitable $K \in \mathbb{N}$. The temporal grid is then $t^n = n \Delta t$, where $n = 0, 1, \dots, K$. Given the ODE $y' = \Phi(t, y)$, one step of this four-stage RK scheme (with y^n approximating $y(t^n)$) is

```

for  $i = 1 \rightarrow 4$  do
     $\tilde{y}^i = y^n + \sum_{j=1}^{i-1} a_{ij} y^{n,j}$ 
     $y^{n,i} = \Phi(t^{n,i}, \tilde{y}^i)$ , evaluated at  $t^{n,i} \equiv t^n + \tau_i \Delta t$ 
end for
 $y^{n+1} = y^n + \Delta t \sum_{j=1}^4 b_j y^{n,j}$ 
    
```

where a_{ij}, τ_i, b_i are given in Table 15. Applying this scheme to (13) and denoting by H^n and U^n the fully discrete approximation in S of $h(\cdot, t^n), u(\cdot, t^n)$, respectively, leads to Algorithm 1.

Given a basis $\{\varphi_i\}$ of S , the implementation of Algorithm 1 requires solving at each time step the following linear systems.

1. Four linear systems with the **time-independent** matrix (φ_i, φ_j) ;
2. Four linear systems with the **time-dependent** matrix $\mathcal{B}_{\text{Serre}}(\varphi_i, \varphi_j; h)$.

All these matrices are cyclic and symmetric due to the periodic boundary conditions. They consist of a seven-diagonal band and two 3×3 triangular blocks on the upper right and lower left corners. To solve these systems, we use the direct method described in [8], which is analogous to the Sherman–Morrison–Woodbury method. To approximate the inner products, we use the Gauss–Legendre quadrature with 5 nodes per Δx .

We note that the above algorithms are almost identical when using other standard boundary conditions, such as Dirichlet or Neumann. However, the convergence and the stability properties can be different on those cases. In the analogous case of the cB system, the convergence of the FEM has optimal rates of convergence in the periodic case, [3], contrary to the suboptimal rates characterized the problem subject to non-periodic boundary conditions, cf. [1,2]. For more information on implementation of the Galerkin / Finite Element methods with other boundary conditions see [35].

Algorithm 1 Time-marching FEM scheme for the IBVP of the Serre system (13)

$$H^0 = \mathcal{P}\{h_0\}$$

$$U^0 = \mathcal{P}\{u_0\}$$

for $n = 0 \rightarrow N - 1$ **do**

for $i = 1 \rightarrow 4$ **do**

$$\tilde{H}^i = H^n + \sum_{j=1}^{i-1} a_{ij} H^{n,j}$$

$$\tilde{U}^i = U^n + \sum_{j=1}^{i-1} a_{ij} U^{n,j}$$

$$(H^{n,i}, \phi) = -((\tilde{H}^i \tilde{U}^i)_x, \phi), \text{ evaluated at } t^{n,i} \equiv t^n + \tau_i \Delta t$$

$$\mathcal{B}_{\text{Serre}}(U^{n,i}, \phi; \tilde{H}^i) = -((\tilde{H}^i (\tilde{H}_x^i + \tilde{U}^i \tilde{U}_x^i), \phi) - \frac{1}{3}(\tilde{H}^i)^3(\tilde{U}^i \tilde{U}_{xx}^i - (\tilde{U}_x^i)^2), \phi_x)$$

end for

$$H^{n+1} = H^n + \Delta t \sum_{j=1}^4 b_j H^{n,j}$$

$$U^{n+1} = U^n + \Delta t \sum_{j=1}^4 b_j U^{n,j}$$

end for

3.3 FEM Scheme for the cB System

Below we briefly outline the corresponding scheme for the cB system (3). See [1,3,4] for details. Making the transformation $h(x, t) \mapsto 1 + \eta(x, t)$, the semi-discrete problem for the cB system (3) is

$$(\tilde{h}_t, \phi) + ((\tilde{h}\tilde{u})_x, \phi) = 0, \tag{16a}$$

$$\mathcal{B}_{\text{cB}}(\tilde{u}_t, \phi) + (\tilde{h}_x, \phi) + (\tilde{u}\tilde{u}_x, \phi) = 0, \tag{16b}$$

where, in this case, the bilinear form is defined as

$$\mathcal{B}_{\text{cB}}(\psi, \phi) \doteq (\psi, \phi) + \frac{1}{3}(\psi_x, \phi_x). \tag{16c}$$

Using the notation as in Algorithm 1 and denoting the fully-discrete solution by $h(x, t) \rightarrow H^n(x)$ and $u(x, t) \rightarrow U^n(x)$, the corresponding fully-discrete algorithm based on the same RK scheme is given by Algorithm 2.

As for the FEM scheme for the Serre problem, we employ cubic splines and the inner products are approximated using the 5-node Gauss–Legendre quadrature. The resulting linear systems are similar with those of Algorithm 1 and are solved using the same numerical method. The key difference from Algorithm 1 is that **all the matrices in Algorithm 2 are time-independent**. Therefore, the matrices are assembled and factorized once and for all at $t = 0$.

3.4 Theoretical Considerations

For the semi-discrete problem (16), it was proven in [3] that, for appropriate initial conditions and for small values of Δx , there is a unique semi-discrete solution, which satisfies the estimate

$$\max_{0 \leq t \leq T} (\|h - \tilde{h}\| + \|u - \tilde{u}\|) \leq C \Delta x^r, \tag{17}$$

Algorithm 2 Time-marching FEM scheme for the IBVP of the cB system (3)

```

 $H^0 = \mathcal{P}\{h_0\}$ 
 $U^0 = \mathcal{P}\{u_0\}$ 
for  $n = 0 \rightarrow N - 1$  do
  for  $i = 1 \rightarrow 4$  do
     $\tilde{H}^i = H^n + \sum_{j=1}^{i-1} a_{ij} H^{n,j}$ 
     $\tilde{U}^i = U^n + \sum_{j=1}^{i-1} a_{ij} U^{n,j}$ 
     $(H^{n,i}, \phi) = -((\tilde{H}^i \tilde{U}^i)_x, \phi)$ , evaluated at  $t^{n,i} \equiv t^n + \tau_i \Delta t$ 
     $\mathcal{B}_{cB}(U^{n,i}, \phi) = -(\tilde{H}_x^i, \phi) - (\tilde{U}^i \tilde{U}_x^i, \phi)$ 
  end for
   $H^{n+1} = H^n + \Delta t \sum_{j=1}^4 b_j H^{n,j}$ 
   $U^{n+1} = U^n + \Delta t \sum_{j=1}^4 b_j U^{n,j}$ 
end for

```

where the constant C is independent of Δx . This result also shows that the numerical solution is stable. Furthermore, the same result is valid for the linear cB system. Since the linearization of the cB and Serre systems is the same, it is implied that the semi-discrete solution of the linearized Serre system is stable. No stability or convergence results are known for the nonlinear (semi- or fully-) discrete schemes for the Serre equations.

4 Validation of the FEM Scheme for the Serre System

In this section we study the spatial and temporal accuracy and efficiency of the FEM scheme for the Serre system, which is presented in Algorithm 1. To do so, we use various metrics of the solitary and cnoidal wave solutions and Hamiltonian conservation (see Sect. 2).

4.1 Spatial Accuracy

To test the spatial accuracy of the scheme, we use the exact solitary wave solution (4) of the Serre system with $\varepsilon = a_0 = \sigma = 1$ and traveling with speed $c_s = 1.5$. The spatial domain is chosen as $x \in [-150, 150]$. This large interval ensures that the solution is practically zero near the endpoints of the interval. To ensure that the errors incurred by the temporal integration are negligible, we take $\Delta t/\Delta x = 0.1$.

Tables 1, 2 and 3 show the normalized errors of the computed solutions evaluated at $T = 100$. These errors are defined as

$$E_s[F] \doteq \frac{\|F(x, T; \Delta x) - F_{\text{exact}}(x, T)\|_s}{\|F_{\text{exact}}(x, T)\|_s}, \tag{18}$$

where $F = F(\cdot; \Delta x)$ is the computed solution, i.e., either $H \approx h(x, T)$ or $U \approx u(x, T)$, F_{exact} is the corresponding exact solitary wave solution with the same parameters (see (4a)–(4b)), and $s = 0, 1, \infty$ corresponds to the L^2, H^1 , and L^∞ norms, respectively.

For the calculations of the L^∞ norm and related variables mentioned in the following sections, we recover **location of the peak amplitude** curve of the solution $h(x, t)$. This

Table 1 Spatial errors and rates of convergence for a solitary wave solution using the L^2 norm [Eq. (18) with $s = 2$]

Δx	Δt	$E_2[H]$	rate for $E_2[H]$	$E_2[U]$	Rate for $E_2[U]$
0.50	0.050	0.1970×10^{-4}	–	0.5669×10^{-4}	–
0.25	0.025	0.7989×10^{-6}	4.6240	0.2153×10^{-5}	4.7183
0.10	0.010	0.1798×10^{-7}	4.1402	0.4973×10^{-7}	4.1124
0.08	0.008	0.7298×10^{-8}	4.0420	0.2018×10^{-7}	4.0403
0.05	0.005	0.1102×10^{-8}	4.0224	0.3043×10^{-8}	4.0257

Table 2 Same as Table 1 using the H^1 norm

Δx	Δt	$E_2[H]$	Rate for $E_2[H]$	$E_2[U]$	Rate for $E_2[U]$
0.50	0.050	0.1951×10^{-3}	–	0.41324×10^{-3}	–
0.25	0.025	0.1873×10^{-4}	3.3805	0.4261×10^{-4}	3.2775
0.10	0.010	0.1111×10^{-5}	3.0830	0.2601×10^{-5}	3.0515
0.08	0.008	0.5660×10^{-6}	3.0236	0.1327×10^{-5}	3.0143
0.05	0.005	0.1374×10^{-6}	3.0122	0.3230×10^{-6}	3.0073

Table 3 Same as Table 1 using the L^∞ norm

Δx	Δt	$E_\infty[H]$	Rate for $E_\infty[H]$	$E_\infty[U]$	Rate for $E_\infty[U]$
0.50	0.050	0.4228×10^{-3}	–	0.5315×10^{-4}	–
0.25	0.025	0.2101×10^{-4}	4.3309	0.2882×10^{-5}	4.2049
0.10	0.010	0.4887×10^{-6}	4.1046	0.7123×10^{-7}	4.0383
0.08	0.008	0.1988×10^{-6}	4.0291	0.2893×10^{-7}	4.0381
0.05	0.005	0.3013×10^{-7}	4.0148	0.4373×10^{-8}	4.0199

curve, denoted by $x^*(t)$, is defined via

$$\left. \frac{d}{dx} h(x, t) \right|_{x=x^*(t)} = 0. \tag{19}$$

To compute $x^*(t)$, we use Newton’s method. As an initial guess, we use the quadrature node at which $H^n(x)$ attains a maximum over all the quadrature nodes. This ensures that $H^n(x^*)$ is the global maximum. Usually, only a few iterations are needed to achieve x^* with a tolerance of 10^{-13} .

Tables 1, 2 and 3 also show the corresponding calculated rates of convergence, defined as

$$\text{rate for } E_s[F] \doteq \frac{\ln(E_s[F(\cdot; \Delta x_{k-1})]/E_s[F(\cdot; \Delta x_k)])}{\ln(\Delta x_{k-1}/\Delta x_k)},$$

where Δx_k is the grid size listed in row k in each table.

These tables show that the rates using the L^2 and L^∞ norms approach 4, whereas, the rates using the H^1 norm approaches 3. These results indicate that the FEM scheme achieves the optimal orders of convergence. Moreover, one might expect the emergence of large errors (in space and/or time) due to the following challenging conditions:

1. The high-order nonlinear dispersive terms in (1).
2. The strongly nonlinear and dispersive regime ($\varepsilon = \sigma = 1$).
3. The use of an explicit RK method.

Yet, in spite of these challenging conditions, Tables 1, 2 and 3 show that the actual errors are very small, even when using relatively large grid sizes. **Hence, these results show that this scheme is highly efficient.**

4.2 Temporal Accuracy and Stability

In order to study the temporal accuracy, we use the same solitary wave solutions as above. Here we take $\Delta t = \Delta x$ for various values of $\Delta x = (b - a)/N$. This choice for Δt and Δx is sufficient to estimate the temporal order of accuracy for the following reason. Since our spatial discretization is 4th-order, we may assume that the scheme’s total error at some final time $t = T$ scales as

$$E_s[F] \doteq \|f - F\| = C' \Delta x^4 + C \Delta t^r, \tag{20}$$

where $f = f(\cdot, T; \Delta x, \Delta t)$ stands for the computed solution, $F(\cdot, T)$ stands for the exact solution, C, C' are constants, and r is the temporal convergence rate. Since our schemes use a 4th-order Runge–Kutta method, it is expected that $r \leq 4$. By choosing $\Delta x = \Delta t \ll 1$ and using (20), the total error scales as

$$E_s[F] = \|f - F\| \approx C \Delta t^r, \tag{21}$$

Choosing two different values of Δt , i.e., Δt_{k-1} and Δt_k , gives

$$E_s[F(\cdot; \Delta t_{k-1})] \approx C \Delta t_{k-1}^r, \quad E_s[F(\cdot; \Delta t_k)] \approx C \Delta t_k^r. \tag{22}$$

Taking the ratio of these two errors and solving for the temporal convergence rate, r , yields

$$\text{rate for } E_s[F] \doteq \frac{\ln(E_s[F(\cdot; \Delta t_{k-1})])/E_s[F(\cdot; \Delta t_k)]}{\ln(\Delta t_{k-1}/\Delta t_k)}. \tag{23}$$

Tables 4, 5 and 6 present the errors defined in (18) and the corresponding rates of convergence, defined by (23), where Δt_k is the grid size listed in row k . These tables show that the FEM scheme achieves the optimal temporal rate of convergence in all three norms. The actual errors are fairly small as well. Moreover, one might expect the scheme to be conditionally stable, due to the complexity of the problem and the use of an explicit RK method. Yet, the scheme converges in spite of these challenging conditions and the large temporal grid size ($\Delta t = \Delta x$). We note that it has been proven that a similar FEM scheme is *unconditionally*

Table 4 Same as Table 1 for the temporal errors. N and M are the number of spatial and temporal grid points, respectively

$\Delta x = \Delta t$	$E_2[H]$	Rate for $E_2[H]$	$E_2[U]$	Rate for $E_2[U]$
0.2	0.1824×10^{-1}	–	0.9673×10^{-2}	–
0.1	0.7114×10^{-3}	4.6808	0.3759×10^{-3}	4.6853
0.05	0.3055×10^{-4}	4.5411	0.1669×10^{-4}	4.4935
0.025	0.1496×10^{-5}	4.3521	0.8296×10^{-6}	4.3303
0.0125	0.8134×10^{-7}	4.2011	0.4518×10^{-7}	4.1986
0.00625	0.4724×10^{-8}	4.1058	0.2633×10^{-8}	4.1008

Table 5 Same as Table 4 using the H^1 norm

$\Delta x = \Delta t$	$E_1[H]$	Rate for $E_1[H]$	$E_1[U]$	Rate for $E_1[U]$
0.2	0.4138×10^{-2}	–	0.1286×10^{-1}	–
0.1	0.1634×10^{-3}	4.6624	0.5075×10^{-3}	4.6631
0.05	0.7126×10^{-5}	4.5192	0.2213×10^{-4}	4.5189
0.025	0.3481×10^{-6}	4.3555	0.1081×10^{-5}	4.3549
0.0125	0.1872×10^{-7}	4.2167	0.5822×10^{-7}	4.2159
0.00625	0.1082×10^{-8}	4.1121	0.3371×10^{-8}	4.1101

Table 6 Same as Table 4 using the L^∞ norm

$\Delta x = \Delta t$	$E_\infty[H]$	Rate for $E_\infty[H]$	$E_\infty[U]$	Rate for $E_\infty[U]$
0.2	0.6037×10^{-2}	–	0.1491×10^{-1}	–
0.1	0.2386×10^{-3}	4.6610	0.5889×10^{-3}	4.6623
0.05	0.1041×10^{-4}	4.5176	0.2569×10^{-4}	4.5183
0.025	0.5106×10^{-6}	4.3507	0.1256×10^{-5}	4.3539
0.0125	0.2787×10^{-7}	4.1953	0.6777×10^{-7}	4.2126
0.00625	0.1713×10^{-8}	4.0239	0.3960×10^{-8}	4.0969

stable for several types of Boussinesq systems [4]. Although we do not have a proof that this fully-discrete problem is unconditionally stable, these results show that **the stability of this scheme does not impose restrictive conditions on Δt** but mild conditions of the form $\Delta t < C \Delta x$ are adequate for the solutions to remain stable. This property is further explored in Sect. 4.4.

4.3 Accuracy in Shape, Phase, and Hamiltonian

The results of the spatial and temporal accuracy show that the FEM scheme is optimally accurate in all the standard norms and also that the actual errors are very small. To further test the accuracy of this scheme, we consider the propagation of a solitary wave as in Sect. 4.1, while using several other norms that are pertinent to solitary waves (cf. [8]).

First, since the solitary wave’s peak amplitude remains constant during propagation, we define the normalized peak amplitude error as

$$E_{amp}[F] \doteq \frac{|F(x^*(t), t) - F_0|}{|F_0|}, \tag{24}$$

where $x^*(t)$ is the curve along which the computed approximate solution $F(x, t)$ achieves its maximum (see Sect. 4.1) and $F_0 \equiv F_{\text{exact}}(x, 0)$ is the initial peak amplitude of the solitary wave. Monitoring E_{amp} as a function of propagation time, we observe that it remains very small and practically constant during propagation, i.e., $E_{amp}[H] \approx 1.5066 \times 10^{-5}$ and $E_{amp}[U] \approx 1.2076 \times 10^{-5}$. Furthermore, recall that the exact solitary wave solution travels with speed $c_s = 1.5$. We recover the solitary wave’s traveling speed as

$$\tilde{c}_s \doteq \frac{x^*(t) - x^*(t - \tau)}{\tau}, \tag{25}$$

where τ is a constant. The results using $\tau = 10$ are such that \tilde{c}_s coincides with c_s within the computed precision, i.e., double precision on a GNU Fortran compiler parallelized using OpenMP. This serves as additional indications of the **high accuracy and non-dissipativity** of this scheme. We note that the value of \tilde{c}_s depends weakly on the choice of τ , which indicates a phase error. This is further studied below.

Two other error norms that are pertinent to solitary waves are the *shape and phase errors*, defined below. We define the normalized shape error as the distance in L^2 between the computed solution at time $t = t^n$ and the family of temporally-translated exact solitary waves (with the same parameters), i.e.,

$$E_{shape}[F] \doteq \min_{\tau} \zeta(\tau), \quad \zeta(\tau) \doteq \frac{\|F(x, t^n) - F_{\text{exact}}(x, \tau)\|}{\|F_{\text{exact}}(x, 0)\|}. \tag{26}$$

The minimum in (26) is attained at some critical $\tau = \tau^*(t^n)$. This, in turn, is used to define the (signed) phase error as

$$E_{phase}[F] \doteq \tau^* - t^n. \tag{27}$$

In order to find τ^* , we use Newton’s method to solve the equation $\zeta''(\tau) = 0$. The initial guess for Newton’s method is chosen as $\tau^0 = t^n - \Delta t$. Having computed τ^* , the shape error (26) is then

$$E_{shape}[F] = \zeta(\tau^*).$$

These error norms are closely related to the *orbit* of the solitary wave. Loosely speaking, they measure “softer” properties of the wave, which are often not well conserved using dissipative schemes, even when the schemes are accurate in all the standard norms.

Table 7 presents the shape and phase errors as functions of propagation time, using $c_s = 1.5$, $\Delta x = 0.1$, $\Delta t = 0.01$. We observe that both errors remain very small. Moreover, the shape error is practically constant during the propagation.

Next, we test the conservation of the Hamiltonian (11) and define the corresponding normalized Hamiltonian error as

$$E_{\mathcal{H}}(t^n) \doteq \left| \frac{\mathcal{H}[F(x, t^n)] - \mathcal{H}(0)}{\mathcal{H}(0)} \right|, \tag{28}$$

where $\mathcal{H}[\cdot] \equiv \mathcal{H}(t)$ denotes the energy functional (11).

Table 8 shows the results using the same wave parameters and grid sizes as in Table 7 in the time interval $t^n \in [0, 200]$. These results show that the Hamiltonian is conserved within at least 8 decimal digits of accuracy in this interval and for the specific values of $\Delta x = 0.1$ and $\Delta t = 0.01$. However, the error increases linearly with time. This is to be expected, since

Table 7 Shape and phase errors [Eqs. (26) and (27)] for a solitary wave as functions of propagation time

t^n	$E_{shape}[H]$	$E_{shape}[U]$	$E_{phase}[H]$	$E_{phase}[U]$
20	0.1779×10^{-7}	0.4543×10^{-7}	-0.6636×10^{-8}	-0.6651×10^{-8}
40	0.1779×10^{-7}	0.4544×10^{-7}	-0.1184×10^{-7}	-0.1187×10^{-7}
60	0.1779×10^{-7}	0.4543×10^{-7}	-0.2353×10^{-7}	-0.2355×10^{-7}
80	0.1779×10^{-7}	0.4543×10^{-7}	-0.3002×10^{-7}	-0.3004×10^{-7}
100	0.1779×10^{-7}	0.4543×10^{-7}	-0.2353×10^{-7}	-0.2355×10^{-7}
200	0.1779×10^{-7}	0.4543×10^{-7}	-0.6899×10^{-7}	-0.6901×10^{-7}

Table 8 Hamiltonian \mathcal{H} [Eq. (11)] and corresponding error (28) for a solitary wave as functions of the propagation time

t^n	\mathcal{H}	$E_H(t^n)$
0	7.4266250954	0.3208×10^{-11}
20	7.4266250944	0.1398×10^{-9}
40	7.4266250933	0.2828×10^{-9}
60	7.4266250922	0.4258×10^{-9}
80	7.4266250912	0.5688×10^{-9}
100	7.4266250901	0.7117×10^{-9}
200	7.4266250848	0.1427×10^{-8}

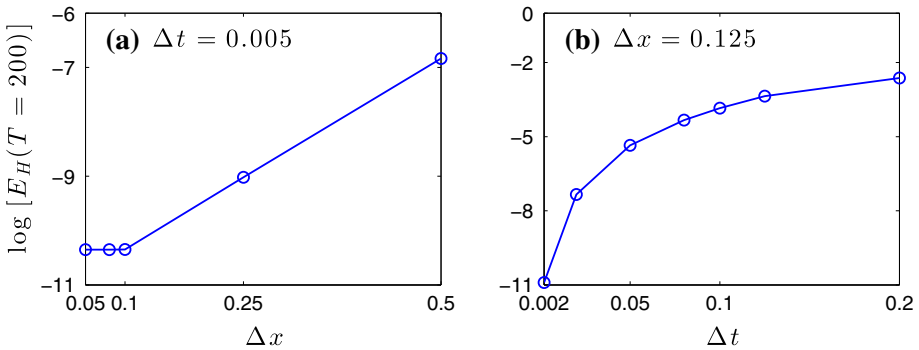


Fig. 3 The log of the Hamiltonian error, $E_H(T = 200)$, as a function of (a) Δx and (b) Δt

the *explicit* RK method for this problem is non-conservative. Figure 3 presents the error in the Hamiltonian as a function of Δx when $\Delta t = 0.005$ fixed and as a function of Δt when the value of $\Delta x = 0.25$. We observe that the logarithm of the error E_H increases linearly with Δx and as $\sqrt{\Delta t}$.

We close this section by computing the same error norms for **cnoidal waves**. Specifically, we consider the cnoidal wave solution (5) with $(a_0, a_1) = (0.3, 0.1)$ and $m \in \{0.05, 0.1, 0.5, 0.99\}$. For this choice of (a_0, a_1) , the cnoidal waves are spectrally unstable for $m > 0.09$ (see [11]). In these computations, we consider a domain of length equal to one period, with very small values for Δx and Δt , i.e., $N = 200$ and $\Delta t = 10^{-3}$. The profiles of the propagation of these cnoidal waves are presented in Fig. 4. Table 9 presents some of the error results. In all cases, the cnoidal waves propagate without significant changes in their amplitude, speed, shape, phase, and Hamiltonian. In particular, the Hamiltonian is conserved very well to within double precision. These results also show that the phase and shape errors increase as m increases, especially as m approaches 1. This is expected, because as m increases, the cnoidal wave becomes steeper and, in the limit $m \rightarrow 1$, it approaches a solitary wave.

4.4 Stability of the FEM Scheme

Here we perform a series of computations to study the stability of the FEM scheme. First, we consider the propagation of a solitary wave with $c_s = 1.5$, using five different CFL ratios,

$$\frac{\Delta t}{\Delta x} \in \{1, 1.5, 2, 2.1, 3\} . \tag{29}$$

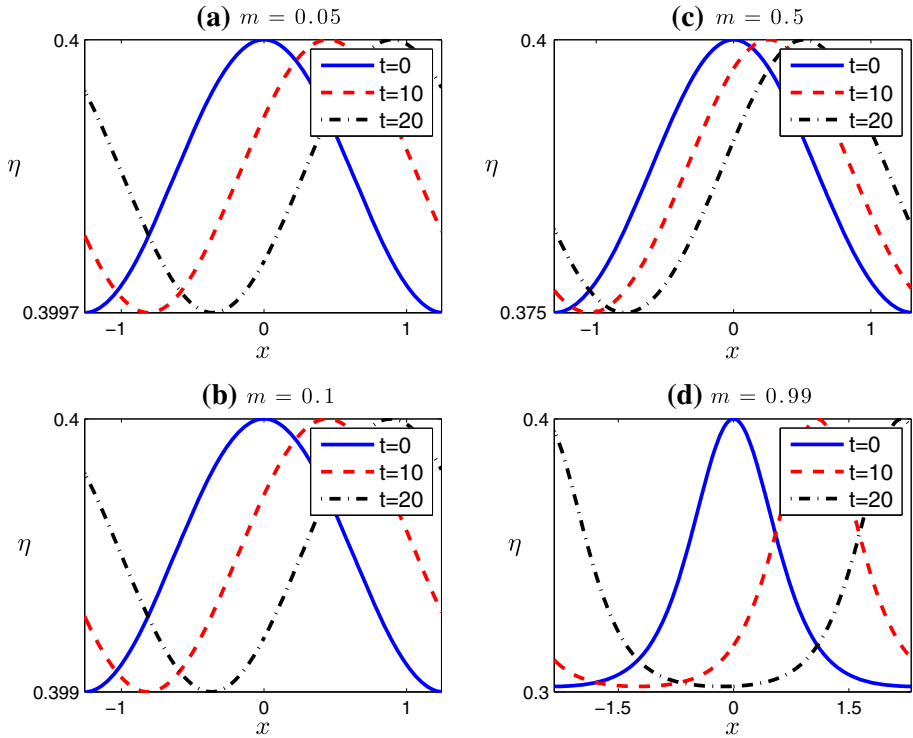


Fig. 4 Propagation of cnoidal waves with $(a_0, a_1) = (0.3, 0.1)$ for four different values of the elliptic modulus m . Shown is the free surface elevation $\eta(x, t)$ at three different propagation times (see legends). The axes are chosen to fit a single period and the amplitude span of the waves

Table 9 Shape and phase errors for the computed solution $H(x, t^n)$ and the Hamiltonian error, $E_{\mathcal{H}}$ at $t^n = 20$, and the conserved values of the Hamiltonian \mathcal{H} for cnoidal waves with $(a_0, a_1) = (0.3, 0.1)$, at four different values of the elliptic modulus m

m	$E_{shape}[H]$	$E_{phase}[H]$	$E_{\mathcal{H}}$	\mathcal{H}
0.05	0.1960×10^{-12}	-0.5471×10^{-11}	0.1005×10^{-16}	0.90543930873897
0.1	0.7843×10^{-12}	-0.1314×10^{-10}	0.2447×10^{-17}	0.90743193859945
0.5	0.2337×10^{-10}	-0.9521×10^{-12}	0.3954×10^{-17}	0.98158535527292
0.99	0.1637×10^{-8}	-0.4267×10^{-8}	0.1268×10^{-16}	2.1009810886027

Table 10 presents the values of the normalized shape and phase errors for the case $\Delta t = 2\Delta x$. For example, when $\Delta t = 2\Delta x$ and $\Delta x = 0.1$, the solitary wave propagates without significant changes in shape and speed. The results in the other cases are comparable except when $\Delta t > 2\Delta x$, where the solution becomes unstable. The fact that the CFL ratio can be chosen greater than 1 indicates that the FEM is **very stable**. This is rather surprising, considering the use of an explicit RK method.

To further test the stability of the numerical method, we consider initial conditions representing a heap of water, i.e.,

$$\eta_0(x) = Ae^{-x^2/\lambda}, \quad u_0(x) = 0, \tag{30}$$

Table 10 Shape and phase errors for a solitary wave as functions of propagation time, using $\Delta x = 0.1$ and $\Delta t = 0.2$

t^n	$E_{shape}[H]$	$E_{shape}[U]$	$E_{phase}[H]$	$E_{phase}[U]$
20	0.3031×10^{-5}	-0.7622×10^{-4}	0.1997×10^{-4}	-0.7840×10^{-4}
40	0.3526×10^{-5}	-0.1787×10^{-3}	0.2191×10^{-4}	-0.1810×10^{-3}
60	0.4134×10^{-5}	-0.3238×10^{-3}	0.2487×10^{-4}	-0.3261×10^{-3}
80	0.4815×10^{-5}	-0.5114×10^{-3}	0.2859×10^{-4}	-0.5139×10^{-3}
100	0.5556×10^{-5}	-0.7423×10^{-3}	0.3263×10^{-4}	-0.7445×10^{-3}
200	0.9518×10^{-5}	-0.2527×10^{-2}	0.5709×10^{-4}	-0.2529×10^{-2}

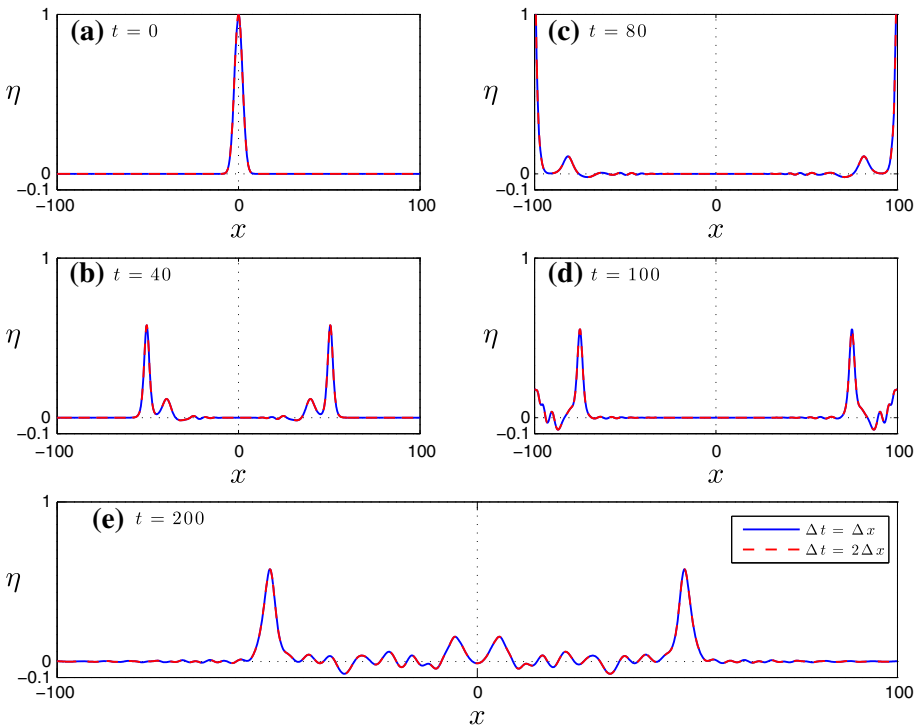


Fig. 5 Breakup of a Gaussian hump [Eq. (30) with initial amplitude $A = 1$ and width $\lambda = 10$] into two solitary waves traveling in opposite directions and dispersive tails. Shown are the results using $\Delta x = 0.1$ and two different values of Δt [see legend in (e)], which are almost indistinguishable

where A and λ are constants. In all cases, the scheme is numerically stable for large values of propagation time t and all the CFL ratios in (29). For example, Fig. 5 presents the solutions using $A = 1$, $\lambda = 10$ and two CFL ratios: $\Delta t = \Delta x$ and $\Delta t = 2\Delta x$ (for $CFL > 2$ the scheme becomes unstable). Here, the initial hump breaks up into two large solitary waves and smaller dispersive tails. These waves and their dispersive tails travel in opposite directions. Figure 6 shows the results using $A = 1$ and $\lambda = 40$. In this case, the solution breaks up into pairs or a larger number of solitary waves, which travel in opposite directions. In all cases, the scheme is stable even when $\Delta t = 2\Delta x$ and the difference between the solutions using the

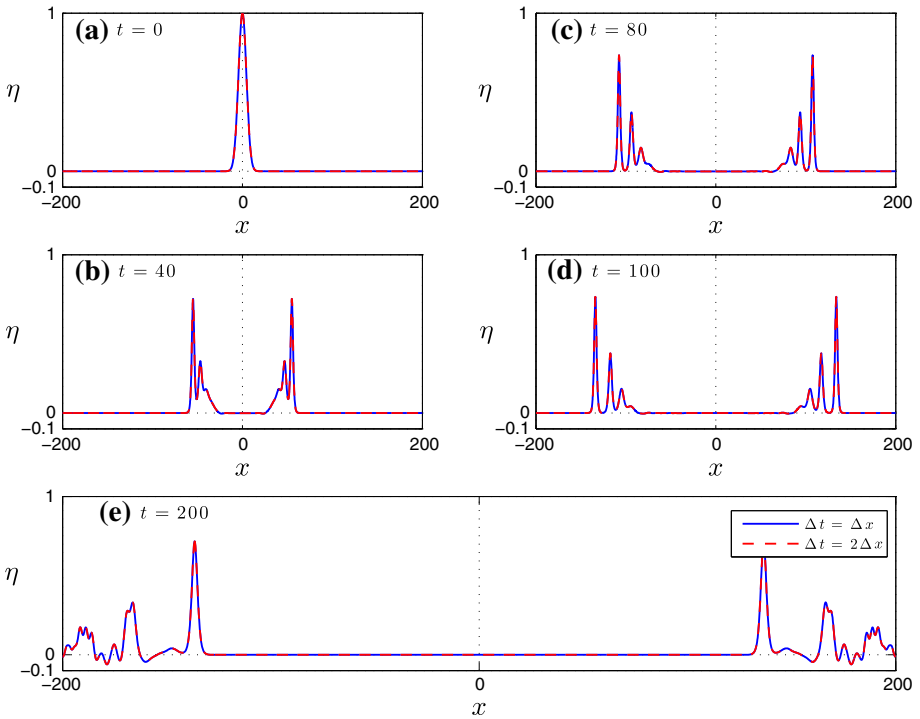


Fig. 6 Same as Fig. 5 using the initial width $\lambda = 40$

two time steps remains negligible. These results give further indication that **the stability of this scheme does not impose a restrictive condition on Δt** , such that $\Delta t \leq C(\Delta x)^r$ for $r > 1$.

5 Numerical Experiments of Solitary Waves and DSWs

In this section we present numerical experiments illustrating the behavior of the solitary waves and **dispersive shock waves (DSWs)** in the Serre and cB system. Systems (1) and (3) are solved using Algorithm 1 and Algorithm 2, respectively. Notwithstanding the apparent stability of both FEM schemes, we use $\Delta t = \Delta x/10$ in order to ensure that the errors resulting from the time discretization are negligible even if the numerical solution is stable for larger values of Δt .

5.1 Interactions of Solitary Waves

When solitary waves interact, they incur a phase shift. In non-integrable systems, such interactions are often accompanied by the generation of small amplitude dispersive tails. Capturing this dynamics accurately requires a highly accurate scheme. Here, we study two kinds of interactions of solitary waves.

1. **Head-on collisions** of counter-propagating solitary waves.
2. **Overtaking collisions** of solitary waves co-propagating at different speeds.

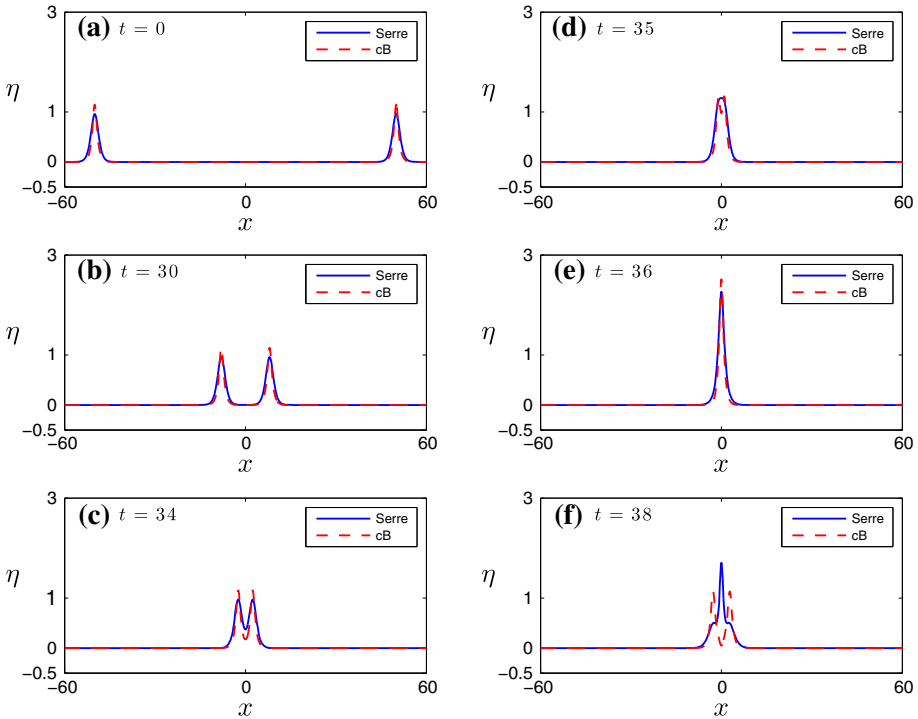


Fig. 7 Head-on collision of two solitary waves in the Serre (*solid*) and cB (*dashes*) systems

For the **head-on collisions**, we generated initial conditions using (4) for each wave with speed $c_s = 1.4$ and amplitude $A = 0.96$. The waves are initially well-separated, i.e., their peak amplitudes are located at $x = \pm 50$. The spatial domain is $x \in (-200, 200)$ and the grid sizes are $\Delta x = 0.1$ and $\Delta t = 0.01$.

In order to make a meaningful comparison with solitary waves in the cB system, the cB solitary waves need to travel at the same speed. Since there is no known exact closed formula for cB solitary waves, we compute them using a fixed-point iterative scheme, in which the wave’s speed, c_s , enters as a parameter (see [19]). One upshot of this is that, for the same $c_s = 1.4$, the corresponding cB solitary wave has a somewhat larger amplitude, i.e., $A \approx 1.14763$. Moreover, the cB and Serre waves have significantly different shapes.

Figures 7 and 8 present the solutions of the Serre and cB systems at different propagation times. As expected, the waves collide and emerge with small dispersive tails. Figure 9 presents the peak amplitude, and the location of the peak amplitude as functions of time, $x^*(t)$, computed via (19). We note that, $t \mapsto x^*(t)$ is not a globally continuous function. Indeed, Fig. 9 shows that for the Serre solution, $x^*(t)$ is discontinuous at $t \approx 38$ and $t \approx 40.5$. To better understand this picture, Figs. 7 and 8 show that, for the Serre system, as the colliding waves separate, the location of the peak amplitude changes abruptly between $t = 38$ and $t = 39$, as two off-center humps grow larger than the center hump. A similar phenomenon occurs at each peak around $t = 40.5$, though with less distinguishable humps.

There are several interesting similarities and differences between the results for the cB and Serre systems.

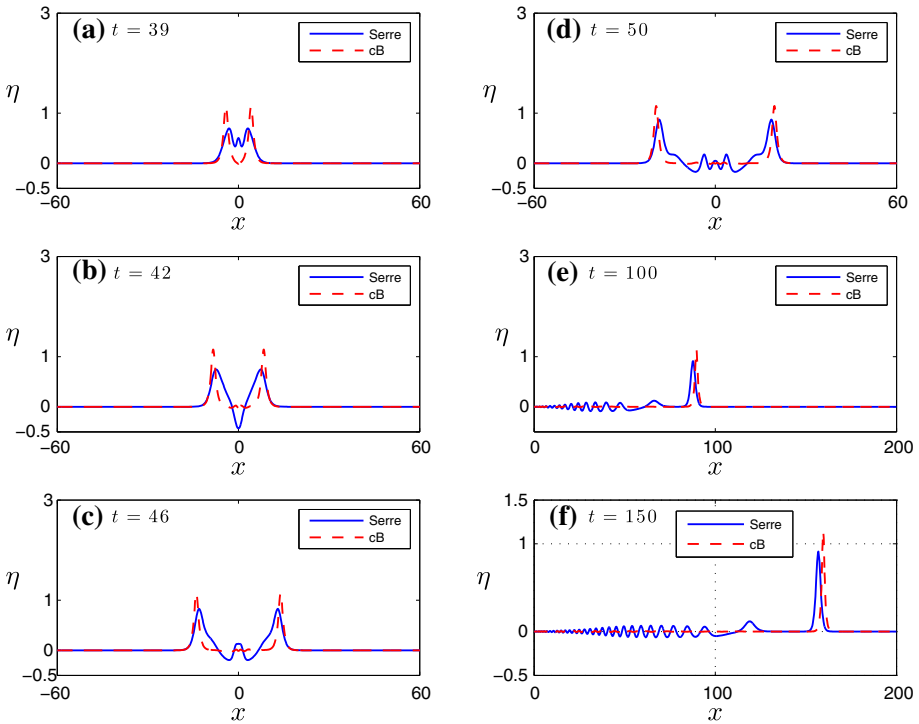


Fig. 8 Continuation of the results in Fig. 7. Note the three humps at $t = 39$. For clarity, (e) and (f) are shown on different scales

- The dispersive tails in the Serre system are considerably larger.
- During the interaction, the peak amplitude reaches approximately the same value in both systems, but at somewhat different times, i.e., a maximum of approximately 2.5 at $t \approx 36.7$ for the Serre system and a maximum of approximately 2.547 at $t \approx 35.9$ for the cB system. Furthermore, the Serre system gives rise to the large off-center humps and associated with the discontinuities in Fig. 9, whereas, the corresponding cB system does not have this phenomenon.
- The collision in the Serre system lasts longer.
- The change in the wave’s long-time amplitude (sufficiently after the collision) compared with the initial amplitude is much larger in the Serre system, i.e., it decreases by approximately 4.9 % in the Serre system, whereas, in the cB system, it decreases by only 0.061 %. For example, the amplitude of the solitary waves after the interaction for the Serre system is $A = 0.9125$ while in the case of cB system is $A = 1.1469$.
- The phase shift is significantly larger in the Serre system. The ensuing wave trajectories are closer together (with respect to linear propagation) in the Serre system.

From these experiments, we conclude that solitary waves behave qualitatively the same in the Serre and cB systems. However, quantitatively, **the interaction of the solitary waves in the Serre system is more inelastic**. These results are consistent with the fact that the Serre system contains nonlinear dispersive terms not present in the cB system. We also note that the Hamiltonian in this experiment is conserved to within 9 decimal digits and its conserved value was $\mathcal{H}(t) = 9.13794051$ for up to $T = 200$.

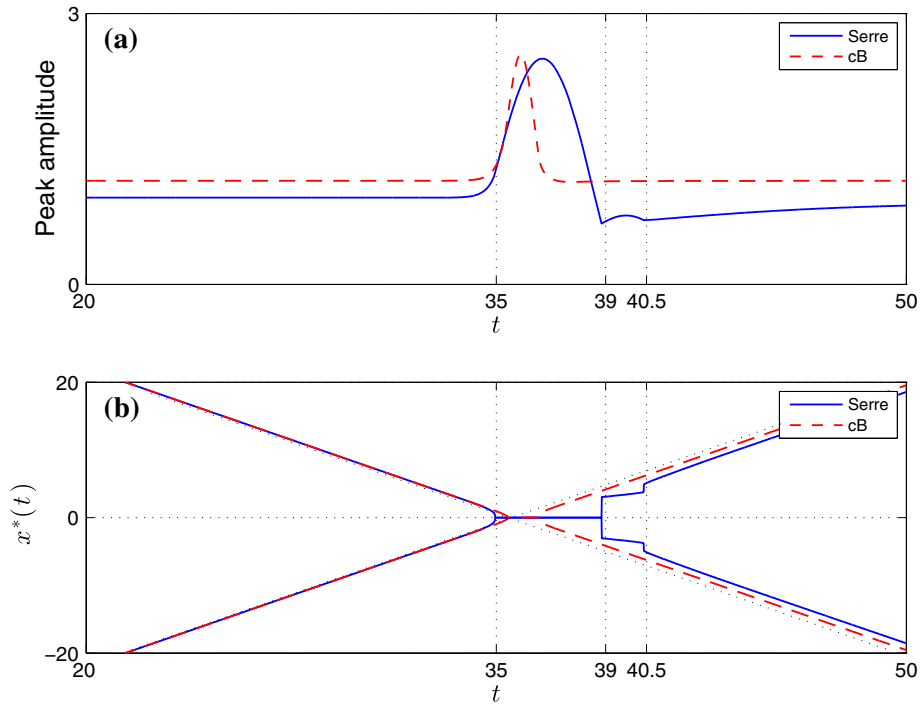


Fig. 9 (a) Peak amplitude of the solutions and (b) the location of the peak amplitude [see (19)] of each wave for the head-on collision presented in Figs. 7 and 8

We note that, given our choice of $\varepsilon = 1$ and solitary wave amplitude of $O(1)$, physically speaking, these solitary waves have very large amplitudes. For this reason, cB system might not be valid in this regime. We also compare the two systems in the small-amplitude regime, in which both systems are valid, i.e., repeat the head-on collision experiments using small-amplitude solitary waves with $A \approx 0.2$ and $c_s = 1.1$. (Specifically the amplitude of the solitary waves in the case of Serre system were $A = 0.21$ and after the interaction the amplitude have been reduced to the value $A = 0.2098659$. In the case of the cB system we have $A = 0.21774185$ before the collision and $A = 0.21774165$ after.) In that case, the results of the two systems are approximately the same before and after the interaction, and we refer to Fig. 10 for details. It is worth noting that the dispersive tail generated by the Serre system is larger compared to the respective tail generated by the cB system. The Hamiltonian in this experiment was $\mathcal{H} = 0.6764072912$ with the conserved digits shown here.

For the study of **overtaking collisions**, we consider two solitary waves traveling in the positive x direction, with speeds $c_{s,1} = 1.4$ and $c_{s,2} = 1.2$, centered at $x = \mp 50$, respectively. The amplitudes of the solitary waves of the Serre system are $A_1 \approx 0.96$ and $A_2 \approx 0.44$; and for the cB system $A_1 \approx 1.14763$ and $A_2 \approx 0.475729$.

Figures 11 and 12 present the solutions $\eta(x, t)$ of the Serre and cB systems at different propagation times. Given their initial positions and speed difference, the two waves (if they were linear) should collide at $t = 100/0.2 = 500$. In reality, the interaction begins at approximately $t = 400$ (see Fig. 11c). During the interaction, the waves appear to exchange mass (similar dynamics has been observed in Boussinesq systems [1, 21] and the Euler equations [16]). The faster wave overtakes the slower one at approximately $t = 490$ (see Fig. 12b). The interaction ends at approximately $t = 600$.

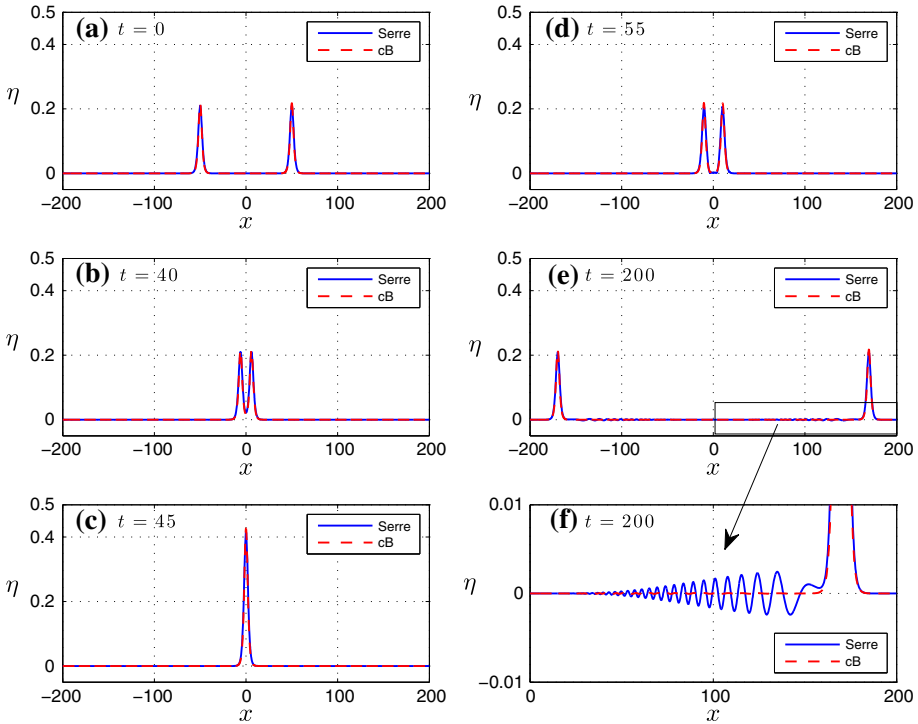


Fig. 10 Same as Fig. 6 using solitary waves with speed $c_s = 1.2$. (f) is a zoomed in version of (e)

A phase shift and a small change in amplitude (compared with the initial amplitudes) is observed. Specifically, in the Serre system, the long-time amplitudes of the two waves decrease by approximately 0.025 % for the larger wave and 0.0705 % for the smaller one. In the cB system, the decrease of the amplitudes is negligible, i.e., 0.00087 and 0.0027 %, respectively. The Hamiltonian in this experiment conserved within 8 decimal digits and it was $\mathcal{H} = 5.7237794$ up to $T = 800$.

Furthermore, Fig. 13 shows in detail the dispersive tails after the interaction. These tails contain N -shaped wavelets. The generation of wavelets has been studied recently for the cB system and other Boussinesq-like systems (cf. [1,4]), as well as for the Euler equations (cf. [15,16]). In [19], a related system based on a Galilean invariant equation, which contains some (but not all) of the nonlinear terms of the Serre system, showed how the wavelets depend on the nonlinear terms.

Here, Fig. 13 shows that the *signs* of the wavelets in both systems are the same, but their *amplitudes* differ, i.e., the wavelet is larger and travels faster in the Serre system. Moreover, the dispersive tail is larger in the Serre system.

These results show that:

- The interaction in the Serre system is significantly stronger.
- Compared with the head-on collisions, the overtaking collision is significantly weaker in terms of the amplitude and phase shifts and the size of the dispersive tails.

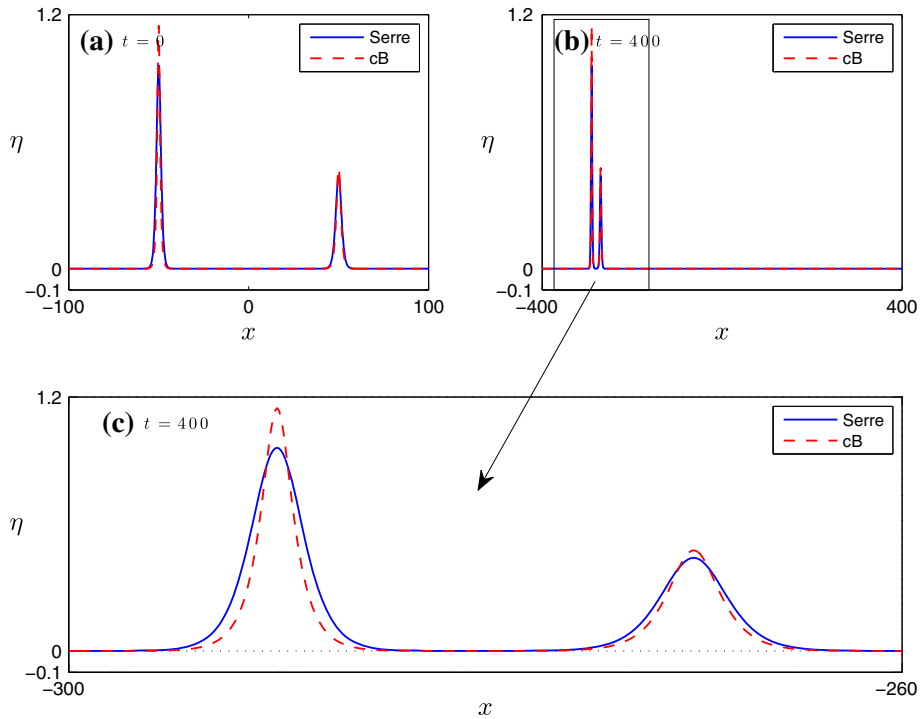


Fig. 11 Overtaking collision of two solitary waves, (c) is zoomed in version of (b) which is the beginning of the interaction

5.2 Dispersive Shock Waves

Here we test the ability of the FEM scheme to compute DSWs with high accuracy. In particular, we test the FEM scheme in two cases, i.e., for a **simple DSW** and for the **dam break problem**. The rapid oscillations in DSWs make it challenging to simulate these problems accurately. Moreover, finite-volume and other methods are prone to adding spurious dissipative effects. This can lead to *viscous-DSWs*, which look like DSWs, but travel more slowly and have smaller-amplitude oscillations [22]. One of the advantages of the FEM scheme is that it is non-dissipative, as shown below.

First, we study **simple DSWs** in the Serre and cB systems. The simulations are carried on the interval $x \in (-700, 700)$ with $\Delta x = 0.1$ and $\Delta t = 0.01$. We choose as initial data for h as a step function that decays to zero as $|x| \rightarrow \infty$, i.e.,

$$h(x, 0) = \frac{1}{2} \eta_0 [1 + \tanh(250 - |x|)], \tag{31}$$

where $\eta_0 = 0.4182$. The initial data for $u(x, 0)$ is chosen as

$$u(x, 0) = 2 \left[\sqrt{h(x, 0)} - 1 \right].$$

These initial data generate a simple DSW with (see (7) and [22, 23])

$$h^+ = 1, \quad h^- = 1 + \eta_0, \quad u^+ = 0, \quad u^- = 2(\sqrt{1 + \eta_0} - 1). \tag{32}$$

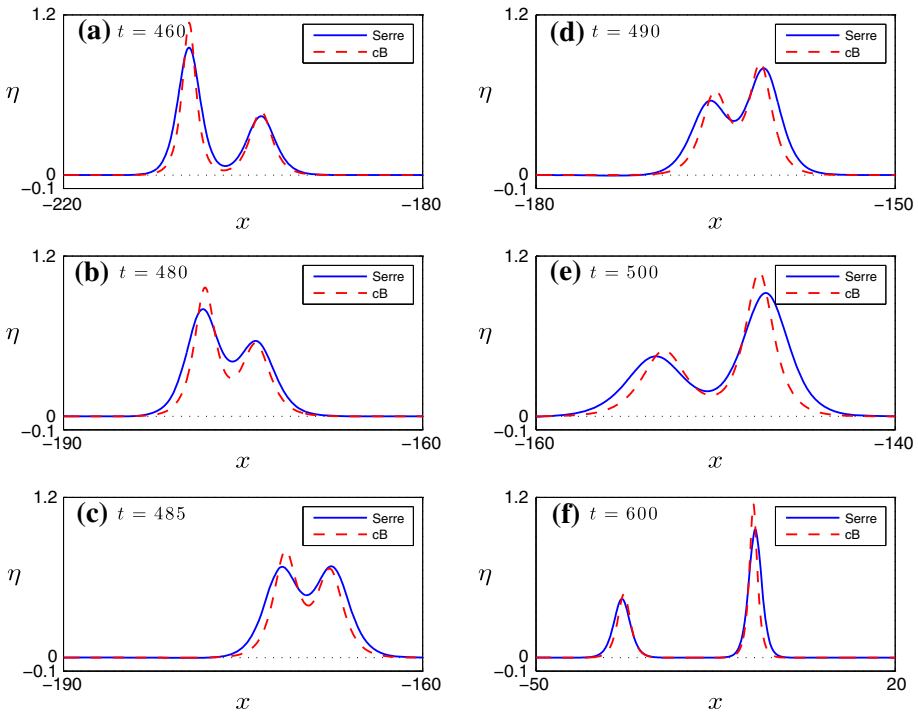


Fig. 12 Continuation of the simulation in Fig. 11, showing the overtaking

Figures 14 and 15 show the results for the Serre and cB systems, respectively. In both systems, a simple DSW is generated, which travels to the right, and a rarefaction wave travels to the left with a small dispersive tail.

To test the non-dissipativity of the FEM scheme, we compare the computational results with the asymptotics of the leading edge solitary wave, whose long-time amplitude and speed are given in Eqs. (9) with the jump [Eq. (8)] $\delta = \eta_0 = 0.4182$. Here, $a_1^{\text{simple}} \approx 0.8656$ and $c_s^{\text{simple}} \approx 1.3453$. Figure 16 shows the peak amplitude and speed of the solitary wave recovered from the computations approach the corresponding asymptotic values (solid horizontal lines). Even though δ is not much smaller than 1, it turns out that the asymptotic values are fairly accurate. These results show that **the FEM is non-dissipative even for DSWs**. We note that the Hamiltonian in these simulations is conserved to within 10 decimal digits of accuracy and it remained $\mathcal{H}(t) = 190.4720453$ even after the interaction of the leading edge with the other parts of the solution and up to $T = 400$, while, Fig. 16 shows that, in the cB system, the DSW travels significantly more slowly and with a larger amplitude.

We also consider the **dam-break problem** (see Sect. 2.2). Here, the initial data for $h(x, 0)$ are the same as (31), but $u(x, 0) = 0$. Figure 17 shows the results of this computation for the Serre system, i.e., two counter-propagating DSWs and rarefaction waves. These initial data generate a simple DSW with (32), whose leading edge solitary wave has amplitude and speed given by (10) with $\delta = \eta_0 = 0.4182$ (Fig. 18). For comparison, Fig. 17d shows the corresponding *non-dispersive* shallow water shock and rarefaction waves, which connect the same flow states. Similar results are obtained for the cB system using the same initial conditions—see Figs. 19 and 20. The discretization parameters are the same as in the previous

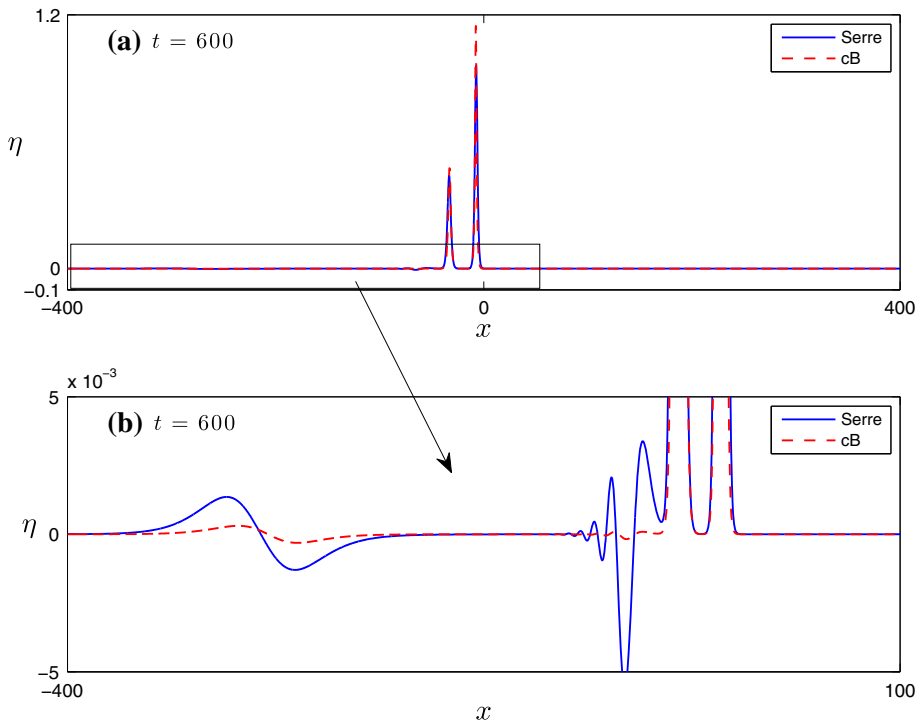


Fig. 13 Same as Fig. 12, showing the dispersive tail in detail

experiment and the Hamiltonian is conserved with 10 decimal digits of accuracy and it was $\mathcal{H}(t) = 87.27888421$.

6 Summary and Conclusions

We present a fully discrete numerical scheme for the Serre system based on the standard Galerkin / finite-element method with smooth periodic splines and on the fourth-order, four-stage, explicit Runge–Kutta method. The computational results show that this numerical scheme is highly accurate and stable. In particular, this scheme achieves the optimal orders of convergence in time and space. Moreover, the actual numerical errors remain fairly small during propagation. In addition, the stability of this scheme does not impose very restrictive conditions on the temporal stepsize.

In addition, we perform a series of highly-accurate numerical experiments of interacting solitary waves in the Serre and ‘classical’ Boussinesq systems. The computational results show that the interactions of solitary waves in the Serre system are more inelastic, i.e., the interaction is significantly longer and incurs a larger amplitude change and larger phase shift. This greater “inelasticity” does not affect the nonlinear stability of the solitary waves. Furthermore, in the Serre system, the dispersive tails generated by the interacting solitary waves have larger amplitude.

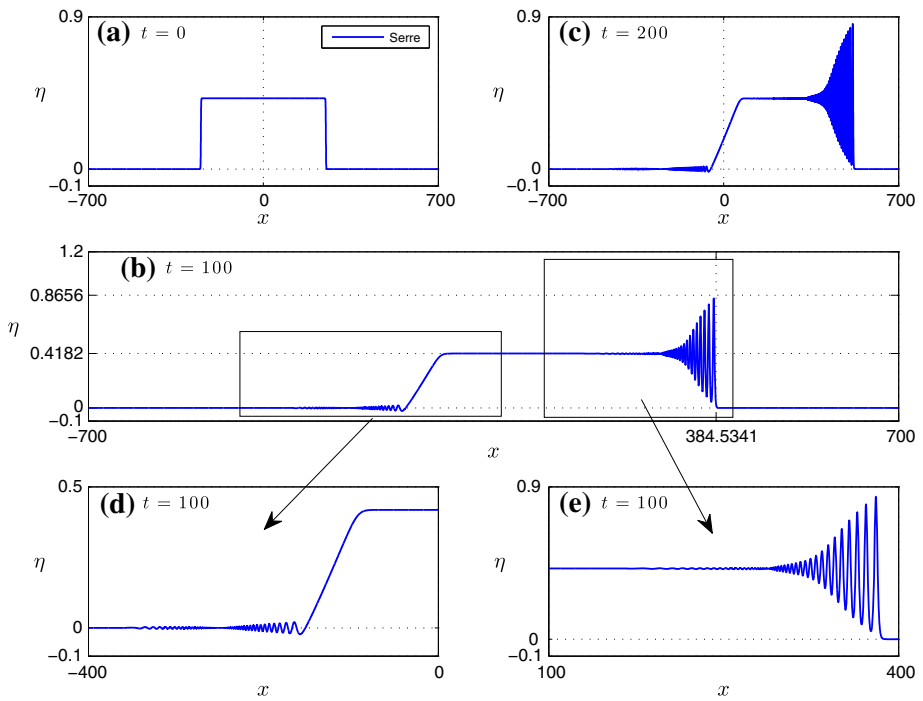


Fig. 14 Simple DSW in the Serre system

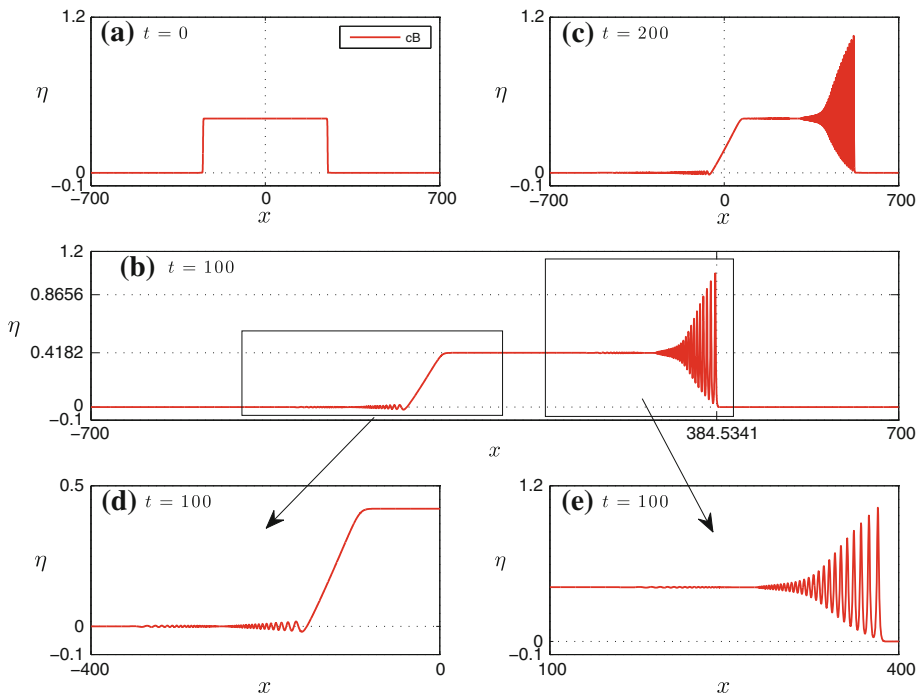


Fig. 15 Same as Fig. 14 in the cB system

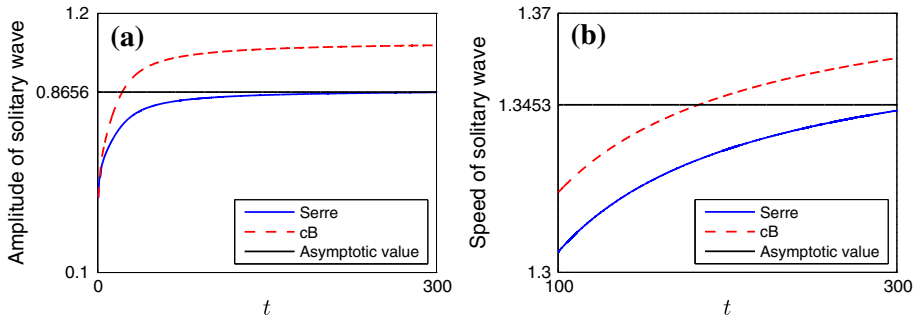


Fig. 16 Amplitude and speed of the leading-edge solitary wave for the simulations in Figs. 14 and 15. Also shown are the asymptotic values for the Serre system [solid horizontal lines, Eqs. (9)]

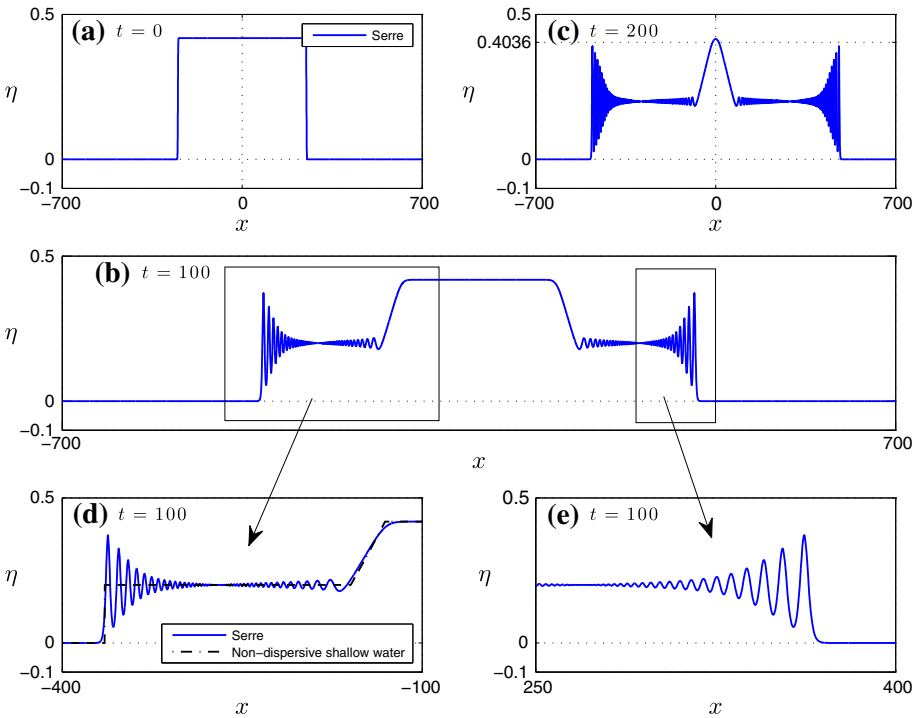


Fig. 17 Dam break in the Serre system. Part (d) also shows the shock and rarefaction waves for the corresponding non-dispersive shallow water problem (dot-dashes)

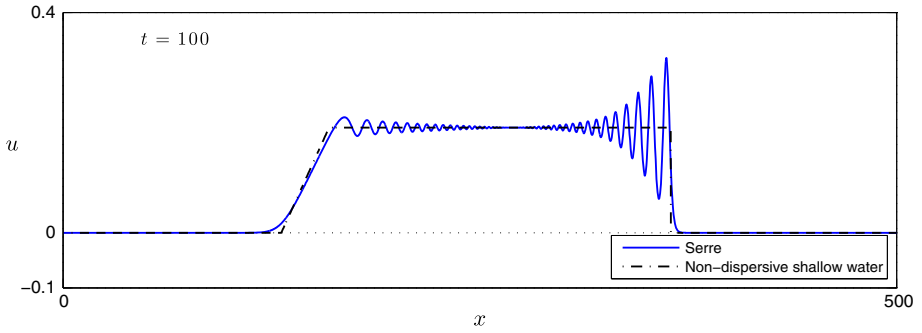


Fig. 18 Velocity profile of the DSW and rarefaction waves corresponding to the free surface elevation shown in the right side of Fig. 17b. Also shown are the shock and rarefaction waves for the corresponding non-dispersive shallow water problem (dot-dashes)

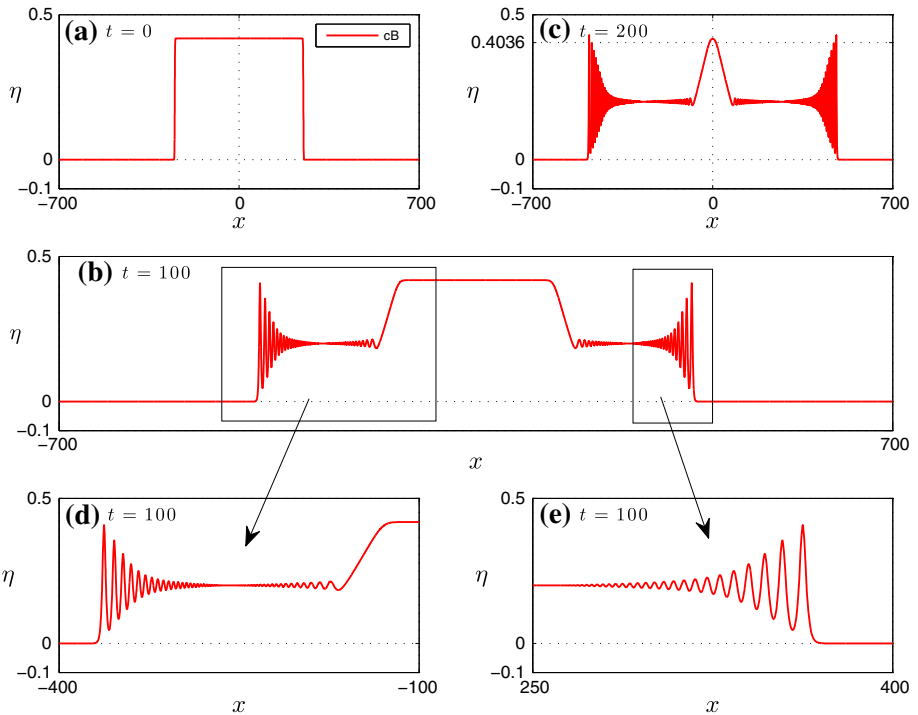


Fig. 19 Same as Fig. 17 in the cB system

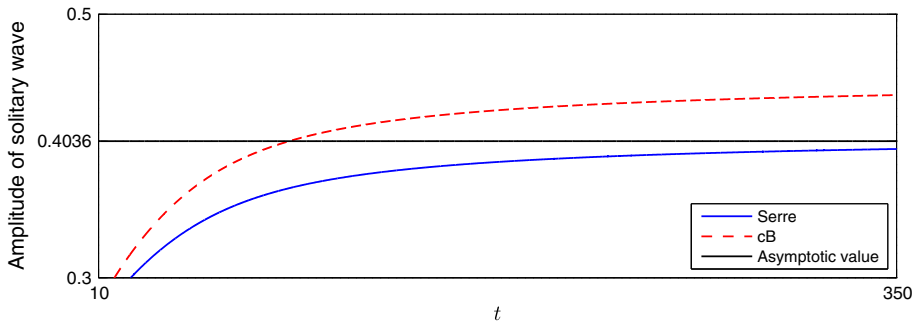


Fig. 20 Amplitude of the leading-edge solitary waves for the simulations in Figs. 17 and 19. Also shown is the asymptotic value for the Serre system [solid horizontal lines, Eq. (10a)]

We also use this scheme to study the generation and propagation of rapidly oscillating dispersive shocks and rarefaction waves. The results show that this scheme can resolve the fine details of the solutions, without inducing numerical (artificial) dissipative effects.

Acknowledgments D. Dutykh would like to acknowledge the hospitality of UC Merced during his visit in April 2013 and the support from ERC under the research project ERC-2011-AdG 290562-MULTIWAVE. D. Mitsotakis would like to thank Prof. Mark Hoefer for his suggestions and his comments and for the fruitful discussions on dispersive waves.

References

1. Antonopoulos, D., Dougalis, V.: Numerical solution of the ‘classical’ Boussinesq system. *Math. Comput. Simul.* **82**, 984–1007 (2012)
2. Antonopoulos, D., Dougalis, V.: Error estimates for Galerkin approximations of the “classical” Boussinesq system. *Math. Comput.* **82**, 689–717 (2013)
3. Antonopoulos, D., Dougalis, V., Mitsotakis, D.: Galerkin approximations of the periodic solutions of Boussinesq systems. *Bull. Greek Math. Soc.* **57**, 13–30 (2010)
4. Antonopoulos, D., Dougalis, V., Mitsotakis, D.: Numerical solution of Boussinesq systems of the Bona–Smith family. *Appl. Numer. Math.* **30**, 314–336 (2010)
5. Barthélémy, E.: Nonlinear shallow water theories for coastal waves. *Surv. Geophys.* **25**, 315–337 (2004)
6. Benjamin, T., Lighthill, J.: On cnoidal waves and bores. *Proc. R. Soc. Lond. Ser. A* **224**, 448–460 (1954)
7. Bona, J., Chen, M., Saut, J.C.: Boussinesq equations and other systems for small-amplitude long waves in nonlinear dispersive media. I: derivation and linear theory. *J. Nonlinear Sci.* **12**, 283–318 (2002)
8. Bona, J., Dougalis, V., Karakashian, O., McKinney, W.: Conservative, high-order numerical schemes for the generalized Korteweg–de Vries equation. *Philos. Trans. R. Soc. Lond. A* **351**, 107–164 (1995)
9. Bonneton, P., Chazel, F., Lannes, D., Marche, F., Tissier, M.: A splitting approach for the fully nonlinear and weakly dispersive Green–Naghdi model. *J. Comput. Phys.* **230**, 1479–1498 (2010)
10. Boussinesq, J.: Théorie des ondes et des remous qui se propagent le long d’ un canal rectangulaire horizontal, en communiquant au liquide contenu dans ce canal des vitesses sensiblement pareilles de la surface au fond. *J. Math. Pure Appl.* **17**, 55–108 (1872)
11. Carter, J., Cienfuegos, R.: The kinematics stability of solitary and cnoidal wave solutions of the Serre equations. *Eur. J. Mech. B: Fluids* **30**, 259–268 (2011)
12. Chazel, F., Lannes, D., Marche, F.: Numerical simulation of strongly nonlinear and dispersive waves using a Green–Naghdi model. *J. Sci. Comput.* **48**, 105–116 (2010)
13. Chen, H., Chen, M., Nguyen, N.: Cnoidal wave solutions to Boussinesq systems. *Nonlinearity* **20**, 1443–1461 (2007)
14. Chen, M.: Solitary-wave and multi pulsed traveling-wave solution of Boussinesq systems. *Appl. Anal.* **7**, 213–240 (2000)
15. Choi, W., Camassa, R.: Exact evolution equations for surface waves. *J. Eng. Mech.* **125**(7), 756 (1999). doi:10.1061/(ASCE)0733-9399(1999)125:7(756)

16. Craig, W., Guyenne, P., Hammack, J., Henderson, D., Sulem, C.: Solitary water wave interactions. *Phys. Fluids* **18**, 57106 (2006)
17. Dias, F., Milewski, P.: On the fully-nonlinear shallow-water generalized Serre equations. *Phys. Lett. A* **374**, 1049–1053 (2010)
18. Dougalis, V., Mitsotakis, D.: Theory and numerical analysis of Boussinesq systems: a review. In: Kampanis, N.A., Dougalis, V.A., Ekaterinaris, J.A. (eds.) *Effective Computational Methods in Wave Propagation*, pp. 63–110. CRC Press, Boca Raton (2008)
19. Duran, A., Dutykh, D., Mitsotakis, D.: On the Galilean invariance of some nonlinear dispersive wave equations. *Stud. Appl. Math.* **131**, 359–388 (2013)
20. Dutykh, D., Clamond, D., Milewski, P., Mitsotakis, D.: Finite volume and pseudo-spectral schemes for fully-nonlinear 1D Serre equations. *Eur. J. Appl. Math.* **24**, 761–787 (2013)
21. Dutykh, D., Katsaounis, T., Mitsotakis, D.: Finite volume schemes for dispersive wave propagation and runup. *J. Comput. Phys.* **230**(8), 3035–3061 (2011)
22. El, G., Grimshaw, R., Smyth, N.: Unsteady undular bores in fully nonlinear shallow-water theory. *Phys. Fluids* **18**, 027104 (2006)
23. El, G., Grimshaw, R., Smyth, N.: Asymptotic description of solitary wave trains in fully nonlinear shallow-water theory. *Phys. D* **237**(19), 2423–2435 (2008)
24. Esler, J.G., Pearce, J.D.: Dispersive dam-break and lock-exchange flows in a two-layer fluid. *J. Fluid Mech.* **667**, 555–585 (2011)
25. Green, A., Naghdi, P.: A derivation of equations for wave propagation in water of variable depth. *J. Fluid Mech.* **78**, 237–246 (1976)
26. Gurevich, A.V., Pitaevskii, L.P.: Nonstationary structure of a collisionless shock wave. *Sov. Phys. JETP* **38**, 291–297 (1974)
27. Israwi, S.: Large time existence for 1D Green–Naghdi equations. *Nonlinear Anal. Theory Methods Appl.* **74**, 81–93 (2011)
28. Johnson, R.S.: Camassa–Holm, Korteweg–de Vries and related models for water waves. *J. Fluid Mech.* **455**, 63–82 (2002)
29. Lannes, D.: *The Water Waves Problem: Mathematical Analysis and Asymptotics*. American Mathematical Society, New York (2013)
30. Lannes, D., Bonneton, P.: Derivation of asymptotic two-dimensional time-dependent equations for surface water wave propagation. *Phys. Fluids* **21**, 016601 (2009)
31. Li, Y.: Hamiltonian structure and linear stability of solitary waves of the Green–Naghdi equations. *J. Nonlinear Math. Phys.* **9**, 99–105 (2002)
32. Lord Rayleigh, J.: On waves. *Philos. Mag.* **1**, 257–279 (1876)
33. Métayer, O.L., Gavriluk, S., Hank, S.: A numerical scheme for the Green–Naghdi model. *J. Comput. Phys.* **229**(6), 2034–2045 (2010)
34. Peregrine, D.: Long waves on beaches. *J. Fluid Mech.* **27**, 815–827 (1967)
35. Schultz, M.: *Spline Analysis*. Prentice Hall, Englewood Cliffs (1973)
36. Seabra-Santos, F., Renouard, D., Temperville, A.: Numerical and experimental study of the transformation of a solitary wave over a shelf or isolated obstacle. *J. Fluid Mech.* **176**, 117–134 (1987)
37. Serre, F.: Contribution à l'étude des écoulements permanents et variables dans les canaux. *La Houille Blanche* **8**, 374–388 (1953)
38. Serre, F.: Contribution à l'étude des écoulements permanents et variables dans les canaux. *La Houille Blanche* **8**, 830–872 (1953)
39. Su, C., Gardner, C.: KdV equation and generalizations. Part III. Derivation of Korteweg–de Vries equation and Burgers equation. *J. Math. Phys.* **10**, 536–539 (1969)
40. Wei, G., Kirby, J.T., Grilli, S.T., Subramanya, R.: A fully nonlinear Boussinesq model for surface waves. Part I. Highly nonlinear unsteady waves. *J. Fluid Mech.* **294**, 71–92 (1995)
41. Whitham, G.B.: Non-linear dispersion of water waves. *J. Fluid Mech.* **27**, 399–412 (1967)
42. Whitham, G.B.: *Linear and Nonlinear Waves*. Wiley, New York (1999)
43. Zhang, Y., Kennedy, A.B., Panda, N., Dawson, C., Westerink, J.J.: Boussinesq–Green–Naghdi rotational water wave theory. *Coastal Eng.* **73**, 13–27 (2013)