Contents

1 Introduction **1**

1.1 Objectives of the Thesis . . . 1

1.2 Original Contribution of the Thesis . . . 1

1.3 Organisation of the Thesis . . . 1

2 The Serre Equations **3**

2.1 The Equations . . . 3

2.1.1 Alternative form of the Serre Equations . . . 5

2.2 Properties of the Serre Equations . . . 6

2.2.1 Conservation Properties . . . 6

2.2.2 Dispersion Properties . . . 7

2.2.3 Analytic Solutions . . . 8

2.2.4 Forced Solutions . . . 11

2.2.5 Asymptotic Results . . . 13

3 Finite Element Volume Method **15**

3.1 Notation for Numerical Grids . . . 15

3.2 Structure Overview . . . 16

(i) Reconstruction . . . 18

(ii) Fluid Velocity . . . 20

(iii) Flux Across the Cell Interfaces . . . 25

(iv) Source Terms . . . 29

(v) Update Cell Averages . . . 30

(vi) Second-Order SSP Runge-Kutta Method . . . 30

3.3 CFL condition . . . 31

3.4 Boundary Conditions . . . 31

3.5 Dry Beds . . . 32

4 Linear Analysis of the Numerical Methods **35**
4.1 Linearised Serre equations with horizontal bed 36
4.2 Evolution Matrix 37
4.2.1 Overview of the Evolution Step 38
(i) Reconstruct the Quantities Inside the Cells 39
(ii) Calculate the Velocity Over the Domain 40
(iii) Calculate All the Fluxes Across the Cell Interfaces 42
(iv) Calculate All the Source Terms for the Cells 45
(v) Update All the Cell Averages Using a Forward Euler Approximation 45
(vi) Update All the Cell Averages Using a Second-Order SSP Runge-Kutta Method 46
4.3 Convergence Analysis 47
4.3.1 Stability 47
4.3.2 Consistency 49
4.4 Dispersion Analysis 54

5 Numerical Validation **63**
5.1 Measuring Convergence and Conservation 63
5.1.1 Measures of Convergence 63
5.1.2 Measures of Conservation 64
5.2 Analytic Solution for Horizontal Bed 65
5.2.1 Results for Solitary Travelling Wave Solution 65
5.3 Analytic Solution for Variable Bathymetry 66
5.3.1 Results for Lake at Rest 66
5.4 Forced Solution For Finite Water Depth 66
5.5 Forced Solution with Dry Beds 66

6 Experimental Validation **77**
6.1 Periodic Waves Over A Submerged Bar 77
6.1.1 Low Frequency Results 78
6.1.2 High Frequency Results 79
6.2 Synolakis 84
6.3 Roeber 84

A Linear Analysis Results **97**

Bibliography **100**

Chapter 1

Introduction

1.1 Objectives of the Thesis

1.2 Original Contribution of the Thesis

1.3 Organisation of the Thesis

Chapter 2

The Serre Equations

2.1 The Equations

There are three primary ways in which the Serre equations have been derived from the Euler equations in the literature; by asymptotic expansion [1, 2], directed fluid sheets [3] and depth integration [4, 5]. In this thesis the depth integration view of the equations is taken, although the derivation is omitted given the extent of literature already available.

From the depth-integration approach the Serre equations describe a free surface defined by its height $h(x,t)$ above a stationary bed profile $b(x)$ and a depth average of its horizontal velocity $u(x,t)$ as in Figure 2.1. The derivation is similar to that of the Shallow Water Wave Equations (SWWE) [], except for the Serre equations we allow the vertical velocity to vary linearly with depth and so we get non-hydrostatic pressure terms and therefore dispersion.

By depth integrating the Euler equations [] with a no-slip condition at the bed and a free surface condition at the free surface we get a depth integrated approximation of the conservation of mass and momentum equations

$$\frac{\partial h}{\partial t} + \frac{\partial (uh)}{\partial x} = 0, \tag{2.1a}$$

$$\frac{\partial (uh)}{\partial t} + \frac{\partial}{\partial x}\left(u^2 h + \frac{gh^2}{2} + \frac{h^2}{2}\Psi + \frac{h^3}{3}\Phi\right) + \frac{\partial b}{\partial x}\left(gh + h\Psi + \frac{h^2}{2}\Phi\right) = 0. \tag{2.1b}$$

Definition 2.1. The Φ and Ψ terms account for the non-hydrostatic part of the

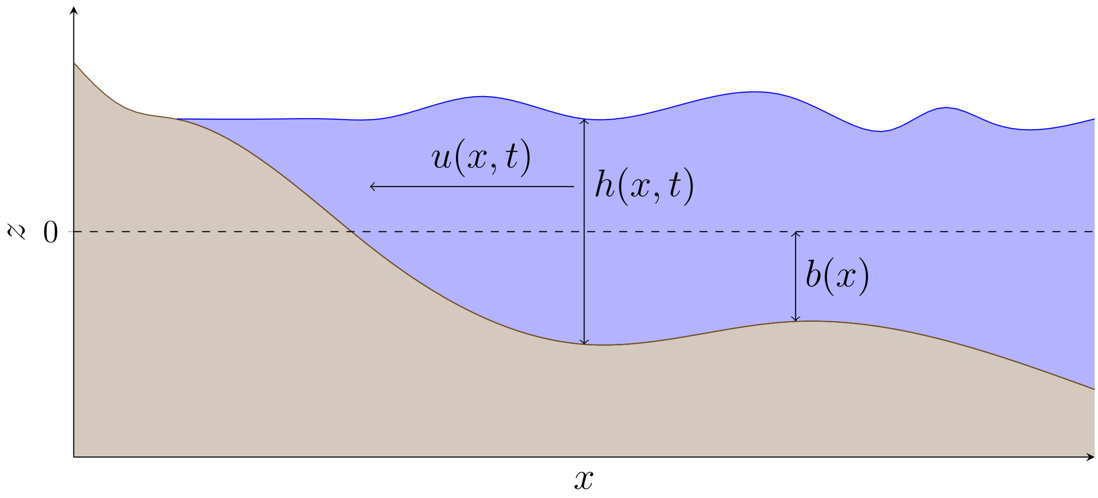


Figure 2.1: Diagram demonstrating the quantities used to describe the fluid (■) and the bed (■) for the Serre equations.

pressure and are

$$\Psi = \frac{\partial b}{\partial x}\left(\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x}\right) + u^2\frac{\partial b}{\partial x}, \tag{2.2a}$$

$$\Phi = \frac{\partial u}{\partial x}\frac{\partial u}{\partial x} - u\frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial x \partial t}. \tag{2.2b}$$

When $\Phi = \Psi = 0$ the Serre equations are equivalent to the SWWE where the vertical velocity is constant in depth, only the hydrostatic pressure is present and there is no dispersion. Due to the presence of the Φ and Ψ terms the Serre equations are much more difficult to solve analytically and numerically than the SWWE. The primary reason for this is that whilst the SWWE are hyperbolic for finite water depth, the Serre equations are neither hyperbolic nor parabolic. Furthermore the Serre equations are not in conservation law form due to the presence of temporal derivatives in Φ and Ψ, although they are derived from conservation equations.

For a horizontal bed $\partial b/\partial x = 0$, thus $\Psi = 0$ and all the source terms drop out and the Serre equations become

$$\frac{\partial h}{\partial t} + \frac{\partial (uh)}{\partial x} = 0, \tag{2.3a}$$

$$\frac{\partial (uh)}{\partial t} + \frac{\partial}{\partial x}\left(u^2 h + \frac{gh^2}{2} + \frac{h^3}{3}\Phi\right) = 0. \tag{2.3b}$$

These equations are still neither hyperbolic nor parabolic and are still not in conservation law from as Φ contains a temporal derivative. As such even for horizontal beds the Serre equations are more difficult to solve analytically and numerically than the SWWE.

2.1.1 Alternative form of the Serre Equations

A major hurdle for developing numerical methods for the Serre equations is the presence of the mixed temporal and spatial derivative in Φ (2.2b). By rewriting the Serre equations and introducing a new conserved quantity G [6, 5, 7], the mixed temporal and spatial derivative can be removed and the Serre equations can be written in conservation law form.

Definition 2.2. The conserved quantity G is

$$G = hu\left(1 + \frac{\partial h}{\partial x}\frac{\partial b}{\partial x} + \frac{1}{2}h\frac{\partial^2 b}{\partial x^2} + \frac{\partial b}{\partial x}^2\right) - \frac{\partial}{\partial x}\left(\frac{1}{3}h^3\frac{\partial u}{\partial x}\right).$$

The Serre equations (2.1) can then rewritten as conservation laws for the conserved variables h and G

$$\frac{\partial h}{\partial t} + \frac{\partial (uh)}{\partial x} = 0, \tag{2.4a}$$

$$\begin{aligned}\frac{\partial G}{\partial t} + \frac{\partial}{\partial x}\left(uG + \frac{gh^2}{2} - \frac{2}{3}h^3\frac{\partial u}{\partial x}^2 + h^2 u\frac{\partial u}{\partial x}\frac{\partial b}{\partial x}\right)& \\ + \frac{1}{2}h^2 u\frac{\partial u}{\partial x}\frac{\partial^2 b}{\partial x^2} - hu^2\frac{\partial b}{\partial x}\frac{\partial^2 b}{\partial x^2} + gh\frac{\partial b}{\partial x} = 0.&\end{aligned} \tag{2.4b}$$

The conserved quantity G resembles h multiplied by the irrotationality [8, 9] and its conservation equation is equivalent to the conservation equation for momentum (2.1b).

This conservation law form makes the Serre equations well suited for a finite volume method for the conservation of mass and G equations, provided one can solve for u given h and G. Even in conservation law from these equations are still considerably more difficult to solve numerically and analytically than the SWWE.

For a horizontal bed $\partial b/\partial x = 0$ the conservation law form of the Serre equations is

$$\frac{\partial h}{\partial t} + \frac{\partial (uh)}{\partial x} = 0, \tag{2.5a}$$

$$\frac{\partial G}{\partial t} + \frac{\partial}{\partial x}\left(uG + \frac{gh^2}{2} - \frac{2}{3}h^3\frac{\partial u}{\partial x}^2\right) = 0, \tag{2.5b}$$

$$G = hu - \frac{\partial}{\partial x}\left(\frac{1}{3}h^3\frac{\partial u}{\partial x}\right). \tag{2.5c}$$

These equations are much simpler to solve than their counterparts that allow for variations in bathymetry, although they still present a number of challenges.

2.2 Properties of the Serre Equations

The Serre equations are significantly more complex than their dispersionless counterparts the SWWE and so fewer analytic solutions can be found in the literature. This increases the need to utilise the known properties of the Serre equations and construct forced solutions to assess our numerical methods. The relevant properties of the Serre equations, analytic solutions and forced solutions are thus presented here, beginning with the conservation properties of the Serre equations.

2.2.1 Conservation Properties

Conservation of a quantity means that in a closed system the total amount of a quantity q remains constant in time.

Definition 2.3. The total amount of a quantity q in a system occurring on the interval $[a, b]$ at time t is

$$\mathcal{C}_q(t) = \int_a^b q(x,t)\, dx.$$

Conservation of a quantity q means that $\mathcal{C}_q(0) = \mathcal{C}_q(t)$ $\forall t$. Given that the Serre equations (2.1) are conservation equations for mass and momentum and that the conservation of momentum equation can be rewritten as a conservation equation for G (2.4), the Serre equations conserve these quantities. Additionally owing to the Hamiltonian structure of the Serre equations, the Serre equations possess a Hamiltonian $\mathcal{H}$ which acts as an energy and is therefore also conserved.

Definition 2.4. The Hamiltonian of the Serre equations is

$$\mathcal{H}(x,t) = \frac{1}{2}\left(hu^2 + \frac{h^3}{3}\left(\frac{\partial u}{\partial x}\right)^2 + gh^2 + 2ghb + u^2h\frac{\partial b}{\partial x} - uh^2\frac{\partial u}{\partial x}\frac{\partial b}{\partial x}\right).$$

For the system to be closed the flux terms of the conservation of mass and momentum equations at both boundaries must with cancel and the integral of the source terms over the domain must be zero.

2.2.2 Dispersion Properties

The dispersion properties of wave equations are primarily studied through linearising the equations, assuming periodic wave solutions and then deriving a relationship between the frequency ω and wave number k of these solutions. For the Serre equations the dispersion relation [] is

$$\omega = Uk \pm k\sqrt{gH}\sqrt{\frac{3}{(kH)^2 + 3}}. \tag{2.6}$$

Barthélemy [10] compared this dispersion relation to that of the linear theory of water waves and demonstrated its utility when k is large. However when k is small the difference between the dispersion relation of the Serre equations and that of water wave theory increases. The dispersion relation of the Serre equations can be modified by introducing terms to reduce this difference [10], but such modifications are beyond the scope of this thesis.

From the dispersion relation (2.6) the phase velocity $v_p = \omega/k$ and the group velocity $v_g = \partial\omega/\partial k$ can be written in terms of wave number as

$$v_p = U \pm \sqrt{gH}\sqrt{\frac{3}{(kH)^2 + 3}}, \tag{2.7a}$$

$$v_g = U \pm \sqrt{gH}\left(\sqrt{\frac{3}{(kH)^2 + 3}} \mp (kH)^2\sqrt{\frac{3}{\left((kH)^2 + 3\right)^3}}\right). \tag{2.7b}$$

Since both the phase and group velocities depend on the wave number, waves of different spatial frequencies travel at different speeds meaning the Serre equations describe dispersive waves.

Fortunately, the phase velocity and the group velocity of waves are bounded, since as $k \to 0$ then $v_p, v_g \to U$ and as $k \to \infty$ then $v_p, v_g \to U \pm \sqrt{gH}$. Therefore we have that

$$U - \sqrt{gH} \leq v_p \leq U + \sqrt{gH}, \tag{2.8a}$$
$$U - \sqrt{gH} \leq v_g \leq U + \sqrt{gH}. \tag{2.8b}$$

2.2.3 Analytic Solutions

Due to the complexity of the Serre equations, few analytic solutions have been discovered. In particular there is a travelling wave solution for horizontal beds and a lake at rest solution for any bathymetry. Both analytic solutions will be used to assess the capabilities of the numerical methods and so we present them here.

Solitary Travelling Wave Solution

The Serre equations admit a travelling wave solution that propagates at a constant speed without deformation due to a balance between nonlinear and dispersive effects. Unlike the Euler equations this travelling wave solution has a closed form

$$h(x,t) = a_0 + a_1 \text{sech}\left(\kappa\left(x - ct\right)\right), \tag{2.9a}$$

$$u(x,t) = c\left(1 - \frac{a_0}{h(x,t)}\right), \tag{2.9b}$$
$$b(x) = 0 \tag{2.9c}$$

with

$$\kappa = \frac{\sqrt{3a_1}}{2a_0\sqrt{(a_0 + a_1)}},$$

$$c = \sqrt{g(a_0 + a_1)}.$$

This wave is has an amplitude of a_1, an infinite wavelength and propagates on water a_0 deep. These solitary waves are not true solitons however, due to their inelastic collisions with one another [11]. These solitary waves can be generalised to a family of periodic travelling wave solutions [12].

From (2.9) we can derive G for the solitary wave solution in terms of h

$$G = c(h - a0) - \frac{ca_0}{3}\left(h\frac{\partial^2 h}{\partial x^2} + \left[\frac{\partial h}{\partial x}\right]\right) \tag{2.10}$$

where the derivatives of h are

$$\begin{aligned}
\frac{\partial h}{\partial x} &= -2a_1\kappa \text{sech}^2\left(\kappa\left(x-ct\right)\right)\tanh\left(\kappa\left(x-ct\right)\right),\\
\frac{\partial^2 h}{\partial x^2} &= 2a_1\kappa^2\left(\cosh\left(2\kappa\left(x-ct\right)\right)-2\right)\text{sech}^4\left(\kappa\left(x-ct\right)\right).
\end{aligned}$$

As we would like to assess the conservation properties of our numerical methods we will require the total mass, momentum, G and Hamiltonian for this solution. In particular we require these totals at the initial time $t = 0$, to allow for various domains we present the integrals in indefinite form

$$\int h(x,0)\;dx = a_0 x + \frac{a_1}{\kappa}\tanh\left(\kappa x\right) + \text{constant}, \tag{2.11a}$$

$$\int u(x,0)h(x,0)\;dx = \frac{a_1 c}{\kappa}\tanh\left(\kappa x\right) + \text{constant}, \tag{2.11b}$$

$$\begin{aligned}
\int G(x,0)\;dx = \frac{ca_1}{3\kappa}\Bigg(&3 + 2a_0^2\kappa^2\text{sech}^2\left(\kappa x\right)\\
&+ 2a_0a_1\kappa^2\text{sech}^4\left(\kappa x\right)\Bigg)\tanh\left(\kappa x\right) + \text{constant},
\end{aligned} \tag{2.11c}$$

$$\begin{aligned}
\int \mathcal{H}(x,0)\;dx = \frac{1}{2}\Bigg(&\int g\left[h(x,0)\right]^2\;dx + \int h(x,0)\left[u(x,0)\right]^2\;dx\\
&+ \int \left[h(x,0)\right]^3\left[\frac{\partial u(x,0)}{\partial x}\right]^2\;dx\Bigg)
\end{aligned} \tag{2.11d}$$

where the integrals of the Hamiltonian are

$$\begin{aligned}\int g\left[h(x,0)\right]^2 \, dx = &\frac{g}{12\kappa}\operatorname{sech}^3(\kappa x)\Big[9a_0^2\kappa x\cosh(\kappa x) + 3a_0^2\kappa x\cosh(3\kappa x)\\ &+ 4a_1\left(3a_0 + 2a_1 + (3a_0+a_1)\cosh(2\kappa x)\right)\sinh(\kappa x)\Big]\\ &+ \text{constant},\end{aligned}$$

$$\begin{aligned}\int h(x,0)\left[u(x,0)\right]^2 \, dx = &\frac{\sqrt{a_1}c^2}{\kappa}\left(-\frac{a0}{\sqrt{a_0+a_1}}\operatorname{arctanh}\left(\frac{\sqrt{a1}\tanh(\kappa x)}{\sqrt{a_0+a1}}\right)\right.\\ &\left.+\frac{\sqrt{a_1}}{\kappa}\tanh(\kappa x)\right) + \text{constant},\end{aligned}$$

$$\begin{aligned}\int \left[h(x,0)\right]^3\left[\frac{\partial u(x,0)}{\partial x}\right]^2 \, dx = &\frac{2a_0^2c^2\kappa}{9\sqrt{a_1}\left(a_0 + a_1\operatorname{sech}^2(\kappa x)\right)}\\ &\times\left(a_0 + 2a_1 + a_0\cosh(2\kappa x)\right)\operatorname{sech}^2(\kappa x)\\ &\times\Bigg[-3a_0\sqrt{a_0+a_1}\operatorname{arctanh}\left(\frac{\sqrt{a1}\tanh(\kappa x)}{\sqrt{a_0+a1}}\right)\\ &+\sqrt{a_1}\left(3a_0 + a_1 - a_1\operatorname{sech}^2(\kappa x)\right)\tanh(\kappa x)\Bigg] + \text{constant}.\end{aligned}$$

Therefore, we have the analytic values of our physical variables h, u, G and b for all t as well as the total amounts of our conserved quantities for the initial conditions when $t = 0s$, as desired.

Lake at Rest

The lake at rest solution is a rudimentary stationary solution of the Serre equations that exists for all bathymetry $b(x)$, due to a balance between hydrostatic pressure and the forcing of the bed slope. The lake at rest solution is

$$h(x,t) = \max\left\{a_0 - b(x), 0\right\}, \tag{2.12a}$$

$$u(x,t) = 0, \tag{2.12b}$$

$$G(x,t) = 0. \tag{2.12c}$$

It represents a quiescent body of water with a horizontal water surface or stage $w(x,t) = h(x,t) + b(x)$ over any bathymetry. The maximum function is included for the water depth to allow for dry regions of the bed when $b(x) > a_0$. We write these solutions in terms of $b(x)$ as this solution holds for all bed profiles.

For these quantities (2.12) the Serre equations (2.4) reduce to

$$\frac{\partial h}{\partial t} = 0,$$

$$\frac{\partial G}{\partial t} + \frac{\partial}{\partial x}\left(\frac{gh^2}{2}\right) + gh\frac{\partial b}{\partial x} = 0.$$

Since we have that $\partial h/\partial x = -\partial b/\partial x$ when $h \neq 0$, then G and h are constant in time and therefore so is u and thus we possess a stationary solution.

The total momentum and G in our system is straightforward to calculate as both are zero everywhere and so we have

$$\int u(x,0)h(x,0)\,dx = 0 + \text{constant}, \tag{2.13}$$

$$\int G(x,0)\,dx = 0 + \text{constant}. \tag{2.14}$$

To calculate the total mass and Hamiltonian in our system we must break up our domain into wet regions where $b(x) < a_0$ and dry regions where $b(x) \geq a_0$. For the dry regions the total mass and energy are 0 and so we have

$$\int h(x,0)\,dx = 0, \tag{2.15a}$$

$$\int \mathcal{H}(x,0)\,dx = 0 \tag{2.15b}$$

whilst in a wet region we have

$$\int h(x,0)\,dx = a_0 x - \int b(x)\,dx, \tag{2.16a}$$

$$\int \mathcal{H}(x,0)\,dx = \frac{g}{2}\left(a_0^2 x - 2a_0\int b(x)\,dx + \int b(x)^2\,dx\right). \tag{2.16b}$$

Therefore as desired we have expressions for all the quantities in terms of the bed profile $b(x)$.

2.2.4 Forced Solutions

To account for the small number of analytic solutions to the Serre equations we also constructed forced solutions to assess our numerical methods. These work by introducing a source term S for both the conservation of mass and G equations that balances these equation to force a solution. For instance by adding the source terms S_{mass} and S_G into (2.4) we obtain the forced Serre equations in conservation law form

$$\frac{\partial h}{\partial t} + \frac{\partial (uh)}{\partial x} + S_{\text{mass}} = 0, \tag{2.17a}$$

$$\frac{\partial G}{\partial t} + \frac{\partial}{\partial x}\left(uG + \frac{gh^2}{2} - \frac{2}{3}h^3\frac{\partial u}{\partial x}^2 + h^2 u\frac{\partial u}{\partial x}\frac{\partial b}{\partial x}\right)$$
$$+\frac{1}{2}h^2 u\frac{\partial u}{\partial x}\frac{\partial^2 b}{\partial x^2} - hu^2\frac{\partial b}{\partial x}\frac{\partial^2 b}{\partial x^2} + gh\frac{\partial b}{\partial x} + S_G = 0. \tag{2.17b}$$

To construct the forced solution we choose some functions for h, u and b for all x and t, which will be our forced analytic solutions to the forced Serre equations. From these chosen functions we determine G, S_{mass} and S_G so that these equations (2.17) possess our chosen h, u and b as solutions.

We demonstrate this for the particular forced solution we employ, a travelling Gaussian wave over a periodic bed profile which has

$$h(x,t) = a_0 + a_1 \exp\left(-\frac{\left((x - a_2 t) - a_3\right)^2}{2a_4}\right), \tag{2.18a}$$

$$u(x,t) = a_5 \exp\left(-\frac{\left((x - a_2 t) - a_3\right)^2}{2a_4}\right), \tag{2.18b}$$

$$b(x) = a_6 \sin\left(a_7 x\right). \tag{2.18c}$$

From these definitions G can be calculated using Def 2.2 while S_{mass} and S_G can be calculated using

$$S_{\text{mass}} = -\frac{\partial h}{\partial t} - \frac{\partial (uh)}{\partial x},$$

$$S_G = -\frac{\partial G}{\partial t} - \frac{\partial}{\partial x}\left(uG + \frac{gh^2}{2} - \frac{2}{3}h^3\frac{\partial u}{\partial x}^2 + h^2 u\frac{\partial u}{\partial x}\frac{\partial b}{\partial x}\right)$$
$$-\frac{1}{2}h^2 u\frac{\partial u}{\partial x}\frac{\partial^2 b}{\partial x^2} + hu^2\frac{\partial b}{\partial x}\frac{\partial^2 b}{\partial x^2} - gh\frac{\partial b}{\partial x}$$

with h, u and b from (2.18) and the corresponding G.

With these values the forced Serre equations (2.17) admit the analytic solutions (2.18). This allows us to assess our numerical solutions for a larger range of scenarios than possible with the current analytic solutions of the Serre equations.

2.2.5 Asymptotic Results

Beyond analytic solutions to the Serre equations there have also been studies of the long term behaviour of the Serre equations for situations that are difficult to treat analytically. One particular scenario of interest is the evolution of a moving discontinuous jump in h known as a bore, which can be observed naturally, for example tidal bores [].

For the non-dispersive SWWE bores propagate at a fixed speed and have a constant shape []. For the Serre equations dispersion causes bores to break up into wave train, which are referred to as undular bores []. This process is more difficult to treat analytically particularly over short time spans and so we do not possess analytic solutions for the Serre equations for bores.

To gain some insight into the behaviour of bores for long time spans Whitham modulation techniques were applied to the Serre equations as $t \to \infty$ [12]. These techniques provided an estimate of the speed S^+ and amplitude A^+ of the front of a bore

$$\frac{\Delta}{\left(A^+ + 1\right)^{1/4}} - \left(\frac{3}{4 - \sqrt{A^+ + 1}}\right)^{21/10} \left(\frac{2}{1 + \sqrt{A^+ + 1}}\right)^{2/5} = 0 \tag{2.19a}$$

$$S^+ = \sqrt{g\left(A^+ + 1\right)} \tag{2.19b}$$

where $\Delta = h_b/h_0$, h_b is the height of the bore and h_0 is the depth of still water in front of the bore. These estimates agreed well with numerical simulations provided that $\Delta < 1.43$ [12].

Chapter 3

Finite Element Volume Method

3.1 Notation for Numerical Grids

We begin by defining the numerical grid in both space and time from a starting location x_0 and a beginning time t^0. For any $j, n \in \mathbb{N}$ we have

$$x_j = x_0 + j\Delta x, \tag{3.1a}$$

$$t^n = t^0 + n\Delta t \tag{3.1b}$$

which produces a uniform grid; x_j in space and t^n in time. The Finite Element Volume Method (FEVM) can be readily adapted to nonuniform grids and we restrict our description of the method to uniform grids for simplicity. We will extend this notation to describe locations and times not on the grid by allowing (3.1) to also have subscripts and superscripts in $\mathbb{R}$.

The grid notation naturally extends to our quantities of interest, for example, for a general quantity q

$$q_j^n = q(x_j, t^n). \tag{3.2}$$

The description of our numerical method focuses on cells which are regions surrounding the grid points. The j^{th} cell is the region $\left[x_{j-1/2}, x_{j+1/2}\right]$ centred around x_j. For each cell we define the average of a quantity

$$\overline{q}_j = \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} q(x,t)\; dx \tag{3.3}$$

in cell j.

In the FEVM we reconstruct quantities at various points inside the cell from the cell average values. We distinguish between two reconstructions that are possible at each cell edge, which both exist due to the overlap of neighbouring cells. To do this we use superscripts so that for the cell edge $x_{j+1/2}$ and a general quantity q, we have $q^-_{j+1/2}$ as the reconstructed value of q at $x_{j+1/2}$ from the j^{th} cell and $q^+_{j+1/2}$ as the reconstructed value of q at $x_{j+1/2}$ from the $(j+1)^{th}$ cell. Since the particular time level is typically obvious from context and hence omitted the use of a superscript will not clutter the notation.

3.2 Structure Overview

To describe the FEVM we first present an overview of the evolution step and then provide the details for each component. We begin our evolution step with all the cell averages for h, w and G at time t^n and all the nodal values of b. We write these as vectors from the 0^{th} cell to the m^{th} in the following way

$$\overline{\boldsymbol{h}} = \begin{bmatrix} \overline{h}^n_0 \\ \overline{h}^n_1 \\ \vdots \\ \overline{h}^n_m \end{bmatrix}, \qquad \overline{\boldsymbol{w}} = \begin{bmatrix} \overline{w}^n_0 \\ \overline{w}^n_1 \\ \vdots \\ \overline{w}^n_m \end{bmatrix}, \qquad \overline{\boldsymbol{G}} = \begin{bmatrix} \overline{G}^n_0 \\ \overline{G}^n_1 \\ \vdots \\ \overline{G}^n_m \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

The evolution step proceeds as follows:

(i) Reconstruction: We reconstruct the quantities h, w, G and b inside every cell at various points. The values of h, w and G in the j^{th} cell are reconstructed at $x_{j-1/2}$, x_j and $x_{j+1/2}$ using the reconstruction operators $\mathcal{R}^+_{j-1/2}$, $\mathcal{R}_j$ and $\mathcal{R}^-_{j+1/2}$ respectively. The bed profile b in the j^{th} cell is reconstructed at $x_{j-1/2}$, $x_{j-1/6}$, $x_{j+1/6}$ and $x_{j+1/2}$ using the reconstruction operators $\mathcal{B}_{j-1/2}$, $\mathcal{B}_{j-1/6}$, $\mathcal{B}_{j+1/6}$ and $\mathcal{B}_{j+1/2}$ respectively. So that

$$\begin{aligned} h^+_{j-1/2} &= \mathcal{R}^+_{j-1/2}\left(\overline{\boldsymbol{h}}\right), & G^+_{j-1/2} &= \mathcal{R}^+_{j-1/2}\left(\overline{\boldsymbol{G}}\right), \\ h_j &= \mathcal{R}_j\left(\overline{\boldsymbol{h}}\right), & G_j &= \mathcal{R}_j\left(\overline{\boldsymbol{G}}\right), \\ h^-_{j+1/2} &= \mathcal{R}^-_{j+1/2}\left(\overline{\boldsymbol{h}}\right), & G^-_{j+1/2} &= \mathcal{R}^-_{j+1/2}\left(\overline{\boldsymbol{G}}\right), \end{aligned}$$

$$\begin{aligned} w^+_{j-1/2} &= \mathcal{R}^+_{j-1/2}\left(\overline{\boldsymbol{w}}\right), & b_{j-1/2} &= \mathcal{B}_{j-1/2}\left(\boldsymbol{b}\right), \\ w_j &= \mathcal{R}_j\left(\overline{\boldsymbol{w}}\right), & b_{j-1/6} &= \mathcal{B}_{j-1/6}\left(\boldsymbol{b}\right), \\ w^-_{j+1/2} &= \mathcal{R}^-_{j+1/2}\left(\overline{\boldsymbol{w}}\right), & b_{j+1/6} &= \mathcal{B}_{j+1/6}\left(\boldsymbol{b}\right), \\ & & b_{j+1/2} &= \mathcal{B}_{j+1/2}\left(\boldsymbol{b}\right). \end{aligned}$$

This generates the vectors of these quantities reconstructed for every cell; $\hat{\boldsymbol{h}}$, $\hat{\boldsymbol{w}}$, $\hat{\boldsymbol{G}}$ and $\hat{\boldsymbol{b}}$ which are

$$\hat{\boldsymbol{h}} = \begin{bmatrix} h^+_{-1/2} \\ h_0 \\ h^-_{1/2} \\ h^+_{1/2} \\ \vdots \\ h_m \\ h^-_{m+1/2} \end{bmatrix}, \quad \hat{\boldsymbol{w}} = \begin{bmatrix} w^+_{-1/2} \\ w_0 \\ w^-_{1/2} \\ w^+_{1/2} \\ \vdots \\ w_m \\ w^-_{m+1/2} \end{bmatrix}, \quad \hat{\boldsymbol{G}} = \begin{bmatrix} G^+_{-1/2} \\ G_0 \\ G^-_{1/2} \\ G^+_{1/2} \\ \vdots \\ G_m \\ G^-_{m+1/2} \end{bmatrix} \quad \text{and} \quad \hat{\boldsymbol{b}} = \begin{bmatrix} b_{-1/2} \\ b_{-1/6} \\ b_{1/6} \\ b_{1/2} \\ \vdots \\ b_{m+1/6} \\ b_{m+1/2} \end{bmatrix}.$$

(ii) Fluid Velocity: The remaining unknown quantity, the depth averaged fluid velocity, u is calculated by solving the elliptic equation (2.2) with a FEM from which we obtain $u_{j-1/2}$, u_j and $u_{j+1/2}$ at every cell. We denote the solution of the FEM by the map $\mathcal{G}$ given $\hat{\boldsymbol{h}}$, $\hat{\boldsymbol{G}}$ and $\hat{\boldsymbol{b}}$ as inputs. So that

$$\hat{\boldsymbol{u}} = \begin{bmatrix} u_{-1/2} \\ u_0 \\ u_{1/2} \\ \vdots \\ u_m \\ u_{m+1/2} \end{bmatrix} = \mathcal{G}\left(\hat{\boldsymbol{h}}, \hat{\boldsymbol{G}}, \hat{\boldsymbol{b}}\right).$$

(iii) Flux Across Cell Interfaces: We calculate the average fluxes $F_{j-1/2}$ and $F_{j+1/2}$ across the cell boundaries $x_{j-1/2}$ and $x_{j+1/2}$ over time using $\mathcal{F}_{j-1/2}$ and $\mathcal{F}_{j+1/2}$ so that

$$F_{j-1/2} = \mathcal{F}_{j-1/2}\left(\hat{\boldsymbol{h}}, \hat{\boldsymbol{G}}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{u}}\right),$$
$$F_{j+1/2} = \mathcal{F}_{j+1/2}\left(\hat{\boldsymbol{h}}, \hat{\boldsymbol{G}}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{u}}\right).$$

(iv) Source Terms: We calculate the source term contribution to the cell average of a quantity over a time step S_j with the operator $\mathcal{S}$ like so

$$S_j = \mathcal{S}_j\left(\hat{\boldsymbol{h}}, \hat{\boldsymbol{w}}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{u}}\right).$$

(v) Update All the Cell Averages Using a Forward Euler Approximation: We update the cell average values from time t^n to the next time level combining a fractional step method and a forward Euler approximation for the flux and source terms.

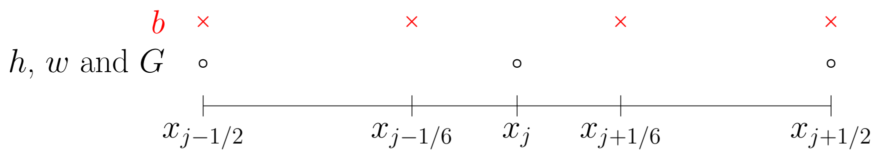


Figure 3.1: Location of the reconstructions for h, w, G and b .

(vi) Update All the Cell Averages Using a Second-Order SSP Runge-Kutta Method: We repeat steps (I)-(V) and use SSP Runge-Kutta time stepping to calculate $\overline{\boldsymbol{h}}$ and $\overline{\boldsymbol{G}}$ at t^{n+1} with second-order accuracy in space and time.

(i) Reconstruction

We now provide the details for how each of the quantities inside the j^{th} cell are calculated at the locations demonstrated in Figure 3.1. For h, w and G the reconstructions is performed from the cell averages. While b is reconstructed from the nodal values.

The height, stage and G

We reconstruct h, w and G with piecewise functions that are linear over a cell with discontinuous jumps at the cell edges. Since h, w and G use the same reconstruction operators we will demonstrate them for a general quantity q. For the j^{th} cell we reconstruct the values of q at $x_{j-1/2}$, x_j and $x_{j+1/2}$ in the following way

$$q^+_{j-1/2} = \mathcal{R}^+_{j-1/2}\left(\overline{\boldsymbol{q}}\right) = \bar{q} - \frac{\Delta x}{2} d_j, \tag{3.4a}$$

$$q_j = \mathcal{R}_j\left(\overline{\boldsymbol{q}}\right) = \bar{q}, \tag{3.4b}$$

$$q^-_{j+1/2} = \mathcal{R}^-_{j+1/2}\left(\overline{\boldsymbol{q}}\right) = \bar{q} + \frac{\Delta x}{2} d_j \tag{3.4c}$$

where

$$d_j = \text{minmod}\left(\theta \frac{\overline{q}_j - \overline{q}_{j-1}}{\Delta x}, \frac{\overline{q}_{j+1} - \overline{q}_{j-1}}{2\Delta x}, \theta \frac{\overline{q}_{j+1} - \overline{q}_j}{\Delta x}\right) \tag{3.5}$$

with $\theta \in [1, 2]$. The choice of the θ parameter changes the diffusion introduced by the reconstruction, with $\theta = 1$ the reconstruction introduces the most diffusion and is equivalent to the basic minmod reconstruction [13] While for $\theta = 2$ the reconstruction is equivalent to the monotized central reconstruction 14 and introduces the least diffusion.

Definition 3.1. The minmod function takes a list of $x_j \in \mathbb{R}$. If all elements have the same sign then minmod returns the element with smallest absolute value, otherwise it returns 0.

$$\text{minmod}\,(x_0, x_1, \dots) := \begin{cases} \min\{x_j\} & x_j > 0\ \forall j \\ \max\{x_j\} & x_j < 0\ \forall j \\ 0 & \text{otherwise} \end{cases}.$$

The nonlinear limiting used to calculate d_j ensures that the reconstruction of h, w and G inside the cell does not introduce non-physical oscillations and is thus Total Variation Diminishing (TVD). It does this by constraining the slope d_j to zero near local extrema, resulting in a first-order approximation, which is always TVD. Away from local extrema d_j will be the gradient with the smallest absolute value, making our reconstruction second-order accurate.

The reconstruction operator $\mathcal{R}_j$ is second-order accurate regardless of the presence of local extrema. This can be seen through the error analysis of the midpoint method for which we get that

$$\overline{q} = \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} q\, dx = q_j + \mathcal{O}\left(\Delta x^2\right). \tag{3.6}$$

The bed profile

To reconstruct b at $x_{j-1/2}$, $x_{j-1/6}$, $x_{j+1/6}$ and $x_{j+1/2}$ we construct the cubic polynomial $C_j(x)$ centred around x_j

$$C_j(x) = c_0\left(x - x_j\right)^3 + c_1\left(x - x_j\right)^2 + c_2\left(x - x_j\right) + c_3. \tag{3.7}$$

By forcing $C_j(x)$ to pass through the nodal values b_{j-2}, b_{j-1}, b_{j+1} and b_{j+2} we get

$$\begin{bmatrix} -8\Delta x^3 & 4\Delta x^2 & -2\Delta x & 1 \\ -\Delta x^3 & \Delta x^2 & -\Delta x & 1 \\ \Delta x^3 & \Delta x^2 & \Delta x & 1 \\ 8\Delta x^3 & 4\Delta x^2 & 2\Delta x & 1 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} b_{j-2} \\ b_{j-1} \\ b_{j+1} \\ b_{j+2} \end{bmatrix}.$$

Solving this we get the polynomial coefficients

$$c_0 = \frac{-b_{j-2} + 2b_{j-1} - 2b_{j+1} + b_{j+2}}{12\Delta x^3},$$

$$c_1 = \frac{b_{j-2} - b_{j-1} - b_{j+1} + b_{j+2}}{6\Delta x^2},$$

$$c_2 = \frac{b_{j-2} - 8b_{j-1} + 8b_{j+1} - b_{j+2}}{12\Delta x},$$

$$c_3 = \frac{-b_{j-2} + 4b_{j-1} + 4b_{j+1} - b_{j+2}}{6}.$$

We require a continuous bed profile across the cell edges and so on top of the reconstruction of the bed given by these coefficients and (3.7) we also average between the two reconstructions at the cell edge so that our reconstruction of the bed profile in the j^{th} cell is

$$b_{j-1/2} = \mathcal{B}_{j-1/2}\left(\boldsymbol{b}\right) = \frac{1}{2}\left(C_j(x_{j-1/2}) + C_{j-1}(x_{j-1/2})\right), \tag{3.8a}$$

$$b_{j-1/6} = \mathcal{B}_{j-1/6}\left(\boldsymbol{b}\right) = C_j(x_{j-1/6}), \tag{3.8b}$$

$$b_{j+1/6} = \mathcal{B}_{j+1/6}\left(\boldsymbol{b}\right) = C_j(x_{j+1/6}), \tag{3.8c}$$

$$b_{j+1/2} = \mathcal{B}_{j+1/2}\left(\boldsymbol{b}\right) = \frac{1}{2}\left(C_j(x_{j+1/2}) + C_{j+1}(x_{j+1/2})\right). \tag{3.8d}$$

(ii) Fluid Velocity

The elliptic equation that relates the conserved variables h and G and the bed profile b to the primitive variable u was given in Def 2.2. To form the FEM we take the weak form of Def 2.2 which is

$$\int_\Omega Gv\;dx = \int_\Omega uh\left(1 + \frac{\partial h}{\partial x}\frac{\partial b}{\partial x} + \frac{1}{2}h\frac{\partial^2 b}{\partial x^2} + \frac{\partial b}{\partial x}^2\right) - \frac{\partial}{\partial x}\left(\frac{1}{3}h^3\frac{\partial u}{\partial x}\right) v\;dx.$$

Integrating by parts with Dirichlet boundary conditions we get

$$\int_\Omega Gv\;dx = \int_\Omega uh\left(1 + \frac{\partial b}{\partial x}^2\right) v\;dx + \int_\Omega \frac{1}{3}h^3\frac{\partial u}{\partial x}\frac{\partial v}{\partial x}\;dx - \int_\Omega \frac{1}{2}h^2\frac{\partial b}{\partial x}u\frac{\partial v}{\partial x}\;dx. - \int_\Omega \frac{1}{2}h^2\frac{\partial b}{\partial x}\frac{\partial u}{\partial x}v\;dx. \tag{3.9}$$

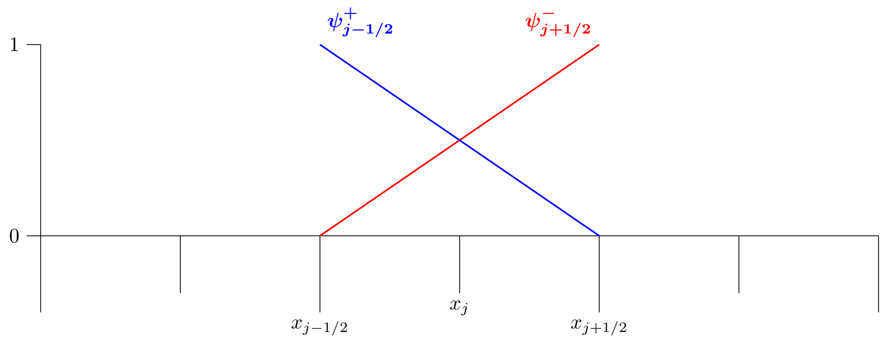


Figure 3.2: Discontinuous linear basis functions over a cell.

This weak formulation implies that if G and h are members of the function space $\mathbb{L}^p$ and b is a member of the Sobolev space $\mathbb{W}^{1,p}$ then u is a member of the Sobolev space $\mathbb{W}^{1,p}$ as well [15]. Since our method requires the derivative of u to be well defined and hence that $u \in \mathbb{W}^{1,p}$ we will restrict ourselves to only allow $h, G \in \mathbb{L}^p$ and $b \in \mathbb{W}^{1,p}$.

We simplify (3.9) by performing the integration over the cells and then summing the integrals together to get the equation for the entire domain

$$\sum_j \int_{x_{j-1/2}}^{x_{j+1/2}} \left[\left(uh \left(1 + \frac{\partial b}{\partial x}^2 \right) - \frac{1}{2} h^2 \frac{\partial b}{\partial x} \frac{\partial u}{\partial x} - G \right) v \right.$$
$$\left. + \left(\frac{1}{3} h^3 \frac{\partial u}{\partial x} - \frac{1}{2} h^2 \frac{\partial b}{\partial x} u \right) \frac{\partial v}{\partial x} \right] dx = 0 \quad (3.10)$$

which holds for all test functions v. The next step is to replace the functions for the quantities h, G, b and u with their basis function approximations.

Basis Function Approximations

For h and G we use the basis functions ψ which are linear inside a cell and 0 everywhere outside the cell as shown in Figure 3.2. This produces a second-order approximation to h and G inside the cell and allows for discontinuities at the cell edges. Since these basis functions are in $\mathbb{L}^p$ our basis function approximations to h and G are in the appropriate function space.

From the basis functions ψ we have the following representation for h and G

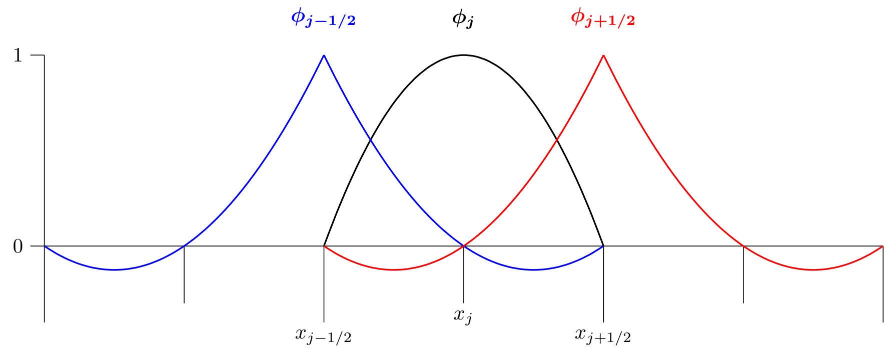


Figure 3.3: Continuous piecewise quadratic basis functions over a cell.

in our finite element method

$$h = \sum_j h^+_{j-1/2}\psi^+_{j-1/2} + h^-_{j+1/2}\psi^-_{j+1/2}, \tag{3.11a}$$

$$G = \sum_j G^+_{j-1/2}\psi^+_{j-1/2} + G^-_{j+1/2}\psi^-_{j+1/2}. \tag{3.11b}$$

For u we require a locally calculated second-order approximation to the first derivative, which forces at least a quadratic representation of u in each cell. Since we require u to be a member of $\mathbb{W}^{1,p}$, u will be continuous. Therefore, we use basis functions ϕ that are continuous at the cell edges $x_{j\pm1/2}$, resulting in the basis functions depicted in Figure 3.3.

From the basis functions ϕ our basis function approximation to u is

$$u = \sum_j u_{j-1/2}\phi_{j-1/2} + u_j\phi_j + u_{j+1/2}\phi_{j+1/2}. \tag{3.12}$$

For the bed term we require a local approximation to the second derivative of the bed that is second-order accurate. To allow for an appropriate second derivative of the bed profile, b must be a member of $\mathbb{W}^{2,p}$ which is smoother than indicated by the elliptic equation (3.9). We choose the cubic basis functions γ which are continuous across the cell edges, as the bed profile will be continuous. These basis functions are shown in Figure 3.4 and from them we get our basis function approximation to b

$$b = \sum_j b_{j-1/2}\gamma_{j-1/2} + b_{j-1/6}\gamma_{j-1/6} + b_{j+1/6}\gamma_{j+1/6} + b_{j+1/2}\gamma_{j+1/2}. \tag{3.13}$$

With $b_{j-1/2}$ and $b_{j+1/2}$ being well defined as our method (3.8) forces both reconstructions at the cell edges to be the identical for adjacent cells.

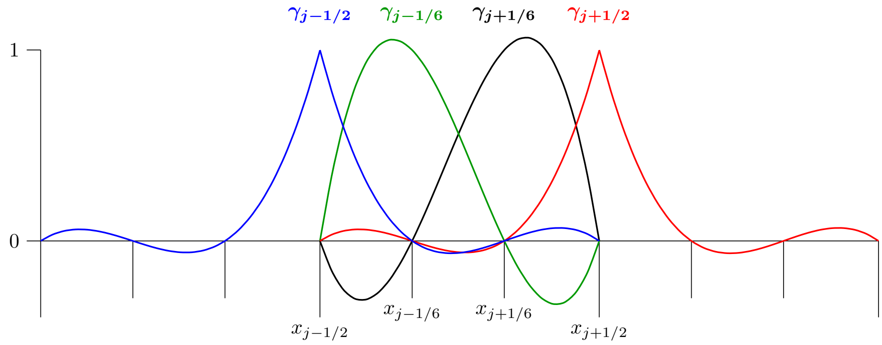


Figure 3.4: Continuous piecewise cubic basis functions over a cell.

Calculation of Element-wise Matrices

The integral equation (3.10) holds for all v. However, since our solution space has the basis functions ϕ it is sufficient to satisfy (3.10) for all ϕ to generate the solution. Since only the basis functions $\phi_{j-1/2}$, ϕ_j and $\phi_{j+1/2}$ are non-zero over the j^{th} cell we can calculate the j^{th} term in the sum (3.10) like so

$$\int_{x_{j-1/2}}^{x_{j+1/2}} \left[\left(uh \left(1 + \frac{\partial b}{\partial x}^2 \right) - \frac{1}{2} h^2 \frac{\partial b}{\partial x} \frac{\partial u}{\partial x} - G \right) \begin{bmatrix} \phi_{j-1/2} \\ \phi_j \\ \phi_{j+1/2} \end{bmatrix} \right.$$
$$\left. + \left(\frac{1}{3} h^3 \frac{\partial u}{\partial x} - \frac{1}{2} h^2 \frac{\partial b}{\partial x} u \right) \frac{\partial}{\partial x} \left(\begin{bmatrix} \phi_{j-1/2} \\ \phi_j \\ \phi_{j+1/2} \end{bmatrix} \right) \right] dx \quad (3.14)$$

where we use our finite element approximations for h (3.11a), G (3.11b), u (3.12) and b (3.13). Because the basis functions over an element are just translations of the other basis functions, this integral can be generalised by moving to the ξ space. The mapping from the x space to the ξ space is

$$x = x_j + \xi \frac{\Delta x}{2}.$$

Making the change of variables from x to ξ in (3.14) we get

$$\frac{\Delta x}{2} \int_{-1}^{1} \left[\left(uh \left(1 + \frac{4}{\Delta x^2} \frac{\partial b}{\partial \xi}^2 \right) - \frac{2}{\Delta x^2} h^2 \frac{\partial b}{\partial \xi} \frac{\partial u}{\partial \xi} - G \right) \begin{bmatrix} \phi_{j-1/2} \\ \phi_j \\ \phi_{j+1/2} \end{bmatrix} \right.$$
$$\left. + \frac{4}{\Delta x^2} \left(\frac{1}{3} h^3 \frac{\partial u}{\partial \xi} - \frac{1}{2} h^2 \frac{\partial b}{\partial \xi} u \right) \frac{\partial}{\partial \xi} \left(\begin{bmatrix} \phi_{j-1/2} \\ \phi_j \\ \phi_{j+1/2} \end{bmatrix} \right) \right] d\xi.$$

We will demonstrate the rest of the process for the uh term as an example and provide the remaining integrals in the Appendix[]. The uh term is

$$\frac{\Delta x}{2}\int_{-1}^{1} uh \begin{bmatrix}\phi_{j-1/2}\\ \phi_j\\ \phi_{j+1/2}\end{bmatrix} d\xi.$$

Since the integral is computed over $\left[x_{j-1/2}, x_{j+1/2}\right]$, there are only a few non-zero contributions from the finite element approximations to h and u, so we have

$$\begin{aligned}
&\frac{\Delta x}{2}\int_{-1}^{1}\left(u_{j-1/2}\phi_{j-1/2}+u_j\phi_j+u_{j+1/2}\phi_{j+1/2}\right)\\
&\qquad\qquad\left(h^{+}_{j-1/2}\psi^{+}_{j-1/2}+h^{-}_{j+1/2}\psi^{-}_{j+1/2}\right)\begin{bmatrix}\phi_{j-1/2}\\ \phi_j\\ \phi_{j+1/2}\end{bmatrix} d\xi\\
&=\frac{\Delta x}{2}\left(h^{+}_{j-1/2}\int_{-1}^{1}\psi^{+}_{j-1/2}\begin{bmatrix}\phi_{j-1/2}\phi_{j-1/2} & \phi_j\phi_{j-1/2} & \phi_{j+1/2}\phi_{j-1/2}\\ \phi_{j-1/2}\phi_j & \phi_j\phi_j & \phi_{j+1/2}\phi_j\\ \phi_{j+1/2}\phi_{j-1/2} & \phi_{j+1/2}\phi_j & \phi_{j+1/2}\phi_{j+1/2}\end{bmatrix} d\xi\right.\\
&\qquad\left.+h^{-}_{j+1/2}\int_{-1}^{1}\psi^{-}_{j+1/2}\begin{bmatrix}\phi_{j-1/2}\phi_{j-1/2} & \phi_j\phi_{j-1/2} & \phi_{j+1/2}\phi_{j-1/2}\\ \phi_{j-1/2}\phi_j & \phi_j\phi_j & \phi_{j+1/2}\phi_j\\ \phi_{j+1/2}\phi_{j-1/2} & \phi_{j+1/2}\phi_j & \phi_{j+1/2}\phi_{j+1/2}\end{bmatrix} d\xi\right)\\
&\qquad\qquad\qquad\qquad\qquad\qquad\begin{bmatrix}u_{j-1/2}\\ u_j\\ u_{j+1/2}\end{bmatrix}.
\end{aligned}$$

Calculating the integrals of all the basis function combinations we get

$$\begin{aligned}
&\frac{\Delta x}{2}\int_{-1}^{1} uh \begin{bmatrix}\phi_{j-1/2}\\ \phi_j\\ \phi_{j+1/2}\end{bmatrix} d\xi =\\
&\frac{\Delta x}{2}\begin{bmatrix}\frac{7}{30}h^{+}_{j-1/2}+\frac{1}{30}h^{-}_{j+1/2} & \frac{4}{30}h^{+}_{j-1/2} & -\frac{1}{30}h^{+}_{j-1/2}-\frac{1}{30}h^{-}_{j+1/2}\\ \frac{4}{30}h^{+}_{j-1/2} & \frac{16}{30}h^{+}_{j-1/2}+\frac{16}{30}h^{-}_{j+1/2} & \frac{4}{30}h^{-}_{j+1/2}\\ -\frac{1}{30}h^{+}_{j-1/2}-\frac{1}{30}h^{-}_{j+1/2} & \frac{4}{30}h^{-}_{j+1/2} & \frac{1}{30}h^{+}_{j-1/2}+\frac{7}{30}h^{-}_{j+1/2}\end{bmatrix}\\
&\qquad\qquad\qquad\qquad\qquad\qquad\begin{bmatrix}u_{j-1/2}\\ u_j\\ u_{j+1/2}\end{bmatrix}.
\end{aligned} \tag{3.15}$$

Assembly of the Global Matrix

By combining all the matrices generated by the integral of each of the u terms we get the j^{th} cells contribution to the stiffness matrix $\mathbf{A}_j$. Likewise all the integrals of the remaining term Gv generate the vector $\boldsymbol{g}_j$. Therefore, (3.10) can be rewritten as

$$\sum_j \mathbf{A}_j \begin{bmatrix} u_{j-1/2} \\ u_j \\ u_{j+1/2}. \end{bmatrix} = \sum_j \boldsymbol{g}_j. \tag{3.16}$$

This is a penta-diagonal matrix equation which can be solved by direct banded matrix solution techniques to obtain

$$\hat{\boldsymbol{u}} = \mathcal{G}\left(\hat{\boldsymbol{h}}, \hat{\boldsymbol{G}}, \hat{\boldsymbol{b}}\right) = \mathbf{A}^{-1}\boldsymbol{g} \tag{3.17}$$

as desired.

(iii) Flux Across the Cell Interfaces

We use Kurganovs method [16] to calculate the flux across a cell interface. This method was employed because; it can handle discontinuities across the cell boundary and only requires an estimate of the maximum and minimum wave speeds. This is precisely the situation for the Serre equations which do not have an expression for the characteristics but do possess estimates on the maximum and minimum wave speeds (2.8a).

We will only demonstrate the calculation of the flux term for $F_{j+1/2}$ as the process to calculate the flux term $F_{j-1/2}$ is identical but with different cells. For a general quantity q, Kurganov's method [16] is

$$F_{j+\frac{1}{2}} = \frac{a^+_{j+\frac{1}{2}} f\left(q^-_{j+\frac{1}{2}}\right) - a^-_{j+\frac{1}{2}} f\left(q^+_{j+\frac{1}{2}}\right)}{a^+_{j+\frac{1}{2}} - a^-_{j+\frac{1}{2}}} + \frac{a^+_{j+\frac{1}{2}}\, a^-_{j+\frac{1}{2}}}{a^+_{j+\frac{1}{2}} - a^-_{j+\frac{1}{2}}} \left[q^+_{j+\frac{1}{2}} - q^-_{j+\frac{1}{2}}\right] \tag{3.18}$$

where $a^+_{j+\frac{1}{2}}$ and $a^-_{j+\frac{1}{2}}$ are given by the wave speed bounds. For the Serre equations we have the wave speed bounds (2.8a) and so we obtain

$$a^-_{j+\frac{1}{2}} = \min\left\{0\ ,\ u^-_{j+1/2} - \sqrt{gh^-_{j+1/2}}\ ,\ u^+_{j+1/2} - \sqrt{gh^+_{j+1/2}}\right\}, \tag{3.19}$$

$$a^+_{j+\frac{1}{2}} = \max\left\{0\ ,\ u^-_{j+1/2} + \sqrt{gh^-_{j+1/2}}\ ,\ u^+_{j+1/2} + \sqrt{gh^+_{j+1/2}}\right\}. \tag{3.20}$$

The flux functions $f(q^-_{j+\frac{1}{2}})$ and $f(q^+_{j+\frac{1}{2}})$ are evaluated using the reconstructed values from the j^{th} and $(j+1)^{th}$ cell respectively. From the continuity equation (2.4a) we have

$$
\begin{aligned}
f\left(h^-_{j+\frac{1}{2}}\right) &= u^-_{j+1/2} h^-_{j+1/2},\\
f\left(h^+_{j+\frac{1}{2}}\right) &= u^+_{j+1/2} h^+_{j+1/2}.
\end{aligned}
$$

For the irrotationality equation (2.4b) we have

$$
\begin{aligned}
f\left(G^-_{j+\frac{1}{2}}\right) &= u^-_{j+1/2} G^-_{j+1/2} + \frac{g}{2}\left(h^-_{j+1/2}\right)^2 - \frac{2}{3}\left(h^-_{j+1/2}\right)^3 \left[\left(\frac{\partial u}{\partial x}\right)^-_{j+1/2}\right]^2 && (3.21a)\\
&+ \left(h^-_{j+1/2}\right)^2 u^-_{j+1/2} \left(\frac{\partial u}{\partial x}\right)^-_{j+1/2} \left(\frac{\partial b}{\partial x}\right)^-_{j+1/2}, && (3.21b)\\
f\left(G^+_{j+\frac{1}{2}}\right) &= u^+_{j+1/2} G^+_{j+1/2} + \frac{g}{2}\left(h^+_{j+1/2}\right)^2 - \frac{2}{3}\left(h^+_{j+1/2}\right)^3 \left[\left(\frac{\partial u}{\partial x}\right)^+_{j+1/2}\right]^2 && (3.21c)\\
&+ \left(h^+_{j+1/2}\right)^2 u^+_{j+1/2} \left(\frac{\partial u}{\partial x}\right)^+_{j+1/2} \left(\frac{\partial b}{\partial x}\right)^+_{j+1/2}. && (3.21d)
\end{aligned}
$$

During the reconstruction process we calculated $h^+_{j-1/2}$, $h^-_{j+1/2}$, $G^+_{j-1/2}$, $G^-_{j+1/2}$ (3.4) while during the calculation of velocity we obtained $u^{\pm}_{j+1/2} = u_{j+1/2}$; because u is continuous across the cell boundaries. Therefore, we further require an approximation to $\left(\frac{\partial b}{\partial x}\right)^{\pm}_{j+1/2}$ and $\left(\frac{\partial u}{\partial x}\right)^{\pm}_{j+1/2}$ to calculate the flux (3.21).

Calculation of Derivatives

To calculate the derivatives in u and b we use the basis function approximation to these quantities in the FEM. For u we have the quadratic $P^u_j(x)$ that passes through $u_{j-1/2}$, u_j and $u_{j+1/2}$ while for b we have the cubic $P^b_j(x)$ that passes through $b_{j-1/2}$, $b_{j-1/6}$, $b_{j+1/6}$ and $b_{j+1/2}$. So we have

$$
\begin{aligned}
P^u_j(x) &= p^u_0 \left(x - x_j\right)^2 + p^u_1 \left(x - x_j\right) + p^u_2, && (3.22a)\\
P^b_j(x) &= p^b_0 \left(x - x_j\right)^3 + p^b_1 \left(x - x_j\right)^2 + p^b_2 \left(x - x_j\right) + p^b_3, && (3.22b)
\end{aligned}
$$

By forcing the polynomials to pass through these reconstructed values we get that for $P_j^u(x)$

$$\begin{aligned}
p_0^u &= \frac{u_{j-1/2} - 2u_j + u_{j+1/2}}{2\Delta x^2}, \\
p_1^u &= \frac{-u_{j-1/2} + u_{j+1/2}}{\Delta x}, \\
p_2^u &= u_j.
\end{aligned}$$

While for $P_j^b(x)$ we get

$$p_0^b = \frac{-9b_{j-1/2} + 27b_{j-1/6} - 27b_{j+1/6} + 9b_{j+1/2}}{2\Delta x^3},$$

$$p_0^b = \frac{9b_{j-1/2} - 9b_{j-1/6} - 9b_{j+1/6} + 9b_{j+1/2}}{4\Delta x^2},$$

$$p_0^b = \frac{b_{j-1/2} - 27b_{j-1/6} + 27b_{j+1/6} - b_{j+1/2}}{8\Delta x},$$

$$p_0^b = \frac{-b_{j-1/2} + 9b_{j-1/6} + 9b_{j+1/6} - b_{j+1/2}}{16}.$$

Taking the derivative of the polynomials (3.22) we get

$$\begin{aligned}
\frac{\partial}{\partial x} P_j^u(x) &= 2p_0^u \left(x - x_j\right) + p_1^u, \\
\frac{\partial}{\partial x} P_j^b(x) &= 3p_0^b \left(x - x_j\right)^2 + 2p_1^b \left(x - x_j\right) + p_2^b.
\end{aligned}$$

This gives a second-order approximation to the derivative of u and b at $x_{j+1/2}$ for the j^{th} cell. The process for the $(j+1)^{th}$ cell is the same and we get

$$\left(\frac{\partial u}{\partial x}\right)^-_{j+1/2} = \frac{\partial}{\partial x} P_j^u(x_{j+1/2}), \tag{3.23a}$$

$$\left(\frac{\partial u}{\partial x}\right)^+_{j+1/2} = \frac{\partial}{\partial x} P_{j+1}^u(x_{j+1/2}), \tag{3.23b}$$

$$\left(\frac{\partial b}{\partial x}\right)^-_{j+1/2} = \frac{\partial}{\partial x} P_j^b(x_{j+1/2}), \tag{3.23c}$$

$$\left(\frac{\partial b}{\partial x}\right)^+_{j+1/2} = \frac{\partial}{\partial x} P_{j+1}^b(x_{j+1/2}). \tag{3.23d}$$

Therefore, we possess all the terms need to calculate the approximation to the flux (3.18) for both the continuity and irrotationality equation, as desired. However, we must make a slight modification to ensure that our scheme is well balanced and recovers the lake at rest steady state solution.

Well Balancing

To recover the lake at rest steady state solution we follow the work of Audusse et al. [17], who accomplished this for the Shallow Water Wave equations. It was demonstrated that this process could also be extended to the Serre equations [18]. To enforce well balancing h is modified at the cell edges in the following way; we first calculate

$$\dot{b}^-_{j+1/2} = w^-_{j+1/2} - h^-_{j+1/2} \qquad \text{and} \qquad \dot{b}^+_{j+1/2} = w^+_{j+1/2} - h^+_{j+1/2}. \tag{3.24}$$

Find the maximum

$$\grave{b}_{j+1/2} = \max\left\{\dot{b}^-_{j+1/2}, \dot{b}^+_{j+1/2}\right\}$$

then we define

$$\grave{h}^-_{j+1/2} = \max\left\{0, w^-_{j+1/2} - \grave{b}_{j+1/2}\right\}, \tag{3.25a}$$

$$\grave{h}^+_{j+1/2} = \max\left\{0, w^+_{j+1/2} - \grave{b}_{j+1/2}\right\}. \tag{3.25b}$$

This generates the vector $\grave{\boldsymbol{h}}$

$$\grave{\boldsymbol{h}} = \begin{bmatrix} \grave{h}^+_{-1/2} \\ \grave{h}^-_{1/2} \\ \grave{h}^+_{1/2} \\ \vdots \\ \grave{h}^-_{m+1/2} \end{bmatrix}$$

which we use to calculate the flux term $F_{j+1/2}$ in (3.18) for both the continuity (2.4a) and irrotationality (2.4b) equation. Applying the same process but with different cells we obtain $F_{j-1/2}$ and we have

$$F_{j-1/2} = \mathcal{F}_{j-1/2}\left(\grave{\boldsymbol{h}}, \hat{\boldsymbol{G}}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{u}}\right), \tag{3.26a}$$

$$F_{j+1/2} = \mathcal{F}_{j+1/2}\left(\grave{\boldsymbol{h}}, \hat{\boldsymbol{G}}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{u}}\right) \tag{3.26b}$$

for both the continuity and irrotationality equation as desired.

(iv) Source Terms

The Serre equations in conservation law form (2.4) contain a flux term and a source term, to treat this numerically we use a first-order splitting method. This requires an approximation to the source term at the cell centre x_j which we denote as S_j. Since the continuity equation (2.4a) has no source term, we will just present the calculation of the source term for the irrotationality equation (2.4b).

Following the work of Audusse et al. [17], we split our approximation to S_j into the centred source term S_{ci} and the corrective interface source terms $S^-_{j+\frac{1}{2}}$ and $S^+_{j+\frac{1}{2}}$. Where S_{ci} is the naive source term approximation and $S^-_{j+\frac{1}{2}}$ and $S^+_{j+\frac{1}{2}}$ are correction terms that ensure that the flux and source term cancel for the lake at rest steady state.

We calculate the centred source term using

$$S_{ci} = -\frac{1}{2}\left(h_j\right)^2 u_j \left(\frac{\partial u}{\partial x}\right)_j \left(\frac{\partial^2 b}{\partial x^2}\right)_j + h_j \left(u_j\right)^2 \left(\frac{\partial b}{\partial x}\right)_j \left(\frac{\partial^2 b}{\partial x^2}\right)_j - g h_j \left(\frac{\partial b}{\partial x}\right)_j .$$

Where we use h_j from the reconstruction process (3.4) and u_j from the solution of the elliptic equation (3.17). To calculate the derivatives we employ our polynomial representations of u and b inside a cell. However, to ensure that the terms cancel properly for a lake at rest we must modify our approximation to $\dfrac{\partial b}{\partial x}$ to use $\dot{b}^-_{j+1/2}$ and $\dot{b}^+_{j+1/2}$ from (3.24). Therefore, the following approximations are used in the calculation of S_{ci}

$$\left(\frac{\partial u}{\partial x}\right)_j = \frac{\partial}{\partial x} P^u_j(x_j), \tag{3.27a}$$

$$\left(\frac{\partial b}{\partial x}\right)_j = \frac{\dot{b}^-_{j+1/2} - \dot{b}^+_{j-1/2}}{\Delta x}, \tag{3.27b}$$

$$\left(\frac{\partial^2 b}{\partial x^2}\right)_j = \frac{\partial^2}{\partial x^2} P^b_j(x_j). \tag{3.27c}$$

The corrective interface source terms are

$$S^-_{j+\frac{1}{2}} = \frac{g}{2}\left(\grave{h}^-_{j+\frac{1}{2}}\right)^2 - \frac{g}{2}\left(h^-_{j+\frac{1}{2}}\right)^2,$$

$$S^+_{j-\frac{1}{2}} = \frac{g}{2}\left(h^+_{j-\frac{1}{2}}\right)^2 - \frac{g}{2}\left(\grave{h}^+_{j-\frac{1}{2}}\right)^2.$$

Which makes use of $h^-_{j+\frac{1}{2}}$ and $h^+_{j+\frac{1}{2}}$ obtained from the reconstruction (3.4) and the modified values $\grave{h}^-_{j+\frac{1}{2}}$ and $\grave{h}^+_{j+\frac{1}{2}}$ from (3.25). Combining the centred and interface source terms our approximation to the source term for the irrotationality equation is

$$S_j = \mathcal{S}_j\left(\hat{\boldsymbol{h}}, \grave{\boldsymbol{h}}, \hat{\boldsymbol{w}}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{u}}\right) = S^-_{j+\frac{1}{2}} + \Delta x S_{ci} + S^+_{j-\frac{1}{2}}. \tag{3.28}$$

(v) Update Cell Averages

We use the forward Euler approximation to approximate the time derivatives and obtain an update formula for the cell averages. Additionally we employ a fractional step method [] to split the flux term and source term part of the Serre equations written in conservation law form (2.4). This results in the following time-stepping method that is first-order accurate in time

$$\begin{aligned}\overline{q}'_j &= \overline{q}^n_j + \frac{\Delta t}{\Delta x}\left(F_{j+\frac{1}{2}} - F_{j-\frac{1}{2}}\right),\\ \overline{q}^{n+1}_j &= \overline{q}'_j + \frac{\Delta t}{\Delta x} S_j\end{aligned}$$

where $F_{j+\frac{1}{2}}$, $F_{j-\frac{1}{2}}$ and S_j are all calculated using the quantities at time t^n. Therefore, this method can be condensed into

$$\overline{q}^{n+1}_j = \overline{q}^n_j + \frac{\Delta t}{\Delta x}\left(F_{j+\frac{1}{2}} - F_{j-\frac{1}{2}} + S_j\right). \tag{3.29}$$

(vi) Second-Order SSP Runge-Kutta Method

To increase the order of accuracy in time we employ the strong stability preserving Runge-Kutta method [19] which is a convex combination of multiple first-order time steps (3.29) in the following way

$$\overline{q}'_j = \overline{q}^n_j + \frac{\Delta t}{\Delta x}\left(F_{j+\frac{1}{2}} - F_{j-\frac{1}{2}} + S_j\right), \tag{3.30a}$$

$$\overline{q}''_j = \overline{q}'_j + \frac{\Delta t}{\Delta x}\left(F'_{j+\frac{1}{2}} - F'_{j-\frac{1}{2}} + S'_j\right), \tag{3.30b}$$

$$\overline{q}^{n+1}_j = \frac{1}{2}\left(\overline{q}^n_j + \overline{q}''_j\right). \tag{3.30c}$$

This results in a time stepping method that preserves the stability of the first-order method (3.29) and is second-order accurate in time. Since all the spatial approximations are second-order accurate, the steps (I-VI) result in a second-order accurate FEVM for the Serre equations, as desired.

3.3 CFL condition

To ensure the stability of our FEVM we use the Courant-Friedrichs-Lewy (CFL) condition [20] which is necessary for stability. The CFL condition ensures that time steps are small enough so that information is only transferred between neighbouring cells. For the Serre equations the CFL condition is

$$\Delta t \leq \frac{Cr}{\max_j \left\{ a^{\pm}_{j+1/2} \right\}} \Delta x \tag{3.31}$$

where $a^{\pm}_{j+1/2}$ are the wavespeed bounds used in Kurganovs flux approximation (3.20) and $0 \leq Cr \leq 1$ is the Courant number. Typically, we use the conservative $Cr = 0.5$ for our numerical experiments.

3.4 Boundary Conditions

To numerically model the Serre equations over finite spatial domains we must enforce boundary conditions at the left and right edge of the domain; $x_{-1/2}$ and $x_{m+1/2}$ respectively. We have only developed Dirichlet boundary conditions for the FEVM, which we enforce using ghost cells located outside the domain boundaries. These ghost cells contain the complete representation of their respective quantities over the cell. For h, w, G and u only one ghost cell at each boundary is required, while for b we require two ghost cells at each boundary. We therefore have ghost cells with the following associated quantities

$$\hat{\boldsymbol{h}}_{-1} = \begin{bmatrix} h^{+}_{-3/2} \\ h_{-1} \\ h^{-}_{-1/2} \end{bmatrix}, \qquad \hat{\boldsymbol{h}}_{m+1} = \begin{bmatrix} h^{+}_{m+1/2} \\ h_{m+1} \\ h^{-}_{m+3/2} \end{bmatrix},$$

$$\hat{\boldsymbol{w}}_{-1} = \begin{bmatrix} w^{+}_{-3/2} \\ w_{-1} \\ w^{-}_{-1/2} \end{bmatrix}, \qquad \hat{\boldsymbol{w}}_{m+1} = \begin{bmatrix} w^{+}_{m+1/2} \\ w_{m+1} \\ w^{-}_{m+3/2} \end{bmatrix},$$

$$\hat{\boldsymbol{G}}_{-1} = \begin{bmatrix} G^{+}_{-3/2} \\ G_{-1} \\ G^{-}_{-1/2} \end{bmatrix}, \qquad \hat{\boldsymbol{G}}_{m+1} = \begin{bmatrix} G^{+}_{m+1/2} \\ G_{m+1} \\ G^{-}_{m+3/2} \end{bmatrix},$$

$$\hat{\boldsymbol{u}}_{-1} = \begin{bmatrix} u_{-3/2} \\ u_{-1} \\ u_{-1/2} \end{bmatrix}, \qquad \hat{\boldsymbol{u}}_{m+1} = \begin{bmatrix} u_{m+1/2} \\ u_{m+1} \\ u_{m+3/2} \end{bmatrix}$$

$$\hat{\boldsymbol{b}}_{-2} = \begin{bmatrix} b_{-5/2} \\ b_{-13/6} \\ b_{-11/6} \\ b_{-3/2} \end{bmatrix}, \quad \hat{\boldsymbol{b}}_{-1} = \begin{bmatrix} b_{-3/2} \\ b_{-7/6} \\ b_{-5/6} \\ b_{-1/2} \end{bmatrix}, \quad \hat{\boldsymbol{b}}_{m+1} = \begin{bmatrix} b_{m+1/2} \\ b_{m+5/6} \\ b_{m+7/6} \\ b_{m+3/2} \end{bmatrix}, \quad \hat{\boldsymbol{b}}_{m+2} = \begin{bmatrix} b_{m+3/2} \\ b_{m+11/6} \\ b_{m+13/6} \\ b_{m+5/2} \end{bmatrix}.$$

To ensure that the solution of u by (3.17) agrees with the boundary conditions $\hat{\boldsymbol{u}}_{-1}$ and $\hat{\boldsymbol{u}}_m$ the element matrices $\mathbf{A}_0$ and $\mathbf{A}_m$ and vectors $\boldsymbol{g}_0$ and $\boldsymbol{g}_m$ must be modified in the following way

$$\mathbf{A}_0 = \begin{bmatrix} 1 & 0 & 0 \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \qquad \boldsymbol{g}_0 = \begin{bmatrix} u_{-1/2} \\ g_1 \\ g_2 \end{bmatrix}, \tag{3.32}$$

$$\mathbf{A}_m = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{bmatrix}, \qquad \boldsymbol{g}_m = \begin{bmatrix} g_0 \\ g_1 \\ u_{m+1/2} \end{bmatrix}. \tag{3.33}$$

These are then combined with the other element contributions to the global matrix (3.16).

3.5 Dry Beds

Dry beds are handled adequately by all steps of our FEVM in their current form, except the FEM solution for u. For the FEM a dry bed presents two issues; when $h = 0$ the stiffness matrix will become singular, and when h is small u may become quite large.

In previous work [21] we employed direct banded matrix solvers such as the Thomas algorithm to solve (3.17), however such methods rely on non-singular matrices and so are unsuitable for dry beds. To deal with this an LU decomposition algorithm by Press et al. [22] was used. This algorithm solves banded matrix problems using an LU decomposition with partial pivoting, which inserts small non-zero pivots when their value is below some tolerance value p_{tol}. It does this while also keeping the banded matrix structure, and so is not as memory intensive as a standard *LU* decomposition. This results in a FEM that could allow for $h = 0$, but still faced the problem of large u values when h was small.

To deal with large u values we restricted our solution of the FEM for u to cells where $\overline{h}_j > h_{tol}$ and modified our approximation to $h^+_{j-1/2}$ and $h^-_{j+1/2}$ in our stiffness matrix. To restrict the domain of our solution we search through the domain at the beginning of each first-order step and identify wet regions where $\overline{h}_j \geq h_{tol}$ and dry regions where $\overline{h}_j < h_{tol}$. In the dry regions we set h, G and u to be zero uniformly over the cell. In the wet regions we solve the FEM with the appropriate boundary conditions and the following modifications to $h^+_{j-1/2}$ and $h^-_{j+1/2}$

$$h^+_{j-1/2} = h^+_{j-1/2} \left(\frac{h^+_{j-1/2} + h_{base}}{h^+_{j-1/2} + h_{tol}} \right), \tag{3.34a}$$

$$h^-_{j+1/2} = h^-_{j+1/2} \left(\frac{h^-_{j+1/2} + h_{base}}{h^-_{j+1/2} + h_{tol}} \right). \tag{3.34b}$$

Where on the right hand side we mean $h^+_{j-1/2}$ and $h^-_{j+1/2}$ as calculated from the reconstruction (3.4). This modification ensures that as $h \to 0$ then $u \to 0$ by increasing the size of h in the stiffness matrix.

Typically we had $p_{tol} = 10^{-20}$, $h_{tol} = 10^{-8}$ and $h_{base} = 10^{-4}$ which allowed for an accurate calculation of u in the presence of dry beds and very small water depths.

Chapter 4

Linear Analysis of the Numerical Methods

An important property of a numerical method is convergence. Convergence guarantees that if we increase the spatial and temporal resolution of a numerical method, then the numerical solution approaches the solution of the partial differential equations they approximate.

For linear partial differential equations the Lax-equivalence theorem states that a numerical method is convergent if and only if it is stable and consistent [23]. A numerical scheme is consistent if the error introduced by the numerical method over a time step approaches zero as the spatial and temporal resolution is increased. While a numerical method is stable if the errors from previous time steps are not amplified in subsequent time steps.

Another important property of a numerical method modelling dispersive wave equations, such as the Serre equations is its dispersion properties. The dispersion relation of a system determines the phase and group velocity of travelling waves in that system. The Serre equations possess a dispersion relation that well approximates the dispersion relation given by linear theory for water waves []. Therefore, how well the dispersion relation of our numerical methods approximate the dispersion relation of the Serre equations is of particular interest.

We analysed the convergence and the dispersion properties of our numerical methods for the linearised Serre equations with horizontal beds. The effect of variations in the bed and nonlinear terms are important when studying the convergence properties of our methods for the full Serre equations. However, these effects greatly increase the complexity of the convergence analysis. We restrict ourselves to the study of the linearised Serre with horizontal beds to offer some insight into the convergence properties of our numerical methods without having to deal with this greater complexity. In general we would expect that a numerical method that has poor convergence properties for the linearised Serre equations with horizontal beds will also have poor convergence properties when the bed and nonlinear terms are included. The dispersion properties are derived from the linearised Serre equations with the no effect from the bed [], therefore this analysis of dispersion properties of the numerical methods is complete.

These linear analyses of convergence and dispersion rely on establishing a relationship of the form

$$\begin{bmatrix} h \\ G \end{bmatrix}_j^{n+1} = \mathbf{E} \begin{bmatrix} h \\ G \end{bmatrix}_j^{n} \tag{4.1}$$

where $\mathbf{E}$ is the evolution matrix relating the conserved quantities h and G at time level t^n with the conserved quantities at time level t^{n+1}. The evolution matrix $\mathbf{E}$ is obtained in the analyses by propagating Fourier modes through the numerical scheme. By analysing the properties of $\mathbf{E}$ we can determine the convergence and dispersion properties of its associated numerical method.

We derive $\mathbf{E}$ in (4.1) for the second-order FEVM and perform the convergence and dispersion analysis. We will then present the results of these analyses for all our numerical methods.

4.1 Linearised Serre equations with horizontal bed

The Serre equations with a horizontal bed (2.3) are linearised by considering waves as small perturbations $\delta\eta$ and $\delta\upsilon$ on a flow with a mean height H and a mean velocity U respectively. So we have

$$h(x,t) = H + \delta\eta(x,t) + \mathcal{O}\left(\delta^2\right), \tag{4.2a}$$

$$u(x,t) = U + \delta\upsilon(x,t) + \mathcal{O}\left(\delta^2\right), \tag{4.2b}$$

where $\delta \ll 1$. These waves are relatively small so terms of order δ^2 are negligible. We substitute (4.2) into the Serre equations and neglect terms of order δ^2 to obtain

$$\frac{\partial\left(\delta\eta\right)}{\partial t} + H\frac{\partial\left(\delta\upsilon\right)}{\partial x} + U\frac{\partial\left(\delta\eta\right)}{\partial x} = 0, \tag{4.3a}$$
$$H\frac{\partial\left(\delta\upsilon\right)}{\partial t} + gH\frac{\partial\left(\delta\eta\right)}{\partial x} + UH\frac{\partial\left(\delta\upsilon\right)}{\partial x} - \frac{H^3}{3}\left(U\frac{\partial^3\left(\delta\upsilon\right)}{\partial x^3} + \frac{\partial^3\left(\delta\upsilon\right)}{\partial x^2\partial t}\right) = 0 \tag{4.3b}$$

and for G

$$G = UH + U\delta\eta + H\delta\upsilon - \frac{H^3}{3}\frac{\partial^2\left(\delta\upsilon\right)}{\partial x^2}. \tag{4.3c}$$

These equations can be reformulated in terms of the conserved quantities η and G

$$\frac{\partial\eta}{\partial t} + \frac{\partial}{\partial x}\left(H\upsilon + U\eta\right) = 0, \tag{4.4a}$$
$$\frac{\partial G}{\partial t} + \frac{\partial}{\partial x}\left(UG + UH\upsilon + gH\eta\right) = 0. \tag{4.4b}$$

where

$$G = UH + U\eta + H\upsilon - \frac{H^3}{3}\frac{\partial^2\left(\upsilon\right)}{\partial x^2}. \tag{4.4c}$$

We have absorbed the δ factor into the corresponding η and υ terms to simplify the notation.

4.2 Evolution Matrix

To derive the evolution matrix we first assume that the solutions of the linearised Serre equations with horizontal beds (4.4) are periodic in space and time. In particular, we assume that η and υ are Fourier modes, which for a general quantity q means

$$q(x,t) = q(0,0)e^{i(\omega t + kx)} \tag{4.5}$$

where k is the wavenumber , ω is the frequency and i is the imaginary number. This is also the assumption made to derive the analytical dispersion relation of the linearised Serre equations []. A consequence of a quantity q being a Fourier mode represented on a uniform temporal and spatial grid is that for any real numbers m and l we have

$$q^{n+m}_{j+l} = q^n_j e^{i(m\omega\Delta t + lk\Delta x)}. \tag{4.6}$$

Because η and υ are Fourier modes then so is G. Furthermore, the cell averages of these quantities $\overline{\eta}$, $\overline{\upsilon}$ and $\overline{G}$ are Fourier modes as well.

4.2.1 Overview of the Evolution Step

We will now present a brief overview of an evolution step of the second-order FEVM. Given the vectors of the cell averages $\overline{\boldsymbol{\eta}}$ and $\overline{\boldsymbol{G}}$ at the current time the second-order FEVM evolution step progresses in the following way:

(i) Reconstruction of Midpoint and Cell Interface Values: We use the second-order accurate operator $\mathcal{M}$ to calculate η and G at the cell midpoint x_j from the cell averages. We also reconstruct η and G at the cell interfaces $x^-_{j+1/2}$ and $x^+_{j+1/2}$ from the cell average values using $\mathcal{R}^-$ and $\mathcal{R}^+$ respectively which are both spatially second-order. So that

$$\begin{aligned} \eta_j &= \mathcal{M}\left(\overline{\boldsymbol{\eta}}\right), & G_j &= \mathcal{M}\left(\overline{\boldsymbol{G}}\right), \\ \eta^-_{j+1/2} &= \mathcal{R}^-\left(\overline{\boldsymbol{\eta}}\right), & G^-_{j+1/2} &= \mathcal{R}^-\left(\overline{\boldsymbol{G}}\right), \\ \eta^+_{j+1/2} &= \mathcal{R}^+\left(\overline{\boldsymbol{\eta}}\right), & G^+_{j+1/2} &= \mathcal{R}^+\left(\overline{\boldsymbol{G}}\right). \end{aligned}$$

(ii) Calculate the Velocity at the Cell Interface: The remaining unknown quantity, $\upsilon_{j+1/2}$ is calculated from a spatially second-order accurate solution of the elliptic equation (4.4c). This calculation is represented by $\mathcal{G}$ as

$$\upsilon_{j+1/2} = \mathcal{G}\left(H, \boldsymbol{G}, \boldsymbol{\eta}\right).$$

(iii) Calculate the Flux Across the Cell Interface: We calculate the average flux $F_{j+1/2}$ across the cell boundary $x_{j+1/2}$ over time using $\mathcal{F}$

$$F_{j+1/2} = \mathcal{F}\left(\eta^-_{j+1/2}, G^-_{j+1/2}, \eta^+_{j+1/2}, G^+_{j+1/2}, \upsilon_{j+1/2}\right).$$

(iv) Time Step Using Forward Euler Approximation: We repeat this process for each cell edge and then apply (**??**) to update the vectors $\overline{\boldsymbol{\eta}}$ and $\overline{\boldsymbol{G}}$ from the current time level to the next time level with first-order accuracy in time.

(v) Complete Second-Order Time Step Using SSP Runge-Kutta Steps: We repeat steps 1-4 and use SSP Runge-Kutta time stepping to calculate $\overline{\boldsymbol{\eta}}$ and $\overline{\boldsymbol{G}}$ at the next time level with second-order accuracy in time.

We will now derive expressions for all the operators in the evolution step of the linear equations. These operators are linear because our assumption that η and υ are Fourier modes. We will then combine these linear operators to derive $\mathbf{E}$ for the second-order FEVM.

(i) Reconstruct the Quantities Inside the Cells

Given $\overline{\boldsymbol{\eta}}$ and $\overline{\boldsymbol{G}}$ at t^n the first step of our numerical method is to calculate η and G at x_j using $\mathcal{M}$ and at $x^-_{j+1/2}$ and $x^+_{j+1/2}$ using $\mathcal{R}^-$ and $\mathcal{R}^+$ respectively. The derivation of these operators is given in terms of a general quantity q, as they are the same for η and G.

Cell average values to nodal values: $\mathcal{M}$

For the second-order FEVM we use the fact that

$$\overline{q}_j = q_j + \mathcal{O}\left(\Delta x^2\right).$$

So to attain second-order accuracy we use

$$q_j = \overline{q}_j = \mathcal{M}\overline{q}_j \tag{4.7}$$

where the factor $\mathcal{M} = 1$ represents the mapping between cell averages and nodal values for the numerical method.

Cell average values to interface values: $\mathcal{R}^-$ and $\mathcal{R}^+$

We reconstruct η and G at $x^-_{j+1/2}$ and $x^+_{j+1/2}$. These quantities can be discontinuous across the cell interfaces in our finite volume method. However, since we are assuming that these quantities are Fourier modes and therefore smooth we do not require non-linear limiters to ensure our scheme is TVD. Therefore, the reconstruction scheme for η and G can be written for a general quantity q as

$$q^-_{j+\frac{1}{2}} = \overline{q}_j + \frac{-\overline{q}_{j-1} + \overline{q}_{j+1}}{4},$$

$$q^+_{j+\frac{1}{2}} = \overline{q}_{j+1} + \frac{-\overline{q}_j + \overline{q}_{j+2}}{4}.$$

Using (4.6) and (4.7) these equations become

$$q^-_{j+\frac{1}{2}} = \overline{q}_j + \frac{-\overline{q}_j e^{-ik\Delta x} + \overline{q}_j e^{ik\Delta x}}{4} = \left(1 + \frac{i\sin\left(k\Delta x\right)}{2}\right)\overline{q}_j = \mathcal{R}^-\overline{q}_j, \tag{4.8a}$$

$$q^+_{j+\frac{1}{2}} = \frac{\overline{q}_j e^{ik\Delta x} + \overline{q}_j + \overline{q}_j e^{2ik\Delta x}}{4} = e^{ik\Delta x}\left(1 - \frac{i\sin\left(k\Delta x\right)}{2}\right)\overline{q}_j = \mathcal{R}^+\overline{q}_j \tag{4.8b}$$

where $\mathcal{R}^\pm$ are the reconstruction factors for both $\eta^\pm_{j+1/2}$ and $G^\pm_{j+1/2}$.

(ii) Calculate the Velocity Over the Domain

To calculate $\upsilon_{j+1/2}$ we use a second-order FEM. We begin our FEM for (4.4c) with its weak formulation, obtained by multiplying (4.4c) by a test function τ and integrating over the domain Ω

$$\int_\Omega G\tau \, dx = UH \int_\Omega \tau \, dx + U \int_\Omega \eta\tau \, dx + H \int_\Omega \upsilon\tau \, dx + \frac{H^3}{3} \int_\Omega \frac{\partial \upsilon}{\partial x}\frac{\partial \tau}{\partial x} \, dx.$$

For G we use the basis functions $\psi^+_{j-1/2}$ and $\psi^-_{j+1/2}$ defined in Chapter [], which means our approximation to G is linear inside a cell with discontinuous jumps at the cell edges. For τ and υ we use the basis functions $\phi_{j-1/2}$, ϕ_j and $\phi_{j+1/2}$ defined in Chapter [], so that τ and our approximation to υ are quadratic functions inside a cell and are continuous across the cell edges. Substituting in the finite element approximations to these quantities and only integrating over a cell as in Chapter [], we get

$$\sum_j \int_{x_{j-1/2}}^{x_{j+1/2}} \left(G^+_{j-1/2}\psi^+_{j-1/2} + G^-_{j+1/2}\psi^-_{j+1/2}\right) \begin{bmatrix}\phi_{j-1/2}\\ \phi_j \\ \phi_{j+1/2}\end{bmatrix} dx =$$

$$\sum_j UH \int_{x_{j-1/2}}^{x_{j+1/2}} \begin{bmatrix}\phi_{j-1/2}\\ \phi_j \\ \phi_{j+1/2}\end{bmatrix} dx + \sum_j U \int_{x_{j-1/2}}^{x_{j+1/2}} \left(\eta^+_{j-1/2}\psi^+_{j-1/2} + \eta^-_{j+1/2}\psi^-_{j+1/2}\right) \begin{bmatrix}\phi_{j-1/2}\\ \phi_j \\ \phi_{j+1/2}\end{bmatrix} dx$$

$$+ \sum_j H \int_{x_{j-1/2}}^{x_{j+1/2}} \left(\upsilon_{j-1/2}\phi_{j-1/2} + \upsilon_j\phi_j + \upsilon_{j+1/2}\phi_{j+1/2}\right) \begin{bmatrix}\phi_{j-1/2}\\ \phi_j \\ \phi_{j+1/2}\end{bmatrix} dx$$

$$+ \sum_j \frac{H^3}{3} \int_{x_{j-1/2}}^{x_{j+1/2}} \left(\upsilon_{j-1/2}\frac{\partial \phi_{j-1/2}}{\partial x} + \upsilon_j\frac{\partial \phi_j}{\partial x} + \upsilon_{j+1/2}\frac{\partial \phi_{j+1/2}}{\partial x}\right) \begin{bmatrix}\frac{\partial \phi_{j-1/2}}{\partial x}\\ \frac{\partial \phi_j}{\partial x} \\ \frac{\partial \phi_{j+1/2}}{\partial x}\end{bmatrix} dx.$$

Calculating all the integrals of the appropriate basis function combinations we get

$$\sum_j \frac{\Delta x}{6}\begin{bmatrix} G^+_{j-1/2} \\ 2G^+_{j-1/2} + 2G^-_{j+1/2} \\ G^-_{j+1/2}\end{bmatrix} = \sum_j UH\frac{\Delta x}{6}\begin{bmatrix}1\\4\\1\end{bmatrix} + \sum_j \frac{\Delta x}{6}U\begin{bmatrix} \eta^+_{j-1/2} \\ 2\eta^+_{j-1/2} + 2\eta^-_{j+1/2} \\ \eta^-_{j+1/2}\end{bmatrix}$$
$$+\sum_j \left(H\frac{\Delta x}{30}\begin{bmatrix} 4 & 2 & -1 \\ 2 & 16 & 2 \\ -1 & 2 & 4\end{bmatrix} + \frac{H^3}{9\Delta x}\begin{bmatrix} 7 & -8 & 1 \\ -8 & 16 & -8 \\ 1 & -8 & 7\end{bmatrix}\right)\begin{bmatrix} \upsilon_{j-1/2} \\ \upsilon_j \\ \upsilon_{j+1/2}\end{bmatrix}.$$

Using (4.6) and (4.8), we obtain

$$\sum_j \frac{\Delta x}{6}\begin{bmatrix} e^{-ik\Delta x}\mathcal{R}^+\overline{G}_j \\ 2e^{-ik\Delta x}\mathcal{R}^+\overline{G}_j + 2\mathcal{R}^-\overline{G}_j \\ \mathcal{R}^-\overline{G}_j\end{bmatrix} = \sum_j UH\frac{\Delta x}{6}\begin{bmatrix}1\\4\\1\end{bmatrix}$$
$$+\sum_j \frac{\Delta x}{6}U\begin{bmatrix} e^{-ik\Delta x}\mathcal{R}^+\overline{\eta}_j \\ 2e^{-ik\Delta x}\mathcal{R}^+\overline{\eta}_j + 2\mathcal{R}^-\overline{\eta}_j \\ \mathcal{R}^-\overline{\eta}_j\end{bmatrix}$$
$$\sum_j \left(H\frac{\Delta x}{30}\begin{bmatrix} 4 & 2 & -1 \\ 2 & 16 & 2 \\ -1 & 2 & 4\end{bmatrix} + \frac{H^3}{9\Delta x}\begin{bmatrix} 7 & -8 & 1 \\ -8 & 16 & -8 \\ 1 & -8 & 7\end{bmatrix}\right)\begin{bmatrix} e^{-ik\frac{\Delta x}{2}}\upsilon_j \\ \upsilon_j \\ e^{ik\frac{\Delta x}{2}}\upsilon_j\end{bmatrix}.$$

After simplifying

$$\sum_j \frac{\Delta x}{6}\begin{bmatrix} e^{-ik\Delta x}\mathcal{R}^+ \\ 2e^{-ik\Delta x}\mathcal{R}^+ + 2\mathcal{R}^- \\ \mathcal{R}^-\end{bmatrix}\overline{G}_j = \sum_j UH\frac{\Delta x}{6}\begin{bmatrix}1\\4\\1\end{bmatrix}$$
$$+\sum_j \frac{\Delta x}{6}\begin{bmatrix} e^{-ik\Delta x}\mathcal{R}^+ \\ 2e^{-ik\Delta x}\mathcal{R}^+ + 2\mathcal{R}^- \\ \mathcal{R}^-\end{bmatrix}\overline{\eta}_j + \sum_j \left(H\frac{\Delta x}{30}\begin{bmatrix} 4e^{-ik\frac{\Delta x}{2}} + 2 - e^{ik\frac{\Delta x}{2}} \\ 2e^{-ik\frac{\Delta x}{2}} + 16 + 2e^{ik\frac{\Delta x}{2}} \\ -e^{-ik\frac{\Delta x}{2}} + 2 + 4e^{ik\frac{\Delta x}{2}}\end{bmatrix}\right.$$
$$\left. + \frac{H^3}{9\Delta x}\begin{bmatrix} 7e^{-ik\frac{\Delta x}{2}} - 8 + e^{ik\frac{\Delta x}{2}} \\ -8e^{-ik\frac{\Delta x}{2}} + 16 - 8e^{ik\frac{\Delta x}{2}} \\ e^{-ik\frac{\Delta x}{2}} - 8 + 7e^{ik\frac{\Delta x}{2}}\end{bmatrix}\right)\upsilon_j.$$

These vectors represent the contribution from the j^{th} cell to the three equations relating the quantities at $x_{j-1/2}$, x_j and $x_{j+1/2}$ respectively. Since the intercell flux only requires the velocity at the cell edges we only need to solve the first and third equations. These equations are equivalent up to a translation and therefore we will only consider the third equation.

So far we have only given the contribution to the equation at $x_{j+1/2}$ from the j^{th} cell, but there is also a contribution from the adjacent $(j+1)^{th}$ cell because $\phi_{j+1/2}$ is non-zero over both cells. Accounting for this we get

$$
\begin{aligned}
\frac{\Delta x}{6}\left(\mathcal{R}^- + \mathcal{R}^+\right)\overline{G}_j = {} & \frac{\Delta x}{6}2UH + \frac{\Delta x}{6}U\left(\mathcal{R}^- + \mathcal{R}^+\right)\overline{\eta}_j \\
& \Bigg(H\frac{\Delta x}{30}\left(-e^{-ik\frac{\Delta x}{2}} + 2 + 4e^{ik\frac{\Delta x}{2}} + e^{ik\Delta x}\left(4e^{-ik\frac{\Delta x}{2}} + 2 - e^{ik\frac{\Delta x}{2}}\right)\right) \\
& + \frac{H^3}{9\Delta x}\left(e^{-ik\frac{\Delta x}{2}} - 8 + 7e^{ik\frac{\Delta x}{2}} + e^{ik\Delta x}\left(7e^{-ik\frac{\Delta x}{2}} - 8 + e^{ik\frac{\Delta x}{2}}\right)\right)\Bigg)\upsilon_j \\
= {} & \frac{\Delta x}{3}UH + \frac{\Delta x}{6}U\left(\mathcal{R}^- + \mathcal{R}^+\right)\overline{\eta}_j \\
& + \left[H\frac{\Delta x}{30}\left(4\cos\left(\frac{k\Delta x}{2}\right) - 2\cos\left(k\Delta x\right) + 8\right)\right. \\
& \left. + \frac{H^3}{9\Delta x}\left(-16\cos\left(\frac{k\Delta x}{2}\right) + 2\cos\left(k\Delta x\right) + 14\right)\right]e^{ik\frac{\Delta x}{2}}\upsilon_j .
\end{aligned}
$$

From (4.6) $\upsilon_{j+1/2} = e^{ik\frac{\Delta x}{2}}\upsilon_j$ then we have

$$
\begin{aligned}
\upsilon_{j+1/2} = {} & \left[\left(\frac{\Delta x}{6}\left(\mathcal{R}^- + \mathcal{R}^+\right)\right)\overline{G}_j - U\left(\frac{\Delta x}{6}\left(\mathcal{R}^- + \mathcal{R}^+\right)\right)\overline{\eta}_j - \frac{\Delta x}{3}UH\right] \\
& \div \left[H\frac{\Delta x}{30}\left(4\cos\left(\frac{k\Delta x}{2}\right) - 2\cos\left(k\Delta x\right) + 8\right)\right. \\
& \left. + \frac{H^3}{9\Delta x}\left(-16\cos\left(\frac{k\Delta x}{2}\right) + 2\cos\left(k\Delta x\right) + 14\right)\right] \\
= {} & \mathcal{G}^G\overline{G}_j + \mathcal{G}^\eta\overline{\eta}_j + \mathcal{G}^c .
\end{aligned} \tag{4.9}
$$

(iii) Calculate All the Fluxes Across the Cell Interfaces

The average intercell flux $F_{j+1/2}$ is approximated using Kurganov's method [16]. For the linearised Serre equations we have the wave speed bounds (2.8a), so that

$$
a^-_{j+1/2} = \min\left\{0, U - \sqrt{gH}\right\} \quad \text{and} \quad a^+_{j+1/2} = \max\left\{0, U + \sqrt{gH}\right\} . \tag{4.10}
$$

This method has three different approximations to $F_{j+1/2}$ depending on the Froude number $Fr = \dfrac{U}{\sqrt{gH}}$; (i) supercritical flow to the left where $Fr < -1$, (ii) critical and subcritical flow in both directions where $-1 \leq Fr \leq 1$ and (iii) supercritical flow to the right where $Fr > 1$. We will derive the flux operators for each of these cases separately.

Left Supercritical Flow $Fr < -1$:

For left supercritical flow; $Fr < -1$ and therefore $U + \sqrt{gH} < 0$ so we have from (4.10) that $a^-_{j+1/2} = U - \sqrt{gH}$ and $a^+_{j+1/2} = 0$. For these values the Kurganov flux approximation for a general quantity q [] reduces to

$$F_{j+\frac{1}{2}} = f\left(q^+_{j+\frac{1}{2}}\right). \tag{4.11}$$

Substituting the flux function from the continuity equation (4.4a) into the Kurganov flux approximation we obtain

$$F^{\eta}_{j+\frac{1}{2}} = H\upsilon_{j+1/2} + U\eta^+_{j+1/2}$$

since υ is continuous and therefore $\upsilon_{j+1/2} = \upsilon^+_{j+1/2} = \upsilon^-_{j+1/2}$. Using the FEM for $\upsilon_{j+1/2}$ (4.9) and the reconstruction (4.8) we have

$$\begin{aligned}
F^{\eta}_{j+\frac{1}{2}} &= H\left(\mathcal{G}^G\overline{G}_j + \mathcal{G}^{\eta}\overline{\eta}_j + \mathcal{G}^c\right) + U\eta^+_{j+1/2} \\
&= \left(H\mathcal{G}^{\eta} + U\mathcal{R}^+\right)\overline{\eta}_j + H\mathcal{G}^G\overline{G}_j + H\mathcal{G}^c \\
&= \mathcal{F}^{\eta,\eta}_{j+\frac{1}{2}}\overline{\eta}_j + \mathcal{F}^{\eta,G}_{j+\frac{1}{2}}\overline{G}_j + \mathcal{F}^{\eta,c}_{j+\frac{1}{2}}
\end{aligned} \tag{4.12}$$

Substituting the flux function for irrotationality equation (4.4b) into the Kurganov flux approximation (4.11) we obtain

$$F^{G}_{j+\frac{1}{2}} = UG^+_{j+1/2} + UH\upsilon_{j+1/2} + gH\eta^+_{j+1/2}$$

Using the FEM (4.9) to calculate $\upsilon_{j+1/2}$ and our interface reconstruction (4.8) we have

$$\begin{aligned}
F^{G}_{j+\frac{1}{2}} &= UG^+_{j+1/2} + UH\left(\mathcal{G}^G\overline{G}_j + \mathcal{G}^{\eta}\overline{\eta}_j + \mathcal{G}^c\right) + gH\eta^+_{j+1/2} \\
&= \left(UH\mathcal{G}^{\eta} + gH\mathcal{R}^+\right)\overline{\eta}_j + \left(U\mathcal{R}^+ + UH\mathcal{G}^G\right)\overline{G}_j + UH\mathcal{G}^c \\
&= \mathcal{F}^{G,\eta}_{j+\frac{1}{2}}\overline{\eta}_j + \mathcal{F}^{G,G}_{j+\frac{1}{2}}\overline{G}_j + \mathcal{F}^{G,c}_{j+\frac{1}{2}}
\end{aligned} \tag{4.13}$$

Subcritical flow $-1 \leq Fr \leq 1$:

When the flow is subcritical we have $-1 \leq Fr \leq 1$, which means that $a^-_{j+1/2} = U - \sqrt{gH}$ and $a^+_{j+1/2} = U + \sqrt{gH}$. Therefore Kurganov flux approximation for a general quantity q [] is

$$F_{j+\frac{1}{2}} = \frac{U}{2\sqrt{gH}} \left[f\left(q^-_{j+\frac{1}{2}}\right) - f\left(q^+_{j+\frac{1}{2}}\right) \right] + \frac{1}{2} \left[f\left(q^-_{j+\frac{1}{2}}\right) + f\left(q^+_{j+\frac{1}{2}}\right) \right] + \frac{U^2 - gH}{2\sqrt{gH}} \left[q^+_{j+\frac{1}{2}} - q^-_{j+\frac{1}{2}} \right]. \tag{4.14}$$

Substituting in the flux function from the continuity equation (4.4a) we get

$$\begin{aligned} F^\eta_{j+\frac{1}{2}} = &\frac{U}{2\sqrt{gH}} \left[H\upsilon_{j+1/2} + U\eta^-_{j+\frac{1}{2}} - H\upsilon_{j+1/2} - U\eta^+_{j+\frac{1}{2}} \right] \\ &+ \frac{1}{2} \left[H\upsilon_{j+1/2} + U\eta^-_{j+\frac{1}{2}} + H\upsilon_{j+1/2} + U\eta^+_{j+\frac{1}{2}} \right] \\ &+ \frac{U^2 - gH}{2\sqrt{gH}} \left[\eta^+_{j+\frac{1}{2}} - \eta^-_{j+\frac{1}{2}} \right]. \end{aligned} \tag{4.15}$$

Using the reconstruction factors (4.8) and the elliptic solver (4.9) we get

$$\begin{aligned} F^\eta_{j+\frac{1}{2}} &= \left(H\mathcal{G}^\eta + \frac{U}{2} \left[\mathcal{R}^- + \mathcal{R}^+ \right] - \frac{\sqrt{gH}}{2} \left[\mathcal{R}^+ - \mathcal{R}^- \right] \right) \overline{\eta}_j + H\mathcal{G}^G \overline{G}_j + H\mathcal{G}^c \\ &= \mathcal{F}^{\eta,\eta}_{j+\frac{1}{2}} \overline{\eta}_j + \mathcal{F}^{\eta,G}_{j+\frac{1}{2}} \overline{G}_j + \mathcal{F}^{\eta,c}_{j+\frac{1}{2}}. \end{aligned} \tag{4.16}$$

For the flux function of the irrotationality equation (4.4b) the flux approximation (4.14) becomes

$$\begin{aligned} F^G_{j+\frac{1}{2}} = &\frac{U}{2\sqrt{gH}} \left[UG^-_{j+\frac{1}{2}} + UH\upsilon_{j+1/2} + gH\eta^-_{j+\frac{1}{2}} - UG^+_{j+\frac{1}{2}} - UH\upsilon_{j+1/2} - gH\eta^+_{j+\frac{1}{2}} \right] \\ &+ \frac{1}{2} \left[UG^-_{j+\frac{1}{2}} + UH\upsilon_{j+1/2} + gH\eta^-_{j+\frac{1}{2}} + UG^+_{j+\frac{1}{2}} + UH\upsilon_{j+1/2} + gH\eta^+_{j+\frac{1}{2}} \right] \\ &+ \frac{U^2 - gH}{2\sqrt{gH}} \left[G^+_{j+\frac{1}{2}} - G^-_{j+\frac{1}{2}} \right]. \end{aligned} \tag{4.17}$$

By using the reconstruction factors (4.8) and the elliptic solver (4.9) we get

$$\begin{aligned} F^G_{j+\frac{1}{2}} &= \left(\frac{U\sqrt{gH}}{2} \left[\mathcal{R}^- - \mathcal{R}^+ \right] + UH\mathcal{G}^\eta + \frac{gH}{2} \left[\mathcal{R}^- + \mathcal{R}^+ \right] \right) \overline{\eta}_j \\ &\quad + \left(UH\mathcal{G}^G + + \frac{U}{2} \left[\mathcal{R}^- + \mathcal{R}^+ \right] - \frac{\sqrt{gH}}{2} \left[\mathcal{R}^+ - \mathcal{R}^- \right] \right) \overline{G}_j + UH\mathcal{G}^c \\ &= \mathcal{F}^{G,\eta}_{j+\frac{1}{2}} \overline{\eta}_j + \mathcal{F}^{G,G}_{j+\frac{1}{2}} \overline{G}_j + \mathcal{F}^{G,c}_{j+\frac{1}{2}}. \end{aligned} \tag{4.18}$$

Right supercritical flow $Fr > 1$:

When the flow is flowing to the right and supercritical we have $Fr > 1$, which means that $a^-_{j+1/2} = 0$ and $a^+_{j+1/2} = U + \sqrt{gH}$. This is very similar to the left supercritical case, except instead of using the $\mathcal{R}^+$ we have $\mathcal{R}^-$ in our flux approximation for a general quantity which reduces to

$$F_{j+\frac{1}{2}} = f\left(q^-_{j+\frac{1}{2}}\right). \tag{4.19}$$

Substituting in the flux function for the continuity equation (4.4a) and the irrotationality equation (4.4b) we obtain

$$\begin{aligned} F^{\eta}_{j+\frac{1}{2}} &= \left(H\mathcal{G}^{\eta} + U\mathcal{R}^-\right)\overline{\eta}_j + H\mathcal{G}^{G}\overline{G}_j + H\mathcal{G}^{c} \\ &= \mathcal{F}^{\eta,\eta}_{j+\frac{1}{2}}\overline{\eta}_j + \mathcal{F}^{\eta,G}_{j+\frac{1}{2}}\overline{G}_j + \mathcal{F}^{\eta,c}_{j+\frac{1}{2}}, \end{aligned} \tag{4.20}$$

$$\begin{aligned} F^{G}_{j+\frac{1}{2}} &= \left(UH\mathcal{G}^{\eta} + gH\mathcal{R}^-\right)\overline{\eta}_j + \left(U\mathcal{R}^- + UH\mathcal{G}^{G}\right)\overline{G}_j + UH\mathcal{G}^{c} \\ &= \mathcal{F}^{G,\eta}_{j+\frac{1}{2}}\overline{\eta}_j + \mathcal{F}^{G,G}_{j+\frac{1}{2}}\overline{G}_j + \mathcal{F}^{G,c}_{j+\frac{1}{2}}. \end{aligned} \tag{4.21}$$

(iv) Calculate All the Source Terms for the Cells

[]

(v) Update All the Cell Averages Using a Forward Euler Approximation

We have obtained the operators for the flux functions for all three cases, supercrticial flow in the left or right direction and subcritical flow. By substituting the appropriate flux approximation for the physical situation into our update scheme (**??**) our second-order FEVM can be written as

$$\begin{aligned} \overline{\eta}^{n+1}_j &= \overline{\eta}^{n}_j - \frac{\Delta t}{\Delta x}\left[\left(\mathcal{F}^{\eta,\eta}_{j+\frac{1}{2}}\overline{\eta}_j + \mathcal{F}^{\eta,G}_{j+\frac{1}{2}}\overline{G}_j + \mathcal{F}^{\eta,c}_{j+\frac{1}{2}}\right) - \left(\mathcal{F}^{\eta,\eta}_{j-\frac{1}{2}}\overline{\eta}_j + \mathcal{F}^{\eta,G}_{j-\frac{1}{2}}\overline{G}_j + \mathcal{F}^{\eta,c}_{j-\frac{1}{2}}\right)\right], \\ \overline{G}^{n+1}_j &= \overline{G}^{n}_j - \frac{\Delta t}{\Delta x}\left[\left(\mathcal{F}^{G,\eta}_{j+\frac{1}{2}}\overline{\eta}_j + \mathcal{F}^{G,G}_{j+\frac{1}{2}}\overline{G}_j + \mathcal{F}^{G,c}_{j+\frac{1}{2}}\right) - \left(\mathcal{F}^{G,\eta}_{j-\frac{1}{2}}\overline{\eta}_j + \mathcal{F}^{G,G}_{j-\frac{1}{2}}\overline{G}_j + \mathcal{F}^{G,c}_{j-\frac{1}{2}}\right)\right]. \end{aligned}$$

Since $\mathcal{F}_{j-\frac{1}{2}} = e^{-ik\Delta x}\mathcal{F}_{j+\frac{1}{2}}$ for all superscripts we have

$$\begin{aligned} \overline{\eta}^{n+1}_j &= \overline{\eta}^{n}_j - \frac{\Delta t}{\Delta x}\left[\left(1 - e^{-ik\Delta x}\right)\left(\mathcal{F}^{\eta,\eta}_{j+\frac{1}{2}}\overline{\eta}_j + \mathcal{F}^{\eta,G}_{j+\frac{1}{2}}\overline{G}_j\right)\right], \\ \overline{G}^{n+1}_j &= \overline{G}^{n}_j - \frac{\Delta t}{\Delta x}\left[\left(1 - e^{-ik\Delta x}\right)\left(\mathcal{F}^{G,\eta}_{j+\frac{1}{2}}\overline{\eta}_j + \mathcal{F}^{G,G}_{j+\frac{1}{2}}\overline{G}_j\right)\right]. \end{aligned}$$

This can be written in matrix form as

$$\begin{aligned}
\begin{bmatrix} \overline{\eta} \\ \overline{G} \end{bmatrix}_j^{n+1} &= \begin{bmatrix} \overline{\eta} \\ \overline{G} \end{bmatrix}_j^{n} - \left(1 - e^{-ik\Delta x}\right) \frac{\Delta t}{\Delta x} \begin{bmatrix} \mathcal{F}^{\eta,\eta} & \mathcal{F}^{\eta,G} \\ \mathcal{F}^{G,\eta} & \mathcal{F}^{G,G} \end{bmatrix} \begin{bmatrix} \overline{\eta} \\ \overline{G} \end{bmatrix}_j^{n} \\
&= (\mathbf{I} - \Delta t \mathbf{F}) \begin{bmatrix} \overline{\eta} \\ \overline{G} \end{bmatrix}_j^{n} \qquad (4.22)
\end{aligned}$$

for a single Euler step which results in a method that is first-order in time and second-order in space. To increase the order of accuracy in time we use SSP Runge-Kutta time stepping which makes use of a convex combination of multiple Euler steps.

(vi) Update All the Cell Averages Using a Second-Order SSP Runge-Kutta Method

The second-order SSP Runge Kutta time stepping uses two forward Euler steps to accomplish a temporally second-order accurate method in the following way

$$\begin{bmatrix} \overline{\eta} \\ \overline{G} \end{bmatrix}_j^{1} = (\mathbf{I} - \Delta t \mathbf{F}) \begin{bmatrix} \overline{\eta} \\ \overline{G} \end{bmatrix}_j^{n}, \qquad (4.23a)$$

$$\begin{bmatrix} \overline{\eta} \\ \overline{G} \end{bmatrix}_j^{2} = (\mathbf{I} - \Delta t \mathbf{F}) \begin{bmatrix} \overline{\eta} \\ \overline{G} \end{bmatrix}_j^{1}, \qquad (4.23b)$$

$$\begin{bmatrix} \overline{\eta} \\ \overline{G} \end{bmatrix}_j^{n+1} = \frac{1}{2}\left(\begin{bmatrix} \overline{\eta} \\ \overline{G} \end{bmatrix}_j^{n} + \begin{bmatrix} \overline{\eta} \\ \overline{G} \end{bmatrix}_j^{2} \right). \qquad (4.23c)$$

Substituting (4.23a) and (4.23b) into (4.23c) we can write this in terms of the flux matrix $\mathbf{F}$ and our cell averages at t^n as

$$\begin{bmatrix} \overline{\eta} \\ \overline{G} \end{bmatrix}_j^{n+1} = \frac{1}{2}\left(\begin{bmatrix} \overline{\eta} \\ \overline{G} \end{bmatrix}_j^{n} + (\mathbf{I} - \Delta t \mathbf{F})^2 \begin{bmatrix} \overline{\eta} \\ \overline{G} \end{bmatrix}_j^{n} \right).$$

Expanding $(\mathbf{I} - \Delta t \mathbf{F})^2$ we get

$$\begin{aligned}
\begin{bmatrix} \overline{\eta} \\ \overline{G} \end{bmatrix}_j^{n+1} &= \left(\mathbf{I} - \Delta t \mathbf{F} + \frac{1}{2}\Delta t^2 \mathbf{F}^2\right) \begin{bmatrix} \overline{\eta} \\ \overline{G} \end{bmatrix}_j^{n} \\
&= \mathbf{E} \begin{bmatrix} \overline{\eta} \\ \overline{G} \end{bmatrix}_j^{n} \qquad (4.24)
\end{aligned}$$

which is in the desired form (4.1).

This is the evolution matrix $\mathbf{E}$ for the second-order FEVM. The matrix $\mathbf{E}$ is dependent on the flux matrix $\mathbf{F}$ and therefore will depend on the Froude number. The Froude number is constant and so we can analyse these flow scenarios individually.

Both the convergence and dispersion analysis then proceed by investigating the properties of the evolution matrix $\mathbf{E}$. We begin with the convergence analysis.

4.3 Convergence Analysis

The linearised Serre equations are linear partial differential equations and therefore we can apply the Lax-equivalence theorem to demonstrate the convergence of our numerical methods by establishing their consistency and stability. We use a Von Neumann stability analysis to demonstrate stability. Consistency is demonstrated for Fourier mode solutions (4.5), which are eigenfunctions of the linearised Serre equations. Together this stability and consistency condition imply convergence of the numerical method under the L_2 norm.

4.3.1 Stability

For a numerical method to be stable we must ensure that errors from previous time steps are not amplified at the current time step. To accomplish this we must ensure

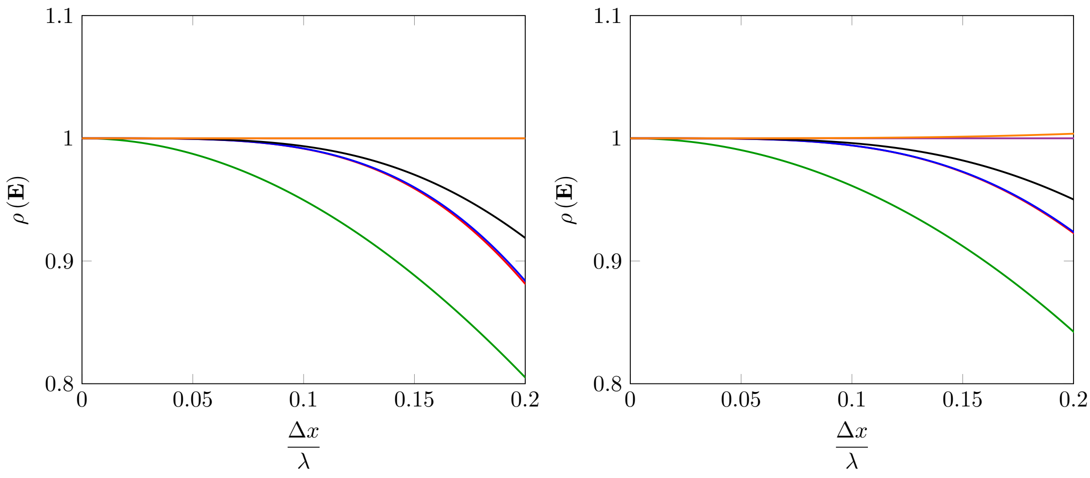
$$\rho\left(\mathbf{E}\right) \le 1 \tag{4.25}$$

where $\rho\left(\mathbf{E}\right)$ is the spectral radius of $\mathbf{E}$. Since $\mathbf{E}$ was derived for our methods by using Fourier modes, this condition implies Von Neumann stability.

We calculated $\rho\left(\mathbf{E}\right)$ numerically for various values of Δx, Δt, k, H and U to check if (4.25) holds. We summarised our results in Figure 4.1 which is a plot of $\rho\left(\mathbf{E}\right)$ against $\Delta x / \lambda$ for representative values of k, H and U. We we used g $= 9.81 m/s^2$ and chose $\Delta t = 0.5 / \left(U + \sqrt{gH}\right) \Delta x$ to satisfy the CFL condition []. This is the common choice of Δt in our numerical experiments.

The values $H = 1m$, $k = \frac{\pi}{10} m^{-1}$ and $U = 0m/s$ and $1m/s$ shown in Figure 4.1 where chosen because they represent the general behaviour of these plots. For these k and H values our shallowness parameter $\sigma = \frac{1}{20}$ and so the Serre equations are applicable [].

In Figure 4.1 it can be seen that all methods have $\rho\left(\mathbf{E}\right) \le 1$ for $U = 0m/s$ and are therefore stable. The two finite difference methods overlap and have



(a) $U = 0$ (b) $U = 1$

Figure 4.1: Spectral radius of $\mathbf{E}$ for first-order FDVM (–), second-order FDVM(–), second-order FEVM (–), third-order FDVM (–), $\mathcal{D}$ (–) and $\mathcal{W}$ (–). With $H = 1m$ and $k = \frac{\pi}{10}$.

$\rho(\mathbf{E}) = 1$ for all Δx values, while the second-order FDVM and the second-order FEVM also overlap. However, when $U \neq 0m/s$ then $\mathcal{W}$ has $\rho(\mathbf{E}) > 1$ for all Δx and is therefore unstable.

The analytic value of $\rho(\mathbf{E})$ is given by using (4.6) to write

$$\begin{bmatrix} \overline{\eta} \\ \overline{G} \end{bmatrix}_j^{n+1} = e^{i\omega\Delta t} \begin{bmatrix} \overline{\eta} \\ \overline{G} \end{bmatrix}_j^{n}.$$

Therefore, the analytic growth factor is

$$\rho(\mathbf{E}) = \left|e^{i\omega\Delta t}\right| = \sqrt{\cos^2(\omega\Delta t) + \sin^2(\omega\Delta t)} = 1 \tag{4.26}$$

since $\omega \in \mathbb{R}$. Therefore numerical methods with $\rho(\mathbf{E})$ closer to 1 are closer to the analytic value. In this sense the two finite difference methods are best, although $\mathcal{W}$ is unstable. While for the FDVM we can see that the higher-order methods better approximate the analytic value, as expected. We can see in Figure 4.1 that $\lim_{\Delta x \to 0} \rho(\mathbf{E}) = 1$ for all methods, as expected.

We observed the same results for a wide range of k, H and U, in particular all methods except $\mathcal{W}$ were stable for any value of these variables. While $\mathcal{W}$ was only stable when $U = 0m/s$.

4.3.2 Consistency

For a numerical method to be consistent the error introduced by the method for a single time step must approach zero as the spatial and temporal resolution is increased. We will demonstrate this only for Fourier mode solutions of the linearised Serre equations. Therefore, we can demonstrate consistency by investigating the evolution matrix $\mathbf{E}$. The error introduced for a single time step from t^n to t^{n+1}, $\mathcal{T}^n$ is

$$\mathcal{T}^n = \mathbf{E}\begin{bmatrix}\overline{\eta}\\\overline{G}\end{bmatrix}_j^n - \begin{bmatrix}\overline{\eta}\\\overline{G}\end{bmatrix}_j^{n+1}. \tag{4.27}$$

To ensure consistency we must have that $\lim_{\Delta x,\Delta t\to 0}\|\mathcal{T}^n\| = 0$ for all n. Taking the norm of both sides of (4.27) we get

$$\|\mathcal{T}^n\| = \left\|\mathbf{E}\begin{bmatrix}\overline{\eta}\\\overline{G}\end{bmatrix}_j^n - \begin{bmatrix}\overline{\eta}\\\overline{G}\end{bmatrix}_j^{n+1}\right\|.$$

Making use of (4.6) we obtain

$$\|\mathcal{T}^n\| = \left\|\mathbf{E}\begin{bmatrix}\overline{\eta}\\\overline{G}\end{bmatrix}_j^n - e^{i\omega\Delta t}\begin{bmatrix}\overline{\eta}\\\overline{G}\end{bmatrix}_j^{n}\right\|.$$

Using the matrix norm induced by the vector norm we have that

$$\|\mathcal{T}^n\| \le \left\|\mathbf{E} - e^{i\omega\Delta t}\mathbf{I}\right\|\left\|\begin{bmatrix}\overline{\eta}\\\overline{G}\end{bmatrix}_j^n\right\|. \tag{4.28}$$

Since $\overline{\eta}_j^n$ and $\overline{G}_j^n$ are finite and independent of Δx and Δt, if $\lim_{\Delta x,\Delta t\to 0}\left\|\mathbf{E} - e^{i\omega\Delta t}\mathbf{I}\right\| = 0$ then $\lim_{\Delta x,\Delta t\to 0}\|\mathcal{T}^n\| = 0$ as desired.

We calculated the Taylor series of $\mathbf{E} - e^{i\omega_+\Delta t}\mathbf{I}$ for all the numerical methods for all flow scenarios. We have reported the lowest order terms of the Taylor series in Tables 4.1 and 4.2 for the first-order FDVM, Table 4.3 for the second-order FDVM, Tables 4.4 and 4.5 for the third-order FDVM, Table 4.6 for the second-order FEVM, Table 4.7 for $\mathcal{D}$ and Table 4.8 for $\mathcal{W}$. To be concise, we only reported the temporal and spatial errors for the supercritical flow scenarios that where different from those when $-\sqrt{gH} \le U \le \sqrt{gH}$, this only occurred for the spatial errors of the odd-order FDVM.

We observe for all of our methods that the Taylor series of all the elements of $\mathbf{E} - e^{i\omega_+\Delta t}\mathbf{I}$ have a factor of Δt. So we have that for all methods

Element	Lowest Order Term of Error	
	Δx	Δt
$E_{00} - e^{i\omega_+ \Delta t}$	$-\frac{1}{2}\sqrt{gH}k^2\Delta t\Delta x$	$\dfrac{\sqrt{3gH\beta}+3U}{\beta}ik\Delta t$
E_{01}	$\dfrac{3+\beta}{4\beta^2}ik^3\Delta t\Delta x^2$	$-\frac{3}{\beta}ik\Delta t$
E_{10}	$-\frac{1}{2}\sqrt{gH}k^2\Delta t\Delta x$	$\left(-gH+\dfrac{3U^2}{\beta}\right)ik\Delta t$
$E_{11} - e^{i\omega_+ \Delta t}$	$-\frac{1}{2}\sqrt{gH}k^2\Delta t\Delta x$	$\dfrac{\sqrt{3gH\beta}-3U}{\beta}ik\Delta t$

Table 4.1: Table of the lowest order term of the Taylor series for the elements of $\mathbf{E}-e^{i\omega_+\Delta t}\mathbf{I}$ for the first-order FDVM with $-\sqrt{gH} \leq U \leq \sqrt{gH}$ and $\beta = 3+k^2H^2$.

$$\begin{aligned}
\left\|\mathbf{E} - e^{i\omega_+\Delta t}\mathbf{I}\right\| &= \left\|\Delta t\left(\mathbf{A}_0 + \begin{bmatrix}\mathcal{O}(\Delta t) & \mathcal{O}(\Delta t)\\ \mathcal{O}(\Delta t) & \mathcal{O}(\Delta t)\end{bmatrix}\right)\right\| \\
&= |\Delta t|\left\|\mathbf{A}_0 + \begin{bmatrix}\mathcal{O}(\Delta t) & \mathcal{O}(\Delta t)\\ \mathcal{O}(\Delta t) & \mathcal{O}(\Delta t)\end{bmatrix}\right\| \\
&\leq |\Delta t|\left(\|\mathbf{A}_0\| + \left\|\begin{bmatrix}\mathcal{O}(\Delta t) & \mathcal{O}(\Delta t)\\ \mathcal{O}(\Delta t) & \mathcal{O}(\Delta t)\end{bmatrix}\right\|\right).
\end{aligned}$$

Choosing a particular vector norm such as the L_1 or L_∞ and its induced matrix norm we can see from Tables 4.1-4.8 that $\mathbf{A}$ is independent of Δt and finite so that as $\Delta t \to 0$ we have $|\Delta t|\left(\|\mathbf{A}_0\| + \left\|\begin{bmatrix}\mathcal{O}(\Delta t) & \mathcal{O}(\Delta t)\\ \mathcal{O}(\Delta t) & \mathcal{O}(\Delta t)\end{bmatrix}\right\|\right) \to 0$ and therefore $\left\|\mathbf{E} - e^{i\omega_+\Delta t}\mathbf{I}\right\| \to 0$. Therefore, for all our numerical methods we have $\lim_{\Delta x,\Delta t\to 0}\|\mathcal{T}^n\| = 0$ and so all our numerical methods are consistent for Fourier mode solutions as desired.

Scheme	Lowest Order Δx Term of Error	
	$U < -\sqrt{gH}$	$\sqrt{gH} < U$
$E_{00} - e^{i\omega_+ \Delta t}$	$\frac{1}{2}k^2 U \Delta t \Delta x$	$-\frac{1}{2}k^2 U \Delta t \Delta x$
E_{01}	$\frac{1}{2}gHk^2 \Delta t \Delta x$	$\frac{1}{2}gHk^2 \Delta t \Delta x$
$E_{11} - e^{i\omega_+ \Delta t}$	$\frac{1}{2}k^2 U \Delta t \Delta x$	$-\frac{1}{2}k^2 U \Delta t \Delta x$

Table 4.2: Table of the lowest order term of the Taylor series for the elements of $\mathbf{E} - e^{i\omega_+ \Delta t}\mathbf{I}$ for the first-order FDVM which are different than those in Table 4.1 with $\beta = 3 + k^2H^2$.

Element	Lowest Order Term of Error	
	Δx	Δt
$E_{00} - e^{i\omega_+ \Delta t}$	$-\dfrac{i\left(27 + 9H^2k^2 + H^4k^4\right)}{12\beta^2} U k^3 \Delta x^2$	$\dfrac{\sqrt{3gH\beta} + 3U}{\beta} ik\Delta t$
E_{01}	$\dfrac{3+\beta}{4\beta^2} ik^3 \Delta t \Delta x^2$	$-\frac{3}{\beta} ik \Delta t$
E_{10}	$-\left(gH + \dfrac{3U^2}{\beta} + \dfrac{9U^2}{\beta^2}\right)\dfrac{k^3}{12}\Delta t \Delta x^2$	$\left(-gH + \dfrac{3U^2}{\beta}\right) ik\Delta t$
$E_{11} - e^{i\omega_+ \Delta t}$	$\dfrac{-9 + H^2k^2\beta}{\beta^2}\dfrac{k^3}{12} iU \Delta t \Delta x^2$	$\dfrac{\sqrt{3gH\beta} - 3U}{\beta} ik\Delta t$

Table 4.3: Table of the lowest order term of the Taylor series for the elements of $\mathbf{E} - e^{i\omega_+ \Delta t}\mathbf{I}$ for the second-order FDVM with $-\sqrt{gH} \leq U \leq \sqrt{gH}$ and $\beta = 3 + k^2H^2$.

Element	Lowest Order Term of Error	
	Δx	Δt
$E_{00} - e^{i\omega_+ \Delta t}$	$-\frac{1}{12}\sqrt{gH}k^4\Delta t\Delta x^3$	$\dfrac{\sqrt{3gH\beta}+3U}{\beta}ik\Delta t$
E_{01}	$\dfrac{\sqrt{gH}}{4\beta}ik^5\Delta t^2\Delta x^3$	$-\frac{3}{\beta}ik\Delta t$
E_{10}	$-\frac{1}{12}\sqrt{gH}k^4\Delta t\Delta x^3$	$\left(-gH+\dfrac{3U^2}{\beta}\right)ik\Delta t$
$E_{11} - e^{i\omega_+ \Delta t}$	$-\frac{1}{12}\sqrt{gH}k^4\Delta t\Delta x^3$	$\dfrac{\sqrt{3gH\beta}-3U}{\beta}ik\Delta t$

Table 4.4: Table of the lowest order term of the Taylor series for the elements of $\mathbf{E}-e^{i\omega_+\Delta t}\mathbf{I}$ for the third-order FDVM with $-\sqrt{gH} \leq U \leq \sqrt{gH}$ and $\beta = 3+k^2H^2$.

Scheme	Lowest Order Δx Term of Error	
	$U < -\sqrt{gH}$	$\sqrt{gH} < U$
$E_{00} - e^{i\omega_+ \Delta t}$	$\frac{1}{12}k^4U\Delta t\Delta x^3$	$-\frac{1}{12}k^4U\Delta t\Delta x^3$
E_{01}	$\dfrac{1}{4\beta}iUk^5\Delta t^2\Delta x^3$	$-\dfrac{1}{4\beta}iUk^5\Delta t^2\Delta x^3$
E_{10}	$\dfrac{1}{12}gHk^4\Delta t^2\Delta x^3$	$-\dfrac{1}{12}gHk^4\Delta t^2\Delta x^3$
$E_{11} - e^{i\omega_+ \Delta t}$	$\frac{1}{12}k^4U\Delta t\Delta x^3$	$-\frac{1}{12}k^4U\Delta t\Delta x^3$

Table 4.5: Table of the lowest order term of the Taylor series for the elements of $\mathbf{E}-e^{i\omega_+\Delta t}\mathbf{I}$ for the third-order FDVM which are different than those in Table 4.1 with $\beta = 3 + k^2H^2$.

Element	Lowest Order Term of Error	
	Δx	Δt
$E_{00} - e^{i\omega_+ \Delta t}$	$-\frac{i\left(54 + 45H^2k^2 + 10H^4k^4\right)}{120\beta^2}Uk^3\Delta t\Delta x^2$	$\frac{\sqrt{3gH\beta} + 3U}{\beta}ik\Delta t$
E_{01}	$\frac{\beta - 3}{\beta^2}\frac{ik^3}{40}\Delta t\Delta x^2$	$-\frac{3}{\beta}ik\Delta t$
E_{10}	$-\left(gH - \frac{15U^2}{\beta} + \frac{9U^2}{\beta}\right)\frac{k^3}{120}\Delta t\Delta x^2$	$\left(-gH + \frac{3U^2}{\beta}\right)ik\Delta t$
$E_{11} - e^{i\omega_+ \Delta t}$	$\frac{126 + 75H^2k^2 + 10H^4k^4}{\beta^2}\frac{k^3}{120}iU\Delta t\Delta x^2$	$\frac{\sqrt{3gH\beta} - 3U}{\beta}ik\Delta t$

Table 4.6: Table of the lowest order term of the Taylor series for the elements of $\mathbf{E} - e^{i\omega_+ \Delta t}\mathbf{I}$ for the second-order FEVM with $-\sqrt{gH} \leq U \leq \sqrt{gH}$ and $\beta = 3 + k^2H^2$.

Element	Lowest Order Term of Error	
	Δx	Δt
$E_{00} - e^{i\omega_+ \Delta t}$	$\frac{ik^3}{3}U\Delta t\Delta x^2$	$\sqrt{\frac{3gH}{\beta}}2ik\Delta t$
E_{01}	$\frac{iHk^3}{3}\Delta t\Delta x^2$	$-2Hik\Delta t$
E_{10}	$\frac{ig\left(3 + \beta\right)}{2\beta^2}k^3\Delta t\Delta x^2$	$-\frac{6igk}{\beta}\Delta t$
$E_{11} - e^{i\omega_+ \Delta t}$	$\frac{ik^3}{3}U\Delta t\Delta x^2$	$\sqrt{\frac{3gH}{\beta}}2ik\Delta t$

Table 4.7: Table of the lowest order term of the Taylor series for the elements of $\mathbf{E} - e^{i\omega \Delta t}\mathbf{I}$ for $\mathcal{D}$ with $\beta = 3 + k^2H^2$.

Element	Lowest Order Term of Error	
	Δx	Δt
$E_{00} - e^{i\omega_+ \Delta t}$	$\frac{ik^3}{6} U \Delta t \Delta x^2$	$\sqrt{\frac{3gH}{\beta}} ik\Delta t$
E_{01}	$\frac{iHk^3}{6} \Delta t \Delta x^2$	$-Hik\Delta t$
E_{10}	$\frac{ig\,(3+\beta)}{2\beta^2} k^3 \Delta t \Delta x^2$	$-\frac{6igk}{\beta}\Delta t$
$E_{11} - e^{i\omega_+ \Delta t}$	$\frac{ik^3}{3} U \Delta t \Delta x^2$	$\sqrt{\frac{3gH}{\beta}} 2ik\Delta t$

Table 4.8: Table of the lowest order term of the Taylor series for the elements of $\mathbf{E} - e^{i\omega_+ \Delta t}\mathbf{I}$ for $\mathcal{W}$ with $\beta = 3 + k^2H^2$.

4.4 Dispersion Analysis

To study the dispersion of our numerical methods we must calculate ω for our numerical methods. Making use of (4.6) in (4.24) we get

$$e^{i\omega\Delta t} \begin{bmatrix} \overline{\eta} \\ \overline{G} \end{bmatrix}_j^n = \mathbf{E} \begin{bmatrix} \overline{\eta} \\ \overline{G} \end{bmatrix}_j^n . \tag{4.29}$$

Assuming that $\mathbf{E}$ has an eigenvalue decomposition $\mathbf{E} = \mathbf{P}^{-1}\mathbf{\Lambda}\mathbf{P}$ and substituting it into (4.29) we get

$$e^{i\omega\Delta t} \begin{bmatrix} \overline{\eta} \\ \overline{G} \end{bmatrix}_j^n = \mathbf{P}^{-1}\mathbf{\Lambda}\mathbf{P} \begin{bmatrix} \overline{\eta} \\ \overline{G} \end{bmatrix}_j^n . \tag{4.30}$$

Left multiplying (4.30) by $\mathbf{P}$ we obtain

$$e^{i\omega\Delta t}\mathbf{P} \begin{bmatrix} \overline{\eta} \\ \overline{G} \end{bmatrix}_j^n = \mathbf{\Lambda}\mathbf{P} \begin{bmatrix} \overline{\eta} \\ \overline{G} \end{bmatrix}_j^n . \tag{4.31}$$

Since Λ is a diagonal matrix we must have that $e^{i\omega_+\Delta t} = \lambda_+$ and $e^{i\omega_-\Delta t} = \lambda_-$ where $\lambda_\pm$ are the eigenvalues of $\mathbf{E}$ and $\omega_\pm$ are the positive and negative branches of the dispersion relation. Therefore the dispersion relation of a numerical method is

$$\widetilde{\omega}_\pm = \frac{1}{i\Delta t} \log\left[\lambda_\pm\right] . \tag{4.32}$$

By comparing $\widetilde{\omega}_\pm$ with the analytic $\omega_\pm$ given by the linearised Serre equations we can determine the error in the dispersion relation for the numerical method. The real part of $\widetilde{\omega}_\pm$ determines the speed of a wave, while the imaginary part determines the change in amplitude. For $\omega_\pm$ the imaginary part is zero and so the amplitude of waves of the linearised Serre equations are constant in time. We only present the results for the positive branch of the dispersion relation as the results for the negative and positive branches are very similar.

The relative error in the dispersion relation was plotted against $\Delta x/\lambda$ for representative values of H, U and k. We used $g = 9.81m/s^2$ and chose $\Delta t = 0.5/\left(U + \sqrt{gH}\right)\Delta x$ to satisfy the CFL condition [].

In Figures 4.2 and 4.3 we present the plots for $kH = \pi/10$ so that the water is shallow as $\sigma = 1/20$ so the Serre equations are appropriate. We present the real and imaginary errors separately as they account for different physical phenomenon and also present the total relative error as a measure of the overall difference of behaviour between waves in the numerical method and the waves of the linearised Serre equations.

From Figures 4.2 and 4.3 we can see that all methods approximate the dispersion relation of the Serre equations well with the approximation improving as $\Delta x \to 0$, as expected.

For the real part of the dispersion error the FEVM and the FDVM outperform the two finite difference methods and therefore will better approximate the speed of waves of the linearised Serre equations. However, for the dilation of waves the roles are reversed with the two finite difference methods either dilating the waves very little ($\mathcal{W}$ for $U > 0$) or not at all. When taking both effects into account with the complete error we see that the first-order FDVM has the largest dispersion error followed by $\mathcal{W}$, $\mathcal{D}$, the second-order FEVM, the second-order FDVM and finally the third-order FDVM has the lowest dispersion error. So that the size of the total dispersion error is mainly determined by the order of accuracy of the numerical scheme. These results justify choosing these FDVM over these finite difference methods for the Serre equations.

Figures 4.2 and 4.3 furthermore demonstrate that the second-order FDVM is superior to the FEVM not just for the complete dispersion error, but its real and imaginary parts individually as well. Therefore the second-order FDVM will do a better job in accurately modelling the speed and amplitude of waves than the second-order FEVM. Interestingly, for the two finite difference methods and the first-order FDVM there seems to be some trade-off between predicting the speed or amplitude of the waves very well.

We observed similar results across a wide array of k, H and U values. However, as kH is increased the distinction between the second-order FDVM and the second-order FEVM becomes less pronounced. This can be seen in Figure 4.4 where $kH = 2.5$ and $\sigma = 5/4\pi > 1/20$ so that the water is no longer shallow.

These kH values are the same as those by Filippini et al. [24], and our results are similar for the real part of the dispersion error. Our FDVM and the FEVM compare favourably with the methods described and analysed by Filippini et al. [24]. Furthermore, we extended their work by allowing for non-zero values of U and examining the imaginary and complete error in the dispersion relation.

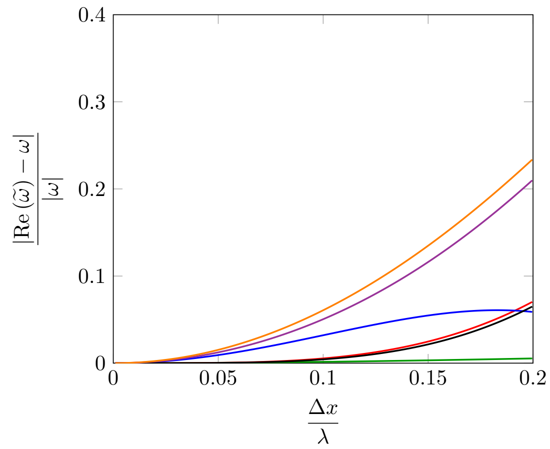
Figure 4.5 demonstrates that the results of the real part of the dispersion error is slightly different if we allow for non-zero values of U. In particular the non-zero value of U changes the real part of the dispersion error for the first-order FDVM, most significantly when $kH = 2.5$. Therefore, for some methods allowing for non-zero values of U can have a significant impact on the conclusions drawn from the dispersion analysis. Furthermore taking the imaginary part of the dispersion error into account is important as ω determines not only the speed of waves but also their amplitude. In particular it is possible that a method like the first-order FDVM performs very well for the real part of the dispersion error and poorly for the imaginary part, leading to false conclusions about the accuracy of the method.

The Taylor series expansion of $\widetilde{\omega}$ was also derived for all the numerical methods. We have compiled the lowest order terms of the Taylor series for $\widetilde{\omega}_+ - \omega_+$ in Table 4.9 when $-\sqrt{gH} \le U \le \sqrt{gH}$ for the FDVM and FEVM. In Table 4.9 it is clear that these schemes estimated ω with the expected order of accuracy in both space and time.

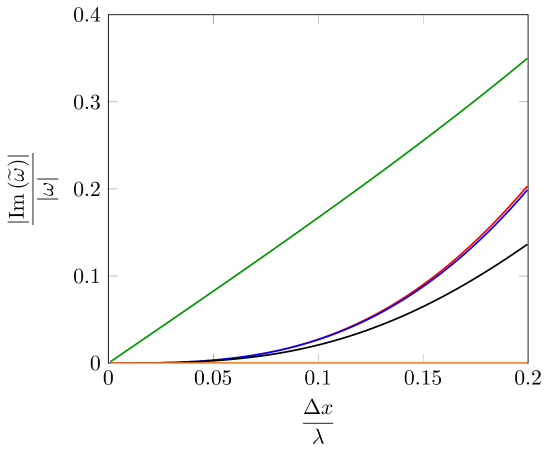
We also present the lowest order terms of the Taylor series for $\widetilde{\omega}_+ - \omega_+$ for both $U < -\sqrt{gH}$ and $U > \sqrt{gH}$ in Table 4.10. We only present the errors that are different from those reported in Table 4.9, this was only the case for the spatial error of the odd-order numerical methods. We can see that for all the flow scenarios that our FDVM and the FEVM have the correct order of accuracy when approximating ω_+.

Finally we present the lowest order terms of the Taylor series for $\widetilde{\omega}_+ - \omega_+$ for the finite difference methods in Table 4.11. These methods do not change depending on the value of the physical quantities. The two finite difference methods both have the correct order of accuracy in both space and time.

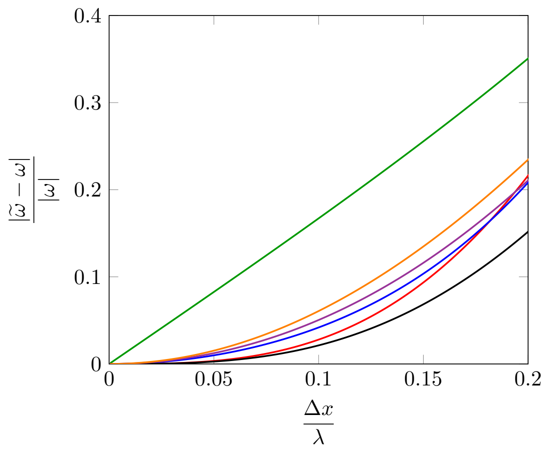
Because all methods were demonstrated to have the expected order of accuracy in approximating ω_+ this implies that for small Δx values the order of accuracy



(a) Real Part

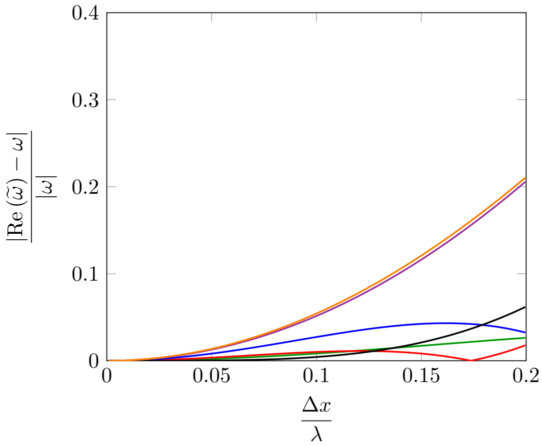


(b) Imaginary Part

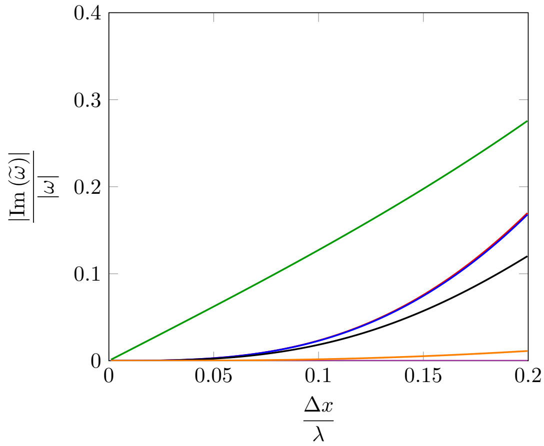


(c) Real and Imaginary Parts

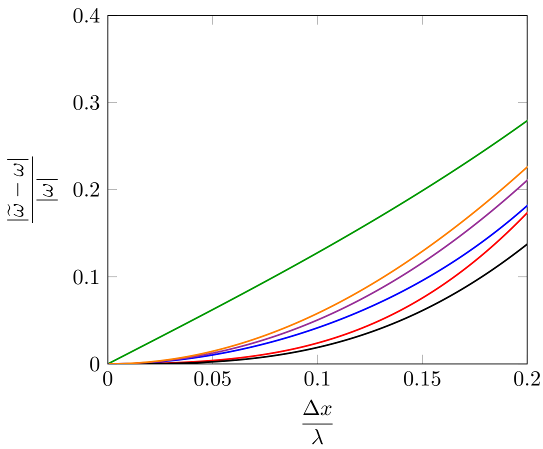
Figure 4.2: Relative dispersion error for first-order FDVM (—), second-order FDVM(—), second-order FEVM (—), third-order FDVM (—), $\mathcal{D}$ (—) and $\mathcal{W}$ (—). With $H = 1m$, $k = \frac{\pi}{10}$ and $U = 0m/s$.



(a) Real part

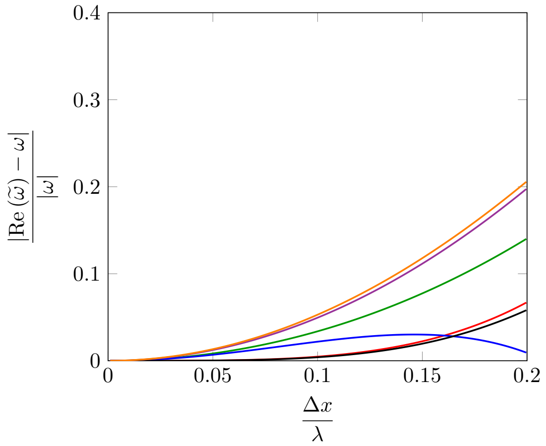


(b) Imaginary part

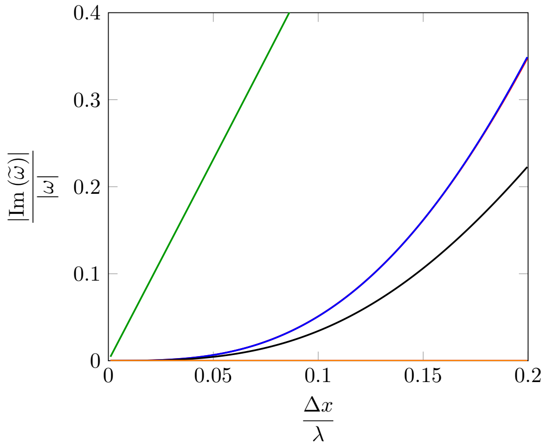


(c) Real and Imaginary Parts

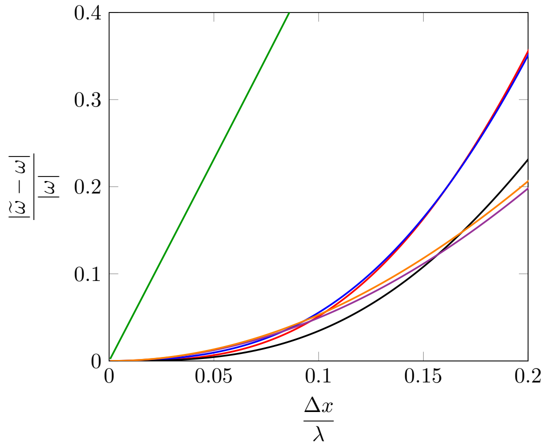
Figure 4.3: Relative dispersion error for first-order FDVM (–), second-order FDVM(–), second-order FEVM (–), third-order FDVM (–), $\mathcal{D}$ (–) and $\mathcal{W}$ (–). With $H = 1m$, $k = \frac{\pi}{10}$ and $U = 1m/s$.



(a) Real part

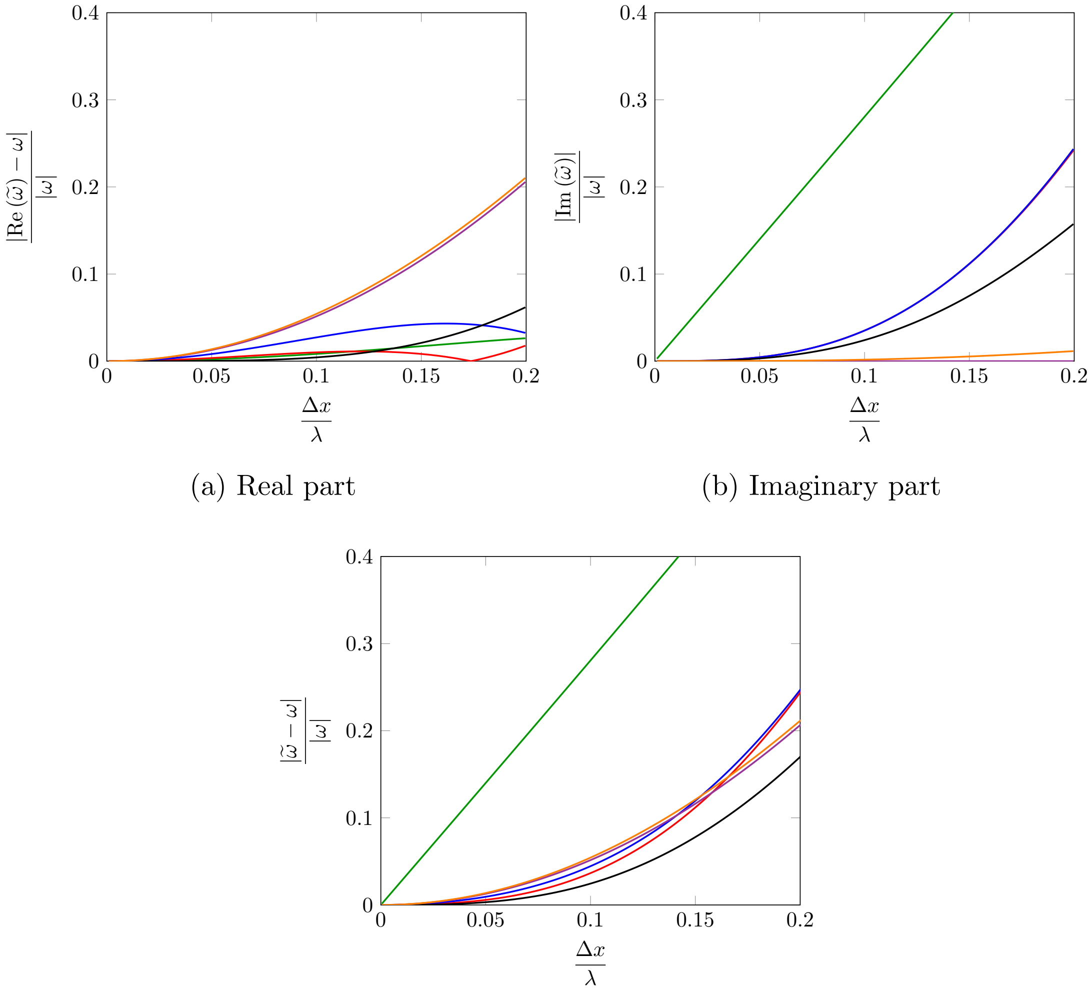


(b) Imaginary part



(c) Real and Imaginary Parts

Figure 4.4: Relative dispersion error for first-order FDVM (—), second-order FDVM(—), second-order FEVM (—), third-order FDVM (—), $\mathcal{D}$ (—) and $\mathcal{W}$ (—). With $H = 1m$, $k = 2.5$ and $U = 0m/s$.



(a) Real part

(b) Imaginary part

(c) Real and Imaginary Parts

Figure 4.5: Relative dispersion error for first-order FDVM (—), second-order FDVM(—), second-order FEVM (—), third-order FDVM (—), $\mathcal{D}$ (—) and $\mathcal{W}$ (—). With $H = 1m$, $k = 2.5$ and $U = 1m/s$.

Scheme	Lowest Order Term of Error	
	Δx	Δt
FDVM$_1$	$-\left(2\sqrt{gH}-\sqrt{\dfrac{3U}{\beta}}\right)\dfrac{ik^2}{4}\Delta x$	$\dfrac{i\omega_+^2}{2}\Delta t$
FDVM$_2$	$\dfrac{2\beta U-3\sqrt{3gH\beta}}{\beta^2}\dfrac{k^3}{24}\Delta x^2$	$-\dfrac{\omega_+^3}{6}\Delta t^2$
FEVM$_2$	$\left(U+\dfrac{\left(42+15k^2H^2\right)\sqrt{3gH\beta}}{20\beta^2}\right)\dfrac{k^3}{12}\Delta x^2$	$-\dfrac{\omega_+^3}{6}\Delta t^2$
FDVM$_3$	$-\left(2\sqrt{gH}-\sqrt{3\beta}U\right)\dfrac{ik^4}{24}\Delta x^3$	$-\dfrac{i\omega_+^4}{24}\Delta t^3$

Table 4.9: Table showing lowest order error term for approximating ω_+ for all FDVM and the FEVM. With $-\sqrt{gH} \leq U \leq \sqrt{gH}$ and $\beta = 3 + H^2k^2$.

will be the primary driver of the dispersion error.

Scheme	Lowest Order Δx Term of Error	
	$U < -\sqrt{gH}$	$\sqrt{gH} < U$
FDVM$_1$	$-\left(2U + \sqrt{\frac{3gH}{\beta}}\right)\frac{ik^2}{4}\Delta x$	$\left(2U + \sqrt{\frac{3gH}{\beta}}\right)\frac{ik^2}{4}\Delta x$
FDVM$_3$	$-\left(2U + \sqrt{\frac{3gH}{\beta}}\right)\frac{ik^4}{24}\Delta x^3$	$\left(2U + \sqrt{\frac{3gH}{\beta}}\right)\frac{ik^4}{24}\Delta x^3$

Table 4.10: Table showing different lowest order spatial error term for approximating ω_+ for all FDVM and the FEVM for different values of U. With $\beta = 3 + H^2k^2$.

Scheme	Lowest Order Term of Error	
	Δx	Δt
$\mathcal{D}$	$-\left(U + \frac{(4 + H^2k^2)\sqrt{3gH\beta}}{4\beta^2}\right)\frac{k^3}{3}\Delta x^2$	$-\frac{\omega_+^3}{3}\Delta t^2$
$\mathcal{W}$	$\left(U + \frac{(4 + H^2k^2)\sqrt{3gH\beta}}{4\beta^2}\right)\frac{k^3}{3}\Delta x^2$	$\Big(\beta U^2\left[9\sqrt{3gH\beta} + 4\beta U\right] + 3gH^2\left[\sqrt{3gH\beta} + 6\beta U\right]\Big)\frac{k^3}{18\beta^2}\Delta t^2$

Table 4.11: Table showing lowest order error term for approximating ω_+ for $\mathcal{D}$ and $\mathcal{W}$.

Chapter 5

Numerical Validation

5.1 Measuring Convergence and Conservation

The numerical methods are assessed in this chapter by investigating their convergence and conservation properties. The convergence of these numerical methods is studied using analytic solutions to the governing equations and forced solutions to the forced Serre equations. While conservation is investigated by comparing the total of a conserved quantity in a numerical solution and comparing with the total of that quantity present in the initial conditions. We introduce notation for these measures and describe their calculation here, beginning with convergence.

5.1.1 Measures of Convergence

By measuring the relative difference between the numerical and analytic solutions as Δx varies, the convergence of the numerical methods can be investigated. To measure the relative difference we use the L_1 vector norm; to compare the numerical and analytic solutions at the numerical grid locations x_j at the end of the simulations. For a quantity q, the vector of its values $\boldsymbol{q}$ at the grid locations x_j and the corresponding numerical solution at those locations $\boldsymbol{q}^*$; the L_1 norm is

$$L_1(\boldsymbol{q}, \boldsymbol{q}^*) = \begin{cases} \dfrac{||\boldsymbol{q}^* - \boldsymbol{q}||_1}{||\boldsymbol{q}||_1} & ||\boldsymbol{q}||_1 > 0 \\ \\ ||\boldsymbol{q}^*||_1 & ||\boldsymbol{q}||_1 = 0 \end{cases} . \tag{5.1}$$

When no analytic solution is present, we can compare the distance between numerical solutions to gain some insight into how a sequence of numerical solutions are behaving. This allows us to demonstrate that a sequence of numerical solutions is convergent to some solution. To do this the L_1 vector norm is again used as in (5.1) except now both vectors are numerical solutions. Since both numerical solutions will have different grid locations, we only take the difference between the two at the common grid points. We have constructed our grids to accommodate for this, varying Δx by successively dividing by 2. This ensures that the grid locations generated by the larger Δx value are all in the grid generated by the smaller Δx value, and so we can compare both numerical solutions at the grid points generated by the larger Δx value.

5.1.2 Measures of Conservation

The conservation properties of the methods are established by calculating the total of a conserved quantity in the numerical solution $\mathcal{C}^*\left(\boldsymbol{q}^*\right)$ at the end of the simulation and comparing it to the total of that quantity for the initial conditions $\mathcal{C}\left(q(x,0)\right)$, derived analytically. We do this again using the relative measure;

$$C_1(q,\boldsymbol{q}^*) = \begin{cases} \dfrac{|\mathcal{C}^*\left(\boldsymbol{q}^*\right) - \mathcal{C}\left(q(x,0)\right)|}{|\mathcal{C}\left(q(x,0)\right)|} & |\mathcal{C}\left(q(x,0)\right)| > 0 \\ \\ |\mathcal{C}^*\left(\boldsymbol{q}^*\right)| & |\mathcal{C}\left(q(x,0)\right)| = 0 \end{cases} . \tag{5.2}$$

$\mathcal{C}^*\left(\boldsymbol{q}^*\right)$ was calculated using 3 point Gaussian quadrature over the j^{th} cell and summing these cell integrals for all j. The three points needed to perform the Gaussian quadrature were calculated by interpolating the j^{th} cell using a quartic that fits the nodal values q_{j-2}, q_{j-1}, q_j, q_{j+1} and q_{j+2}. The Gaussian quadrature using three points is 5^{th} order accurate and interpolation by quartics is 5^{th} order accurate for the quantity q and 4^{th} order accurate for its spatial derivative $\partial q/\partial x$. Since all methods are third-order accurate or less, the error introduced by the calculation of $\mathcal{C}^*\left(\boldsymbol{q}^*\right)$ for the mass, momentum, G and $\mathcal{H}$ will be dominated by the error introduced by the numerical solvers.

In some cases $\mathcal{C}\left(q(x,0)\right)$ may be difficult to derive analytically. In this case we compare $\mathcal{C}^*\left(\boldsymbol{q}^*\right)$ with $\mathcal{C}^*\left(\boldsymbol{q}^0\right)$; where $\boldsymbol{q}^0$ is the vector of the quantity at the grid locations used as the initial conditions of our numerical method. Comparing these we get

$$C^*_1(\boldsymbol{q}^0,\boldsymbol{q}^*) = \begin{cases} \dfrac{|\mathcal{C}^*\left(\boldsymbol{q}^*\right) - \mathcal{C}^*\left(\boldsymbol{q}^0\right)|}{|\mathcal{C}^*\left(\boldsymbol{q}^0\right)|} & |\mathcal{C}^*\left(\boldsymbol{q}^0\right)| > 0 \\ \\ |\mathcal{C}^*\left(\boldsymbol{q}^*\right)| & |\mathcal{C}^*\left(\boldsymbol{q}^0\right)| = 0 \end{cases} . \tag{5.3}$$

5.2 Analytic Solution for Horizontal Bed

To assess the ability of our numerical methods to solve the Serre equations with a horizontal bed (2.5) we use the solitary travelling wave solution (2.9) described in Chapter 2. This is a particular member of the family of periodic travelling wave solutions [12], but all these solutions except the trivial one provide a similar test for the numerical methods and so it is sufficient to only use the the solitary travelling wave solution.

For the solitary wave analytic solution all the terms in (2.5) must be adequately approximated by the numerical method to properly reproduce the analytic solution. Therefore this analytic solution serves as a very good benchmark for the ability of the numerical methods to accurately solve the Serre equations with a horizontal bed for smooth solutions.

For our numerical tests we used the solitary travelling wave solution we used (2.9) with $a_0 = 1m$, $a_0 = 0.7m$ and $g = 9.81m/s^2$ at $t = 0s$ as initial conditions in our numerical methods. The spatial domain was $[-250m, 250m]$ and the problem was solved until $t = 50s$. This was done for a range of Δx values that had the following form; $\Delta x = 100/2^k m$ with $k \in [6, 7, \dots, 19]$. We satisfied the CFL condition with a CFL number of $Cr = 0.5$ by setting $\Delta t = Cr/\sqrt{g\left(a_0 + a_1\right)}$. For FDVM_2 and FEVM_2 we used $\theta = 1.2$ as the limiting parameter in the generalised minmod limiter (3.5).

The parameters $a_0 = 1m$ and $a_0 = 0.7m$ so that the non-linearity parameter $\epsilon = a_1/a_0 = 0.7$ was large but beneath the breaking threshold for water waves []. Because ϵ is large the nonlinear effects are large and therefore so are the dispersive effects making this analytic solution a rigorous test of the numerical methods. For this spatial domain and a final time $t = 50s$ there is no interaction of the wave and the boundary, therefore Dirichlet boundary conditions are appropriate.

5.2.1 Results for Solitary Travelling Wave Solution

An example numerical solution with $\Delta x = 100/2^{11}m$ from all methods was plotted in Figure 5.1 against the analytic solution at $t = 50s$. We have only plotted an illustrative amount of the points in the numerical solution. From these plots it is clear that FDVM_1 performs significantly worse than the higher-order methods at reproducing the analytic solution, even for relatively fine grids where the wave is captured by more than 200 cells. This is primarily due to the numerical diffusion introduced by the method, which has caused the wave in the numerical solution to decrease in amplitude and widen significantly. The higher-order numerical methods all accurately replicate the analytic solution, with insignificant differences in these plots due to the high resolution of the grid.

The L_1 norm was calculated for h, u and G for all numerical solutions and was plotted against Δx for all numerical methods in Figure 5.2. From these plots it is clear that all numerical methods are convergent. The rate at which the numerical solutions converge to the analytic solution over Δx is determined by the order of accuracy of the numerical scheme. All methods demonstrate the expected order of accuracy from the order of accuracy of the approximations used and their order of accuracy determined by linear analysis in Chapter 4.

All methods more accurately reproduced the analytic solution for h than either G or u across all Δx values. This is due to the simplicity of the continuity equation (2.4a) compared to the irrotationality equation (2.4b) and the error in u being dominated by the error in G.

Increasing the order of accuracy of our numerical methods leads to smaller errors when comparing two methods for the same Δx value, as Figure 5.2 clearly demonstrates. This is consistent with the example numerical solution in Figure 5.1, where the lowest order accuracy scheme, FDVM_1 had the poorest reproduction of the analytic solution. However, even though the third-order accurate FDVM_3 is an improvement over its second-order counterparts, this improvement is less pronounced than the improvement between first and second-order methods.

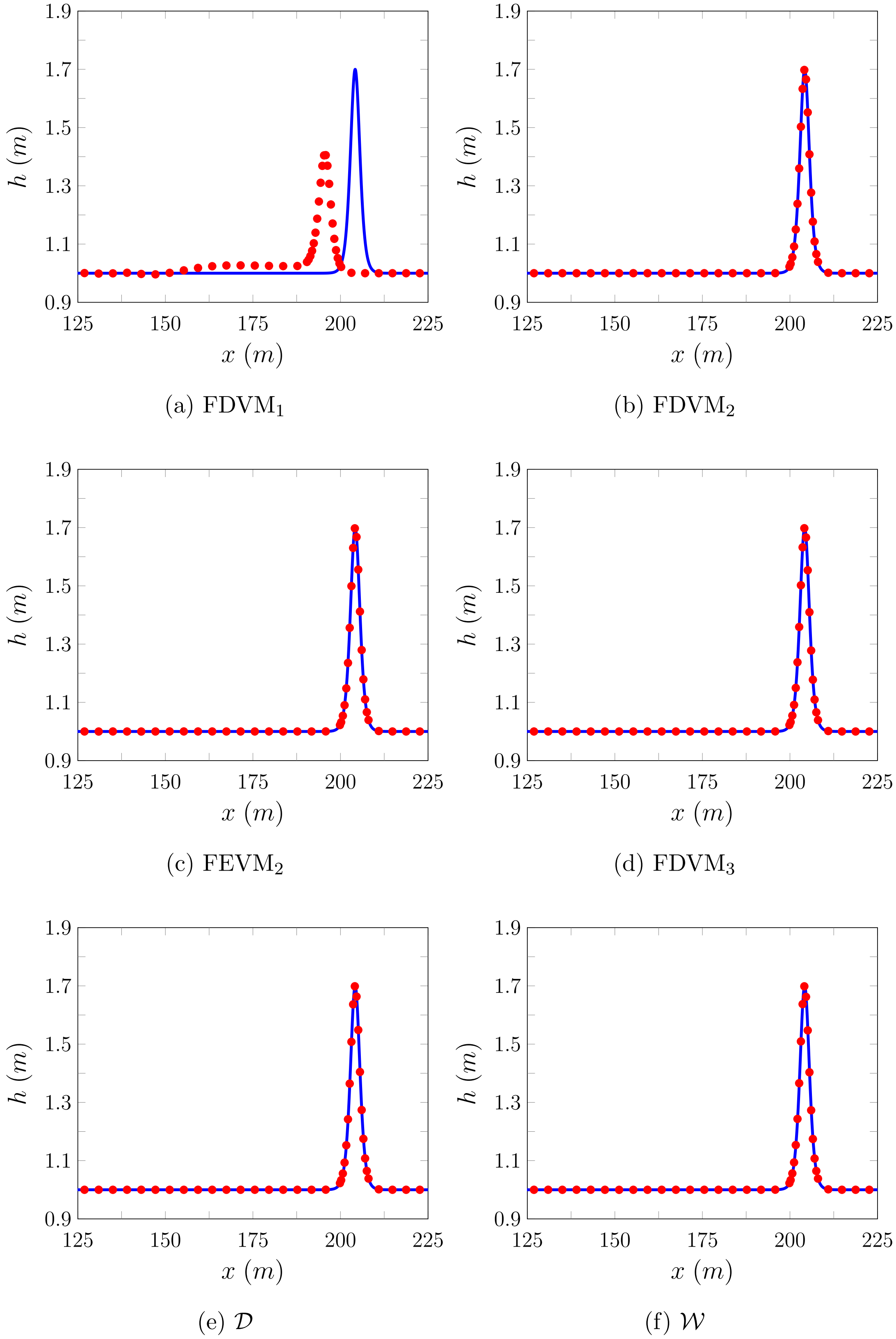
For the second-order methods we find that FDVM_2 consistently produces the smallest L_1 error followed by FEVM_2, $\mathcal{W}$ and $\mathcal{D}$. The difference between the FDVM_2 and FEVM_2 is significant with errors of FEVM_2 being 2 to 4 times larger than FDVM_2. Both finite difference methods produce very similar errors which are about twice as large as the errors from FEVM_2.

5.3 Analytic Solution for Variable Bathymetry

5.3.1 Results for Lake at Rest

5.4 Forced Solution For Finite Water Depth

5.5 Forced Solution with Dry Beds

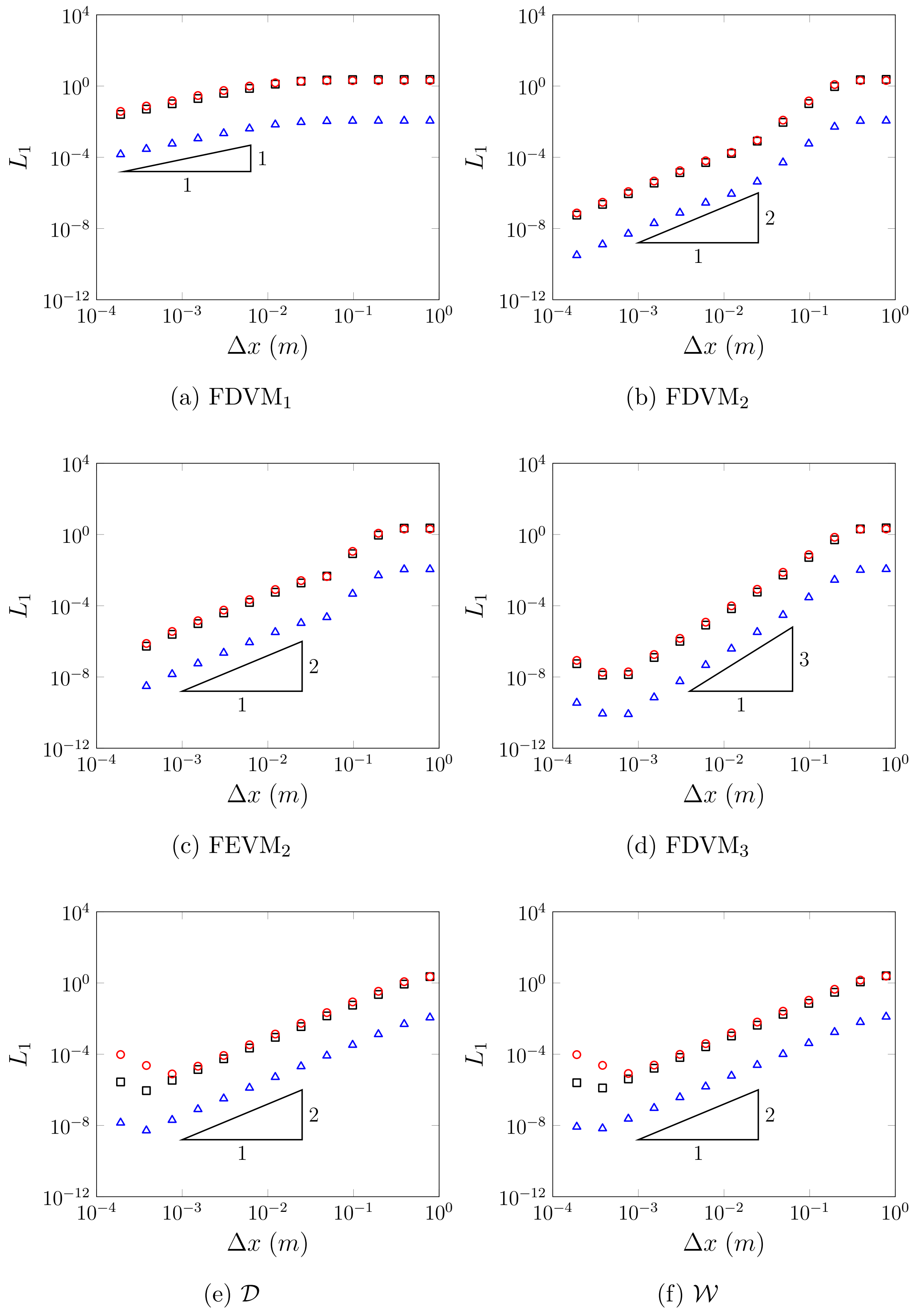


(a) FDVM_1 (b) FDVM_2

(c) FEVM_2 (d) FDVM_3

(e) $\mathcal{D}$ (f) $\mathcal{W}$

Figure 5.1: Comparison of the analytic solution (−) and numerical solution with $\Delta x = 100/2^{11}m$ (•) for the soliton problem at $t = 50s$ for all methods.



(a) FDVM_1

(b) FDVM_2

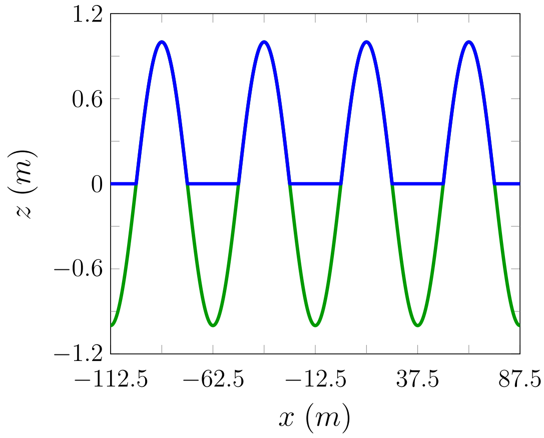
(c) FEVM_2

(d) FDVM_3

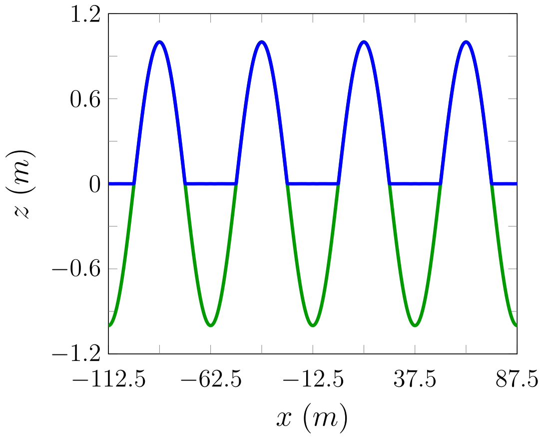
(e) $\mathcal{D}$

(f) $\mathcal{W}$

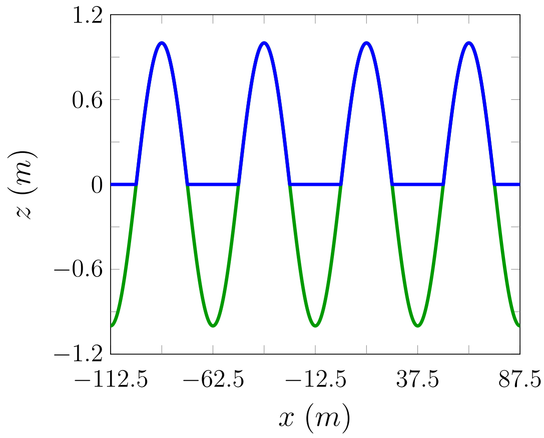
Figure 5.2: Convergence plots as measured by the L_1 norm for h ($\triangle$), u ($\square$) and G ($\Diamond$) for the soliton problem for all methods.



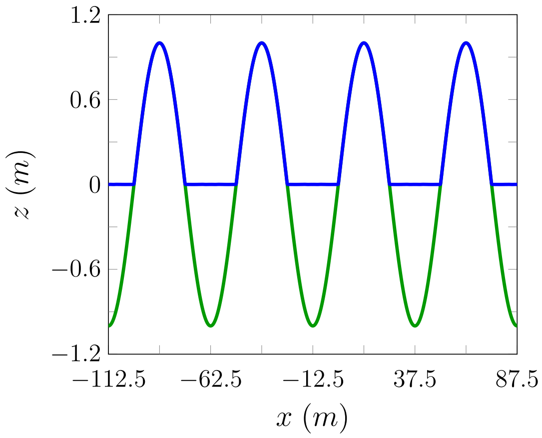
(a) FEVM_2 well balanced



(b) FEVM_2 not well balanced

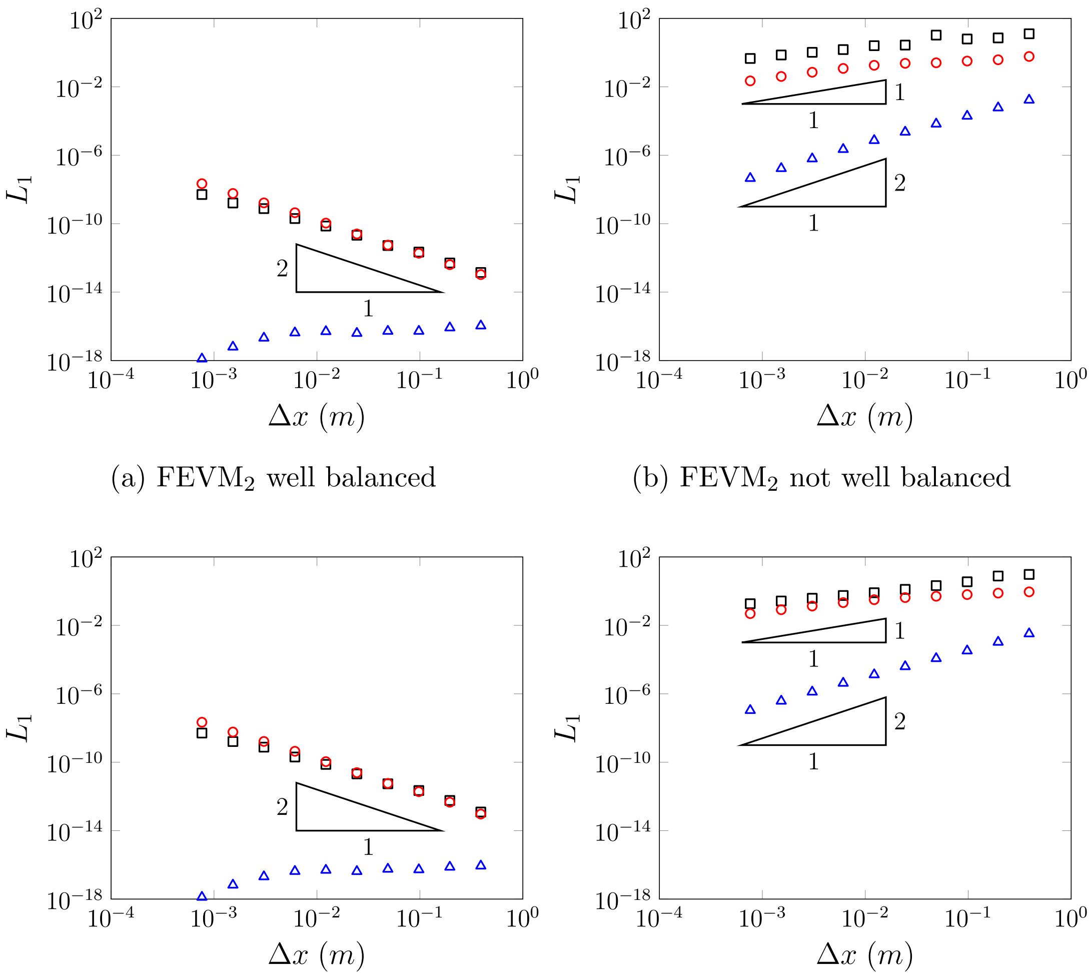


(c) FDVM_2 well balanced



(d) FDVM_2 not well balanced

Figure 5.3: Comparison of the analytic solution (−) and numerical solution with $\Delta x = 200/2^{10}m$ (•) for the lake at rest problem at $t = 10s$ for all methods.



(a) FEVM$_2$ well balanced

(b) FEVM$_2$ not well balanced

(c) FDVM$_2$ well balanced

(d) FDVM$_2$ not well balanced

Figure 5.4: Convergence plots as measured by the L_1 norm for h ($\triangle$), u ($\square$) and G ($\Diamond$) for the lake at rest problem at $t = 10s$ for all methods.

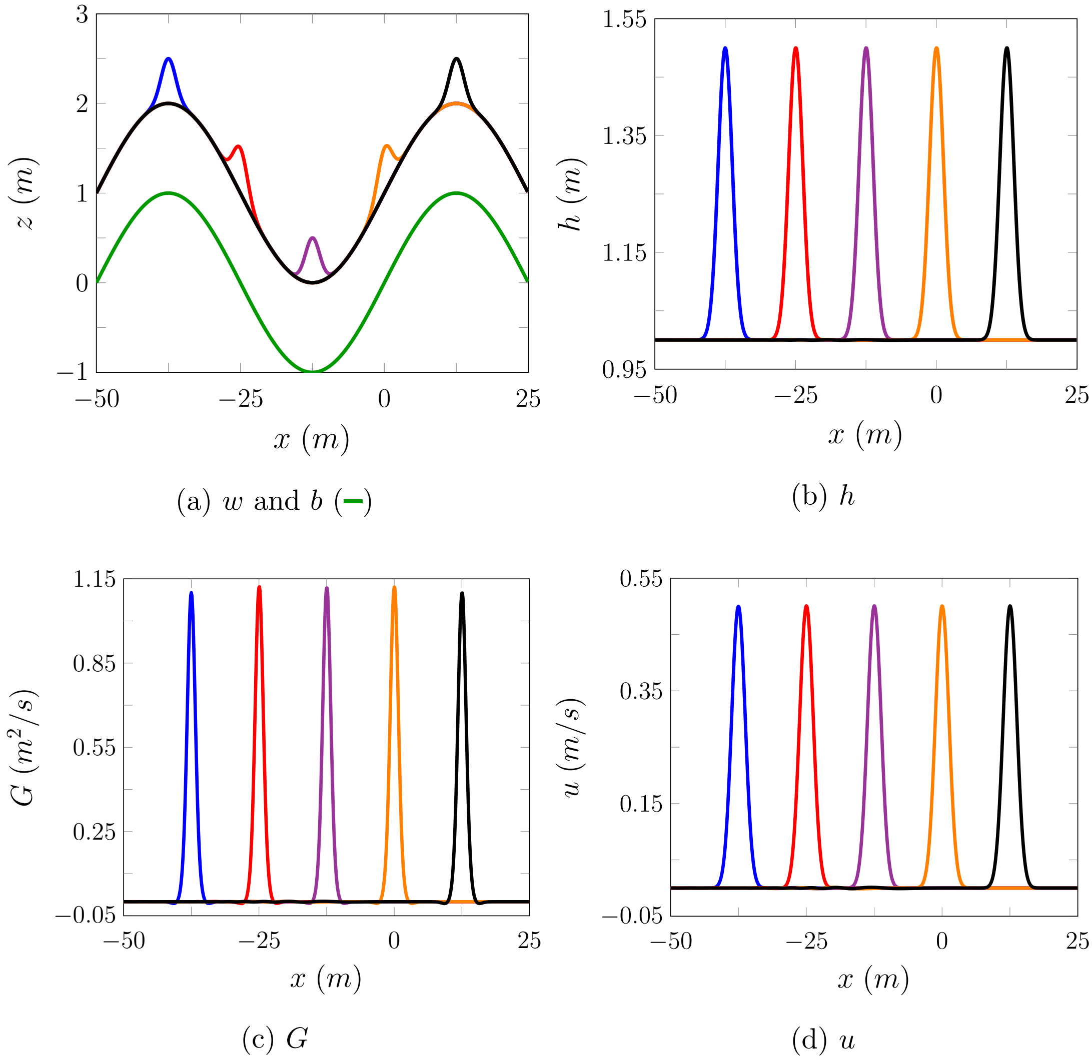
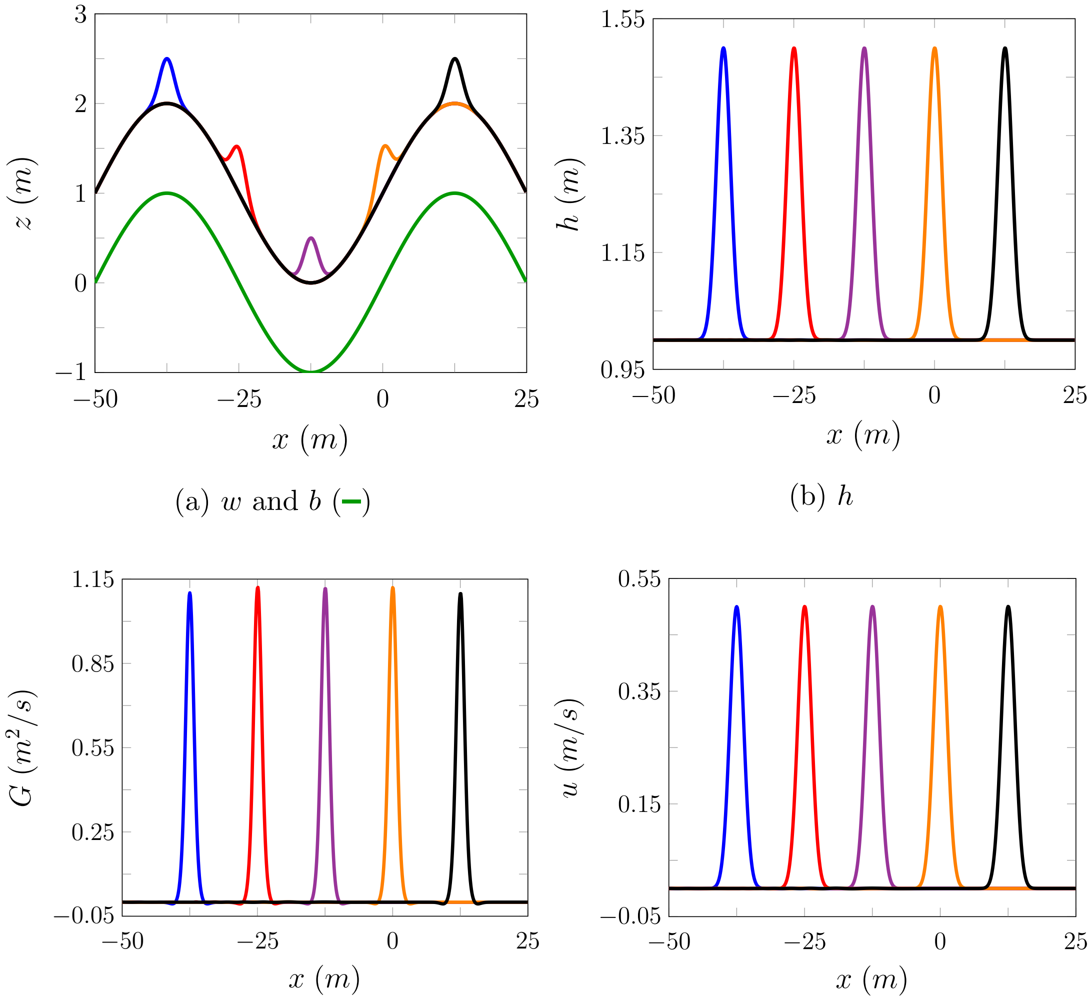


Figure 5.5: Plots of various quantities for the FEVM numerical solution at $0s$ (—), $2.5s$ (—), $5.0s$ (—), $7.5s$ (—), $10.0s$ (—) of the forced Serre equations.



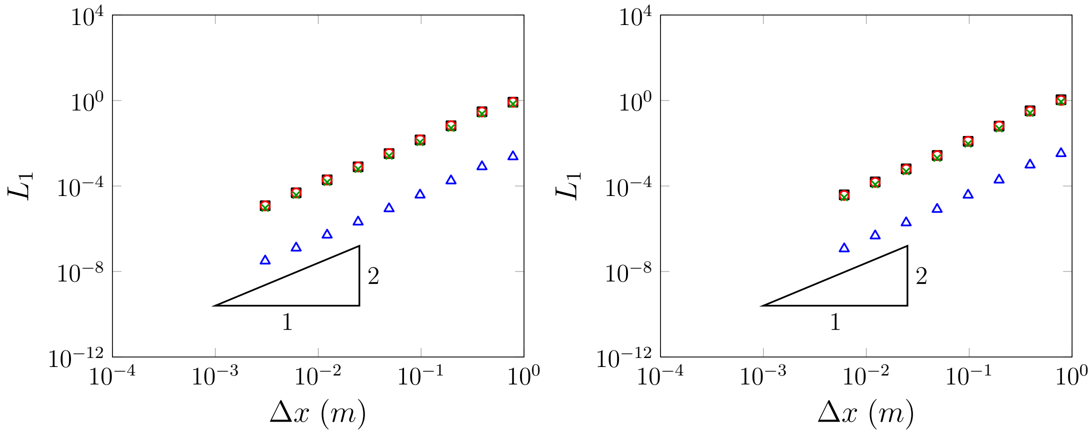
(a) w and b (—)

(b) h

(c) G

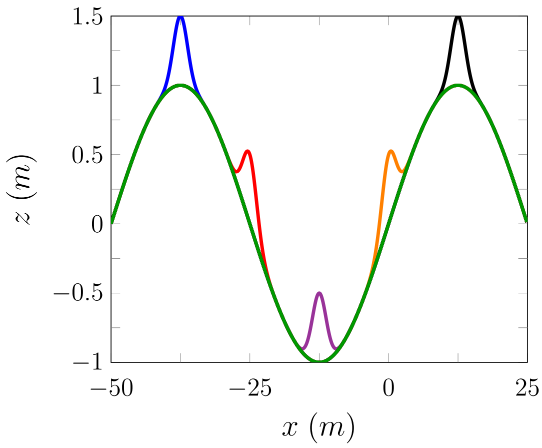
(d) u

Figure 5.6: Plots of various quantities for the FDVM numerical solution at $0s$ (—), $2.5s$ (—), $5.0s$ (—), $7.5s$ (—), $10.0s$ (—) of the forced Serre equations.

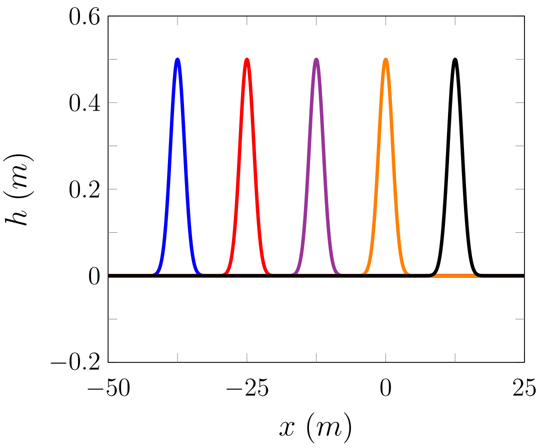


(a) FEVM_2 (b) FDVM_2

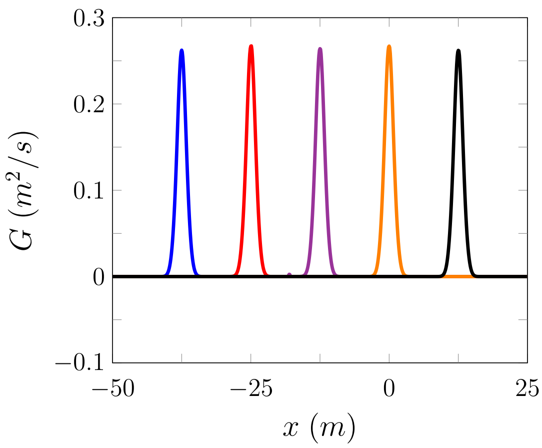
Figure 5.7: Convergence plots as measured by the L_1 norm for h (△), u (□), uh (×) and G (○) for the forced solution problem for FEVM and FDVM at $t = 10s$.



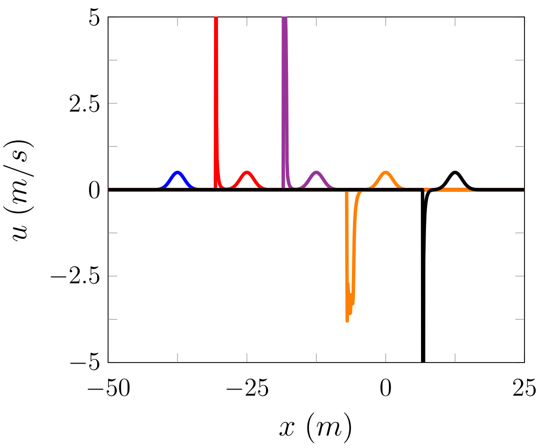
(a) w and b (—)



(b) h



(c) G



(d) u

Figure 5.8: Plots of various quantities for the FEVM numerical solution at $0s$ (—), $2.5s$ (—), $5.0s$ (—), $7.5s$ (—), $10.0s$ (—) of the forced Serre equations.

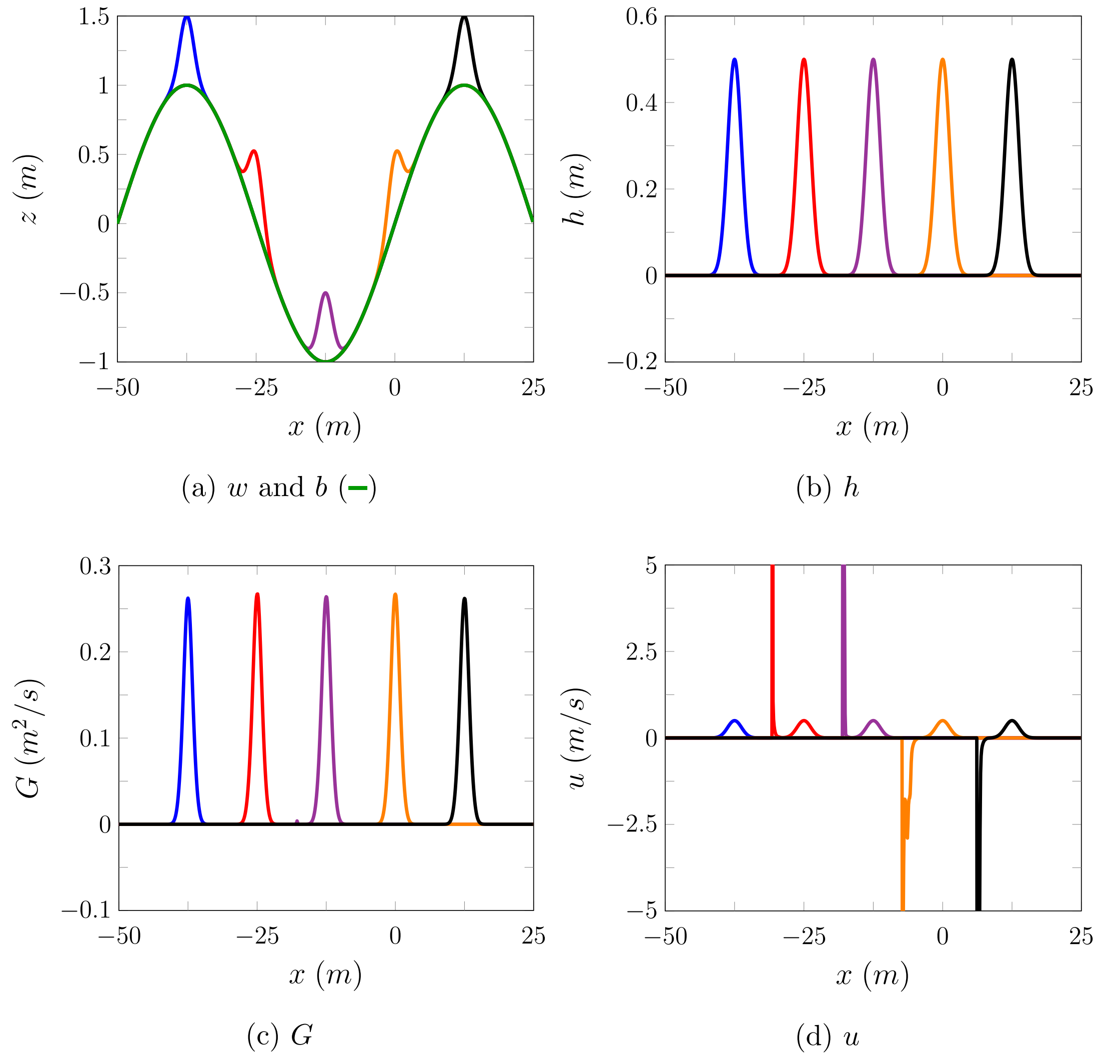
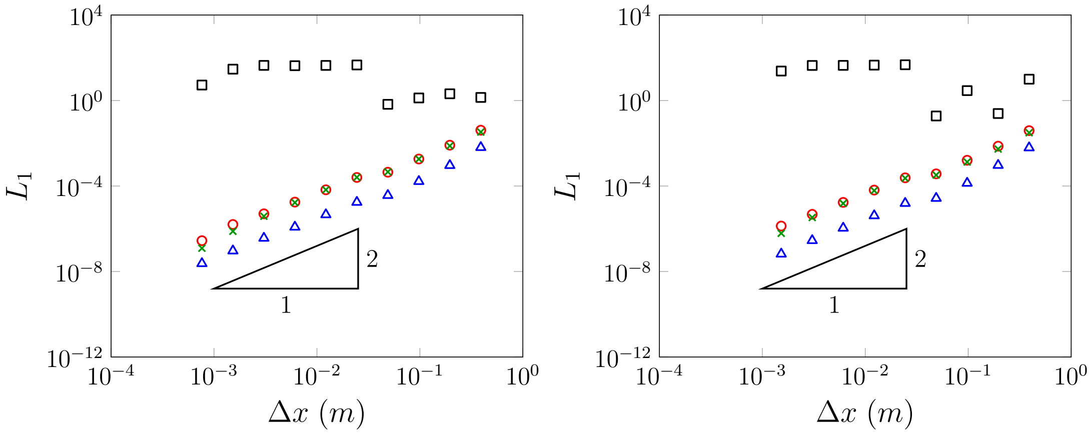
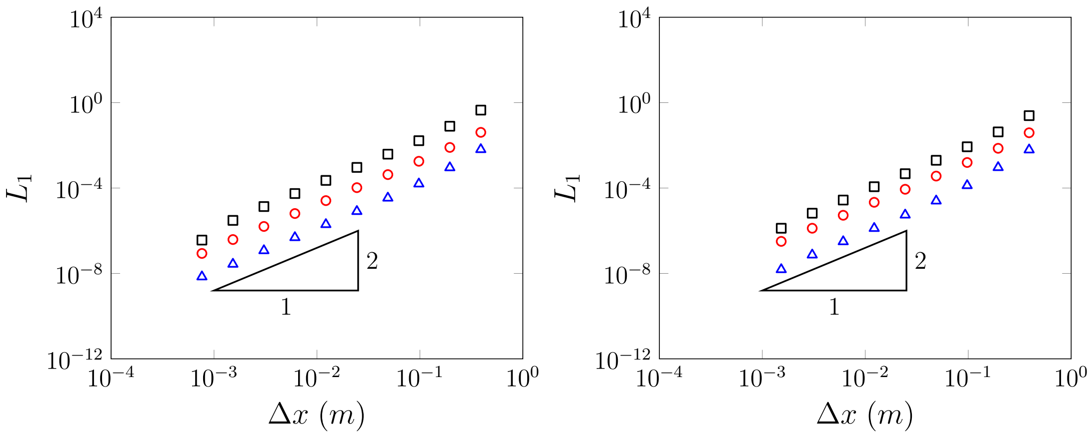


Figure 5.9: Plots of various quantities for the FDVM numerical solution at $0s$ (−), $2.5s$ (−), $5.0s$ (−), $7.5s$ (−), $10.0s$ (−) of the forced Serre equations.



(a) FEVM_2 (b) FDVM_2

Figure 5.10: Convergence plots as measured by the L_1 norm for h ($\triangle$), u ($\square$), uh ($\times$) and G ($\circ$) for the forced solution problem for FEVM and FDVM at $t = 10s$.



(a) FEVM_2 (b) FDVM_2

Figure 5.11: Convergence plots as measured by the L_1 norm around the peak for h ($\triangle$), u ($\square$) and G ($\circ$) for the forced solution problem for FEVM and FDVM at $t = 10s$.

Chapter 6

Experimental Validation

6.1 Periodic Waves Over A Submerged Bar

Beji and Battjes conducted a series of experiments investigating the effect of submerged bars on the propagation of periodic waves [25, 26]. The behaviour of these experiments were mainly driven by the dispersion properties of the waves and their interaction with variations in bathymetry. Therefore, these experiments serve as a good benchmark for our numerical schemes abilities to accurately model both the effect of variable bathymetry and dispersive waves. For our purposes we will focus on the monochromatic wave experiments of Beji and Battjes [26].

The experiments of Beji and Battjes [26] were conducted in a wave tank $37.7m$ long, $0.8m$ wide and $0.75m$ high. A diagram of the longitudinal section of the wave tank is given in Figure 6.1. There are seven wave gauges at the following locations; $5.7m$, $10.5m$, $12.5m$, $13.5m$, $14.5m$, $15.7m$ and $17.3m$. Waves are generated from a piston-type wave maker located at $0m$ and travel on the initially still water to the right, over the submerged trapezoidal bar towards a wave absorbing sloped beach.

Two sinusoidal monochromatic non-breaking wave experiments were conducted. A low frequency one with a wavelength $\lambda \approx 3.69m$ and a period of $T = 2s$, and a high frequency one with $\lambda \approx 2.05m$ and a period of $T = 1.25s$. Both experiments had a wave amplitude of $0.01m$ and so both had the same non-linearity parameter $\epsilon = 0.01/0.4 = 0.025$.

We numerically simulated these experiments over the spatial domain $[5.7m, 150m]$ with $\Delta x = 0.1/2^4 m$ and $\Delta t = Sp/2^5$ where $Sp = 0.039$ is the experimental sampling period. These Δx and Δt values satisfy the CFL condition (3.31) for these experiments. In our numerical experiments only the submerged trapezoidal bar

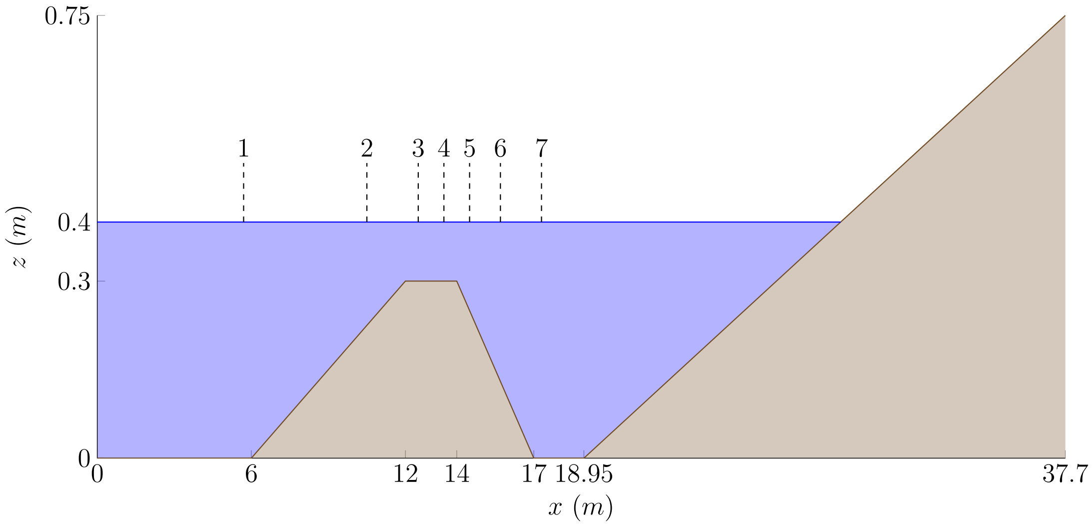


Figure 6.1: Diagram demonstrating the water (■) and the ground (■) for the Beji experiments, with the wave gauge locations marked.

is present, and the sloping beach is replaced with a very long horizontal bed that ensures that we do not observe any boundary effects in our results.

To simulate the incoming waves at the upstream boundary we used the first wave gauge as our left boundary condition and used linear extrapolation to calculate the other required h values in the left ghost cell. The velocity boundary conditions were calculated from the height values by solving the continuity equation (2.1a) assuming u and h are travelling wave solutions

$$u(x,t) = \sqrt{gh_0}\,\frac{h(x,t) - h_0}{h(x,t)}.$$

Finally the boundary conditions for G were calculated using the boundary conditions for h and u. We shall now present our numerical results for the low and high frequency experiments.

6.1.1 Low Frequency Results

A comparison of wave heights of the experimental and numerical results are located in Figures 6.2 and 6.3 for FEVM$_2$ and Figures 6.4 and 6.5 for FDVM$_2$. These numerical schemes both produce identical results for all wave gauges and so this benchmark does not help us discriminate between these two methods.

These results demonstrate the ability of these numerical methods to recreate the experimental results, particularly for wave gauge 1 to 5 where the agreement between experimental and numerical results is best. This validates the numerical schemes for simulating shoaling of dispersion waves as these wave gauges are all located on the windward side of the submerged bar where shoaling occurs in the experiment.

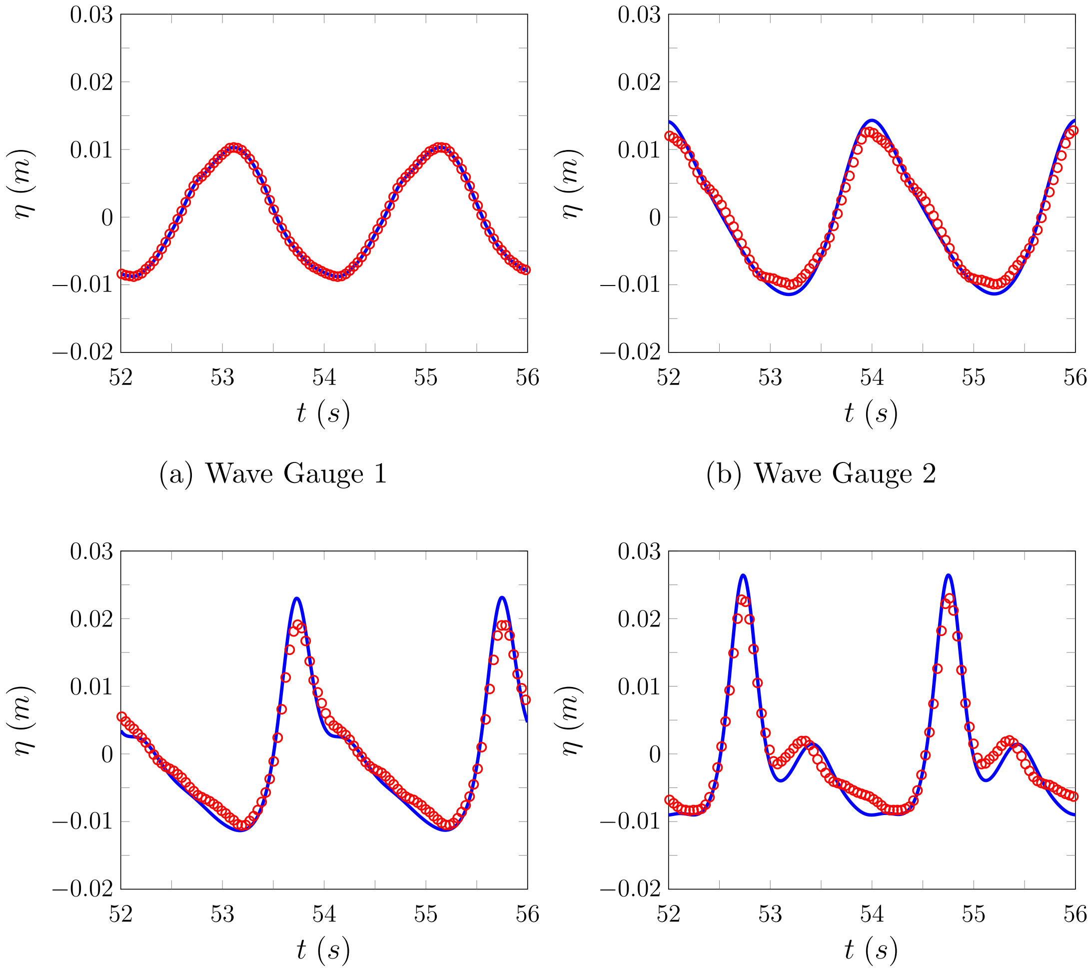
The numerical results for wave gauges 6 and 7 on the leeward side capture some of the wave behaviour but their agreement with the experiments results is much worse. The inadequacy of the numerical results here appears to be due to the discrepancy between the dispersion properties of the Serre equations and water waves, as the numerical solutions of improved dispersion equations [26, 27] accurately reproduce the experimental results on the leeward side.

The dispersion of the Serre equations is vital to recreating the experimental results for wave gauges 2 to 5, as non-dispersive equations such as the SWWE do a very poor job at simulating this experiment [28]. However, to properly reproduce the experimental results on the leeward side of the slope at wave gauges 6 and 7 would require improving the dispersion characteristics of the underlying Serre equations as done by Barthélemy [10]. Such an improvement can be incorporated into the hybrid FDVM and FEVM numerical methods [5] but is beyond the scope of this thesis.

6.1.2 High Frequency Results

The wave heights of the experimental and numerical results are given in Figures 6.6 and 6.7 for FEVM_2. While the results for FDVM_2 are given in Figures 6.8 and 6.9. As for the low frequency experiment, these numerical schemes FEVM_2 and FDVM_2 produce identical results for all wave gauges at this scale and so this benchmark does not discriminate between these two methods.

As in the low frequency experiment we observe that the numerical results perform well on the windward side of the slope for wave gauges 1 to 4 but perform poorly for the leeward side of the slope for wave gauges 5 to 7. With the high frequency experiment we see the divergence between the numerical and experimental results earlier than the low frequency experiment, so that now wave gauge 5 which is on the leeward side exhibits a significant difference between the numerical and experimental results. As in the low frequency example improving the dispersion properties of the governing partial differential equations lead to a much better agreement between the numerical and experimental results [26, 27]. Because the difference between the dispersion relation of the Serre equations and water waves is largest for higher frequency and therefore for shorter waves Barthélemy [10] the earlier divergence between experimental and numerical results is expected.



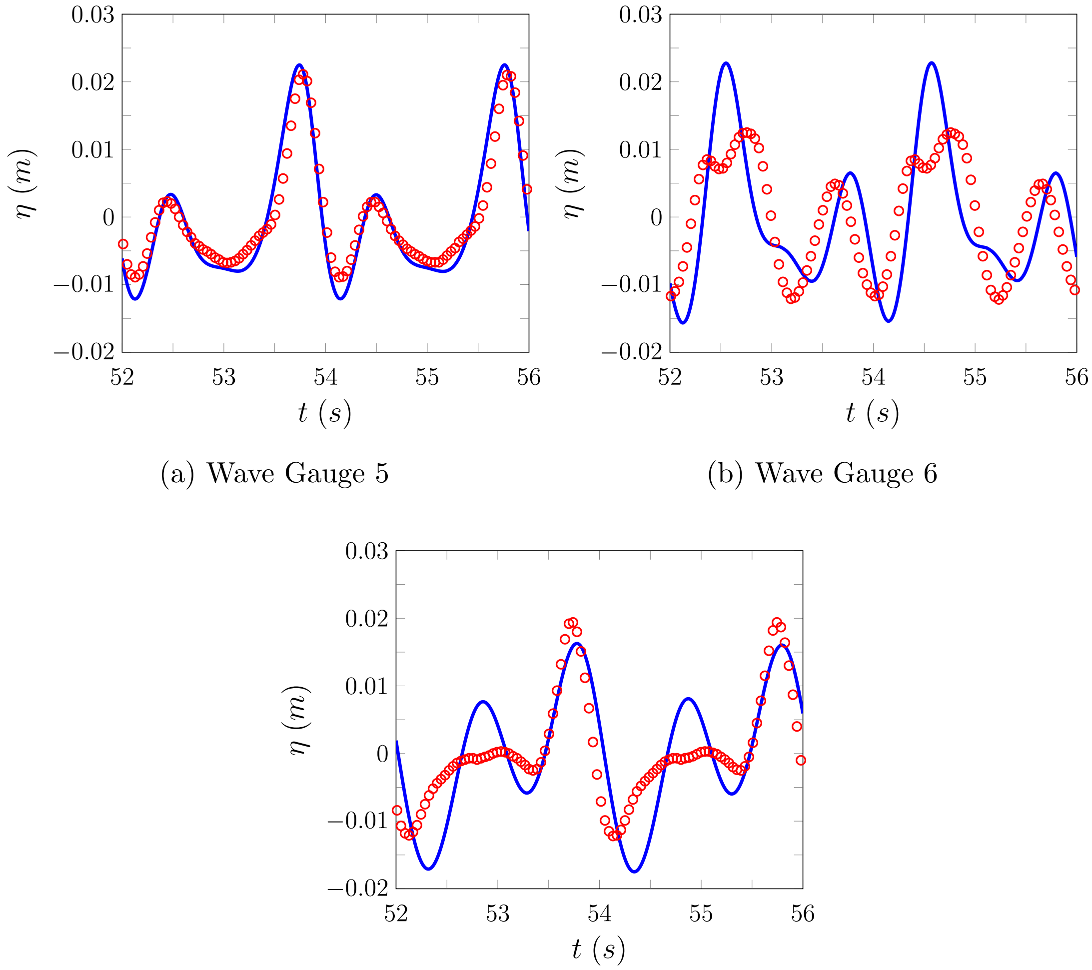
(a) Wave Gauge 1

(b) Wave Gauge 2

(c) Wave Gauge 3

(d) Wave Gauge 4

Figure 6.2: Comparison of the wave heights η of the numerical results for the FEVM_2 (−) and the experimental results (○) for wave gauges 1 - 4 for the low frequency experiment.



(a) Wave Gauge 5

(b) Wave Gauge 6

(c) Wave Gauge 7

Figure 6.3: Comparison of the wave heights η of the numerical results for the FEVM$_2$ (−) and the experimental results (○) for wave gauges 5 - 7 for the low frequency experiment.

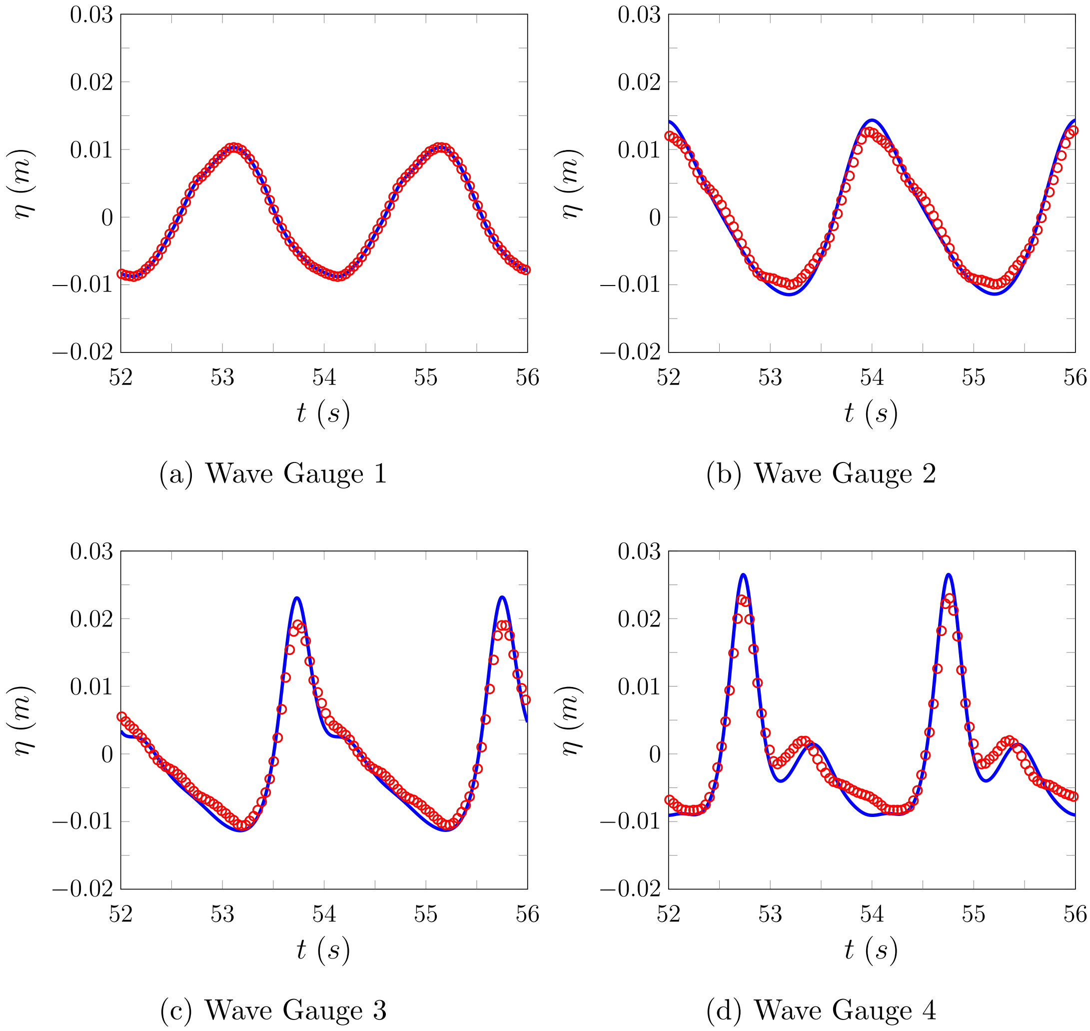
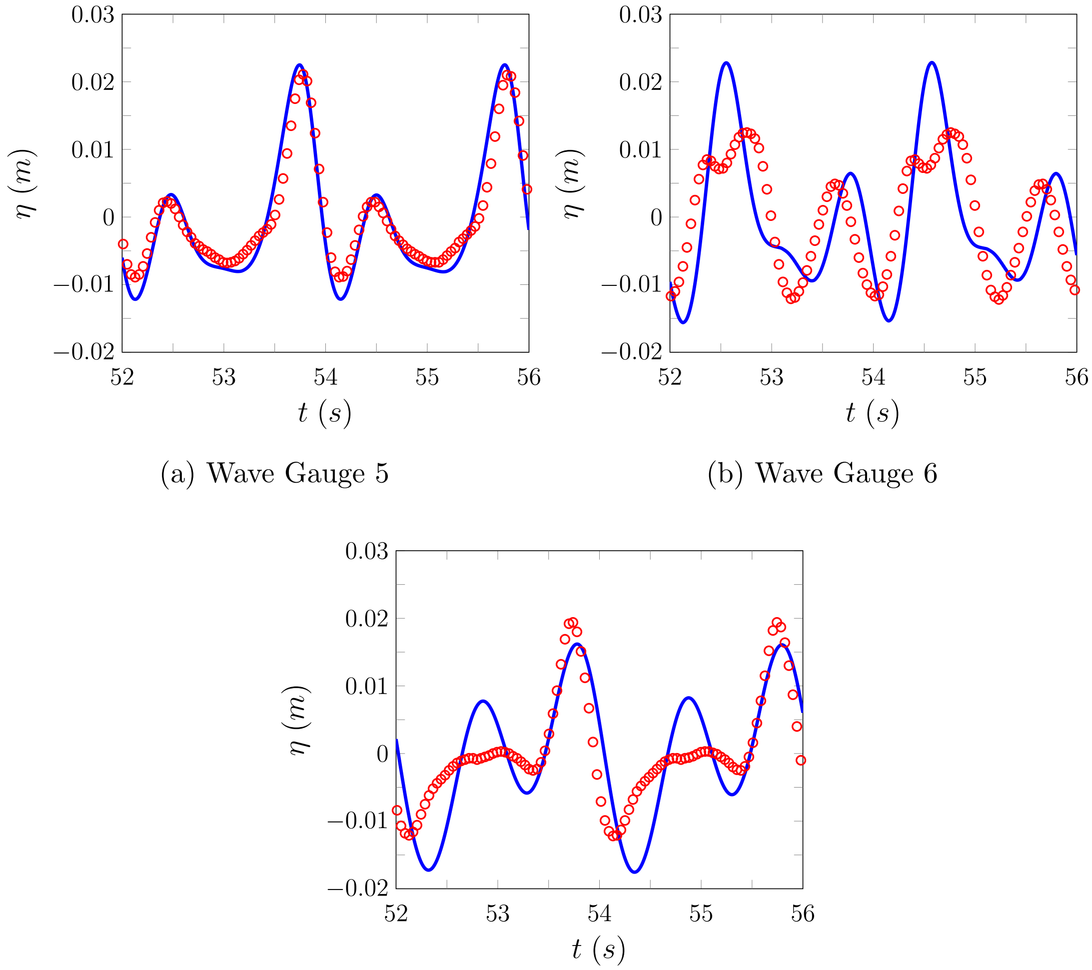


Figure 6.4: Comparison of the wave heights η of the numerical results for the FDVM$_2$ (–) and the experimental results (○) for wave gauges 1 - 4 for the low frequency experiment.



(a) Wave Gauge 5

(b) Wave Gauge 6

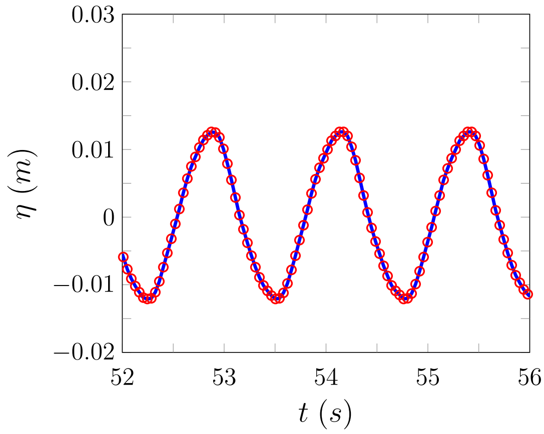
(c) Wave Gauge 7

Figure 6.5: Comparison of the wave heights η of the numerical results for the FDVM$_2$ (–) and the experimental results (○) for wave gauges 5 - 7 for the low frequency experiment.

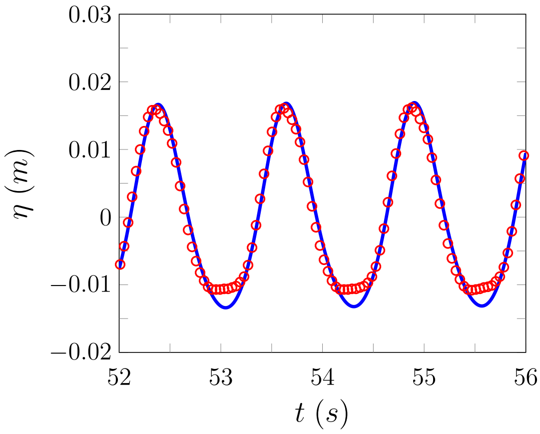
These numerical results for the $FDVM_2$ and $FEVM_2$ agree well with other numerical results for weakly dispersive equations without improved dispersion properties for the simulation of periodic waves over a submerged bar in the literature [26, 27, 7, 29]. Therefore, without changing the underlying partial differential equations, our numerical methods as well as these numerical schemes at recreating the experimental results of Beji and Battjes [26].

6.2 Synolakis

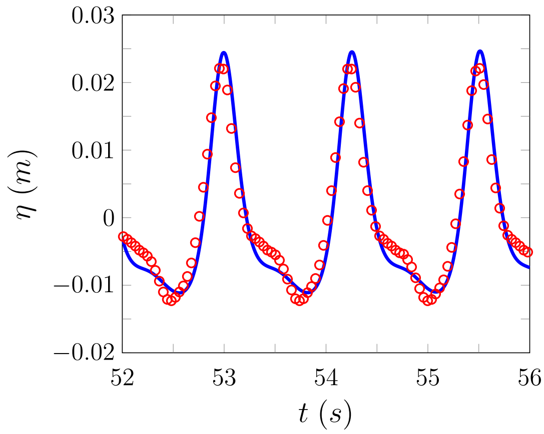
6.3 Roeber



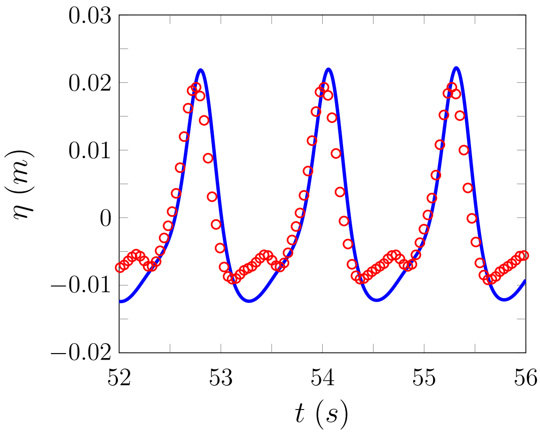
(a) Wave Gauge 1



(b) Wave Gauge 2

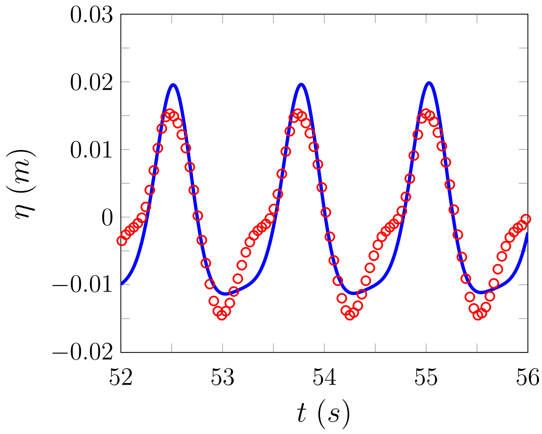


(c) Wave Gauge 3

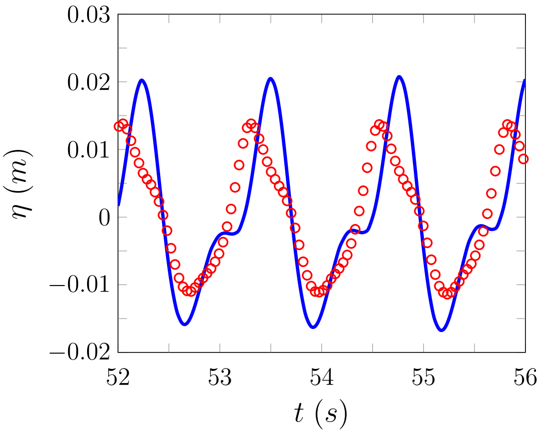


(d) Wave Gauge 4

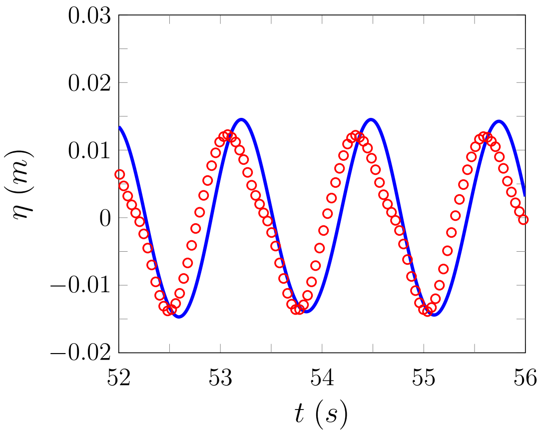
Figure 6.6: Comparison of the wave heights η of the numerical results for the FEVM_2 (—) and the experimental results (○) for wave gauges 1 - 4 for the low frequency experiment.



(a) Wave Gauge 5

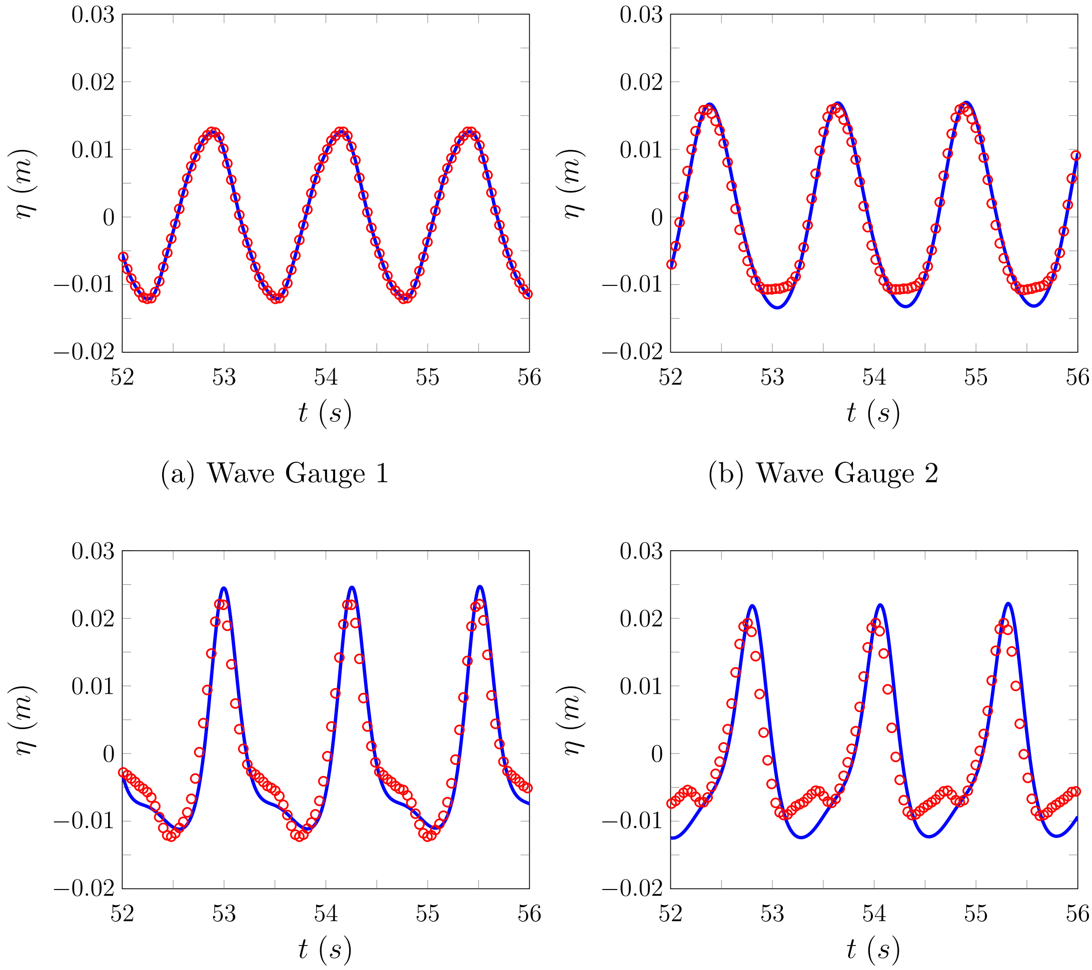


(b) Wave Gauge 6



(c) Wave Gauge 7

Figure 6.7: Comparison of the wave heights η of the numerical results for the FEVM$_2$ (–) and the experimental results (○) for wave gauges 5 - 7 for the high frequency experiment.



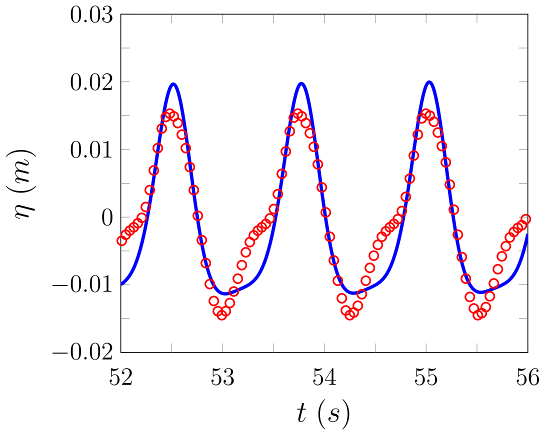
(a) Wave Gauge 1

(b) Wave Gauge 2

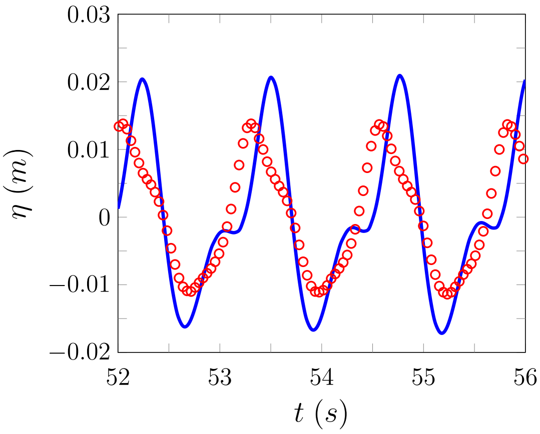
(c) Wave Gauge 3

(d) Wave Gauge 4

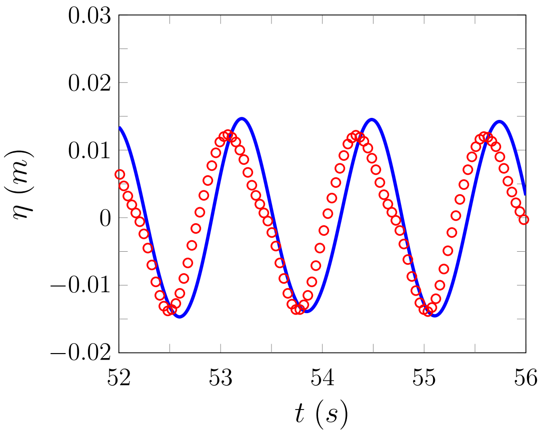
Figure 6.8: Comparison of the wave heights η of the numerical results for the FDVM_2 (−) and the experimental results (○) for wave gauges 1 - 4 for the high frequency experiment.



(a) Wave Gauge 5

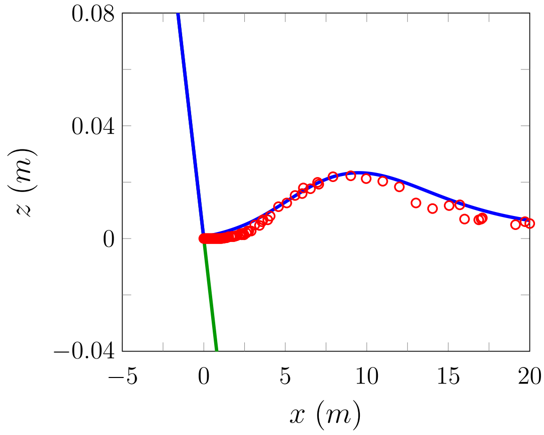


(b) Wave Gauge 6

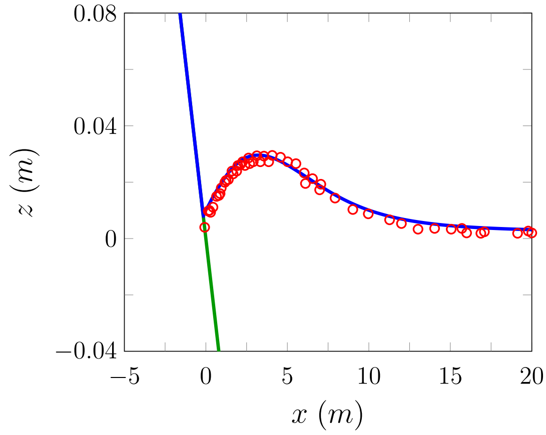


(c) Wave Gauge 7

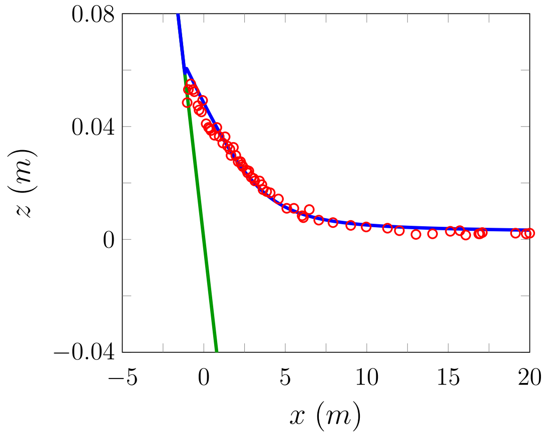
Figure 6.9: Comparison of the wave heights η of the numerical results for the FDVM$_2$ (−) and the experimental results (○) for wave gauges 5 - 7 for the high frequency experiment.



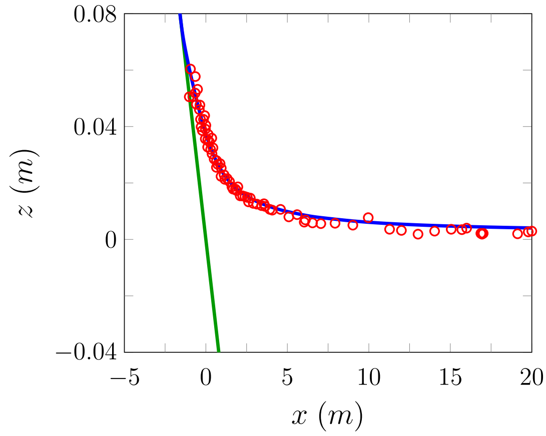
(a) $t = 30s$



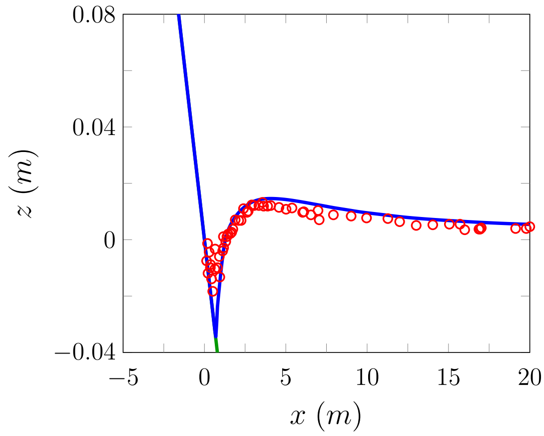
(b) $t = 40s$



(c) $t = 50s$

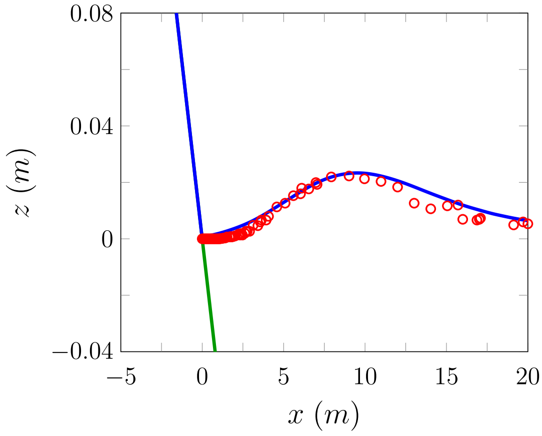


(d) $t = 60s$

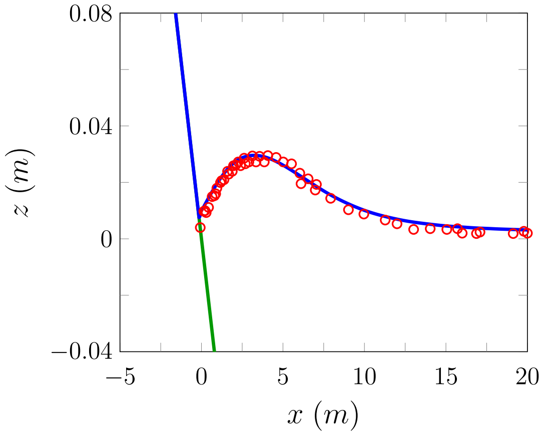


(e) $t = 70s$

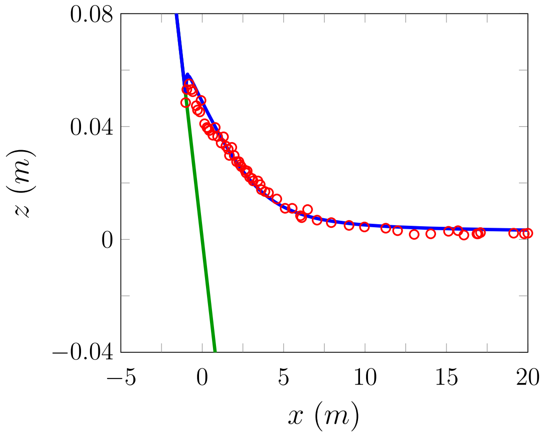
Figure 6.10: FEVM nonbreak times



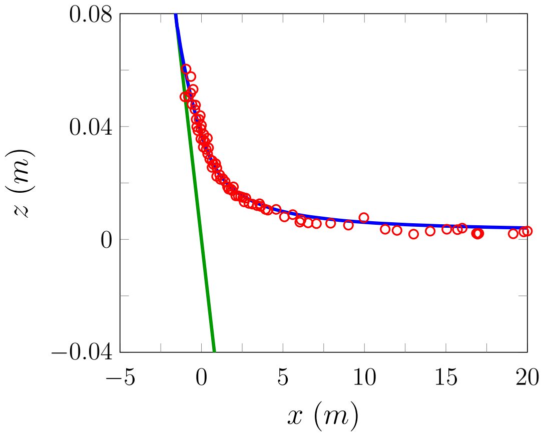
(a) $t = 30s$



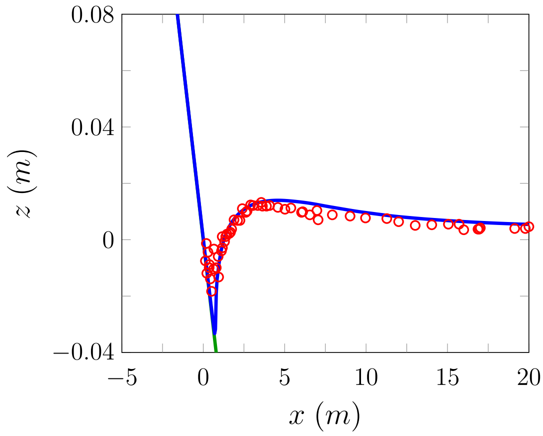
(b) $t = 40s$



(c) $t = 50s$



(d) $t = 60s$



(e) $t = 70s$

Figure 6.11: FDVM nonbreak times

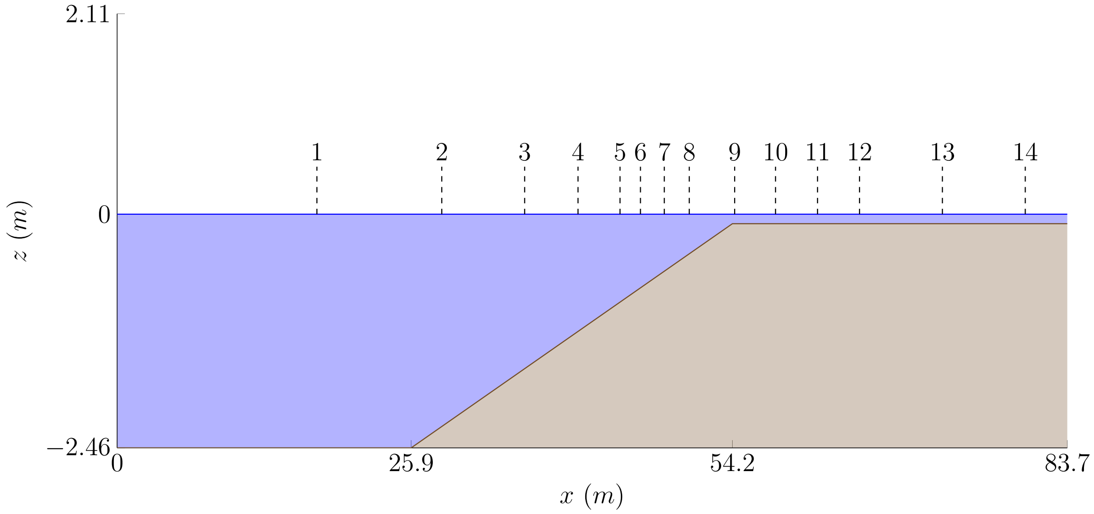
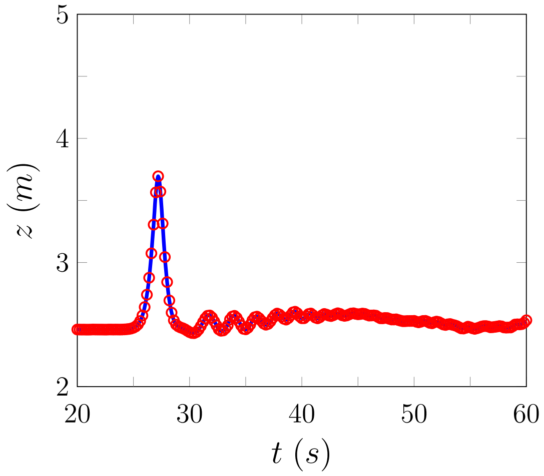
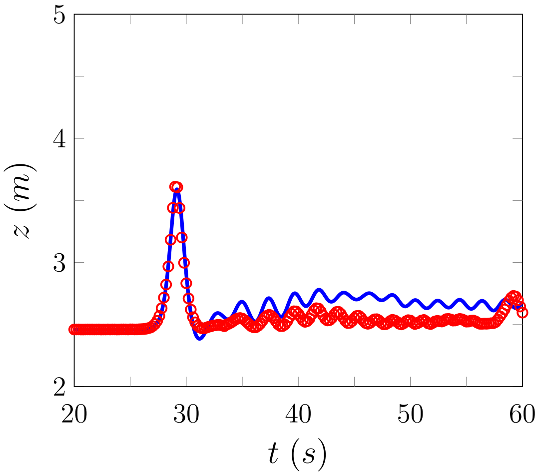


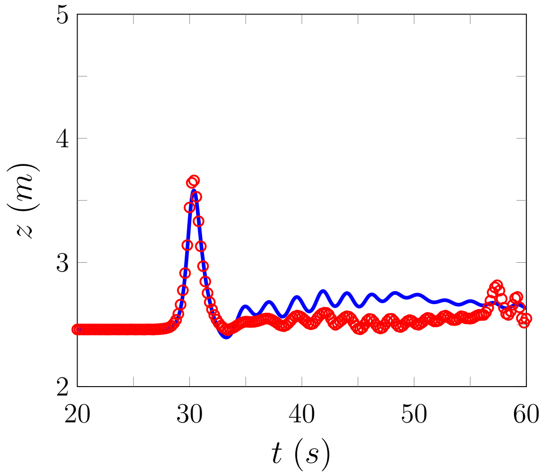
Figure 6.12: Diagram demonstrating the water (■) and the ground (■) for the Beji experiments, with the wave gauge locations marked.



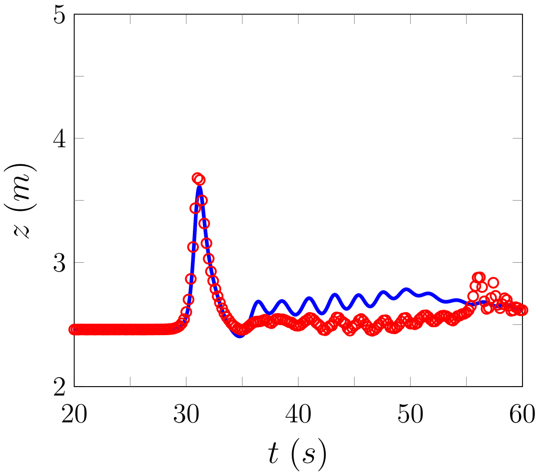
(a) Wave Gauge 1



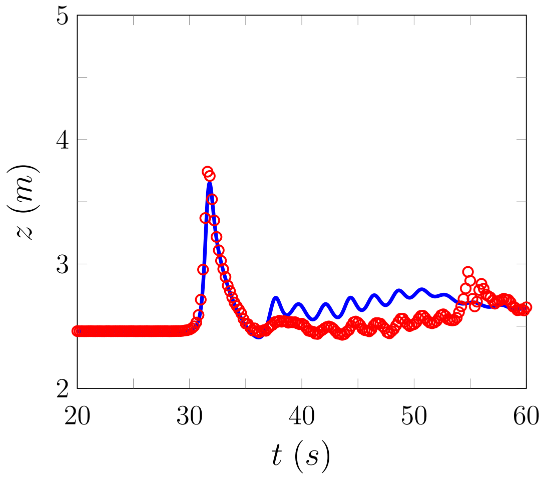
(b) Wave Gauge 2



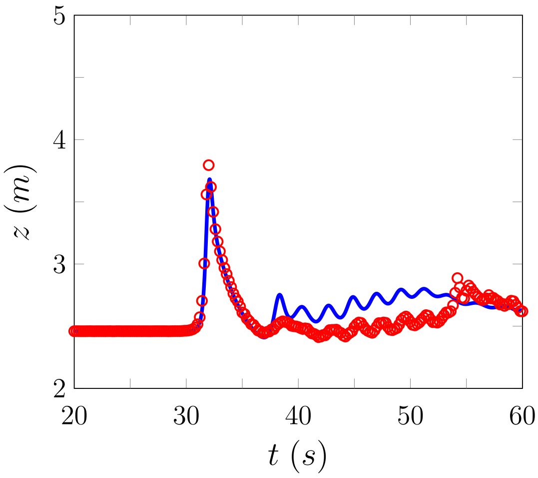
(c) Wave Gauge 3



(d) Wave Gauge 4

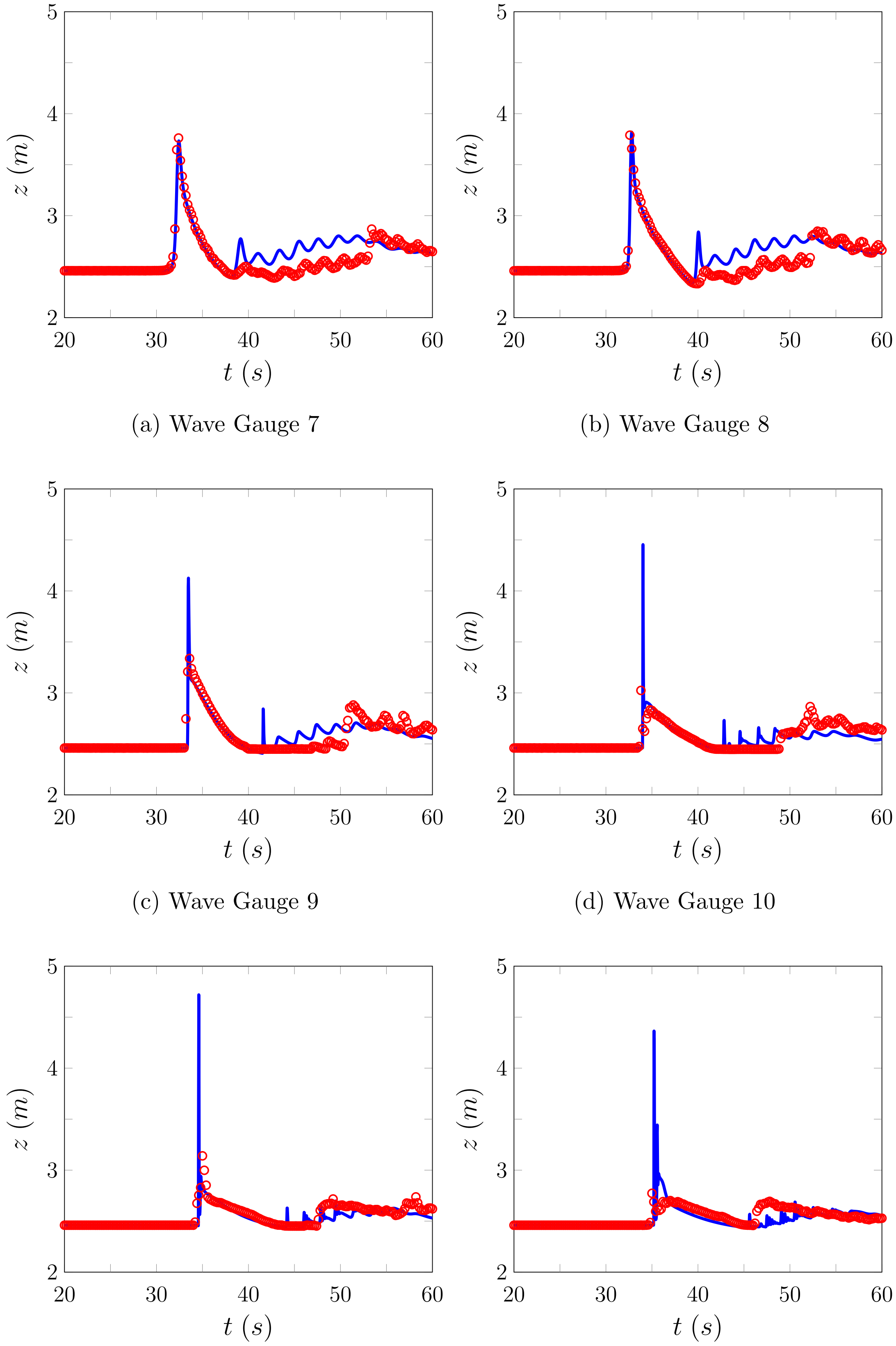


(e) Wave Gauge 5



(f) Wave Gauge 6

Figure 6.13: .



(a) Wave Gauge 7

(b) Wave Gauge 8

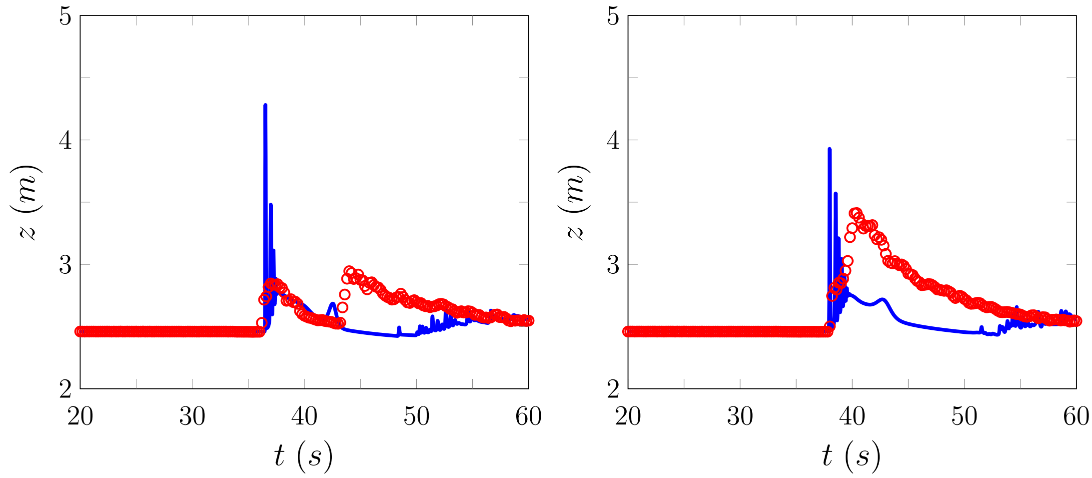
(c) Wave Gauge 9

(d) Wave Gauge 10

(e) Wave Gauge 11

(f) Wave Gauge 12

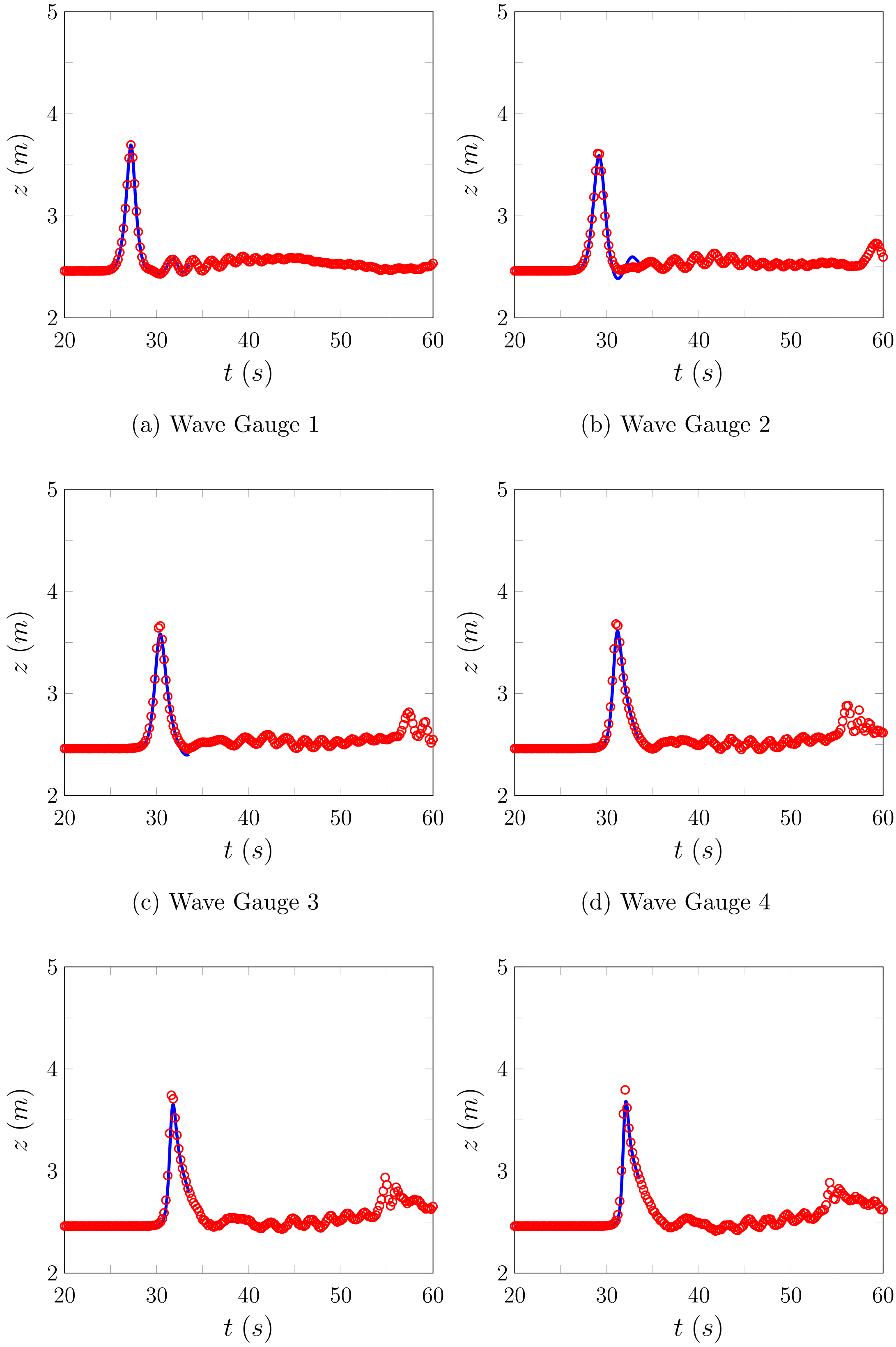
Figure 6.14: .



(a) Wave Gauge 13

(b) Wave Gauge 14

Figure 6.15: .



(a) Wave Gauge 1

(b) Wave Gauge 2

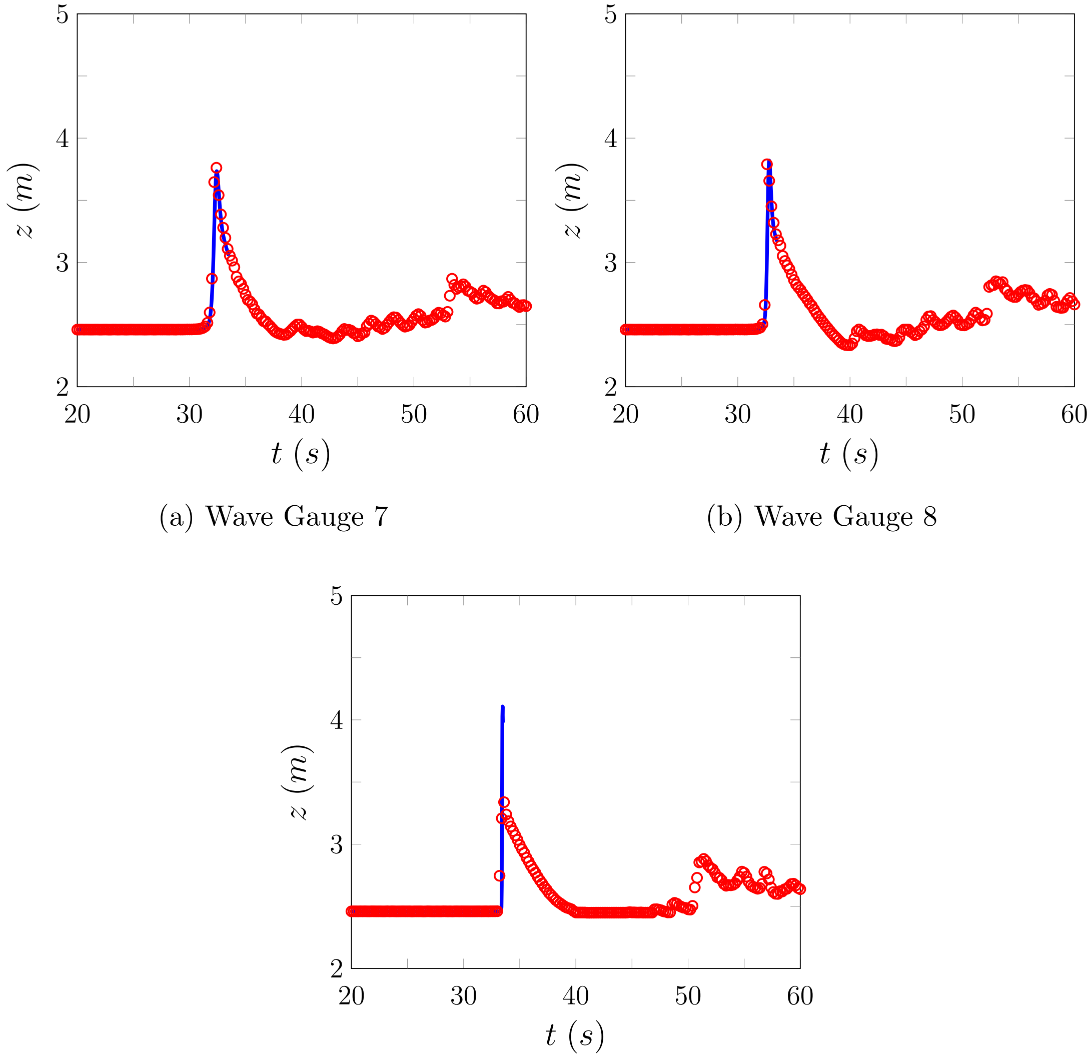
(c) Wave Gauge 3

(d) Wave Gauge 4

(e) Wave Gauge 5

(f) Wave Gauge 6

Figure 6.16: .



(a) Wave Gauge 7

(b) Wave Gauge 8

(c) Wave Gauge 9

Figure 6.17: .

Appendix A

Linear Analysis Results

[]

$\mathcal{D}$

$$\mathbf{E} = \begin{bmatrix} -\frac{2i\Delta t}{\Delta x} U \sin{(k\Delta x)} & -\frac{2i\Delta t}{\Delta x} H \sin{(k\Delta x)} & 1 & 0 \\ -\frac{6gi\Delta x\Delta t}{3\Delta x^2 - 2H^2\left(\cos{(k\Delta x)} - 1\right)} \sin{(k\Delta x)} & -\frac{2i\Delta t}{\Delta x} U \sin{(k\Delta x)} & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

$\mathcal{W}$

$$\mathbf{E} = \begin{bmatrix} E^{0,0} & E^{0,1} & 0 & -\frac{\Delta t}{\Delta x} H \frac{i\sin(k\Delta x)}{2} \\ E^{1,0} & -\frac{2i\Delta t}{\Delta x} U \sin{(k\Delta x)} & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{A.1}$$

with

$$E^{0,0} = 1 - \frac{\Delta t}{\Delta x} \left(-\frac{6gi\Delta x\Delta t}{3\Delta x^2 - 2H^2\left(\cos{(k\Delta x)} - 1\right)} \sin{(k\Delta x)} \right) H \frac{i\sin{(k\Delta x)}}{2}$$
$$- \frac{\Delta t}{\Delta x} U \left(\left(i\sin{(k\Delta x)}\right) - \frac{\Delta t}{\Delta x} U \left(\cos{(k\Delta x)} - 1\right) \right),$$
$$E^{0,1} = -\frac{\Delta t}{\Delta x} \left[H \frac{i\sin{(k\Delta x)}}{2} \left(1 - \frac{2i\Delta t}{\Delta x} U \sin{(k\Delta x)} \right) - U \left(\frac{\Delta t}{\Delta x} H \left(\cos{(k\Delta x)} - 1\right) \right) \right],$$
$$E^{1,0} = -\frac{6gi\Delta x\Delta t}{3\Delta x^2 - 2H^2\left(\cos{(k\Delta x)} - 1\right)} \sin{(k\Delta x)}.$$

Scheme	Expression	Lowest Order Term of Error
FDVM_1	1	$-\frac{1}{24}k^2\Delta x^2$
FDVM_2 and FEVM_2	1	$-\frac{1}{24}k^2\Delta x^2$
FDVM_3	$\frac{26 - 2\cos(k\Delta x)}{24}$	$-\frac{3}{640}k^4\Delta x^4$

Table A.1: Factor $\mathcal{M}$ from transformation between nodal and cell average values. Where the analytic value is $\mathcal{M} = \dfrac{k\Delta x}{2\sin\left(k\frac{\Delta x}{2}\right)}$.

Scheme	Formula	Lowest Order Term of Error
FDVM_1	$e^{ik\Delta x}$	$\frac{i}{2}k\Delta x$
FDVM_2 and FEVM_2	$e^{ik\Delta x}\left(1 - \frac{i\sin(k\Delta x)}{2}\right)$	$\frac{1}{12}k^2\Delta x^2$
FDVM_3	$\frac{e^{ik\Delta x}}{6}\left(5 + 2e^{-ik\Delta x} - e^{ik\Delta x}\right)$	$\frac{i}{12}k^3\Delta x^3$

Table A.2: Factor $\mathcal{R}^+$ from reconstruction of η and G at $x^+_{j+1/2}$. Where the analytic value is $\mathcal{R}^+ = e^{ik\Delta x/2}\dfrac{k\Delta x}{2\sin\left(\frac{k\Delta x}{2}\right)}$.

Scheme	Expression	Lowest Order Term of Error
FDVM_1	1	$-\frac{i}{2}k\Delta x$
FDVM_2 and FEVM_2	$1+\frac{i\sin(k\Delta x)}{2}$	$\frac{1}{12}k^2\Delta x^2$
FDVM_3	$\frac{1}{6}\left(5-e^{-ik\Delta x}+2e^{ik\Delta x}\right)$	$-\frac{i}{12}k^3\Delta x^3$

Table A.3: Factor $\mathcal{R}^-$ from reconstruction of η and G at $x^-_{j+1/2}$. Where the analytic value is $\mathcal{R}^- = e^{ik\Delta x/2}\dfrac{k\Delta x}{2\sin\left(\frac{k\Delta x}{2}\right)}$.

Scheme	Expression	Lowest Order Term of Error
FDVM_1	$\dfrac{3\Delta x^2\left(\frac{1+e^{ik\Delta x}}{2}\right)}{3\Delta x^2 H - H^3\left(2\cos(k\Delta x)-2\right)}$	$-\dfrac{6+H^2k^2}{4H\left(3+H^2k^2\right)^2}k^2\Delta x^2$
FDVM_2	$\dfrac{3\Delta x^2\left(\frac{1+e^{ik\Delta x}}{2}\right)}{3\Delta x^2 H - H^3\left(2\cos(k\Delta x)-2\right)}$	$-\dfrac{6+H^2k^2}{4H\left(3+H^2k^2\right)^2}k^2\Delta x^2$
FEVM_2	$\left(\frac{\Delta x}{6}\left(1+\frac{i\sin(k\Delta x)}{2}+e^{ik\Delta x}\left(1-\frac{i\sin(k\Delta x)}{2}\right)\right)\right) \div \left(H\frac{\Delta x}{30}\left(2\left(2\cos\left(\frac{k\Delta x}{2}\right)-\cos(k\Delta x)+4\right)\right) + \frac{H^3}{9\Delta x}\left(-16\cos\left(\frac{k\Delta x}{2}\right)+2\cos(k\Delta x)+14\right)\right)$	$\dfrac{12+5H^2k^2}{40H\left(3+H^2k^2\right)^2}k^2\Delta x^2$
FDVM_3	$\dfrac{36\Delta x^2\left(\frac{-e^{-ik\Delta x}+9e^{ik\Delta x}-e^{2ik\Delta x}+9}{16}\right)}{36\Delta x^2 H - H^3\left(32\cos(k\Delta x)-2\cos(2k\Delta x)-30\right)}$	$-\dfrac{243+49H^2k^2}{960H\left(3+H^2k^2\right)^2}k^4\Delta x^4$

Table A.4: Factor $\mathcal{G}$ from solving the elliptic equation (**??**) for $\upsilon_{j+1/2}$. Where the analytic value is $\mathcal{G} = \dfrac{3}{3H+H^3k^2}\dfrac{1}{e^{-ik\Delta x/2}}\dfrac{k\Delta x}{2\sin\left(\frac{k\Delta x}{2}\right)}$.

Bibliography

[1] F. Serre. Contribtion à l'étude des écoulements permanents et variables dans les canaux. *La Houille Blanche*, 6:830–872, 1953.

[2] D. Lannes and P. Bonneton. Derivation of asymptotic two-dimensional time-dependent equations for surface water wave propagation. *Physics of Fluids*, 21(1):16601–16610, 2009.

[3] A. E. Green and P. M. Naghdi. A derivation of equations for wave propagation in water of variable depth. *Journal of Fluid Mechanics*, 78(2):237–246, 1976.

[4] C. H. Su and C. S. Gardner. Korteweg-de Vries equation and generalisations. III. Derivation of the Korteweg-de Vries equation and Burgers equation. *Journal of Mathematical Physics*, 10(3):536–539, 1969.

[5] C. Zoppou. *Numerical Solution of the One-dimensional and Cylindrical Serre Equations for Rapidly Varying Free Surface Flows.* PhD thesis, Australian National University, Mathematical Sciences Institute, College of Physical and Mathematical Sciences, Australian National University, Canberra, ACT 2600, Australia, 2014.

[6] O. Le Métayer, S. Gavrilyuk, and S. Hank. A numerical scheme for the Green-Naghdi model. *Journal of Computational Physics*, 229(6):2034–2045, 2010.

[7] M. Li, P. Guyenne, F. Li, and L. Xu. High order well-balanced CDG-FE methods for shallow water waves by a GreenNaghdi model. *Journal of Computational Physics*, 257(1):169–192, 2014.

[8] W. Choi and R. Camassa. Fully nonlinear internal waves in a two-fluid system. *Journal of Fluid Mechanics*, 396:1–36, 1999.

[9] J.D Carter and R. Cienfuegos. The kinematics and stability of solitary and cnoidal wave solutions of the serre equations. *European Journal of Mechanics-B/Fluids*, 30(3):259–268, 2011.

[10] E. Barthélemy. Nonlinear shallow water theories for coastal waves. *Surveys in Geophysics*, 25(3):315–337, 2004.

[11] D. Dutykh, D. Clamond, P. Milewski, and D. Mitsotakis. Finite volume and pseudo-spectral schemes for the fully nonlinear 1D Serre equations. *European Journal of Applied Mathematics*, 24(5):761–787, 2013.

[12] G.A. El, R. H. J. Grimshaw, and N. F. Smyth. Unsteady undular bores in fully nonlinear shallow-water theory. *Physics of Fluids*, 18(2):027104, 2006.

[13] P.L Roe. Characteristic-based schemes for the euler equations. *Annual Review of Fluid Mechanics*, 18(1):337–365, 1986.

[14] B. Van Leer. Towards the ultimate conservative difference scheme. iv. a new approach to numerical convection. *Journal of Computational Physics*, 23(3): 276–299, 1977.

[15] L.C. Evans and American Mathematical Society. *Partial Differential Equations*. Graduate Studies in Mathematics. American Mathematical Society, 1998.

[16] A. Kurganov, S. Noelle, and G. Petrova. Semidiscrete central-upwind schemes for hyperbolic conservation laws and Hamilton-Jacobi equations. *Journal of Scientific Computing, Society for Industrial and Applied Mathematics*, 23(3):707–740, 2002.

[17] E. Audusse, F. Bouchut, M. Bristeau, R. Klein, and B. Perthame. A fast and stable well–balanced scheme with hydrostatic reconstruction for shallow water flows. *Journal of Scientific Computing, Society for Industrial and Applied Mathematics*, 25(6):2050–2065, 2004.

[18] J. Pitt. *A Second Order Well Balanced Hybrid Finite Volume and Finite Difference Method for the Serre Equations*. PhD thesis, Australian National University, Mathematical Sciences Institute, College of Physical and Mathematical Sciences, Australian National University, Canberra, ACT 2600, Australia, 2014.

[19] S. Gottlieb, C. Shu, and E. Tadmor. Strong stability-preserving high-order time discretization methods. *Society for Industrial and Applied Mathematics Review*, 43(1):89–112, 2001.

[20] R. Courant, K. Friedrichs, and H. Lewy. On the partial difference equations of mathematical physics. *IBM Journal of Research and Development*, 11(2): 215–234, 1967.

[21] C. Zoppou, J. Pitt, and S. Roberts. Numerical solution of the fully non-linear weakly dispersive serre equations for steep gradient flows. *Applied Mathematical Modelling*, 48:70–95, 2017.

[22] W.H Press, S.A Teukolsky, W.T Vetterling, and B.P Flannery. *Numerical Recipes in C*, volume 2. Cambridge University Press, 1996.

[23] P. Lax and R. Richtmyer. Survey of the stability of linear finite difference equations. *Communications on Pure and Applied Mathematics*, 9(2):267–293, 1956.

[24] A. G. Filippini, M. Kazolea, and M. Ricchiuto. A flexible genuinely nonlinear approach for nonlinear wave propagation, breaking and run-up. *Journal of Computational Physics*, 310:381–417, 2016.

[25] S Beji and J.A. Battjes. Experimental investigation of wave propagation over a bar. *Coastal Engineering*, 19(1):151–162, 1993.

[26] S Beji and J.A. Battjes. Numerical simulation of nonlinear wave propagation over a bar. *Coastal Engineering*, 23(1):1–16, 1994.

[27] D. Lannes. *The Water Waves Problem: Mathematical Analysis and Asymptotic*, volume 1 of *American Mathematical Society*. Mathematical Surveys and Monographs, 2013.

[28] J. Pitt, C. Zoppou, and S.G Roberts. Importance of dispersion for shoaling waves. *Modelling and Simulation Society of Australia and New Zealand*, 22 (1):1725–1730, 2017.

[29] Y. Zhang, A.B. Kennedy, N. Panda, C. Dawson, and J.J. Westerink. Boussinesq-green-naghdi rotational water wave theory. *Coastal Engineering*, 73(1):13–27, 2013.